

W i e s ł a w B a b i k

S Ł O W A
K L U C Z O W E



Wydawnictwo Uniwersytetu Jagiellońskiego

S Ł O W A
KLUCZOWE

Publikacja dofinansowana przez Uniwersytet Jagielloński ze środków Wydziału Zarządzania i Komunikacji Społecznej oraz Instytutu Informacji Naukowej i Bibliotekoznawstwa

RECENZENT

Prof. dr hab. Bożenna Bojar

PROJEKT OKŁADKI

Jadwiga Burek

© Copyright by Wiesław Babik & Wydawnictwo Uniwersytetu Jagiellońskiego
Wydanie I, Kraków 2010
All rights reserved

Książka, ani żaden jej fragment nie może być przedrukowywana bez pisemnej zgody Wydawcy. W sprawie zezwoleń na przedruk należy zwracać się do Wydawnictwa Uniwersytetu Jagiellońskiego

ISBN 978-83-233-2866-7



www.wuj.pl

Wydawnictwo Uniwersytetu Jagiellońskiego
Redakcja: ul. Michałowskiego 9/2, 31-126 Kraków
tel. 12-631-18-81, 12-631-18-82, fax 12-631-18-83
Dystrybucja:
tel. 12-631-01-97, tel./fax 12-631-01-98
tel. kom. 0506-006-674, e-mail: sprzedaz@wuj.pl
Konto: PEKAO SA, nr 80 1240 4722 1111 0000 4856 3325

Spis treści

Metodologiczne aspekty rozważań nad słowami kluczowymi	9
Część pierwsza	
Słowa kluczowe w różnych kontekstach	23
1.1. Pole znaczeniowe, czyli pojemność terminu „słowo kluczowe”	24
1.1.1. Odmiany słów kluczowych: kontekst leksykograficzny	24
1.1.2. Słowa kluczowe w nauce o informacji	27
1.1.3. Funkcje słów kluczowych	34
1.2. Słowa kluczowe w systemie WWW	36
1.2.1. Słowa kluczowe w wyszukiwarkach internetowych	36
1.2.2. Słowa kluczowe w systemach hipertekstowych i hipermedialnych	42
1.3. Psycholingwistyczne determinanty słów kluczowych	45
1.3.1. Modele pamięci semantycznej	46
1.3.2. Leksykon umysłowy (słownik mentalny) jako wzorzec/model organizacji systemu leksykalnego i słownikowego języków słów kluczowych	53
1.4. Językowe determinanty słów kluczowych	56
1.4.1. Termin jako jednostka leksykalna języka naturalnego i języka słów kluczowych	57
1.4.1.1. Status terminu w systemach leksykalnych języków słów kluczowych	58
1.4.1.2. Znaczenie leksykalne terminów jako jednostek leksykalnych języka słów kluczowych	60
1.4.2. Pole semantyczne języka słów kluczowych a języka naturalnego	62
1.4.2.1. Charakterystyka pola semantycznego języka naturalnego	62
1.4.2.2. Charakterystyka pola semantycznego języków słów kluczowych	63
1.4.3. Modelowanie języka słów kluczowych na bazie języka naturalnego	66
1.4.3.1. Relacje kojarzeniowe w języku naturalnym i języku słów kluczowych	66
1.4.3.2. Kompetencja językowa użytkowników języka słów kluczowych i jej determinanty	70
1.4.3.3. Struktury języka słów kluczowych jako efekt modelowania struktur języka naturalnego	73

Część druga

Słowo kluczowe jako jednostka leksykalna języka słów kluczowych	79
2.1. Geneza języka słów kluczowych	80
2.1.1. Prehistoria języka słów kluczowych.....	80
2.1.2. Historia języka słów kluczowych	84
2.2. Język słów kluczowych – problemy definicyjne	85
2.2.1. Język słów kluczowych jako element systemu informacyjno- -wyszukiwawczego.....	86
2.2.2. Jeden czy wiele języków słów kluczowych, czyli o naturze języka słów kluczowych.....	86
2.3. Elementy strukturalno-funkcjonalne określające tożsamość języka słów kluczowych	88
2.3.1. Funkcje języka słów kluczowych	89
2.3.2. Struktura języka słów kluczowych	90
2.3.2.1. Wymiar wertykalny – poziomy języka słów kluczowych	90
2.3.2.2. Wymiar horyzontalny – elementy języka słów kluczowych.....	93
2.3.2.2.1. Słownictwo języka słów kluczowych	93
2.3.2.2.1.1. System leksykalny języka słów kluczowych	93
2.3.2.2.1.2. System słownikowy języka słów kluczowych.....	95
2.3.2.2.2. System semantyczny języka słów kluczowych.....	100
2.3.2.2.3. System gramatyczny języka słów kluczowych.....	105
2.4. Inżynieria języka naturalnego na potrzeby języka słów kluczowych.....	106
2.4.1. Język słów kluczowych a język naturalny	106
2.4.2. Koncepcja wielofunkcyjnych słowników otwartych.....	109
2.5. Zróżnicowanie języka słów kluczowych	111
2.5.1. Kryteria podziału i odmiany języka słów kluczowych	111
2.5.2. Zalety i wady języka słów kluczowych	112
2.6. Klasyczny język słów kluczowych a inne języki paranaturalne.....	114
2.6.1. Język słów kluczowych a język deskryptorowy	114
2.6.2. Język słów kluczowych a język haseł przedmiotowych	117

Część trzecia

Słowa kluczowe w systemach wyszukiwania informacji	121
3.1. Analiza porównawcza języków słów kluczowych opartych na słowniku jawnym	121
3.1.1. Charakterystyka materiału badawczego	122
3.1.2. Cele badawcze	123
3.1.3. Wyniki badań własnych	124
3.1.3.1. Słowa kluczowe systemów tradycyjnych	124
3.1.3.1.1. Klasyczne słowniki słów kluczowych	124
3.1.3.1.2. Słowniki słów kluczowych z systemami odesłań	131
3.1.3.1.3. Słowniki słów kluczowych z elementami fasetyzacji.....	139

3.1.3.1.4. Słowniki słów kluczowych w układzie gniazdowym	143
3.1.3.1.5. Wielojęzyczne słowniki słów kluczowych	150
3.1.3.1.6. Słowa kluczowe w indeksach rzeczowych, systemach pełnotekstowych, katalogach bibliotecznych (OPAC) i bibliotekach cyfrowych	152
3.1.3.2. Klasyczny język słów kluczowych a słowa kluczowe w systemie WWW.	157
3.1.4. Wnioski z przeprowadzonych analiz	159
3.1.5. Propozycja zestawu kryteriów opisu języków słów kluczowych	164
3.2. Wybrane problemy pragmatyki słów kluczowych	167
3.2.1. Pragmatyka języka słów kluczowych: mentalna i formalna	167
3.2.2. Pragmatyka indeksowania za pomocą słów kluczowych	172
3.2.3. Pragmatyka wyszukiwania za pomocą słów kluczowych	186
3.3. Wyszukiwanie w języku naturalnym. Czy to już koniec języków słów kluczowych?.....	194
3.3.1. Technologia semantyczna – wyszukiwanie za pomocą języka naturalnego..	195
3.3.2. W jakim kierunku zmierza język słów kluczowych?.....	199
Od rozważań teoretycznych do praktyki.....	205
Literatura	217
Indeks osobowy	227
Indeks rzeczowy	231
Aneks. Wykaz słowników słów kluczowych poddanych analizie	235
Spis rysunków	239
Summary	241

Metodologiczne aspekty rozważań nad słowami kluczowymi

Przedmiotem rozprawy są słowa kluczowe używane w indeksowaniu i wyszukiwaniu informacji. Jak wszystkie języki informacyjno-wyszukiawcze¹ stanowią one lingwistyczny instrument służący do organizacji wiedzy o świecie, utrwalonej w postaci dokumentów i udostępnianej za pomocą dokumentacyjnych systemów informacyjno-wyszukiawczych, zarówno w tradycyjnych dokumentacyjnych systemach wyszukiwania informacji² online, jak i w zasobach Internetu. Zostały przedstawione na szerokim tle ich wykorzystywania w różnych dziedzinach, nazwanych tu kontekstami. Rozważania dotyczą przede wszystkim ich statusu jako jednostek leksykalnych języka informacyjno-wyszukiawczego.

Języki informacyjno-wyszukiawcze, nazywane też językami informacyjnymi, w systemach informacji o piśmiennictwie zawsze znajdowały się w centrum zainteresowań badawczych nauki o dokumentacji, a następnie informacji naukowej

¹ Języki informacyjno-wyszukiawcze i ich rola w wyszukiwaniu informacji stanowią ważny problem badawczy, dlatego ciągle poświęca się im sporo uwagi w literaturze przedmiotu. Stanowią bowiem, oprócz użytkownika, najważniejszy element systemu informacyjno-wyszukiawczego. Podejmowane rozważania dotyczą ich struktury, cech oraz funkcji jako języków sztucznych. Termin „język informacyjno-wyszukiawczy” ma stosunkowo krótki rodowód: nazwa ta jest używana zaledwie od kilkudziesięciu lat. Jednak historia tych języków jest niemal tak długa, jak historia bibliotek, gdyż sięga czasów starożytnych. Najstarszym rodzajem języka informacyjno-wyszukiawczego jest klasyfikacja piśmiennictwa, używana już w bibliotece króla Asurbanipala w Niniwie (VII w. p.n.e.), gdzie stosowano ją do porządkowania zbioru glinianych tabliczek zapisanych pismem klinowym. Klasyfikacja piśmiennictwa była wykorzystywana również w katalogu największej biblioteki starożytności – Biblioteki Aleksandryjskiej (IV–III w. p.n.e.). Obecnie ten typ języka reprezentują liczne klasyfikacje biblioteczno-bibliograficzne, zwane też klasyfikacjami dokumentacyjnymi. Pojawienie się drugiego typu – języka hasel przedmiotowych – datuje się na okres średniowiecza, kiedy w katalogach i inwentarzach bibliotecznych zaczęto umieszczać opisy książek pod hasłami, którymi były najpierw pierwsze wyrazy tytułu, a następnie tzw. wyrazy główne tytułu. Taka forma katalogu oraz inwentarza dała początek katalogom przedmiotowym. W średniowieczu tworzono już indeksy przedmiotowe do różnego rodzaju książek. Szczególną klasę języków stanowią powstałe na początku lat pięćdziesiątych XX wieku języki słów kluczowych, które dzięki związkowi z językami etnicznymi zaliczane są w stosowanej do tej pory terminologii języków informacyjno-wyszukiawczych i przyjętej w Polsce typologii do klasy języków paranaturalnych.

² Dokumentacyjne systemy informacyjno-wyszukiawcze to systemy udostępniające użytkownikowi na wyjściu metainformacje, zapisane w metajęzyku (ang. *meta-language*) w odróżnieniu od języków przedmiotowych (ang. *object-language*), służących do odwzorowywania rzeczywistości pozajęzykowej.

i nauki o informacji. Narzędzia te wywodzą się z metodyki działalności bibliotecznej i bibliograficznej, ale dopiero w ramach dokumentacji i nauki o informacji nabrały szczególnego znaczenia, a to z uwagi na ich decydujący wpływ na skuteczność zapewniania użytkownikom dostępu do zawartości informacyjnej dokumentów. Poza zakresem rozważań pozostało wykorzystywanie słów kluczowych w takich dziedzinach, jak m.in. informatyka albo stylistyka. Zagadnienia te tylko zasygnalizowałem w sposób ogólny w celu dopełnienia całościowego obrazu oraz wielowymiarowości spojrzenia na słowa kluczowe.

Słowa kluczowe i język słów kluczowych były i są rozumiane bardzo różnie, dlatego w teorii wyszukiwania informacji ciągle pojawiają się wątpliwości, czy słowa kluczowe, nazywane też językami słów kluczowych³, są jednostkami języka sztucznego czy języka naturalnego, a także czy słowa kluczowe używane w Internecie to jednostki leksykalne języka słów kluczowych. Moim zamiarem, a zarazem główną tezą i zasadniczym celem niniejszego opracowania, jest próba zdefiniowania tego języka poprzez ustalenie jego cech konstytutywnych, określających jego tożsamość i rolę w obecnym heterogenicznym sieciowym środowisku informacyjnym.

Funkcje słownictwa języka naturalnego w systemach wyszukiwania informacji były i są wciąż badane (Buckland 1999). Ich istotę stanowi przekształcanie wyrażzeń języka naturalnego w formy przystosowane dla indeksów i słowników poszczególnych typów języków informacyjnych, zwłaszcza paranaturalnych. Odpowiedni dobór słownictwa w indeksowaniu i wyszukiwaniu informacji o tekstach w języku naturalnym i językach sztucznych (np. chemia) oraz w różnych typach dokumentów (fotografie oraz inne obiekty ikoniczne i/lub multimedialne, także dźwiękowe, gdzie indeksuje się całą warstwę ikoniczną) decyduje o efektywności procesów wyszukiwawczych. Osobną trudność stanowi częsta wieloznaczność słów: to samo słowo może mieć różne znaczenia w różnych kontekstach, a to samo znaczenie może być wyrażone za pomocą wielu różnych słów.

Zajęcie się problematyką słów kluczowych jest ważne z kilku powodów. Wiele problemów słów kluczowych jest rozpoznanych w niewielkim stopniu: brakuje całościowego opracowania oraz dogłębnej analizy i syntezy wiedzy, a więc i głębszej refleksji teoretycznej o tych językach oraz określenia kierunków rozwoju. Zmiany w środowisku informacyjnym i w technologii komputerowej wymuszają podjęcie naukowej refleksji nad ich dalszymi losami, zwłaszcza w kontekście coraz szerszego stosowania języków naturalnych w procesach wyszukiwania informacji. Na tej kanwie zrodziło się pytanie o tożsamość⁴ języków słów kluczowych i o sposoby jej określenia. Do tej pory nie udało się wypracować jednolitego stanowiska. Skutkiem jest bałagan terminologiczny w tej dziedzinie, zarówno w Polsce, jak i za granicą. W związku z tym podjęcie próby uporządkowania jest pożądane. Ponadto istnieje

³ Tam gdzie było to możliwe, starałem się konsekwentnie stosować terminologię wypracowaną w ramach warszawskiej szkoły teorii języków informacyjno-wyszukiwawczych. Jej źródłem jest *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych* (Bojar, oprac. 2002). Na temat tej szkoły zob. Kurek-Kokocińska 1994.

⁴ W tekście tożsamość jest rozumiana jako zespół cech, które odróżniają język słów kluczowych od innych.

pilna potrzeba rewizji poglądów na istotę, mechanizmy i narzędzia opracowania rzeczowego oraz sposoby zapisu charakterystyk rzeczowych. Zmiany w tym zakresie są potrzebne i nieuniknione. Jestem przekonany, że należy podjąć taką próbę. Zawarte tu opinie stanowią kontynuację i podsumowanie moich dotychczasowych zainteresowań badawczych, umiejscowionych w nurcie teorii języków informacyjno-wyszukiwawczych, a także refleksji teoretycznej dotyczącej lingwistycznych problemów nauk o informacji, o książce i o bibliotece.

W latach pięćdziesiątych XX wieku znacząco rozwinęła się teoria informacji. W jej ramach uwzględniono też rolę mózgu człowieka w przetwarzaniu danych. Zaczęto mianowicie dostrzegać, że w umyśle następuje scalanie przyjmowanych informacji. Mogą temu służyć słowa kluczowe, ponieważ pozwalają na odrzucanie olbrzymiej części informacji zbędnej i na pozostawienie tylko tej, która ma dużą wartość poznawczą.

Współczesne badania nad słowami kluczowymi nabrały charakteru interdyscyplinarnego. W szerokim zakresie zaczyna się wykorzystywać wyniki nowej strategii naukowej, zwanej nauką o poznawaniu lub kognitywistyką (ang. *cognitive science*), łączącej m.in. teorie z zakresu psychologii, filozofii, lingwistyki, informatyki, neurobiologii oraz antropologii. Stało się bowiem jasne, że wyizolowane koncepcje języków informacyjno-wyszukiwawczych nie wystarczają do znalezienia zadowalających wyjaśnień złożonego fenomenu, jaki stanowią procesy indeksowania i wyszukiwania informacji. Uwzględnienie m.in. wyników badań psycholingwistycznych może pozwolić na dokładniejsze i pełniejsze wyjaśnienie wielu problemów natury infologicznej⁵. Jest to także obszar komunikacji językowej, w którym można upatrywać tożsamość eksplorowanej problematyki. Ta wielorakość perspektyw znajduje odzwierciedlenie w prezentowanej tu rozprawie.

W sensie badawczym języki słów kluczowych stanowią zjawisko stosunkowo nowe. Problem pojawił się w latach pięćdziesiątych XX wieku, kiedy to powstawały pierwsze systemy informacyjno-wyszukiwawcze stosujące słowa kluczowe do opisu zbiorów. W konsekwencji pojawiła się w piśmiennictwie z tego zakresu refleksja terminologiczna, która sprzyja wyodrębnianiu się nowego pola badawczego. W bogatym światowym i polskim piśmiennictwie naukowym z tego zakresu refleksja teoretyczna na temat słów kluczowych jest jednak skromna i odnosi się raczej do spostrzeżeń natury ogólnej. Dominują rozważania teoretyczne lub badania fragmentaryczne, często o profilu instruktazowym, nie zawsze poprawne pod względem metodologicznym. Wbrew pozorom język ten nie został jeszcze poddany dogłębnej analizie z lingwistycznego punktu widzenia. Jest wprawdzie wiele opracowań, ale prawie wyłącznie o charakterze przyczynkarskim. Dominują teksty pisane na dożądanie, najczęściej konferencyjne. Oczywiście mają one pewną wartość, dzięki aktualności i związkom z praktyką, ale brak w nich warstwy porównawczej

⁵ Infologiczna interpretacja informacji powstała w drugiej połowie XX wieku i jest wciąż rozwijana w krajach skandynawskich. Za jej twórcę uważa się szwedzkiego profesora Bo Sundgrena oraz Börje Langeforsa. Stworzona przez nich teoria infologii zmierza do wyjaśnienia znaczenia informacji w aspekcie użytkowym poprzez badania jej własności, zwłaszcza z punktu widzenia wymagań stawianych przez jej użytkowników.

i odniesienia do oryginalnych wyników badań naukowych. W tej dziedzinie odczuwa się wyraźny brak analizy kompleksowej, dotkliwy niedobór prac „fundamentalnych” oraz ustalonej terminologii.

Językami słów kluczowych zajmuję się od 1986 roku, uczestnicząc w licznych pracach nad ich tworzeniem oraz w pracach badawczych. Niniejsza rozprawa zawiera wyniki badań nad słowami kluczowymi, które prowadziłem metodą analizy i krytyki istniejącego piśmiennictwa, opublikowanego przede wszystkim w języku polskim i angielskim. U podstaw tych badań tkwił zamiar podsumowania istniejącej wiedzy dotyczącej języków słów kluczowych, a tam gdzie było to możliwe – także jej uzupełnienia. W celu skompletowania literatury stanowiącej podstawę źródłową oraz dla ustalenia stanu badań nad podjętymi zagadnieniami przeprowadziłem kwerendy bibliograficzne w istniejących bazach danych, w bibliografii narodowej bieżącej oraz Polskiej Bibliografii Bibliologicznej. Analiza uzyskanych tą drogą publikacji potwierdziła, że brakuje całościowego opracowania tej klasy języków.

Przy wykonaniu tego zadania pojawiło się sporo różnych trudności. Oto bowiem już sam termin „słowo kluczowe” posiada wiele znaczeń i dążąc do jasności, musiałem ograniczyć rozważania na temat innych ujęć tego terminu. Nie ma też ostro zarysowanych granic języka słów kluczowych, którego jednostkami leksykalnymi są słowa kluczowe, dlatego zredukowałem wywód jedynie do przedstawienia tych zagadnień, które – moim zdaniem – pozwalają określić jego tożsamość.

Książka stanowi próbę syntetycznego, a zarazem wielostronnego zmierzenia się z problematyką tożsamości tego typu języka informacyjno-wyszukiwawczego, eksplorowaną w ramach lingwistyki strukturalnej i funkcjonalnej, nasyconą nową faktografią oraz nowymi problemami badawczymi, z jednoczesnymi odniesieniami do problematyki informacyjnej. O językach słów kluczowych pisano bowiem głównie w kontekście samych systemów informacyjno-wyszukiwawczych, w których się je stosuje, a nie ich struktury i budowy. Często też traktowano ten język jako język wyjściowy do budowy języków deskryptorowych. Dotychczasowe wyniki badań nad słowami kluczowymi w systemach wyszukiwania informacji zostały opisane w „Journal of Documentation” (Voorbij 1998). Zauważono, że za rozpowszechnieniem ich używania nie poszły badania tych języków, a przynajmniej nie w takim stopniu, w jakim następowało ich upowszechnianie i nie z taką dynamiką, z jaką następuje wykorzystanie. Na temat słów kluczowych, ale głównie w kontekście budowy tezaurusów, bardzo dużo pisano w latach sześćdziesiątych i siedemdziesiątych XX wieku, przede wszystkim w ZSRR.

Na świecie problematyką języków słów kluczowych najczęściej zajmowano się w USA, Wielkiej Brytanii, Niemczech i ZSRR. Badania obserwuje się od początku lat pięćdziesiątych XX wieku. Wśród ważniejszych są analizy Mortimera Taubego (Taube et al. 1953), Hansa Petera Luhna (Luhn 1960), Petera Hanicke (Hanicke 1966) i Josefa Koblitz (Koblitz 1968). Należy wspomnieć również o przeprowadzonych w latach sześćdziesiątych XX wieku w Wielkiej Brytanii tzw. dwóch eksperymentach (badaniach) cranfieldzkich (The ASLIB Cranfield Project), kiedy to badano jednowyrazowe słowa kluczowe w postaci unitermów (Cleverdon 1967). Wprawdzie nie były one „rewolucyjne” dla słów kluczowych, ale okazały się bardzo znaczące dla dojrze-

wania pragmatyki tych języków oraz dla badań porównawczych (Lancaster & Mills 1964). Ujawniły, że najważniejszy jest proces indeksowania, a nie cechy danego języka (Richmond 1963). Obecnie prowadzone badania za granicą dotyczą efektywności słów kluczowych (Voorbij 1998; Taghva, Borsack, Nartker, Condit 2004; Gross, Taylor 2005), ich wykorzystywania w systemach informacyjno-wyszukiwawczych (Hartley, Kostoff 2003), słów kluczowych używanych w metatagach w systemach internetowych (Craven 2004, 2005), metod automatycznej ekstrakcji słów kluczowych (Borger, Kuflik, Shoal, Shapira 2001; Jones, Paynter 2002), a także porównywania ich źródeł (Gil-Leiva, Alonso-Arroyo 2007). Jednak zasygnalizowane tu badania do tej pory nie znalazły odzwierciedlenia w uogólnieniach teoretycznych.

Zagadnienia te pojawiły się także w Polsce, po raz pierwszy w publikacjach Jądwigi Choroszuszy (Choroszuszyńska 1968, 1970), a następnie artykułach Anny J. Bańkowskiej i Wandy Sokołowskiej (Bańkowska, Sokołowska 1969), Teresy Cieszewskiej, Eugeniusza Ścibora i Felicji Żeromskiej (Cieszevska, Ścibor, Żeromska 1969), w rozprawach Józefa Robowskiego (Robowski 1974) i Olgierda Unguriana (Ungurian 1976) oraz artykułach Witolda Marciszewskiego (Marciszewski 1976, 1978) i Wiesława Szczepanka (Szczepanek 1981). Językami tymi zajmowano się także przy okazji badań prowadzonych nad różnymi typami języków informacyjno-wyszukiwawczych (Ścibor, Tomasik-Beck 1970; Robowski 1974; Ungurian 1976) oraz ich projektowania dla różnych dziedzin (Babik 1989, 1996a, 1996b, 1999a, 1999b, 2007, 2010; Kurek-Kokocińska 1998a, 1998b, 2004; Rudnik-Karwatowa 2002; Rudnik-Karwatowa, Mikos 2007). Niewątpliwie szersze zainteresowanie naukowe, głównie projektowe, tą klasą języków wzbudził Ośrodek Informacji Naukowej PAN (OIN PAN), a przede wszystkim prace jego krakowskiej Pracowni, uwieńczone publikowaniem w latach 1988–1993 kilkunastu słowników słów kluczowych⁶.

W Polsce problematyka języków słów kluczowych była przedstawiana z reguły w ogólnych opracowaniach dotyczących języków informacyjno-wyszukiwawczych, m.in. w tekstach Lucyny Anny Bielickiej i Eugeniusza Ścibora (Bielicka, Ścibor 1982; Ścibor 1999). Punktem odniesienia były badania prowadzone w USA i w ZSRR (lata siedemdziesiąte i osiemdziesiąte XX wieku)⁷. Najczęściej jednak zagadnienia te tylko wzmiankowano, na przykład przy okazji omawiania języków deskryptorowych lub jako materiał ilustracyjny dla zagadnień teoretycznych z zakresu języków informacyjnych. Pewne ożywienie w badaniach nad tą klasą języków, a przede wszystkim w ich tworzeniu, można zauważyć w latach osiemdziesiątych XX wieku, od kiedy to istnieją już w znacznej mierze skryształizowana teoria języ-

⁶ Szersze omówienie tak bardzo ważnych dla języków słów kluczowych prac zawiera część trzecia rozprawy.

⁷ Warto zauważyć, że w latach siedemdziesiątych XX wieku została opracowana teoria wyszukiwania informacji, w ramach której wyodrębniono teorię języków informacyjno-wyszukiwawczych (Czerny 1978). Trzeba tu również wspomnieć pracę o informacji dynamicznej i opracowaniu bibliotecznym amerykańskiego uczonego Gerarda Saltona (1927–1995), pioniera w zakresie zastosowań metod statystycznych w wyszukiwaniu informacji i automatycznym indeksowaniu, autora wielu prac dotyczących teorii i praktyki indeksowania dokumentów oraz procesów wyszukiwania informacji, który dostrzega dynamikę systemu bibliotecznego poprzez pryzmat dynamiki stosowanego w nim języka, sugerując wprowadzanie do słownika systemu słów używanych przez użytkowników, to jest słów kluczowych (Salton 1975).

ków informacyjno-wyszukiwawczych i teoria wyszukiwania informacji, kształtującą nową epokę dla słów kluczowych.

Empiryczne badania nad słowami kluczowymi prowadziła w Polsce na przełomie lat osiemdziesiątych i dziewięćdziesiątych XX wieku Ewa Głowacka z Instytutu Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu im. Mikołaja Kopernika w Toruniu. Badania te miały na celu porównanie efektywności wyszukiwania haseł przedmiotowych i słów kluczowych w dwóch eksperymentalnych bazach danych, utworzonych na potrzeby pracy w programie Mikro CDS/ISIS. Pierwsza z nich zawierała informacje o wydawnictwach zwartych z dziedziny bibliotekoznawstwa i informacji naukowej. Opisy były przejmowane w całości z „Przewodnika Bibliograficznego” i zaindeksowane w języku haseł przedmiotowych Biblioteki Narodowej. Uzupełniano je opisami za pomocą słów kluczowych. Druga baza danych, utworzona przez studentów, to bibliografia artykułów z dziedziny bibliotekoznawstwa i informacji naukowej zarejestrowanych w „Bibliografii Zawartości Czasopism” (BZCz) z lat 1986–1990, indeksowanych w obu wymienionych językach przez samych studentów. Badania były prowadzone metodą oceny efektywności wyszukiwania informacji zaproponowaną przez Fredericka Wilfrida Lancastera. Zdaniem autorki, wyniki przeprowadzonych badań wykazały, że język haseł przedmiotowych Biblioteki Narodowej lepiej nadaje się do zwiększania kompletności wyszukiwania, zaś język słów kluczowych jest nastawiony na dokładność wyszukiwania (Głowacka 1994). Badania nad polepszeniem efektywności wyszukiwania informacji w systemie stosującym swobodne słowa kluczowe prowadziła także Dorota Głazek z Biblioteki Głównej i Ośrodka Informacji Naukowo-Technicznej Politechniki Wrocławskiej (Głazek 1987, 1989).

Słowami kluczowymi, jako jednym z elementów systemów informacyjno-wyszukiwawczych znajdujących zastosowanie w Internecie, zajmują się obecnie także specjalistyczne firmy, jak stworzona przy współudziale naukowców z University of Colorado Alexa Resarch (<http://alexa.com>) czy Greg Notess (<http://www.searchengineshowdown.com/>), oraz firmy tworzące internetowe narzędzia wyszukiwawcze, które są mocno zaangażowane w badania nad słownictwem stosowanym w Internecie i rozwojem języków wspomagających wyszukiwanie dokumentów, przede wszystkim w celu optymalizacji wyników wyszukiwania. Stąd badania słów kluczowych są aktualnie zdominowane przez sferę marketingu internetowego.

Powstanie Internetu i możliwości wykorzystania w nim słów kluczowych wzmoгло zainteresowanie i dynamikę badań nad tą klasą języków, zwłaszcza że słowa kluczowe stały się jednym z elementów metadanych. Pojawiły się nowe problemy natury zarówno teoretycznej, jak i praktycznej. Dotyczą one przede wszystkim efektywności stosowania słów kluczowych w etykietowaniu i wyszukiwaniu informacji w wielkim uniwersalnym i słabo kontrolowanym systemie rozproszonym, a także teoretycznego problemu tożsamości obecnie używanych słów kluczowych z dawnymi klasycznymi i nieklasycznymi językami słów kluczowych i ich relacji do języka naturalnego w funkcji metainformacyjnej. Analiza piśmiennictwa z tego zakresu pozwala zauważyć zmianę akcentu w studiach poświęconych temu językowi. W teorii języków informacyjno-wyszukiwawczych zorientowanej na system informacyjno-wyszukiwawczy

położono jednocześnie nacisk na użytkownika tego języka, a więc na jego interakcję i powiązane z nią kognitywne, afektywne i sytuacyjne (społeczne) konteksty indeksowania i wyszukiwania informacji. W polskim i zagranicznym piśmiennictwie z zakresu teorii języków informacyjno-wyszukiwawczych brak szczegółowej analizy krytycznej tej klasy języków w postaci opracowania monograficznego, nie ma też studiów w zaproponowanym tu ujęciu, chociaż wiele pojedynczych aspektów i problemów jest już reprezentowanych w literaturze przedmiotu.

Teoria języków informacyjno-wyszukiwawczych pozostaje obecnie pod wpływem paradygmatu systemowego, zrodzonego w XX wieku, który jest jednym z największych i najpłodniejszych z dotychczasowych sposobów konceptualizacji rzeczywistości i problematyki naukowej. Określa ludzką percepcję i rozumienie otaczającej rzeczywistości oraz wpływa na sposób jej przekształcania. To dominujące podejście systemowe bazuje na hierarchiczności⁸ i ma swoje źródło w przedkomputerowych możliwościach tradycyjnych technologii przetwarzania informacji. Pod koniec XX wieku pojawiła się jednak interesująca odmiana tej koncepcji, która wyznacza nowy sposób myślenia i działania, a mianowicie sieć. Nie jest ona wynalazkiem dwudziestowiecznym⁹, jej genezy można doszukiwać się aż w starożytności. Obecnie metafora sieci wywiera coraz większy wpływ na współczesne postrzeganie, konceptualizację i sposoby objaśniania świata, wyznaczając nowy paradygmat myślenia i zgłębiania problemów (por. Börner, Sanyal, Vespignani 2007). Po części wynika też z rozwoju Internetu i telefonii komórkowej. Tak jak do niedawna na świat patrzono przez systemowe okulary, tak teraz coraz częściej w użyciu są okulary sieciowe. Szereg współczesnych zjawisk łatwiej przedstawiać i analizować jako procesy zachodzące w sieciach złożonych z węzłów i łączących je zależności niż traktować je jako systemy. Sieć pozwala lepiej rozumieć dynamikę tych zjawisk i ich złożoność, łatwiej też dzięki niej proponować ulepszenia modyfikacje (Muraszkiewicz 2005).

Właśnie dzięki przyjęciu paradygmatu sieciowego, polegającego na traktowaniu słów kluczowych jako elementów sieci, stało się możliwe kompleksowe ujęcie problematyki tych języków. Sama analiza języka informacyjno-wyszukiwawczego jako sieci jednostek leksykalnych połączonych różnorodnymi relacjami nie jest czymś nowym w teorii języków informacyjno-wyszukiwawczych, gdyż na takim podejściu opiera się metoda organizacji tezaury, ale jej zastosowanie do słów kluczowych można uznać za rozwiązanie oryginalne.

⁸ W strukturach zorganizowanych hierarchicznie wielką rolę odgrywają często same struktury, gdyż informacja jest przekazywana od najniższego w hierarchii punktu do najwyższego szczebla i odwrotnie.

⁹ Sam pomysł nie jest zupełnie nowy, gdyż już w 1945 roku Vannevar Bush, w artykule pt. *As we may think* (Bush 1945), przedstawił koncepcję maszyny o nazwie MEMEX, która potrafiłaby kojarzyć idee i symulować procesy mentalne, a której zadaniem było usprawnienie pracy badaczy poprzez wykorzystanie relacji asocjacyjnych w organizowaniu zbioru informacyjnego. Koncepcja Busha została oparta na strukturze ludzkiej pamięci, która (według osiągnięć wiedzy) operuje skojarzeniami, tworząc strukturę będącą w istocie siecią powiązanych informacji zawartych w komórkach mózgu. Pomysł Busha opierał się na organizacji dokumentów bazującej nie na abstrakcyjnych relacjach systemowych, lecz na związkach o charakterze asocjacyjnym tworzonych w procesie percepcji informacji. Po MEMEX-ie pozostała ważna koncepcja organizacji informacji w systemie wyszukiwania informacji, oparta na wykorzystaniu relacji asocjacyjnych w określaniu połączeń między wydzielonymi „porcjami” informacji.

Trawestując opinię Anny Sitarskiej (Sitarska 2006) na temat systemu bibliotecznego, można zauważyć, że jednym z podstawowych kierunków zmienności języków informacyjnych jest kierunek narastania i wzbogacania więzi łączących jednostki leksykalne (i elementy treści dokumentów) w coraz liczniejsze i coraz bardziej skomplikowane układy. Więzi te pozwalają przełamywać statyczne bariery hierarchiczne między jednostkami leksykalnymi i ułatwiają dynamiczne krążenie zawartej w nich informacji o dokumentach, podobnie jak w systemach hipertekstowych. Powodem przyjęcia nowego podejścia jest m.in. fakt, że języki te od samego początku nie mieszczą się w obecnym, hierarchicznym paradygmacie organizacji systemu leksykalnego, a do ich opisu użyteczny staje się nowy paradygmat sieciowy, który stanowi rozszerzenie paradygmatu systemowego. Wspomniane więzi wywodzą się nie tylko z samego systemu leksykalnego języka słów kluczowych, który ma charakter otwarty, ale też z autonomicznych układów rozmaitego przeznaczenia dołączanych do systemu leksykalnego tego języka. Z tego względu interesująca może być naukowa refleksja i zbadanie charakteru „ładów leksykalnych” w tym języku oraz jego rodzajów¹⁰.

Język słów kluczowych, podobnie jak inne języki informacyjno-wyszukiwawcze, stanowi źródło wielu niejasności teoretycznych i kłopotów praktycznych. Istnieje więc potrzeba badań holistycznych, w których język byłby postrzegany w kontekście określonego, teoretycznego i/lub praktycznego modelu funkcjonowania. Jednym z takich obszarów badawczych jest sfera tworzenia podstaw pragmatyki, rozumianej jako ta część teorii języków informacyjno-wyszukiwawczych, która dotyczy języka w działaniu, czyli (zgodnie z definicją pragmatyki Charlesa W. Morrisa, do której odwołuje się większość współczesnych teorii) relacji między językami a ich użytkownikami, a więc m.in. potrzebami użytkowników i wpływem użytkowników na strukturę semiotyczną tych języków.

Omawiając słowa kluczowe, przyjąłem podejście kognitywne, stanowiące w nauce o informacji od ponad trzydziestu lat jeden z głównych jej nurtów, w ramach którego sytuuje się teoria języków informacyjno-wyszukiwawczych. Przyjęcie tej perspektywy badawczej jako kontekstu rozważań, zwłaszcza nad tą klasą języków, jest uzasadnione. Rozwój wiedzy o poznawczych uwikłaniach procesów informacyjnych (m.in. w umyśle człowieka) nie może być niezauważony. Podstawowe założenia tej orientacji poznawczej i główne kierunki jej wykorzystania w nauce o informacji były wielokrotnie omawiane w piśmiennictwie fachowym (m.in. Belkin 1984, 1990; Ingwersen 1996; Ingwersen, Järvelin 2005; Spink, Cole 2005). W polskim piśmiennictwie charakterystykę podejścia kognitywnego w informacji przedstawiły m.in. Jadwiga Woźniak-Kasperek (Woźniak 1997, 2000) oraz Barbara Sosińska-Kalata (Sosińska-Kalata 1999).

¹⁰ W sensie intelektualnym pierwszy myślał w ten sposób już w latach sześćdziesiątych XX wieku G. Salton, formułując ideę „biblioteki dynamicznej”, choć robił to z innych pozycji, myśląc nie tyle o *ładzie leksykalnym*, ile o *ładzie systemowym*. Uważał on, że „wnoszenie” do systemu na bieżąco słów używanych przez użytkownika w czasie poszukiwań oraz – co bardzo ważne – robienie w systemie informacyjno-wyszukiwawczym z tymi nowymi elementami ŁADU w języku informacyjnym czyni system dynamicznym odpowiednio do potrzeb.

W rozważaniach nad tą klasą języków za niezbędną, z kognitywnego punktu widzenia, uznałem potrzebę odwołania się do analiz psycholingwistycznych, związanych z badaniami leksykologicznymi, w tym z teoriami i modelami przetwarzania informacji językowych przez człowieka, a szczególnie z koncepcją funkcjonowania leksykonu umysłowego (słownika mentalnego) wraz z drogami dostępu do informacji leksykalnej, co w szerokim ujęciu wiąże się z procesami pamięciowymi (na przykład Kurcz 1987; Gabryś-Barker 2005). Niektóre psychologiczne teorie pamięci semantycznej, na przykład sieciowa czy rozprzestrzeniania się aktywacji (Collins, Loftus 1975; Lindsay, Norman 1991) oraz funkcjonowania tzw. analizatora semantycznego, mają wiele odniesień do lingwistycznej teorii pól semantycznych wykorzystywanej w teorii tych języków. Przydatne są również wyniki ilościowych badań psycholingwistycznych nad współczesnym słownictwem, zwłaszcza dotyczące subiektywnej częstości użycia wyrazów oraz analizy specjalistycznych baz danych. W tym zakresie prowadzi się też badania skojarzeń werbalnych: swobodnych i kierowanych (na przykład Kurcz 2000; Łobacz, Mikołajczak-Matyja 2002; Gabryś-Barker 2005). Podejście psycholingwistyczne jest bardzo użyteczne w badaniach zachowań językowych użytkowników języka, a przede wszystkim lingwistycznych badaniach językowego obrazu świata oraz mechanizmów myślenia użytkowników języka o rzeczywistości pozajęzykowej. Perspektywa psycholingwistyczna, ze względu na stosowane w niej takie metody badawcze jak testy skojarzeń swobodnych i łańcuchy asocjacyjne, introspekcja symultaniczna i retrospekcja, ułatwia udzielenie odpowiedzi na pytania o sposoby myślenia użytkownika języka o rzeczywistości pozajęzykowej, sposoby porządkowania świata oraz sposoby wyboru jednostek leksykalnych na oznaczanie elementów treści i/lub formy dokumentów. Bardzo przydatne dla teorii i praktyki słów kluczowych są też rozwiązania metodologiczne związane z badaniem kompetencji językowych i komunikacyjnych oraz ich roli w teorii dyskursu.

Badanie tak specyficznego obszaru słownictwa wymagało zajęcia się nie tylko aspektami językoznawczymi przydatnymi z punktu widzenia teorii języków informacyjno-wyszukiwawczych, ale i sięgnięcia do najnowszych osiągnięć i wyników badań psycholingwistycznych dotyczących reklamy, a więc zajęcia się aspektami psycholingwistycznymi funkcjonowania języka. Dla języków słów kluczowych bardzo użyteczne są wyniki badań skojarzeń swobodnych. Dla niektórych języków już opracowano odpowiednie programy służące do obliczania siły bezpośrednich i pośrednich skojarzeń. Na początku lat dziewięćdziesiątych XX wieku skonstruowano dla języka niemieckiego sieć skojarzeniową, wykorzystując teoretyczne osiągnięcia konekcjonizmu. Do opisu procesów skojarzeniowych zachodzących w umyśle człowieka wykorzystuje się też obserwacje sieci neuronowych.

W badaniach nad językami słów kluczowych, podobnie jak w badaniach nad systemami wyszukiwawczymi, uważano, że doskonalenie skuteczności korzystania z tych języków wymaga przystosowania do nich użytkownika poprzez wykształcenie odpowiednich umiejętności posługiwania się tymi językami. Możliwość przystosowania słów kluczowych do korzystającego z nich człowieka, do jego ograniczeń percepcyjnych i struktur poznawczych pozostawała poza obszarem zainteresowania badaczy tych języków.

W rozprawie zadałem sobie też pytanie, dlaczego języki słów kluczowych (nazywane najczęściej po prostu słowami kluczowymi) są aż tak rozpowszechnione, jakie są tego przyczyny, jakie generuje to problemy i co można zrobić, aby je zoptymalizować. Przyjąłem, że prawdopodobnie ich organizacja jest najbardziej zbliżona do organizacji pamięci semantycznej w umyśle człowieka, czyli użytkownika. Inne pytania dotyczą tego, co nowego wnoszą języki słów kluczowych do metod opisu i wyszukiwania dokumentów oraz jaka jest obecnie wartość tego typu narzędzia lingwistycznego w porównaniu z innymi i dlaczego ono samo staje się „wystarczające” w Internecie.

Za perspektywę poznawczą uznałem traktowanie naturalnych struktur poznawczych (relacji zachodzących w umyśle człowieka i funkcji poznawczych) jako swobodnego modelu w budowie struktur i funkcji systemów wyszukiwania informacji, dlatego w systemie informacyjno-wyszukiwawczym i w stosowanym w nim języku powinno następować odwzorowanie organizacji wiedzy w ludzkim umyśle w taki sposób, jak to sugerują modele psychologiczne, a więc systemy i języki wyszukiwania informacji powinny być dostosowane do naturalnych struktur poznawczych ich użytkowników.

Analiza modeli zapisu wiedzy w umyśle może stanowić podstawę oceny przystawalności modelu organizacji wiedzy realizowanej za pomocą słów kluczowych w dokumentacyjnych systemach wyszukiwania informacji, a z kolei z niej powinny wynikać wnioski prowadzące do możliwej optymalizacji struktury i funkcji słów kluczowych, a także narzędzi i metod opisu oraz wyszukiwania informacji za pomocą tych języków. Wprawdzie języki te powstały na początku lat pięćdziesiątych XX wieku, to jest w okresie początków automatyzacji procesów informacyjnych, ale funkcjonują do dziś i stopniowo stają się podstawowym narzędziem wyszukiwania informacji.

Aby zrealizować główny cel swojego opracowania, posłużyłem się ogólnonaukowymi metodami analizy i krytyki piśmiennictwa, metodą porównawczą, metodami analizy leksykalnej i leksykograficznej oraz metodą analizy dokumentacji (m.in. wstępów do słowników słów kluczowych). Wykorzystałem również metodę psychologiczną, dokonując obserwacji zachowań użytkowników tego języka. Na wybranych przykładach spróbowałem też dokonać analizy procesu ewolucji języków słów kluczowych, dążąc do ustalenia elementów konstytutywnych definiujących ich tożsamość. W tym celu podjąłem próbę stworzenia na tyle pojemnej formuły słów kluczowych, aby mogły się w niej zmieścić zarówno tradycyjne słowa kluczowe, jak też słowa kluczowe używane w Internecie.

Biorąc to pod uwagę, sformułowałem następujące tezy badawcze:

- słowa kluczowe są wyrażeniami języka naturalnego, ale użyte w jednej (lub wielu) specjalnej funkcji. Oznacza to, że są elementami systemów języka naturalnego, z którego czerpią semantykę i leksykę (otoczenie semantyczne i leksykalne), ale odnoszą je do innych rzeczywistości (literackiej, dziedzinowej, dokumentacyjnej);
- słowa kluczowe są elementarnymi jednostkami leksykalnymi sztucznego języka słów kluczowych, gdy jego konstrukcja poprzedza użycie;

- słowa kluczowe w swoich funkcjach prymarnych to klucze do elementów treści dokumentu;
- natura języka słów kluczowych nie jest jednolita, z czym wiąże się relatywizm jej postrzegania;
- istnieje wzajemny wpływ poszczególnych typów języków paranaturalnych;
- granice poszczególnych języków słów kluczowych w ich konkretnych realizacjach są płynne.

Problem niejasnych granic między słowami kluczowymi a deskryptorami i hasłami przedmiotowymi sygnalizowali wprawdzie J.C. Gardin (Gardin 1966), O. Ungurian (Ungurian 1976), W.J. Hutchins (Hutchins 1978), E. Ścibor (Ścibor 1982, 1999), U. Ribbert (Ribbert 1992), J. Woźniak (Woźniak 1999), B. Sosińska-Kalata (Sosińska-Kalata 1999), J. Bertram (Bertram 2005) i B. Bojar (Bojar 2005), ale nie zajęto się jego kompleksowym rozwiązaniem, poprzestając na próbach doraźnych ocen, m.in. poprzez eksplikację kryteriów ich rozróżniania.

Sposoby realizowania funkcji metainformacyjnej i wyszukiwawczej języka naturalnego nabierają szczególnego znaczenia w zautomatyzowanych systemach wyszukiwawczych. Oferują one szeroki dostęp do zasobów w sieciach komputerowych, także w związku ze zmianami sposobów korzystania z zasobów informacyjnych przez użytkowników, którzy coraz częściej sami, bez pośredników, poszukują informacji, wykorzystując obecne możliwości, co nie jest bez wpływu na sposoby stosowania w nich tych języków. Systemy informacyjno-wyszukiwawcze powinny być przyjazne dla użytkownika, a można to osiągnąć dzięki znajomości języka użytkownika i form subiektywnej percepcji, poprzez wykorzystanie tej wiedzy właśnie w konstrukcji systemu odsyłaczy w celu podnoszenia efektywności i skuteczności realizowanych przez język naturalny funkcji języka informacyjno-wyszukiwawczego.

Jako pola sygnalizowanych tu badań przyjąłem: relację odpowiedniości (ekwiwalencji) wyszukiwawczej, relację relewancji¹¹ i relację kojarzeniową¹². Podstawą było przekonanie o ich znaczącej, lecz ciągle niedocenianej roli w określaniu i kształtowaniu tożsamości języków słów kluczowych w systemie komputerowym. W związku z tym niezbędne okazało się pokazanie odmienności modelu organizacji wiedzy zawartego *implicite* w językach słów kluczowych oraz stopnia jego zbieżności i przystawalności do kognitywnych modeli struktur poznawczych, co stanowi rodzaj wyjaśnienia szerokiego zakresu stosowania słów kluczowych w systemach wyszukiwania informacji, w tym w Internecie.

Za środki prowadzące do osiągnięcia głównego celu uznałem:

- pokazanie jak funkcjonują słowa kluczowe w różnego typu systemach, co ma prowadzić do pogłębionej charakterystyki języków słów kluczowych na przełomie XX i XXI wieku, związanej z wyodrębnieniem ich cech charakterystycznych (dystynktywnych);

¹¹ Niewątpliwie relacja ta wymaga dookreślenia, zwłaszcza dla takich rozmytych systemów, z jakimi mamy do czynienia w działalności informacyjnej, co zostanie uczynione w dalszej części pracy.

¹² W piśmiennictwie dotyczącym języków informacyjno-wyszukiwawczych niesłusznie czasem utożsamia się relację kojarzeniową z relacją pokrewieństwa. Są to jednak różne relacje. Relacja pokrewieństwa jest jednym z rodzajów relacji kojarzeniowych.

- zaprezentowanie jak się zmieniały języki słów kluczowych w zależności od potrzeb informacyjnych użytkowników, co ma ujawnić wewnętrzne zróżnicowania struktur, funkcji i odmian tej klasy języków;
- wskazanie związków (podobieństw i różnic) między językami słów kluczowych a innymi językami paranaturalnymi;
- określenie kierunków rozwoju i dalszych perspektyw tej klasy języków, co ma ujawnić refleksja teoretyczna towarzysząca wymienionym czynnikom i zmianom. W szczególności wymagało to:
 - przedstawienia podobieństw i różnic w sposobach prezentacji treści dokumentów przez poszczególne paranaturalne języki informacyjno-wyszukiwawcze, w tym za pomocą słów kluczowych;
 - określenia podobieństw i różnic między słownikiem słów kluczowych a tezauresem i słownikiem języka haseł przedmiotowych;
 - wskazania najważniejszych zasad tworzenia tekstów (charakterystyk systemowych) w języku słów kluczowych, a więc pragmatyki (użycia) słów kluczowych w indeksowaniu i wyszukiwaniu informacji.

Praktyczną realizację tych zadań znacznie utrudnia niejednorodność terminologiczna badanego i opisywanego języka.

Krytyczna analiza wymienionych tu zagadnień, składająca się na monograficzne ujęcie słów kluczowych, nie byłaby możliwa bez odwołania się do teoretycznego dorobku w zakresie teorii języków informacyjno-wyszukiwawczych, przede wszystkim Bożenny Bojar i Barbary Sosińskiej-Kalaty oraz odniesienia poczynionych przez nie konstatacji do języków słów kluczowych. Stąd częste powoływanie się na ich opracowania.

Struktura moich rozważań kształtuje się następująco. Określiwszy we wstępie problem badawczy, dokonuję niezbędnych ustaleń terminologicznych i sformułowania tez stanowiących osnowę pracy. Część pierwsza stanowi prezentację i uzasadnienie nowej, właściwej dla tej klasy języków, perspektywy badawczej oraz określa różnice w stosunku do perspektywy dotychczasowej. Rozważania rozpoczynam od omówienia podstawowej terminologii słów kluczowych (Rozdział 1.1). W rozdziale 1.2 charakteryzuję i uzasadniam za pomocą słów kluczowych wykorzystywanych w systemie WWW¹³ podejście sieciowe do słów kluczowych. Przedstawiam też wybrane teorie organizacji i mechanizmów działania systemu poznawczego człowieka oraz ich wpływu na strukturę i funkcje słów kluczowych, w tym wybrane modele organizacji pamięci, ze wskazaniem ich przydatności w badaniach i w opisie języków słów kluczowych (Rozdział 1.3). Za ważny element tego opisu uznaję leksykon umysłowy, który został potraktowany jako wzorzec organizacji systemu leksykalnego i słownikowego omawianej klasy języków. Rozdział 1.4 poświęcam analizie relacji między językiem słów kluczowych a językiem naturalnym. Istotną rolę w organizacji systemu leksykalnego języka słów kluczowych odgrywa terminologia. W rozdziale tym omawiam także pole semantyczne języka naturalnego jako strukturę wyjściową w modelowaniu systemu semantycznego języków słów kluczowych, akcentując

¹³ W tekście będę utożsamiać system WWW z Internetem.

szczególnej rolę relacji kojarzeniowych w organizacji poszczególnych systemów języków słów kluczowych, modelowanych na języku naturalnym oraz wpływ kompetencji językowej w zakresie języka naturalnego na język słów kluczowych. Celem tej części pracy jest prezentacja przyjętego kontekstu teoretycznego rozprawy, stanowiącego uzasadnienie przyjęcia sieciowego paradygmatu w omawianiu słów kluczowych.

Zasadniczym członem rozważań stanowi część druga, która zawiera charakterystykę funkcjonalno-strukturalną języka słów kluczowych jako struktury abstrakcyjnej i formalnej oraz prezentuje ją na tle innych typów paranaturalnych języków wyszukiwania informacji. Rozważania rozpoczynam od przedstawienia genezy tej klasy języków (Rozdział 2.1) i od omówienia podstawowej terminologii oraz definicji tych języków, w tym ich natury i rodzajów, a także miejsca w systemie informacyjno-wyszukiwawczym (Rozdział 2.2). Następnie charakteryzuję funkcje tych języków oraz ich strukturę (Rozdział 2.3). Charakterystyka inżynierii języka naturalnego na potrzeby języka słów kluczowych (Rozdział 2.4) stanowi podstawę charakterystyki rodzajów i cech tych języków (Rozdział 2.5). W końcu następuje porównanie języków słów kluczowych z innymi typami języków wyszukiwania informacji (Rozdział 2.6). Do problematyki tej części zostały też włączone zagadnienia tworzenia języków słów kluczowych. Celem tej części pracy jest z kolei identyfikacja i charakterystyka elementów określających tożsamość, to jest istotę klasycznych języków słów kluczowych.

Część trzecia została poświęcona przykładowym implementacjom słów kluczowych w systemach tradycyjnych i w Internecie oraz trudnościom z określeniem granic tego języka. Przedstawiam w niej też wyniki badań własnych polegających na analizie porównawczej słowników słów kluczowych systemów tradycyjnych oraz zastosowań słów kluczowych w Internecie. Stanowi to punkt wyjścia do opracowania zestawu kryteriów opisu tej klasy języków, pokazującego ich zróżnicowanie (Rozdział 3.1). Rozdział 3.2 zawiera omówienie pragmatyki ogólnej i szczegółowej języków słów kluczowych. Zwieńczenie tej części rozprawy stanowi Rozdział 3.3, w którym omawiam ściśle związki słów kluczowych z konkretnym językiem naturalnym. Celem tej części pracy jest pokazanie czynników definiujących granice tego języka, pochodzących ze sfery pragmatyki. Rozważania teoretyczne uzupełniam materiałem egzemplifikacyjnym z badań własnych.

Opracowanie ujawnia w sposób kompleksowy złożoność i dynamiczny charakter uwarunkowań obecnego stanu, a także perspektyw rozwoju języków słów kluczowych jako klasy języków informacyjno-wyszukiwawczych. Języki te zostały usytuowane w kontekście rozwoju współczesnych technologii informacyjno-komunikacyjnych, jak też rozwoju teorii i metodologii nauki o informacji, z uwzględnieniem wyników badań nad zachowaniami informacyjnymi człowieka, które stanowią podstawę do budowy nowych narzędzi lingwistycznych oraz do modyfikacji i nowocześniejszego dotychczas stosowanych narzędzi świadczenia usług informacyjnych czy kształtowania umiejętności informacyjno-komunikacyjnych.

W tekście skupiłem uwagę na językach słów kluczowych w taki sposób, żeby dostrzec ich genezę w okresie przedkomputerowym oraz rozwój na wczesnym etapie

zastosowań komputerów, przy czym szczególnie interesuje mnie faza opracowywania, wyszukiwania oraz udostępniania informacji w epoce pełnego już zastosowania technologii sieciowych.

Rozważania oparłem na tych orientacjach badawczych i koncepcjach praktycznych, które stawiają użytkownika informacji i jego potrzeby w centrum procesów informacyjnych, bo to te orientacje zdominowały obecnie myślenie o procesach i o działalności informacyjnej. Dostrzegając przemiany, poszczególne zagadnienia dotyczące słów kluczowych analizuję nie w autonomicznym aspekcie dokumentacyjnym działalności informacyjnej, lecz przede wszystkim w kontekście zagadnień dotyczących sfery poznawczej człowieka.

Wobec wielości i różnorodności występujących stanowisk moją refleksję należy traktować jako głos w dyskusji nad językiem słów kluczowych, służącej porządkowaniu sposobów myślenia o tym, czym są obecnie słowa kluczowe i języki słów kluczowych. Porządkowanie i charakterystyka istoty słów kluczowych to działania ważne dla budowania ich wizerunku i spójności metodologicznej. Pojawiające się pytanie „czy to już koniec języków informacyjno-wyszukiwawczych” (Bojar 2007) sygnalizuje nieodzowność opracowania m.in. koncepcji spinającej w jedną całość różnorodne aspekty teorii i praktyki słów kluczowych oraz stworzenie spójnej wizji ich przyszłości.

Do rozprawy dodaję rejestr wykorzystanych publikacji, cytowanych w tekście lub w przypisach. Odwołania do literatury (nazwisko, rok). Odesłania do stron internetowych i dokumentów elektronicznych dotyczą ich wersji aktualnych 1 lipca 2009 r. Transliteracji dokonano według PN-ISO 9:2000 „Informacja i dokumentacja. Transliteracja znaków cyrylickich na znaki łacińskie”. Wyrażenia języków słów kluczowych podaję z zachowaniem oryginalnych, jakkolwiek bardzo zróżnicowanych, typografii. Publikacja jest przeznaczona dla szerokiego grona użytkowników słów kluczowych, w tym przede wszystkim dla specjalistów z zakresu języków i systemów informacyjno-wyszukiwawczych, dokumentalistów i bibliotekarzy oraz studentów informacji naukowej i bibliotekoznawstwa, a także dla tych wszystkich, którzy zajmują się teorią informacji naukowej, języków informacyjno-wyszukiwawczych i praktyką systemów wyszukiwania informacji.

Część pierwsza

Słowa kluczowe w różnych kontekstach

Celem tej części pracy jest analiza słów kluczowych w różnych kontekstach ich stosowania. Chodzi o kontekst (ang. *approach*) dziedzinowy, dotyczący określonej dziedziny wiedzy lub działalności praktycznej oraz kontekst funkcjonalny, dotyczący funkcji w procesach informacyjnych. Kontekst dziedzinowy stanowią informatyka, stylistyka i działalność informacyjna, zaś jako kontekst funkcjonalny przyjąłem kontekst organizacji informacji w systemach dokumentacyjnych.

Chciałbym przede wszystkim odpowiedzieć na pytania: czym jest słowo kluczowe, jak różne znaczenia ma ten termin i do czego się odnosi. Pokazuję to na przykładzie funkcjonowania słów kluczowych w wybranych dziedzinach, przede wszystkim w wyszukiwaniu informacji; poszukuję cech wspólnych słów kluczowych, niezależnych od kontekstu, a jednocześnie tego, co je różnicuje. Wspomniane konteksty generują różne funkcje słów kluczowych. Najważniejszym i podstawowym przyjętym tu kontekstem funkcjonalnym jest kontekst wyszukiwania informacji w systemach dokumentacyjnych; w jego ramach wyróżniam kontekst indeksowania i kontekst wyszukiwania informacji. Konteksty te mają wpływ na wieloznaczność i nieostrość terminu „słowo kluczowe” i stanowią jedną z przyczyn niejednoznacznego rozumienia i traktowania słów kluczowych w działalności informacyjnej, a w konsekwencji wpływają na sposób określenia tożsamości języka słów kluczowych.

1.1. Pole znaczeniowe, czyli pojemność terminu „słowo kluczowe”

1.1.1. Odmiany słów kluczowych: kontekst leksykograficzny

Określenie, czym jest słowo kluczowe, rozpoczynam od sformułowania definicji etymologicznej. Słowo kluczowe jest mianowicie kluczem do informacji, a więc narzędziem otwierania komunikatu, dostępu do zawartości dokumentu, umożliwiającym odbiór informacji, przy czym to właśnie słowo jest istotne, ważne, czyli akurat w tekście kluczowe.

Odwołując się do serwisów społecznościowych (choć ich waga jest nieporównywalna w stosunku do źródeł drukowanych), zauważyłem, że Wikipedia.pl (http://pl.wikipedia.org/wiki/S%C5%82slovo_kluczowe) słusznie zwraca uwagę, iż „słowo kluczowe” to termin istotny w wielu dziedzinach:

- w informatyce, gdzie słowem kluczowym jest nazywane słowo, mające szczególne znaczenie i oznaczające określony nakaz, instrukcję lub deklarację w programie komputerowym;
- w językoznawstwie, gdzie jest to słowo o ponadprzeciętnej frekwencji w danym tekście;
- w teorii informacji, gdzie słowo kluczowe ułatwia klasyfikację treści.

To ostatnie znaczenie zostało w 2008 roku nieco zmodyfikowane (<http://pl.wikipedia.org/wiki/Keyword>). Słowem kluczowym, wyrazem kluczowym nazwano bowiem „charakterystyczny ciąg znaków stosowany jako wyróżnik w dokumentach, wspomagający proces wyszukiwania informacji”. Znalazły się tu też informacje o posługiwaniu się słowem kluczowym (wyrazem kluczowym) w dokumencie HTML, gdzie jest ono ciągiem znaków umieszczonym w części nagłówkowej dokumentu: `<META NAME="Keywords" CONTENTS="komputer, Internet, edytor, HTML">`. Tego rodzaju polecenie informuje programy indeksująco-wyszukiwawcze o wyrazach kluczowych dokumentu. Wyrazy takie są umieszczane w specjalnej sekcji bazy danych wyszukiwarki, co zwiększa szansę znalezienia strony przez innych użytkowników sieci. To autor decyduje, jakie są istotne elementy strony.

Więcej znaczeń terminu „słowo kluczowe” podaje Wikipedia w języku angielskim (<http://en.wikipedia.org/Wiki/Keyword>). Oprócz podstawowego, wymienionego na pierwszym miejscu, którym jest wyrażenie o specjalnym znaczeniu, a szczególnie wyrażenie użyte jako klucz do kodu lub użyte w działalności informacyjnej do łączenia wyrażen bądź informacji, podaje, że jest to wyrażenie wykorzystywane do znajdowania relewantnych i użytecznych stron internetowych. Wymienia też obszary używania słów kluczowych:

- języki programowania, gdzie słowo kluczowe oznacza specyficzne polecenie;
- językoznawstwo, gdzie oznacza słowo, które ma w tekście ponadprzeciętną częstotliwość. Oprócz tego sygnalizuje dziedzinę kultury, w której używa się też terminu „słowo kluczowe”.

Odwołam się jeszcze do słownika przekładowego angielsko-polskiego *Oxford Wordpower*, który podaje trzy znaczenia terminu angielskiego „słowo kluczowe”: 1. wyrażenie wyrażające główną ideę lub przedmiot czegoś; 2. wyrażenie, które identyfikuje rekord lub dokument w systemie komputerowym; 3. (używane w programowaniu komputerowym) wyrażenie, które jest stosowane jako polecenie. Języki programowania zawierają zbiór słów kluczowych, z których każde używane jest w sposób specyficzny, na przykład w celu otwierania pliku, i może być użyte tylko w tym znaczeniu (Oxford Wordpower 2002)¹⁴.

W językoznawstwie przyjmuje się, że słowo kluczowe (także wyraz kluczowy, słowo klucz, wyraz klucz) to wyraz charakterystyczny dla danego tekstu (autora, stylu, gatunku). Listę takich słów kluczy uzyskuje się przez porównanie frekwencji słów w tekście z ich frekwencją w języku w ogóle. Słowa o zawyżonej frekwencji to właśnie słowa klucze, na przykład wyrazy klucze w *Weselu* Stanisława Wyspiańskiego to m.in. serce, dusza, krew, sen, duch, koń, złoty, spać, wicher, słuch (Stachurski 1998). Pojęcie słów kluczy znajduje zastosowanie w analizie stylu i poetyki poszczególnych tekstów bądź autorów/epok, a także w językoznawstwie komputerowym, chociażby w przykładzie automatycznym. Blisko spokrewniony z pojęciem wyrazu klucza jest termin „wyraz tematyczny” (por. <http://pl.wikipedia.org/>).

Doszukując się podobieństw słów kluczowych do stosowanych w literaturze pięknej wyrazów/słów kluczy, trzeba pamiętać, że twórcą bardzo użytecznego w badaniach stylistycznych terminu „wyraz klucz” jest semiotyk i językoznawca francuski Pierre Guiraud, który w 1954 roku, przez porównanie list frekwencyjnych słownictwa utworów sześciu poetów reprezentujących symbolizm francuski z listami frekwencyjnymi słownictwa zawartego w ogólnym słowniku frekwencyjnym języka francuskiego, ustalił dla każdego badanego tekstu listę dwudziestu wyrazów kluczy (Guiraud 1954, 1966). Metoda słów kluczy została przeniesiona na grunt polski przez Kazimierza Wykę. Wyka w pracy *Słowa-klucze* (Wyka 1973) prezentuje najpierw metodę klasyczną Pierre’a Guirauda z odwołaniem do Gastona Bachelarda i Carla Gustava Junga. Metoda ta polega na analizie doboru słowa, który to wybór wymyka się spod kontroli danego podmiotu, łączy się ze skojarzeniami określonego podmiotu, a także archetypami myślowymi. Guiraud obliczał, a następnie porównywał trzy częstotliwości występowania wyrazu: pierwsza dotyczyła częstotliwości użycia słów w danym systemie w ogóle, drugą była lista funkcyjna zasobu słów u danego poety lub w danym utworze. Trzecia stanowiła podstawę jego własnej listy. Wyrazy obecne na liście trzeciej Guiraud nazwał właśnie słowami kluczami (Guiraud 1954). O możliwościach i niebezpieczeństwach metod statystycznych w poetyce i stylistyce pisali m.in. Kazimierz Budzyk i Maria Renata Mayenowa. Poza Wyką słownictwem kluczowym zajmowali się Elżbieta Rudnicka-Fira i Edward Stachurski. Omówienie ich badań zawiera praca E. Stachurskiego (Stachurski 1998).

¹⁴ Keyword: 1. a word that tells you about the main idea or subject of sth; 2. a word that identifies a specific record or document in a computer system. A keyword may be used when searching the Internet; 3. (used in computer programming) a word that is used as a command. Programming languages have a set of keywords and each keyword has a specific use, for example, for opening files, and can only be used for this purpose (Oxford Wordpower 2002, p. 428).

W polskim piśmiennictwie naukowym niektórzy badacze przyjęli za Wyką termin „słowo klucz”¹⁵ (S. Cieślikowski; E. Stachurski). Większość badaczy albo stosuje termin „wyraz klucz” (J. Sambor; W. Pisarek; M. Zembaty-Michalakowa), albo używa zamiennie obu terminów (M. Rachwałowa; E. Rudnicka-Fira). W niektórych tekstach w użyciu są równoważniki znaczeniowe tych terminów: „wyraz typowy” (W. Kuraskiewicz; W. Pisarek), „wyraz autosemantyczny (temat) charakterystyczny” (M. Zarębina), „wyraz tematyczny charakterystyczny” lub „wyraz tematyczny typowy” (M. Rachwałowa). Niektórzy europejscy językoznawcy nazywają wyrazy faworyzowane przez pisarzy „słowa uprzywilejowanymi” (por. Sigurd 1975).

Słowa kluczowe danego tekstu literackiego to wyrazy o mniejszych numerach na liście rankingowej leksemów tego tekstu (co wynika z ich wysokiej frekwencji) w porównaniu z odpowiednimi danymi epoki, w której utwór powstał. Przy ustalaniu listy wyrazów kluczowych dla tekstów literackich z niektórych epok nie dysponujemy ogólną płaszczyzną porównawczą. Można wyróżnić specjalne podjęzyki słów kluczy (nazywane także słowami kluczowymi) stosowane w literaturze pięknej przez pisarzy i/lub poetów, na przykład Adama Mickiewicza, Mikołaja Reja i innych.

Niewątpliwie słowa kluczowe należą do słownictwa tematycznego języka naturalnego. Przeważnie są wykładnikami głównych tematów tekstu. Stąd pełnią funkcję metainformacyjną w stosunku do danego tekstu lub korpusu tekstów, bądź – jak w informatyce – funkcję impresywną służącą do sterowania operacjami realizowanymi przez komputer, wyspecjalizowanych w tej funkcji języków programowania (języki sztuczne). Wykładniki tej funkcji są często identyczne z wykładnikami funkcji metainformacyjnej, ale nie zawsze mają takie samo znaczenie. Stanowią rodzaj/ odmianę funkcjonalną słów kluczowych. Są charakterystyczne dla twórczości poszczególnych autorów. Słowa kluczowe to przede wszystkim rzeczowniki, które pełnią funkcje stylistyczne. W literaturoznawstwie i literaturze pięknej funkcjonujące tam słowa kluczowe pełnią czasami funkcję wyrażenia metaforycznych i opisowych. Są one wynikiem nadania przez autora (lub literaturoznawcę) danemu wyrażeniu języka naturalnego specjalnego znaczenia oraz rangi.

Słowa kluczowe o ogólnej wartości znaczeniowej mogą też być jednostkami leksykalnymi języka semantycznego lub języka myślenia, wyrażającymi dominanty znaczeniowe określonych idei i pojęć.

Przytoczone definicje stanowią ilustrację istnienia różnych odmian i funkcji wyrażań zwanych słowami kluczowymi. Jedna z tych odmian służy do wyszukiwania informacji. Szczególnie tą odmianą zajmiemy się w dalszej części rozprawy. Istniejące różnice w podejściu do słowa kluczowego przez poszczególne dyscypliny naukowe i ich różne funkcje wprowadzają pewne zamieszanie terminologiczne i utrudniają ich identyfikację w wyszukiwaniu informacji.

¹⁵ Słowo klucz – wyraz pojawiający się z dużą częstością w tekście dokumentu, utworu, tematu. Frekwencja ta sprawia, że to słowo ma duże znaczenie, dużą wartość. Na bazie tak wykształconego pojęcia słowa klucza pojawiły się w systemie informacyjno-wyszukiwawczym słowa kluczowe.

1.1.2. Słowa kluczowe w nauce o informacji

W nauce o informacji na podstawie literatury przedmiotu można wyróżnić trzy odmienne podejścia do słów kluczowych. W pierwszym z nich podkreśla się odrębność w stosunku do wyrażen języka naturalnego i w tym celu wykorzystuje pojęcie systemu językowego. To system leksykalny sztucznego języka słów kluczowych stanowi kontekst słów kluczowych. Drugie podejście akcentuje bezpośrednio związki integracyjne z językiem naturalnym. Uważa się, że to język naturalny nadaje kontekst słowom kluczowym. Można wyróżnić również podejście, w którym to system leksykalny języka słów kluczowych stanowi kontekst poszczególnych słów kluczowych, zaś język naturalny (jego poszczególne podsystemy) uzupełnia deficyty tego kontekstu.

W polskim piśmiennictwie z zakresu nauki o informacji przeglądu definicji słowa kluczowego jako pierwsza dokonała Jadwiga Choroszuszyzna (Choroszuszyzna 1970), twierdząc, że istnieje wiele definicji tego terminu oraz że jego zakres znaczeniowy nie jest jednoznacznie ustalony. Są autorzy, którzy ograniczają pojęcie słowa kluczowego do jednego wyrazu głównego, derywowanego z danego tekstu. Tak zdefiniowano słowo kluczowe w *Słowniku terminologicznym informacji naukowej* (1979. s. 105), gdzie określono je jako „wyraz lub wyrażenie wybrane z tytułu lub tekstu dokumentu, charakteryzujące jego treść”. Marguerite Fischer (1966), autorka amerykańska, identyfikowała słowo kluczowe z hasłem lub wyrazem głównym, które mogą być wzięte z tytułu, z analizy lub z tekstu i zastosowane w indeksie. Josef Koblitz (1968), badacz niemiecki, określił słowo kluczowe jako „samodzielnie występujący wyraz języka fachowego, który w pierwszej fazie indeksowania, tzn. w fazie odtwarzania zawartości treściowej dokumentu, występuje w języku naturalnym”, utożsamiając słowo kluczowe z jednowyrazowym terminem fachowym (Koblitz 1968), zaś Józef Robowski (1968), specjalista od języków informacyjno-wyszukiwawczych w Polsce, poprzez słowo kluczowe rozumiał terminy fachowe jedno- lub wielowyrazowe, zmieniając odpowiednio nazwy „uniterm” i „słowo kluczowe” na „termin kluczowy”. Mikołaj Poletyło (1968) za słowa kluczowe uważał charakterystyczne dla danej dziedziny wyrażenia w postaci nazwy jedno- lub kilkuwyrazowej, zaś A. I. Czerny (1969) dokonał rozróżnienia na: słowa kluczowe właściwe, czyli unitermy, oraz grupy wyrazowe lub frazy kluczowe, przy czym za słowa kluczowe przyjmował wyrazy o zwiększonej wartości informacyjnej i nie ograniczał ich do terminów fachowych. To tylko wybrane definicje słowa kluczowego.

Próbie określenia pola znaczeniowego terminu „słowo kluczowe” podjęła Jadwiga Choroszuszyzna na podstawie „Nachrichten für Dokumentation” (1968, nr 4, s. 144). Ustaleniem zasięgu znaczeniowego terminu „słowo kluczowe” zajmował się też Niemiecki Komitet Terminologii i Zagadnień Językowych przy Deutsche Gesellschaft für Dokumentation, który na podstawie przeprowadzonych badań opublikował w 1967 roku wyniki prac terminologicznych, inicjując jednocześnie w „Nachrichten für Dokumentation” nr 8 z 1968 roku (s. 143–144) dyskusję na ten temat. Przy opracowywaniu pola znaczeniowego tego terminu wzięto pod uwagę sformułowania związane ściśle z zakresem terminu „słowo kluczowe”, takie jak hasło, hasło przedmiotowe oraz wyraz główny. Opracowano odpowiedni schemat graficzny, na którym

przedstawiono hierarchię i powiązania funkcjonalne oraz formalne terminu „słowo kluczowe” z terminami pokrewnymi, m.in. z hasłem przedmiotowym, z wyrazem głównym i z deskryptorem. Choroszuszyzna wykorzystwała schemat opracowany przez ten Komitet, pokazując funkcjonalne i formalne powiązania terminu „słowo kluczowe” z terminami pokrewnymi (rys. 1).

Na schemacie zaznaczyła następujące pola znaczeniowe terminu „słowo kluczowe”:

A1. SŁOWO KLUCZOWE – Schlüsselwort – keyword – słowo do oznaczania zawartych w dokumencie pojęć;

B1. HASŁO PRZEDMIOTOWE – Schlagwort – subject heading – słowo kluczowe z funkcją porządkującą;

B.1.1. WYRAZ GŁÓWNY = HASŁO – Stichwort – clue word – hasło przedmiotowe wzięte bezpośrednio z dokumentu;

B.1.1.1. WYRAZ GŁÓWNY Z TYTUŁU = hasło tytułowe – Titelstichwort – wyraz główny wzięty z tytułu dokumentu;

B.1.1.2. WYRAZ GŁÓWNY Z TEKSTU = hasło tekstowe – Textstichwort – wyraz główny wzięty z tekstu;

C.1. HASŁO PRZEDMIOTOWE ZWIĄZANE – Gebundenes Schlagwort – hasło przedmiotowe pochodzące z obowiązującego wykazu hasel;

C.2. HASŁO PRZEDMIOTOWE SWOBODNE – Freies Schlagwort – open ended term – hasło przedmiotowe niepochodzące z obowiązującego wykazu hasel;

SŁOWO INDEKSOWE – Registerwort – każde słowo zawarte w indeksie;

SŁOWO Z TEZAURUSA – Thesauruswort – każde słowo zawarte w teaurusie;

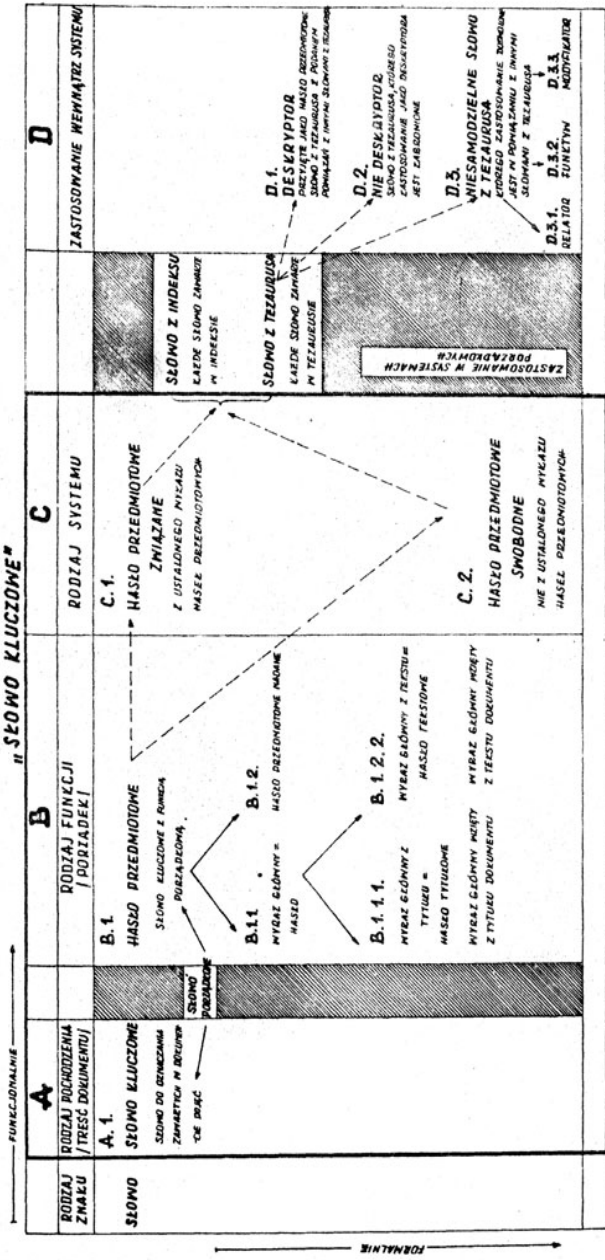
D.1. DESKRYPTOR – Deskryptor – Descriptor – przyjęte jako hasło przedmiotowe słowo z tezaury z zaznaczeniem powiązań z innymi słowami z tezaury;

D.2. NIEDESKRYPTOR – Nicht-Deskryptor – non descriptor – słowo z tezaury, którego użycie jako deskryptora jest zabronione;

D.3. NIESAMODZIELNE SŁOWO Z TEZAURUSA – Unselbständiges Thesauruswort – słowo z tezaury, którego użycie dozwolone jest tylko z jednym lub kilkoma słowami z tezaury.

Komentując ten schemat, zwracam uwagę, że słowa kluczowe różnią się pod względem formalnym i funkcjonalnym od deskryptorów i tematów w języku hasel przedmiotowych. Nie są to więc pojęcia równoznaczne, chociaż wyrazy wybrane z tekstu jako słowa kluczowe mogą pełnić zarówno funkcję deskryptorów, jak i tematów. Słowa kluczowe wybiera się dla konkretnego jednostkowego dokumentu, tematy zaś tworzy się dla pewnej grupy dokumentów. Terminu „deskryptor” używa się na określenie wyrazów, które zostały wprowadzone jako obowiązujące do słownika, czyli tezaury. Mogą nimi być nie tylko terminy, lecz także wyrażenia naturalnego języka ogólnego. Tworzą zbiór znormalizowany zwany teaurusem. Tak jak słowo kluczowe jest kwintesencją znaczeń słów w tekście i niesie treść tekstu, tak deskryptor odwzorowuje treść dziedziny o wyznaczonych granicach. Zatem deskryptor jest słowem kluczowym w dziedzinie, a słowo kluczowe istnieje dzięki tekstowi. Słowa kluczowe, podobnie jak tematy w językach hasel przedmiotowych, są nazwami wywoławczymi, symbolizującymi i przypominającymi dany dokument. Często jest to wyraz wyjęty z tytułu.

SCHEMAT POLE ZNACZENIOWE TERMINU II SŁOWO KLUCZOWE



Wyjaśnienie: Pole znaczeniowe terminu "słowo kluczowe" ograniczono funkcjonalnie w schemacie do pól A, B i C w stosunku do innych pól znaczeniowych. Powiązania funkcjonalne zaznaczono na schemacie poziomo, powiązania formalne pionowo. Pojęcia objęte polem znaczeniowym można oznaczać nazwą według ich funkcji lub formy.

Rys. 1. Pole znaczeniowe terminu „słowo kluczowe” (źródło: „Nachrichten für Dokumentation” 1968, nr 4, s. 144). Za: Choroszyńska 1970, s. 51.

W indeksach KWIC¹⁶ słowo kluczowe jest w zasadzie unitermem, a więc jednym wyrazem wziętym bezpośrednio z tytułu z zachowaniem nawet formy gramatycznej, podczas gdy tematem mogą być wyrazy niewystępujące w danym tekście. Są to tzw. hasła nadane. Warto zauważyć, że słowami kluczowymi nazywa się także jednostki leksykalne w językach syntagmatycznych, na przykład SYNTOL. Rozważania nad relacją słów kluczowych do innych paranaturalnych jednostek leksykalnych będą kontynuował w rozdziale 2.5.

Zdaniem Jadwigi Ćwiekowej można przyjąć, że: „Słowo kluczowe jest to zazwyczaj wyrażenie rzeczownikowe, najczęściej jednowyrazowe, zaczerpnięte z tytułu lub z analizy dokumentacyjnej, charakteryzujące treść danego dokumentu w postaci nazwy tematu” (Ćwiekowa 1988, s. 50). Wyrażanie treści dokumentu za pomocą tematów sugeruje pewien związek z językiem haseł przedmiotowych. Obecnie słowa kluczowe służą do wyrażania w różny sposób dowolnych, najczęściej najważniejszych elementów treści dokumentów relewantnych dla autora lub użytkownika, nie ograniczając się do ich prezentacji w postaci tematów. Słowa kluczowe są wykładnikami tych elementów treści dokumentu, które są relewantne dla autora lub użytkownika systemu.

Termin „słowo kluczowe” oznacza przede wszystkim słowo znaczące, czyli „wyraz języka naturalnego o dużej wartości znaczeniowej dla określonej dziedziny wiedzy, pełniący funkcję semantyczną” (*Słownik terminologiczny informacji naukowej* 1979, s. 106), mający jeszcze tę wagę, że może być wykorzystywany do sporządzania indeksów. Według cytowanego słownika słowo znaczące (ang. *significant word*) to termin szerszy od słowa kluczowego (ang. *key word, keyword, keyterm, catchword*). To ostatnie jest zdefiniowane jako „wyraz lub wyrażenie wybrane z tytułu lub tekstu dokumentu charakteryzujące jego treść” (tamże), bądź „treść zapytania informacyjnego”, jak dodaje *Słownik encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych* (Bojar, red. 1993, s. 138), i zarazem wyróżnia dodatkowe znaczenie, definiując słowo kluczowe jako „elementarną jednostkę leksykalną języka słów kluczowych”.

Użycie terminu „słowo kluczowe” w kontekście wyszukiwania informacji (w roku 1948) przypisuje się amerykańskiemu matematykowi Calvinowi N. Mooersowi. Słowa kluczowe zostały też wykorzystane przez Mortimera Taubego w 1952 roku w systemie UNITERM. Nadał on formalne reguły systemom leksykalnym przeznaczonym do wyszukiwania dokumentów. Taube uważał, że zarówno treść dokumentu, jak i treść pytania użytkownika mogą z pełną dokładnością, zależnie od celów wyszukiwania, być przenoszone przez słowa kluczowe złożone jedynie z leksykalnych komponentów języka naturalnego (Frants, Shapiro, Voiskunskii 1997). Taube nazwał te słowa *uniters* (unit; term), czyli jednostkami mającymi znaczenie terminologiczne, najlepiej odpowiadającymi zarówno treści dokumentu, jak i pytaniu użytkownika. Trochę później termin ten został zastąpiony przez wyrazy *keywords* i *descriptor*, określane jako słowa lub kombinacje słów zawarte w dokumencie, ekwiwalentne do

¹⁶ Indeks KWIC (ang. *Key Words in Context*) to indeks, w którym wyrażenie pełniące funkcję hasła jest umieszczane w stałym miejscu w kontekście innych wyrażań, opatrzone adresem dokumentu.

pytania użytkownika. Opracowany i udoskonalony przez Taubego i jego współpracowników system UNITERM był początkowo przeznaczony do selekcji ręcznej, potem maszynowej. W systemie tym opis dokumentu składał się z szeregu unitermów, które jako elementarne jednostki leksykalne nie ujawniały relacji między pojęciami.

Za cenną i wartą przytoczenia uważam definicję słowa kluczowego zamieszczoną w słowniku języka angielskiego *The Oxford English Dictionary*. Jedno ze znaczeń słowa kluczowego (ang. *keyword*) to: „wyraz lub przedmiot mające wielką wagę i znaczenie; spec. w systemach wyszukiwania informacji (ang. *information retrieval*), zawierający informacyjne znaczenie, wyraz w tytule lub tekście dokumentu etc., wybierany jako wskazujący główną zawartość dokumentu” (Simpson, Weiner 1989, ed., s. 406). Niewątpliwą zaletą tego słownika są odsyłacze do źródeł informujących o tym, gdzie dany termin został po raz pierwszy zastosowany. I tak, w odniesieniu do słowa kluczowego – w znaczeniu terminu wyszukiwawczego słownik oksfordzki przytacza artykuł E.W.F. Tomlina, który ukazał się w czasopiśmie „*The New Outline of Modern Knowledge*” w 1956 roku. Kolejnym źródłem jest artykuł N.S.M. Coxa i Michaela Grose’a zamieszczony w „*Organization in Bibliography Record by Computer*” z 1967 roku. W 1971 roku, w czasopiśmie „*New Scientist and Science Journal*”, w artykule George’a Smitha i Jima Lyncha pt. *Information by the yard* słowo kluczowe zostało zdefiniowane jako termin przeznaczony do opisywania zawartości dokumentu tak ściśle i dokładnie, jak to tylko możliwe¹⁷. Potwierdza to moje wcześniejsze spostrzeżenie, że słowa kluczowe służą w zasadzie do wyrażania dowolnej treści (nie tylko dokumentu) relewantnej dla użytkownika w sposób zależny od jego potrzeb.

Uważam, że warto jeszcze odnotować definicję słowa kluczowego podaną w *Le Grande dictionnaire terminologique de l’Office de la langue français*, gdzie zdefiniowano je jako słowo lub grupę słów, wybrane z tytułu lub z tekstu dokumentu, charakteryzujące zawartość i pozwalające na wyszukanie tego dokumentu w systemie informacyjno-wyszukiwawczym i w bazach danych. Definicja ta lokuje słowo kluczowe w kontekście systemu wyszukiwania informacji.

Analizując różne definicje słowa kluczowego, odniosę się jeszcze do definicji zamieszczonej w słynącym z dobrego poziomu *ODLIS: Online Dictionary for Library and Information Science* (http://lu.com/odlis/odlis_k.cfm), w którym słowo kluczowe zdefiniowano jako słowo znaczące lub frazę (i tu wymienia się źródła słów kluczowych) użyte jako termin wyszukiwawczy w wyszukiwaniu pełnotekstowym, w celu wyszukania wszystkich rekordów zawierających to słowo¹⁸. Jest to wyrażenie (języ-

¹⁷ W artykule tym opisany został jeden z pierwszych eksperymentów nad zastosowaniem słów kluczowych, który odbył się w latach 1967–1969, a przeprowadzony został przy pomocy systemu Chemical Titles (CT). Dziedzina, dla której zestawiono słownictwo, to zastosowanie mikroskopów elektronowych. W tym czasie była to stosunkowo wąska specjalność i zarazem dobrze opracowana. Początkowo uzyskano rezultaty niezadowolające z powodu zbyt wielu odrzuconych terminów. Po wprowadzeniu zmian udało się jednak osiągnąć dziewięćdziesięcioprocentową relewancję, co uznano za wielki sukces, ponieważ większość wyników innych badań sytuowała się w przedziale od 40 do 60% relewancji. Jako przykład podobnego projektu autorzy podają eksperyment przeprowadzony w laboratorium komputerowym Uniwersytetu w Newcastle, który dotyczył systemu MEDLARS Data i zawierał próbkę 40 tysięcy artykułów.

¹⁸ Komputerowy słownik *ODLIS: Online Dictionary for Library and Information Science* podaje następującą definicję: “A significant word or phrase in the title, subject headings (descriptors), contents note,

ka naturalnego) użyte w specjalnej funkcji, a mianowicie w funkcji metainformacyjnej (język indeksowania) lub w funkcji wyszukiwawczej (język wyszukiwawczy).

W piśmiennictwie polskim termin „słowo kluczowe” najpełniej zdefiniował Olgierd Ungurian, podając, że jest to „wyraz lub wyrażenie (występujące w tekście lub nie), które zdaniem indeksatora w sposób najbardziej precyzyjny określa zawartość tekstu dokumentu (jego treść, przedmioty, o których mowa), a także często cechy formalne dokumentu. Słowa kluczowe stanowią zazwyczaj zasób nazw języka naturalnego, które mogą być przydatne w procesach opracowania i wyszukiwania informacji, a zebrane w procesie indeksowania dokumentów” (Ungurian 1976, s. 186).

Przegląd definicji terminu „słowo kluczowe” kończą definicją opublikowaną w 1985 roku w normie ISO 5963, gdzie słowo kluczowe to „wyraz lub grupa wyrazów, możliwie leksykograficznie ujednoczonych, wybranych z tytułu lub tekstu dokumentu charakteryzujących jego zawartość i służących do jego wyszukiwania”¹⁹. Z definicje tej wynika, że za słowo kluczowe może być uznane każde wyrażenie (wyraz), wybrane z tytułu lub z tekstu dokumentu, wykorzystane do reprezentowania treści, lub dowolne wyrażenie z pytania informacyjnego.

We wszystkich poddanych analizie węższych lub szerszych definicjach terminu „słowo kluczowe”, powtarzają się pewne elementy, które są istotne dla tego pojęcia. Omawiam je w kolejności występowania:

1. Słowo kluczowe to wyraz lub połączenie wyrazowe, czyli wyrażenie języka.
2. Słowo kluczowe to wyraz lub grupa wyrazów o ustalonym konwencjonalnie znaczeniu, odpowiadającym pojęciu, z czego powinna wynikać ostrość granicy semantycznej, a więc jednoznaczność słowa kluczowego.
3. Słowo kluczowe należy do określonej dziedziny wiedzy lub działalności praktycznej. Najczęściej wymienia się naukę i technikę.
4. Niektóre definicje podkreślają fakt nadawania wyrazom języka ogólnego specjalnych znaczeń. Nie precyzuje się jednak konstrukcji stanowiących połączenia wyrazowe, które są wyjątkowo częstym sposobem wzbogacania zasobu słów kluczowych.
5. W poszczególnych definicjach wymienia się zwykle takie funkcje słów kluczowych, jak funkcja etykietyzacji dokumentu/tekstu, funkcja imperatywna oraz funkcja klucza wyszukiwawczego. Dzięki tym funkcjom stają się one środkiem dostępu do treści/zawartości dokumentu.

Analiza leksykalna przytoczonych definicji słowa kluczowego ujawniła, że w jednych opisuje się, jak „powstaje” słowo kluczowe, w innych określa się, jakie powinno ono być, a tylko w nielicznych autorzy próbują ustalić, czym jest słowo kluczowe i jakie są jego cechy. Zazwyczaj wymienia się te same cechy, które powinien posia-

abstract, or text of a record in an online catalog or bibliographic database that can be used as a search term in a free-text search to retrieve all the records containing it. See also: stopword”. Lista słów kluczowych (*keyword index*) to: “A type of subject index in which significant words, usually from the titles of the works indexed, are used as headings. When a string of keywords is rotated, such an index is said to be permuted. See also: KWAC, KWIC and KWOC” (http://lu.com/odlis/odlis_k.cfm).

¹⁹ A keyword(s) is “a word or group of words, possibly in lexicographically standardized form, taken out of a title or of the text of a document characterizing its content and enabling its retrieval” (ISO 5963: 1985).

dać termin, co prowadzi do wniosku, że w przeważającej części słowo kluczowe to termin pełniący określoną rolę. Słowami kluczowymi najczęściej są więc terminy, czyli wyrażenia o ściśle ustalonym znaczeniu w danej dziedzinie nauki lub techniki. Terminom stawia się przede wszystkim wymaganie dokładnego określania znaczenia oraz stałość. W zasadzie terminy nie powinny mieć synonimów, jednak w wyniku oddziaływania różnych czynników zewnętrznych na język nauki, na wyrażenie tego samego przedmiotu lub pojęcia używa się nieraz wielu terminów. W ten sposób synonimia staje się zjawiskiem właściwym także w stosunku do języków specjalistycznych, w tym specjalistycznej terminologii naukowej i technicznej. Z tego wynika, że słowo kluczowe jest terminem, ale termin nie musi być słowem kluczowym. W literaturze przedmiotu bardzo często utożsamia się słowa kluczowe z niekontrolowaną leksyką języków informacyjno-wyszukiwawczych.

Przytoczone definicje słowa kluczowego wskazują też, że termin ten występuje w dwóch znaczeniach, często niezbyt wyraźnie uświadomianych. W pierwszym znaczeniu – co jest widoczne w przeważającej liczbie definicji – za słowo kluczowe uważa się każde wyrażenie (wyraz), wybrane z tytułu lub z tekstu dokumentu, wykorzystywane do odtwarzania (kodowania) treści, lub dowolne słowo z zapytania informacyjnego. W tym znaczeniu słowa kluczowe to wyrażenia języka naturalnego. W drugim znaczeniu, na co zwracają uwagę tylko nieliczne definicje, słowo kluczowe rozumiane jest jako jednostka leksykalna języka słów kluczowych, czyli języka sztucznego, rozumianego jako język, którego słownictwo tworzą wyrażenia zwane słowami kluczowymi, wybierane z tekstu dokumentu lub z treści pytania informacyjnego, ale przynajmniej w części lub w pełni kontrolowane, zaś gramatykę stanowią reguły indeksowania współrzędnego.

Słowo kluczowe, jako jednostka leksykalna języka słów kluczowych, powinno być z informacyjnego punktu widzenia słowem znaczącym, posiadać zdolność selekcyjną i spełniać przyjęte wymogi formalne. W języku polskim warunek ten spełniają rzeczowniki i przymiotniki w mianowniku liczby pojedynczej lub tzw. pluralia tantum, a nie spełniają go spójniki, zaimki oraz przyimki. Określenie, że są to słowa szczególnie ważne, może oznaczać zarówno ważność wewnętrzną, wyznaczoną przez miejsce, jakie pełni dane pojęcie w całokształcie aparatury pojęciowej tekstu, jak i ważność zewnętrzną – z punktu widzenia określonego odbiorcy tekstu. Zdarza się, że pewne zagadnienie uznane za drugoplanowe jest istotne (bardzo ważne) dla jakiegoś odbiorcy informacji. We współcześnie stosowanych językach słów kluczowych funkcję jednostki leksykalnej, zwanej słowem kluczowym, może pełnić zarówno pojedynczy rzeczownik, jak i fraza rzeczownikowa (na przykład „systemy o kontrolowanej jakości”).

Niezależnie od dziedziny i zakresu użycia w dalszej części niniejszego opracowania przyjąłem, że słowa kluczowe są rezultatem analizy treści dokumentu, stając się swoistego rodzaju reprezentantem treści dokumentu oraz jego etykietą. W stosowanej obecnie w Polsce terminologii języków informacyjno-wyszukiwawczych – jakkolwiek zauważa się to podwójne znaczenie – słowo kluczowe najczęściej jest definiowane wąsko, zgodnie z tradycyjnym ujęciem dokumentacyjnym, wywodzą-

cym się jeszcze z początku lat pięćdziesiątych XX wieku i z terminologii, która powstawała na użytek języków unitermowych.

Słowa kluczowe definiuje się zwykle jako terminy jedno- lub wielowyrazowe, wybrane swobodnie, tj. bez posługiwania się jakimkolwiek normatywnym wykazem terminów z tytułu lub z tekstu charakteryzowanego dokumentu. W istocie są to swobodne (niekontrolowane) słowa kluczowe. Są to więc wyrazy lub wyrażenia przejęte z języka naturalnego.

1.1.3. Funkcje słów kluczowych

W dalszym ciągu tej części zajmę się funkcjami, z jakimi łączy się używanie języków słów kluczowych przede wszystkim w procesach informacyjnych oraz w systemach informacyjno-wyszukiwawczych, ale nie tylko. Jest to odmienne podejście w stosunku do zaprezentowanego wcześniej podejścia dziedzinowego. Funkcję słów kluczowych najlepiej oddaje porównanie ich do kluczy. Powinny otwierać dostęp do treści tekstów (por. słowa klucze w literaturze, konkordancje do Biblii).

Poza – już wspomnianym – specyficznym wykorzystywaniem nazwy „słowo kluczowe”, w niektórych dziedzinach, jak na przykład literaturoznawstwo, stylistyka, informatyka, i przypisywaniem ich do poszczególnych dziedzin, słowa kluczowe są przede wszystkim lingwistycznym narzędziem wyszukiwania dokumentów i informacji. Obserwowanym powszechnie zjawiskiem jest wykorzystywanie słów kluczowych do selekcjonowania informacji w obrębie pól haseł przedmiotowych.

Zakres używania słów kluczowych uległ znacznemu rozszerzeniu przede wszystkim o nowy rodzaj systemów, jaki stanowią strony internetowe będące elementem strukturalnym WWW. Słowa kluczowe są stosowane w Internecie do określania – na poziomie meta – treści dokumentów elektronicznych oraz potrzeb informacyjnych użytkowników. Jest to też narzędzie służące webmasterom do nadawania stronom internetowym (pozycjonowania) odpowiedniej pozycji (rangi) w cyberprzestrzeni. Dla pełnienia tej funkcji istotne są relewancja słów kluczowych w stosunku do zawartości danej strony, ich aktualność, atrakcyjność oraz związana z tą ostatnią konkurencyjność. Szczegółowo zajmę się tym w następnym rozdziale.

W przypadku funkcjonowania w sferze komunikowania naukowego słowa kluczowe jako nazwy są umotywowane naukowo: pełnią funkcję informacyjną (odwozowanie rzeczywistości dokumentacyjnej) i wyszukiwawczą. W obu tych sferach istotne jest przestrzeganie kryteriów tworzenia dobrej nazwy, m.in. zwięzłości, prostoty i zrozumiałości.

Słowa kluczowe mogą też tworzyć specjalną grupę słownictwa. Poszczególnym słowom kluczowym towarzyszy wówczas odpowiednia motywacja²⁰. I tak, słowa

²⁰ „Motywacja” jako termin językoznawczy oznacza „zjawisko niepełnej arbitralności (konwencjonalności) znaku językowego, w tym zdeterminowanie formy i znaczenia znaku językowego przez czynniki pozajęzykowe (...) lub zdeterminowanie formy i/lub znaczenia jednostki językowej – zwanej wówczas motywowaną – przez inne jednostki tego samego języka – zwane motywującymi” (EJO 1993, s. 347). „Motywacja nazewnicza to wyjaśnienie, co nadawca chciał osiągnąć przez użycie w nazwie określonego

kluczowe mogą odwzorowywać system terminologiczny danej dziedziny. Przykładem może być publikacja Haliny Wojnowskiej-Dawiskiby pt. *Wybrane słowa kluczowe i standardy kwalifikacji zawodowych pielęgniarki* (1999) zawierająca wykaz wybranych słów kluczowych mających zastosowanie w kształceniu zawodowym. W rzeczywistości pod tą nazwą kryje się alfabetyczny wykaz terminów z całej dziedziny (w tym wypadku pielęgniarstwa) wraz z definicjami/określeniami. Ich źródłem, jak pisze autorka, były odpowiednie akty prawne i wydawnictwa leksykograficzne z zakresu nauk pedagogicznych, jak *Encyklopedia Pedagogiczna*, *Słownik Pedagogiczny*, *Słownik Pedagogiki Pracy*, oraz istniejąca literatura przedmiotu. Niektóre terminy adaptowano na użytek tego opracowania dotyczącego standardów kwalifikacji dla zawodu pielęgniarka. Wspomniany wykaz, jakkolwiek używa nazwy „słowo kluczowe”, to w intencji autorki nie był i nie jest przeznaczony do indeksowania dokumentów i pytań informacyjnych, a więc dla dokumentacyjnych systemów wyszukiwania informacji; odwzorowuje on jedynie system terminologiczny wspomnianej dziedziny.

Słowo kluczowe, oprócz pełnienia funkcji informacyjnej, może być jednostką leksykalną języka marketingowego, w którym pełni również funkcję perswazyjną, reklamową²¹. W tej sferze słowa kluczowe niosą też odpowiednio silny ładunek perswazyjny, język marketingowy stanowi bowiem lingwistyczne narzędzie manipulacji odbiorcami informacji językowych. W obecnej rzeczywistości medialnej słowa kluczowe w systemie WWW, obok dostarczania meta/informacji, bardzo często pełnią rolę promującą. W tej sytuacji słowo kluczowe, jako nazwa, niejednokrotnie staje się ważniejsze od słowa kluczowego jako jednostki leksykalnej języka, będącej określeniem zakresu jej funkcjonowania. Wiele słów kluczowych, pełniących funkcję promującą, może budzić zastrzeżenia językoznawców co do adekwatności, poprawności językowej oraz łatwości zapamiętania. Słowa kluczowe we wspomnianej funkcji pełnią również funkcję wyszukiwawczą, opartą na tym samym mechanizmie kojarzeniowym co języki swobodnych słów kluczowych. Stąd na słowa kluczowe, jako na

słowa czy wyrażenia” (Łobacz, Mikołajczak-Matyja 2002, s. 117). Tego typu motywacja odnosi się do morfologii wyrazu.

²¹ Część słownictwa kluczowego może mieć nacechowanie emocjonalne pozytywne lub negatywne bądź obojętne. Badania psycholingwistyczne ujawniły, że wykorzystywanie/używanie wyrazów o niskiej częstotliwości występowania w języku jest związane z dłuższym czasem dostępu do leksykonu umysłowego. A więc słowa kluczowe w języku słów kluczowych i języku marketingowym różnią się rodzajem nacechowania. W językach słów kluczowych jest to nacechowanie obojętne, zaś w języku marketingowym – emocjonalne. Słowa kluczowe mogą być również rodzajem komunikatu marketingowego. (Zanikowi) funkcji informacyjnej nazwy towarzyszy zwykle nasilenie jej funkcji ekspresyjnej. W języku marketingowym słowo kluczowe jest komunikatem informacyjno-perswazyjnym. Teksty w językach słów kluczowych mają charakter wyłącznie komunikatów informacyjnych, ewentualnie metainformacyjnych. Teksty reklamowe i nazwy odwołują się najczęściej do skojarzeń sytuacyjnych i werbalnych. Istnieją też skojarzenia formalne i skojarzenia semantyczne. „Jeżeli nazwa odczuwana jest jako obca, to fakt rzadkiego używania (w codziennej komunikacji) tworzących ją wyrazów przez polskiego użytkownika zmniejsza możliwość budowania licznych skojarzeń i konotacji opartych na ich związku z pierwotnymi desygnatami, co może z kolei wzmacniać siłę związku między nazwą a desygnatem (...). Nazwy tego typu mogą zatem zyskiwać wtórnie dodatkową siłę informacyjną, polegającą na szybkim wywoływaniu przez nie reakcji związanej z zespołem, a nie z pierwotnymi desygnatami tworzących je wyrazów” (Łobacz, Mikołajczak-Matyja 2002, s. 110).

obszar specyficznego słownictwa, trzeba także spojrzeć z punktu widzenia polityki komunikowania się (na przykład przedsiębiorstwa z rynkiem), a także psychologii stosowanej (w zakresie promocji i manipulacji), językoznawstwa (z punktu widzenia poprawności, czytelności nazwy, i co za tym idzie – jej siły perswazyjnej) oraz nauk o kulturze (moda na określony typ nazw, odniesień do symboli kulturowych). Tym językiem nie będę się zajmować, zasygnalizowałem tylko jego istnienie.

1.2. Słowa kluczowe w systemie WWW

Słowa kluczowe są podstawowym narzędziem leksykalnym w wyszukiwaniu informacji w Internecie. Dzięki szerokiemu stosowaniu przeżywają prawdziwy renesans²². Ich wykorzystywanie w Internecie obszernie omawia w piśmiennictwie zachodnim Gobinda G. Chowdhury (Chowdhury 1999). Jej uwagi uwzględnię w trakcie dalszych rozważań²³.

1.2.1. Słowa kluczowe w wyszukiwarkach internetowych

W systemie WWW słowo kluczowe to „słowo lub wyrażenie umieszczane w kodzie HTML strony internetowej w celu jego umieszczenia/zaindeksowania w bazie danych wyszukiwarki internetowej w odpowiedniej kategorii tematycznej”²⁴. Nie ma więc wątpliwości, że używane w Internecie słowa kluczowe to jednostki leksykalne języka naturalnego, służącego do opisu (funkcja metainformacyjna) i wyszukiwania informacji (funkcja wyszukiwawcza).

Oprócz zdefiniowania, czym jest słowo kluczowe, ważne jest też porównanie słowa kluczowego z wyrażeniami Webu. Element ten jest istotny, zwłaszcza w HEAD, niewidzialnej części języka HTML, a także wśród pozostałych elementów strony internetowej. Czasem jednak słowa kluczowe bywają źle dobrane, gdyż autorzy stron nie muszą być profesjonalnymi indeksatorami. Generatorami najczęściej używanych przez internautów słów kluczowych są też wyszukiwarki i multiwyszukiwarki. Najbardziej popularne słowa kluczowe w postaci list słów kluczowych prezentują takie wyszukiwarki jak: Altavista, HotBot, Lycos, Sprinks, Yahoo czy Google.

Prosta gramatyka i otwarta postać słownika języka słów kluczowych idealnie pasują do struktury, jaką tworzy światowa pajęczyna. Ideą Internetu jest nieustanny wzrost we wszystkich kierunkach i umożliwienie każdemu dołożenie do niego własnej części. Nie chodzi tu tylko o kwestię rozrostu fizycznego, lecz o zawartość treściową

²² Badania Marcina Roszkowskiego (2008) ujawniły, że ponad 50% (16/30) przebadanych przez niego dziedzinowych systemów hipertekstowych wykorzystuje kontrolowane i niekontrolowane słowa kluczowe, przy czym ich proporcje są następujące: kontrolowane słowa kluczowe (7/16), zaś niekontrolowane słowa kluczowe (9/16).

²³ To teoretyczne studium jest na tyle ważne, że nie można go nie zauważyć, omawiając stosowanie słów kluczowych w wyszukiwarkach internetowych, zwłaszcza w systemie WWW (tzw. *Keyword Search*).

²⁴ Por. http://www.futuro.pl/pomoc_slowniczek.php/haslo/49.

sieci. Wszyscy mogą być zarówno autorami, jak i odbiorcami dokumentów publikowanych w Internecie.

Język słów kluczowych, przy otwartej strukturze słownika, pozwala na dokładanie wciąż nowych wyrażenń czerpanych ze stale rosnącego zasobu słownictwa języka naturalnego oraz najbogatszego ich źródła – właśnie sieci. Stąd w języku słów kluczowych indeksowanie dokumentów nawet w takich językach jak arabski czy japoński nie stanowi trudności, ponieważ wyrazy z tych języków są tak samo swobodnie wprowadzane do słownika, jak słowa w języku polskim czy niemieckim.

Aby dane wyrażenie mogło stać się słowem kluczowym, powinno być (powtórzę) z metainformacyjnego punktu widzenia słowem znaczącym. Opracowanie leksykograficzne słów kluczowych (ujednoczenie ich form gramatycznych) jest ważne w przypadku języków fleksyjnych, a więc także języka polskiego. Nie ma natomiast większego znaczenia dla języków naturalnych bez fleksji, na przykład języka angielskiego.

Dostępne w sieci zbiory słów kluczowych mają postać:

- słowników tradycyjnych w statycznym formacie tekstowym;
- słowników w formacie HTML, ale nadal statycznych, bez aktywnych hiperłączy;
- słowników w formacie HTML dynamicznych, z pełnymi możliwościami nawigowania za pomocą hiperłączy;
- słowników z zaawansowanym interfejsem graficznym i wizualną prezentacją informacji;
- słowników w formacie XML.

Niektóre z nich stanowią integralny element systemu informacyjnego, inne mają charakter słowników samoistnych, dostępnych w sieci i możliwych do wykorzystania.

Podstawą doboru słów kluczowych jest wiedza na temat tego, czego szukają internauci i w jaki sposób to czynią, tj. w jaki sposób wpisują pytania w wyszukiwarkach. Najczęściej polega to na wpisywaniu słów i fraz kluczowych, które internautom przychodzą na myśl (skojarzenia mentalne). Słowa kluczowe w sieci, oprócz ich funkcji oznaczania/identyfikacyjnej, często pełnią – co sygnalizowałem już wcześniej – funkcję promocyjną. W tej drugiej funkcji są jednym z narzędzi promowania serwisu internetowego, produktów i usług, na przykład bankowych. Dobór słów kluczowych ma na celu dotarcie do właściwej grupy docelowej i konwersję sprzedaży odwiedzających witrynę internetową. W tym celu przy doborze słów kluczowych i fraz kluczowych brane są pod uwagę następujące ich cechy:

- ogólne słowa kluczowe dają mniejsze prawdopodobieństwo zdobycia klienta;
- specyficzne, specjalistyczne, niszowe słowa kluczowe generują mniejszy ruch internetowy;
- ogólne słowa kluczowe zwiększają koszty pozycjonowania i reklamy; lepsze efekty daje wykorzystywanie tzw. długiego ogona słów kluczowych.

Stworzono już wiele narzędzi przydatnych przy szukaniu słów kluczowych na potrzeby promocji i reklamy w sponsorowanych linkach. Należą do nich m.in.:

- Google Keyword External Tool – ujawnia konkurencyjność słowa kluczowego i częstotliwość jego wyszukiwania;
- KeywordDiscovery.com, Keyword Research Tool – narzędzia pobierania słów kluczowych z najważniejszych światowych wyszukiwarek;

– Google Trends – pokazuje częstotliwość wyszukiwań słów kluczowych w wyszukiwarce Google z uwzględnieniem kryterium geograficznego.

Typowanie kluczowych słów i/lub fraz, którymi posługują się użytkownicy Internetu w czasie wyszukiwania informacji, jest jednym z elementów pozycjonowania. Obecnie na rynku funkcjonuje już szereg programów, jak AdWords, Analytics czy Pay Per Click Google'a, które oferują niezbędne informacje o tym, które słowa kluczowe są skuteczne, a które nie.

Najpopularniejszym formatem dokumentów występujących w sieci jest HTML. HTML jest językiem, w którego znacznikach umieszczana jest zawartość treściowa dokumentu i który organizuje jego stronę formalną. Pozwala to na umieszczenie w jego strukturze różnej formy dokumentów, czyli grafiki, tekstu, dźwięku i połączenie ich w jeden multimedialny dokument. Kolejną ważną zaletą jest hipertekst, czyli możliwość przenoszenia się poprzez odnośniki do innych dokumentów. Dokumenty pisane w języku HTML przyciągają uwagę użytkownika bogatą i ciekawą oprawą graficzną, co powoduje, że wynik oglądany na ekranie monitora w postaci już gotowej strony znacznie różni się od jego strony źródłowej. Język HTML jest metajęzykiem pozwalającym opisać również inne rodzaje dokumentów, ustawić poszczególne fragmenty w zgodnym z wolą webmastera porządku, a co najważniejsze – umożliwić użytkownikowi wyszukanie dokumentu poprzez umieszczenie w nagłówku słów kluczowych.

Słowa kluczowe są bardzo ważnym elementem systemu WWW jako całości oraz jego poszczególnych elementów zwanych stronami (Babik 2007). Oprócz słów kluczowych czasami stosuje się tzw. frazy językowe. Istnieją też systemy, które każde słowo z tekstu, tytułu lub streszczenia traktują jako słowo kluczowe. System WWW, a w zasadzie jego nieustrukturalizowana i rozproszona baza danych, składa się z dokumentów utworzonych za pomocą języka HTML. Dokumenty te nazywa się stronami internetowymi. Strony te na ekranie komputera są efektem interpretacji dokumentu (znajdującego się na serwerze WWW) przez odpowiednią przeglądarkę zainstalowaną na komputerze użytkownika. Strony różnią się między sobą zawartością, typem (tekst, obraz, dźwięk, wideo), formatem (tekst, HTML, PDF, postscript) oraz językiem. Z informacyjnego punktu widzenia ważne są nie tylko informacje o zawartości strony, lecz także informacje o samych stronach, to jest o ich pochodzeniu (autorstwie), częstotliwości aktualizacji, częstotliwości cytowania.

Strona WWW w formacie HTML składa się zwykle z właściwego tekstu dokumentu, zwanego BODY oraz części zwanej nagłówkiem (HEAD). Nagłówek strony zawiera wiele różnych rodzajów informacji, a właściwie metainformacji. Dotyczą one: sposobu zakodowania polskich znaków diakrytycznych, języka, autora, tytułu, zawartości strony, słów kluczowych. Istotne z punktu widzenia wyszukiwania informacji są przede wszystkim te części strony, które poprzedzają metaznaczniki: TITLE, DESCRIPTION oraz KEYWORDS. Umieszczone tam informacje są traktowane przez wyszukiwarki jako ważniejsze od innych, co przeważnie ma wpływ na kolejność cytowania (ranking) informacji wyświetlanych w odpowiedzi na pytanie użytkownika. Źródło słów kluczowych stanowią następujące pola: <HEAD>

<BODY>, <META NAME>, <DESCRIPTION>, <KEYWORDS>²⁵. Strony są indeksowane przede wszystkim na podstawie zawartości pola KEYWORDS. Jeżeli twórca strony nie przyporządkował jej słów kluczowych lub w ogóle nie sformułował nagłówka HEAD (nie jest on obowiązkowym elementem strony), wówczas strona jest indeksowana na podstawie tekstu zawartego w części BODY. Tekst ten jest indeksowany i wyszukiwany za pomocą swobodnych wyrażen, derywowanych z tekstu języka naturalnego (ang. *free-text searching*), ze wszystkimi mankamentami tego typu indeksowania. Szczegółowe omówienie tego zagadnienia, jakkolwiek jest bardzo interesujące, wykracza jednak poza ramy niniejszej pracy.

Jak zauważyła J. Woźniak-Kasperek: „(...) każdej stronie jest przypisywana jedna i ta sama wartość, którą przez analogię można by nazwać wagą. Wartość wagi jest następnie dzielona przez liczbę słów kluczowych przypisanych stronie przez jej twórcę lub ustanowionych na mocy indeksowania swobodnego sekcji BODY. Wynikałoby stąd, że im większa jest liczba słów kluczowych, tym mniejsza ich waga, co w konsekwencji powoduje dalszą pozycję danej strony na liście rankingowej” (Woźniak-Kasperek 2004, s. 32). Z punktu widzenia użytkownika najważniejsza jest jednak relewantność użytych słów i tekstów samych dokumentów HTML. Wyniki obserwacji pokazują, że autorzy przyporządkowują dokumentom wiele słów kluczowych o wątpliwej relewancji i małej (ze względu na liczbę słów) wadze wyszukiwawczej. Zdarza się też, że słowa kluczowe są tak dobierane, aby jak największa liczba osób odwiedziła daną stronę. Sprawia to, że relewantność słów kluczowych staje się drugoplanowa, a w niektórych wypadkach nawet marginalna.

Serwery wyszukujące indeksują zasoby sieci przeważnie na podstawie zawartości pola <KEYWORDS> dokumentu HTML systemu WWW lub za pomocą swobodnych wyrażen derywowanych z tekstu języka naturalnego w części <BODY>²⁶. Wszystkie wyszukiwarki deklarują, że biorą pod uwagę całą zawartość <BODY> strony WWW, omijając jedynie stop-słowa, dla których tworzą specjalne listy. Te tzw. stop-listy zawierają słowa najczęściej występujące w danym języku, przez co mające małą wartość informacyjną.

Do dostarczania słów kluczowych wyszukiwarce używa się przede wszystkim META-kodów, które z różnych powodów nie wystąpiły w tekście dokumentu. Przykładem może być ich zastosowanie na splash i frame stronach. Splash strony to witryny wejściowe, na których znajduje się logo, często z napisem „Enter”, i które kierują do właściwego (nierzadko zupełnie innego) adresu. Z kolei frame strony określają formę, w jakiej przedstawiona zostanie witryna w oknie przeglądarki. Wprowadzenie znaczników META często powoduje skok danej strony na pierwsze miejsca w rankin-

²⁵ Dokładnie słowa kluczowe znajdują się w tytule strony <TITLE>, metatagach, tj. w opisie strony <DESCRIPTION>, polu słów kluczowych <KEYWORDS>, treści strony <CONTENT>, nagłówkach h1-h6, między tagami i , anchorach, opisach obrazków <ALT>, tytułach linków <TITLE> i w domenie.

²⁶ Warto odnotować, że w systemach, w których wykorzystuje się mechanizmy automatycznego indeksowania, tj. automatycznej charakterystyki treściowej, tego typu opisy pełnią jedynie funkcję uzupełniającą, a wyrażenia wygenerowane w ten sposób są graficznie wyróżniane jako pochodzące z nadania automatycznego.

tation) on "Generating Information Retrieval Languages from Terminological Dictionaries," Faculty of Neophilology of the University of Warsaw, 1996.

[...]

<hr size = + 2 noshade width = 30% align = center >

<center >

e-mail: wbabik@inib.uj.edu.pl </center >

<hr size = + 2 noshade width = 30% align = center > </BODY >

</HTML >

</head >

Rys. 2. Przykład opisu dokumentu w języku HTML

Niewątpliwie słowa kluczowe są wyrazami, frazami lub tzw. quasi-zdaniami języka naturalnego, które internauci wpisują w wyszukiwarkach internetowych w celu znalezienia tego, co ich interesuje, dlatego bardzo ważny jest ich dobór oraz odpowiednie zagęszczenie (ang. *keyword density*). W praktyce większość współczesnych wyszukiwarek ignoruje zawartość pól META, a jeśli je wykorzystuje, to zawsze konfrontuje ze słowami kluczowymi zidentyfikowanymi w całym dokumencie, ewentualnie w Internecie. Do tego zagadnienia jeszcze powrócę w rozdziale 3.2.2, ale w nieco innym aspekcie.

Charakterystyczną cechą słów kluczowych w Internecie, a także w ogóle, jest ich nieostrość oraz różnorodność kryteriów derywacji. Pozwala to na tworzenie wielu punktów dostępu w procesie indeksowania informacji, co obecnie uważa się za źródło siły wyszukiwawczej Internetu, chociaż często prowadzi do szumu lub ciszy informacyjnej. Mimo, często uzasadnionej, krytyki dotychczasowych sposobów przeszukiwania zasobów Internetu, słowa kluczowe są najbliższe sposobom myślenia użytkowników i wyszukiwania przez nich informacji. Zwykle wykorzystywana jest metoda organizacji i przeszukiwania zasobów tego systemu oparta na indeksowaniu i porządkowaniu alfabetycznym. Wyszukiwanie za pomocą słów kluczowych umożliwiają obecnie wszystkie wyszukiwarki. Są efektywnymi narzędziami wyszukiwania informacji według precyzyjnie wskazanych w opisie nazw przedmiotów lub zagadnień, ale często generują odpowiedzi zawierające znaczną liczbę informacji dla użytkownika nierелеwantnej albo o niskiej relewancji.

W systemie WWW używane są przeważnie wyrażenia stanowiące słowa kluczowe, wygenerowane z języka naturalnego na podstawie relacji kojarzeniowych. Ich zbiór odwzorowuje/zawiera język swobodnych słów kluczowych. Język ten stanowi część kompetencji użytkownika w zakresie języka naturalnego, kompetencji leksykalnej²⁷ i kompetencji semantycznej²⁸.

²⁷ Kompetencja leksykalna jest częścią kompetencji językowej pozwalającej użytkownikowi danego języka sprawnie posługiwać się systemem leksykalnym danego języka.

²⁸ Kompetencja semantyczna jest rodzajem kompetencji językowej pozwalającej użytkownikowi danego języka kodować i dekodować (rozpoznawać) informację.

Słowa kluczowe w Internecie tworzą zasób leksykalny języka indeksatora i użytkownika swobodnych słów kluczowych. Pochodzą z tekstu indeksowanego dokumentu lub są wybierane przez indeksatora z jego własnego słownika mentalnego. Nie zawsze *explicite* prezentują system języka, a więc słownik i gramatykę, pozwalającą generować zdania i teksty spójne. Nie są podawane *explicite* reguły semantyki pozwalające na przyporządkowanie wyrażenia języka elementom odwzorowywanej rzeczywistości pozajęzykowej oraz reguły pragmatyki określające reguły poprawnego użycia języka, tj. reguły tworzenia komunikatów w danej sytuacji komunikacyjnej.

Dąży się obecnie do tego, aby systemy leksykalno-semantyczne języków słów kluczowych miały jak najbardziej naturalny charakter, tzn. aby były zbliżone w jak największym stopniu do sposobów myślenia i wyszukiwania informacji przez człowieka, czym zajmujemy się w dalszej części rozprawy. Nieograniczony dostęp do Internetu sprawia, że owa naturalność musi być rozumiana umownie, nie ma bowiem jednego „naturalnego” obrazu świata wszystkich użytkowników systemu. Zatem nie w naturalności, ani w poprawności systemu leksykalno-semantycznego tkwi moc wyszukiwawcza języków słów kluczowych, lecz w gotowości użytkownika systemu stosującego ten typ języka do akceptacji proponowanych słów kluczowych. Są one jednocześnie leksykalnym narzędziem kategoryzacji informacji, za pomocą którego można odnaleźć potrzebną informację. Słowa kluczowe, tworzące charakterystykę słowną dokumentu, pełnią jednocześnie funkcje języków informacyjno-wyszukiwawczych, tj. funkcję metainformacyjną i funkcję wyszukiwawczą. Relewantność i siła semantyczna (waga wyszukiwawcza) używanych w indeksowaniu słów kluczowych mają decydujące znaczenie z punktu widzenia efektywności wyszukiwania. Słowa kluczowe stanowią raczej narzędzie tradycyjne – w porównaniu z nowoczesnymi narzędziami – semantycznej organizacji zasobów Internetu.

Mimo rozwoju Internetu świat Webu nie podąża za osiągnięciami nauki o informacji. Dla większości internautów nie ma różnicy między słowem kluczowym jako wyrażeniem reprezentującym dany dokument a słowem poszukiwanym. Również niewielu użytkowników zwraca uwagę na analizy dokumentacyjne, w których znajdują się słowa kluczowe. W nauce o informacji słowa kluczowe to słowa charakteryzujące zawartość dokumentu, służące przede wszystkim jako klucze wyszukiwawcze w systemach informacyjno-wyszukiwawczych.

1.2.2. Słowa kluczowe w systemach hipertekstowych i hipermedialnych

Słowa kluczowe są stosowane również w systemach wykorzystujących język hipertekstu. Obecnie cecha ta jest jedną z konstytutywnych cech systemu WWW, które przez to stały się systemami hipertekstowymi i hipermedialnymi. Ba, aktualnie wszystkie systemy informacyjno-wyszukiwawcze są systemami hipertekstowymi. Koncepcja organizacji informacji w środowisku hipertekstowym zakłada strukturalizację zbioru informacji poddawanego organizowaniu z wykorzystaniem elementów sieci, które są konstytutywne dla hipertekstu, czyli węzłów i powiązań hiperteksto-

wych (Rada 1991). „Podstawowa idea hipertekstu sprowadza się do opracowania metod strukturalizacji i prezentacji informacji, zawartej w komunikatach językowych (dokumentach), które zapewniają wzbogacenie jej linearnej organizacji, charakterystycznej dla tekstów tradycyjnych dokumentów piśmienniczych” (Sosińska-Kalata 1999, s. 282). Hipertekst pozwala na komputerowe naśladowanie zdolności ludzkiego mózgu do szybszego przyswajania informacji i intuicyjnego wykorzystania odnośników oraz odsyłaczy²⁹. Jest to więc taki sposób zapisywania tekstu w systemie komputerowym, w którym, dzięki odpowiedniemu zakodowaniu informacji, dokonano powiązania wszystkich ważniejszych pojęć relewantnych. Ten sposób organizacji informacji zapewnia elastyczność wyszukiwania informacji i stosowanie w nim indywidualnych strategii określanych na bieżąco przez danego użytkownika.

Systemy hipertekstowe umożliwiają wyszukiwanie informacji nie tylko w zbiorze informacji dokumentacyjnych, lecz także w zbiorze tekstów dzięki wykorzystaniu odpowiednich relacji intra- i intertekstualnych, tworzących sieć łączącą teksty i/lub ich fragmenty powiązane z sobą określonymi relacjami semantycznymi.

Słowa kluczowe w systemach hipertekstowych stanowią narzędzie organizacji i reprezentacji struktury semantycznej wiedzy dziedzinowej, ujawnionej w semantycznej strukturze systemu leksykalnego języka informacyjno-wyszukiwawczego lub zbiorze wyrażen naturalnych, derywowanych z tekstów dokumentów, ich tytułów lub abstraktów. Znaczenie i forma jednostek leksykalnych w tych systemach są reprezentowane przez węzły sieci. Rodzaj relacji paradygmatycznych wiążących określone jednostki określają linki między węzłami. W najprostszym przypadku węzły sieci semantycznej reprezentują jednostki leksykalne, a łączącymi je relacjami są związki hierarchiczne, ewentualnie rozszerzone o relacje kojarzeniowe, specyfikowane w strukturze paradygmatycznej języków informacyjno-wyszukiwawczych. Taki sposób wykorzystania języka słów kluczowych ilustruje opisany przez Barbarę Sosińską-Kalatę system ALLOY (Sosińska-Kalata 1999). Jest to prototypowy system automatycznego indeksowania i wyszukiwania dokumentów oparty na programie typu Shell i przeznaczony do współpracy z bazami danych z różnych dziedzin.

Sygnalizowany system automatycznego generowania opisów dokumentów w postaci list słów kluczowych opiera się na systemie przetwarzania języka naturalnego. Związki między słowami kluczowymi są reprezentowane za pomocą sieci. Język słów kluczowych zapewnia łatwą nawigację po strukturze systemu pojęciowego wiedzy dziedzinowego systemu i selekcję leksyki języka słów kluczowych wyrażającej określone pojęcia. System ALLOY korzysta ze słownika słów kluczowych oraz z sieci semantycznej wyrażanych przez nie pojęć. Struktura systemu pojęciowego dziedziny jest odwzorowywana w języku słów kluczowych. „ALLOY automatycznie łączy węzeł sieci semantycznej reprezentujący określone pojęcie z listą słów i fraz kluczowych wyrażających to pojęcie. Następnie słowa i frazy kluczowe, związane z określonym węzłem sieci semantycznej, są wykorzystywane do generowania syste-

²⁹ W językach informacyjno-wyszukiwawczych termin „odsylacz” bywa używany co najmniej w czterech następujących znaczeniach: jako element, do którego się odsyła; element, od którego się odsyła; wyrażenie symbolizujące relacje odsyłające; wszystkie wskazane elementy łącznie.

mowych reprezentacji dokumentów. Po zidentyfikowaniu wyrażeń reprezentujących istotne pojęcia indeksowanego tekstu, do jego opisu są wprowadzane wszystkie wyrażenia wskazane w węzłach tych pojęć. Poza wskazaniem związków danego pojęcia z innymi pojęciami, każdy węzeł zawiera wyjaśnienie reprezentowanego w nim pojęcia, informacje o wyrażających je słowach kluczowych oraz odesłania do dokumentów, w których charakterystykach słowa te zostały umieszczone. Wyszukiwanie informacji o dokumentach zawierających pojęcia istotne dla wiedzy użytkownika o problemie wyszukiwawczym jest możliwe poprzez przeglądanie samej sieci. Wybierając opcję przeglądania wiedzy zawartej w węźle, użytkownik najpierw otrzymuje dane o reprezentowanym pojęciu i jego związkach semantycznych, następnie może wyświetlić informacje szczegółowe o dokumentach, w których charakterystykach występuje słowo kluczowe reprezentujące to pojęcie, a także przejść do przeglądania informacji o słowach kluczowych związanych z węzłem i zażądać wyświetlenia informacji o dokumentach. W ten sposób wyszukiwanie informacji przyjmuje postać swego rodzaju wędrówki po bazie danych, nawigowanej poprzez relacje semantyczne reprezentowane w sieci semantycznej słów kluczowych” (Sosińska-Kalata 1999, s. 225–226).

Hipertekstowe internetowe systemy wyszukiwania informacji swoje pole informacyjne strukturalizują poprzez pole semantyczne tworzone przez słowa kluczowe wyszukiwarek połączone gęstą siecią odsyłaczy (tzw. linków). Linki te odwzorowują i ustanawiają relacje organizujące poszczególne pola w sposób umożliwiający użytkownikowi-internaucie wyszukiwanie w zbiorze informacyjnym potrzebnych mu informacji. Sposób i stopień organizacji takiego pola informacyjnego, dokumentacyjnego, tematycznego, semantycznego i leksykalnego³⁰ bezpośrednio wpływa na potencjalne możliwości strategii wyszukiwawczych i ich efektywność.

Systemy hipertekstowe i hipermedialne często mają charakter autorski. Każdy węzeł jest specjalnie opracowany. Nie mają więc regularnej struktury, jak systemy baz danych. W związku z tym trudno tu mówić o zapytaniach takich jak w bazie danych. Podstawowymi metodami dostępu do danych jest swobodne przeglądanie oraz wyszukiwanie przez słowa kluczowe. Aby zapobiec zagubieniu się użytkownika, swobodne przeglądanie bywa ograniczane gotowym zestawem ścieżek przygotowanych przez autora (ang. *guided tour*) oraz przez gotowe, przeglądowe diagramy.

Z dotychczasowych rozważań widać, jak ważne są słowa kluczowe w organizacji informacji w ludzkim umyśle, dlatego przedmiotem dalszych analiz uczyniłem procesy mentalne, ze szczególnym zwróceniem uwagi na słownik mentalny człowieka.

³⁰ Z punktu widzenia systemu informacyjno-wyszukiwawczego i jego użytkownika istnieje równoznaczność i izomorfizm pomiędzy polem informacyjnym, polem dokumentacyjnym i polem tematycznym, bowiem zakres tych pól i ich struktura są wyznaczone przez język systemu. Ich struktura jest izomorficzna ze strukturą pola semantycznego języka systemu (por. Bojar 2004).

1.3. Psycholingwistyczne determinanty słów kluczowych

W literaturze poświęconej zarówno metodologii opracowania rzeczowego dokumentów, jak i projektowaniu systemów i języków informacyjno-wyszukiwawczych można spotkać opinie, że struktury organizujące wiedzę w systemach dokumentacyjnych w ograniczonym stopniu korespondują ze sposobem, w jaki ludzie z nich korzystający porządkują wiedzę w swoich umysłach. Dotyczy to również metod i narzędzi organizacji wiedzy, które czasami zawodzą jako instrumenty skutecznego odwzorowywania fragmentów wiedzy publicznej, o których informacji poszukują użytkownicy. Także alfabetyczne wykazy słów kluczowych, z uwagi na formalne kryterium organizacji haseł słownikowych, stanowią w tym względzie znikomą pomoc. Również i słowniki języków informacyjnych, których organizacja jest oparta na kryterium semantycznym, są niewystarczające. Skuteczną metodą pozyskiwania informacji niezbędnej do adekwatnego formułowania problemu wyszukiwawczego jest korzystanie ze źródeł pozasystemowych, takich jak encyklopedie, słowniki terminologiczne, a zwłaszcza specjalistyczna wiedza użytkownika (por. Vickery, Vickery 1987; Saracević 1991; Gödert 1996).

Sposób definiowania pojęć i organizacji ich systemu, za pomocą leksyki, struktur paradygmatycznych i syntagmatycznych języków informacyjno-wyszukiwawczych, w opisach rzeczowych dokumentów, stanowiących podstawę wyszukiwania, często nie jest zgodny ze sposobem, w jaki ludzie porządkują wiedzę w pamięci. Użytkownicy często uznają za najistotniejsze dla problemu wyszukiwawczego inne kategorie pojęć, niż te, które wskazywane są w opisie tematyki dokumentów zgodnie z tradycyjnymi zasadami indeksowania i klasyfikowania. Warunkiem zgodności pojęć istotnych dla fragmentów wiedzy przedstawionych w dokumentach oraz pojęć istotnych dla problemu wyszukiwawczego użytkownika jest stosowanie analogicznych kryteriów relewancji (Bean, Green 2001). Teorie cech semantycznych, podobnie jak teorie sieci semantycznych, stanowią podstawę modeli organizacji wiedzy semantycznej, którą zasadniczo można utożsamiać z organizacją słownika umysłowego na poziomie strukturalnych cech znaczenia leksykalnego.

W kontekście krytycznych uwag o niezgodności organizacji wiedzy, wyrażanej w strukturach dziedzicznych systemów informacyjno-wyszukiwawczych i w strukturach kognitywnych organizacji wiedzy w pamięci człowieka, uważam za zasadne rozważenie, w jakim stopniu koncepcje organizacji wiedzy w ludzkiej pamięci, sformułowane przez psychologię, znajdują lub mogą znaleźć zastosowanie w językach słów kluczowych.

Do tej pory autorzy zajmujący się w Polsce słowami kluczowymi, w swoich wypowiedziach koncentrują się raczej na aspektach technologicznych niż mentalnych, a więc odwrotnie niż na Zachodzie. Podjęte rozważania próbuję zatem połączyć z poznawczym nurtem badań informacyjnych i w dalszych rozważaniach skupię się na tej sferze.

Za przydatne w określeniu i wyjaśnieniu psycholingwistycznych determinant języka słów kluczowych uważam modele pamięci semantycznej, zwane również

modelami mentalnymi. Ich prezentacja pozwoli nie tylko wyjaśnić pewne zjawiska językowe dotyczące słów kluczowych, na przykład w systemach hipertekstowych, ale i uzasadnić potrzebę ujęcia sieciowego języków słów kluczowych i większej zbieżności modeli realizowanych za pomocą języka słów kluczowych ze strukturami kognitywnymi człowieka jako postulowanego kierunku „rozwoju” tych języków.

1.3.1. Modele pamięci semantycznej

W piśmiennictwie naukowym powszechnie przyjmuje się, że człowiek magazynuje informacje i wiedzę w formie odpowiednich struktur. Rezultatem psycholingwistycznych badań, odkrywających strukturę leksykalno-semantycznych aspektów języka, są ustalenia dotyczące budowy pamięci semantycznej. Piśmiennictwo psychologiczne i psycholingwistyczne zawiera opisy wielu modeli pamięci semantycznej, dlatego dalsze rozważania zostaną ograniczone do analizy i dyskusji nad modelami przydatnymi z punktu widzenia potrzeb tego opracowania.

Termin „pamięć semantyczna” istnieje od lat sześćdziesiątych XX wieku. Oznacza wiedzę o świecie, która stanowi istotny element magazynu pamięci długotrwałej. Tak rozumiana pamięć semantyczna jest przedmiotem zainteresowania także psychologii, ale nie tylko poznawczej (kognitywnej). Jednostkami tej części magazynu pamięci są znaczenia wyrazów lub części – pojęcia (np. Quillian 1968).

Pierwowzorem wielu modeli pamięci semantycznej jest sieć powiązań między poszczególnymi jednostkami, odkrywanych w badaniach nad skojarzeniami swobodnymi. Sieci o charakterze czysto skojarzeniowym odzwierciedlają jedynie częstość wzajemnego aktywizowania wyrazów i nie stanowią klasycznych struktur o charakterze hierarchicznym. Opierając się na interpretacjach B. Sosińskiej-Kalaty (Sosińska-Kalata 1999, s. 77–84) oraz Lynn Westbrook (Westbrook 2006), przyjmuję, że teorie sieci semantycznych³¹ stanowią najwcześniejszą koncepcję organizacji pamięci semantycznej. Powstały na przełomie lat sześćdziesiątych i siedemdziesiątych XX wieku. Najbardziej znane są modele opracowane przez M. Rossa Quilliana, Allana M. Collinsa i M. Rossa Quilliana, Johna R. Andersona i Gordona H. Bowera. Modele te wyrosły na bazie asocjacionizmu, przyjmującego, że cała wiedza człowieka podlega strukturalizacji za pomocą związków kojarzeniowych między pojęciami. W podobny sposób definiowane jest znaczenie w lingwistycznych teoriach pola semantycznego (D. Buttler, W. Miodunka, R. Tokarski). Sieciowa organizacja wiedzy językowej w pamięci uważana jest za fundament mechanizmu organizacji wiedzy w umyśle człowieka. Jednostkami takiej wiedzy są m.in. pojęcia, traktowane jako znaczenia wyrażen języka naturalnego, z którymi mogą być powiązane zarówno mentalne reprezentacje ich form wyrażeniowych (na przykład nazw), jak i mentalne wyobrażenia ich desygnatów. Dotyczy to tylko kompetencji (wiedzy) językowej, a nie wiedzy w ogóle. Próbe wyjaśnienia procesów myślenia za pomocą praw kojarzeniowych przez styczność czasową i przestrzenną stanowi empiryczna filozofia Davida Hume’a.

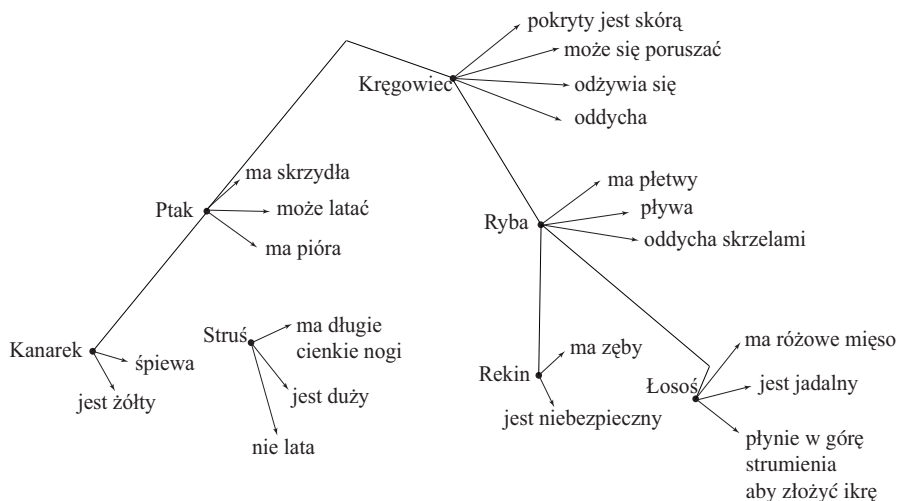
³¹ Termin „sieć semantyczna” (ang. *semantic network*) został użyty po raz pierwszy w 1961 roku. Stosuje się także określenie „semantyczna skojarzeniowa sieć pamięci” (Sowa 1991).

Do koncepcji sieci skojarzeniowych nawiązują teorie sieciowe w ujęciu kognitywnym, ale różnią się od nich interpretacją relacji konstytuujących sieć. Przyjmują, że związki między pojęciami są ukierunkowane i semantycznie wyspecyfikowane. Pogląd, że struktura wiedzy człowieka ma postać asocjacyjnej sieci semantycznej, należy do najpowszechniejszych, stąd z reguły nawiązują do niego intuicyjne modele organizacji wiedzy użytkowników systemu informacyjno-wyszukiwawczego, leżące u podstaw tradycyjnych koncepcji organizacji pojęć odwzorowywanej w semantycznej strukturze leksyki języka informacyjno-wyszukiwawczego. Nie są to jednak poglądy powszechnie przyjęte w psychologii.

Pojęcia są w pewien sposób związane z językiem. Komunikacja językowa opiera się na wspólnej wiedzy pojęciowej uczestników dyskursu. Niektórym słowom odpowiadają jednostkowe pojęcia (tzw. pojęcia leksykalne), inne pojęcia mogą być wyrażane złożonymi konstrukcjami językowymi. Przekaz językowy stanowi przekaz wiedzy pojęciowej, uzupełniający w stosunku do doświadczenia. Przekaz za pomocą słów kluczowych stanowi również przekaz wiedzy pojęciowej. Zdaniem wielu psychologów, filozofów i lingwistów systemy językowy i pojęciowy nie mogą być rozdzielone. Klasyczne ujęcie tego twierdzenia jest znane jako hipoteza Sapira-Whorfa.

Wczesne metody badania struktury pamięci semantycznej polegały na traktowaniu pamięci człowieka jako hierarchicznej struktury pojęć. „Pojęcie jest podstawową strukturą umysłowo-poznawczą stanowiącą uogólniony model jakiejś klasy przedmiotów (zdarzeń), cech lub relacji podobnych do siebie pod pewnym względem” (Chlewiński 1991, s. 21). Stanowią one określony sposób porządkowania rzeczywistości. Dzięki pojęciom możliwa jest percepcja przedmiotów (zdarzeń) jako reprezentantów różnych klas i uwzględnianie w nich przede wszystkim tych cech, które są wspólne danym klasom, z pominięciem ich cech konkretnych, jednostkowych. Podstawowym mechanizmem przyswajania pojęć jest poznawanie nowych słów w różnych kontekstach werbalno-pojęciowych. Definicje służą przede wszystkim uściśleniu zakresu przyswojonych pojęć, a nie wytwarzaniu pojęć zupełnie nowych. W języku naturalnym – jeżeli nazwę rozumie się jako formę – za jedną nazwą używaną w różnych kontekstach często kryje się wiele różnych pojęć. Identyfikacja pojęć zachodzi pod wpływem procesu różnicowania i procesu generalizacji (Chlewiński 1991). Pojęcia, reprezentowane przez oddzielne węzły, tworzą strukturę hierarchiczną. Najbardziej popularnym tego typu modelem jest sieć semantyczna A.M. Collinsa i M.R. Quilliana (Quillian 1968). Model ten ilustruje rys. 3.

Autor jako inspirację tego pomysłu wskazał sposób organizacji wiedzy w typowym słowniku, leksykonie lub encyklopedii, w których sieć pojęciowa ujawnia się w strukturze odesłań porównawczych między hasłami. Zgodnie z hierarchicznym modelem Collinsa i Quilliana, aby opisać jakieś pojęcie, niezbędne jest odwołanie się do reprezentacji pojęciowej znajdującej się na wyższym poziomie hierarchii niż ta, którą się opisuje. Znalazło to wyraz w modelu definicji klasycznej. Konieczność odwoływania się do innych poziomów sieci semantycznej wydłuża czas udzielania odpowiedzi. Powodem zasygnalizowania tej klasy modeli jest fakt ich istnienia i odwoływanie się do nich w teorii języków informacyjno-wyszukiwawczych.



Rys. 3. Model pamięci semantycznej według A.M. Collinsa i M.R. Quilliana (1969).
Za: Nęcka, Orzechowski, Szymura 2007, s. 153

Równoległe z teorią Collinsa i Quilliana pojawiły się inne modele pamięci semantycznej. Należy do nich m.in. model skonstruowany przez George'a A. Millera i Philipa N. Johnsona-Lairda (1976). Rolę relacji semantycznych podkreśla także jedna z wersji modelu Waltera Kintscha (1974), w którym każda jednostka pamięci semantycznej jest określona przez listę relacji semantycznych, w jakie wchodzi z innymi jednostkami (Kintsch 1974).

Założenie zróżnicowanej dostępności poszczególnych powiązań typu hiponimicznego można znaleźć również w koncepcji Jean Aitchinson (2003)³², która powstała na bazie niezgodności wyników badań z założeniem silnie zhierarchizowanego modelu Collinsa i Quilliana (Aitchinson 2003). Koncepcja ta opiera się na rozróżnieniu stałych i czasowych powiązań hierarchicznych między elementami pamięci semantycznej.

Cechą wspólną modeli hierarchicznych pamięci semantycznej, wykorzystywaną w budowie języków wyszukiwania informacji, jest organizowanie jednostek sieci za pomocą poziomów ogólności, a więc za pomocą relacji hipero- i hiponimii, z uwzględnieniem tzw. drzewa Porfiriusza³³ (Sowa 2000). Taka organizacja umożliwia zastosowanie dziedziczenia cech i relacji poprzez poziomy hierarchiczne. Jako najwyższy i najniższy poziom hierarchii można wykorzystywać kategorie niezdefiniowane, tematyczne (Sowa 1991, 2000).

³² Jean Aitchinson jest profesorem językoznawstwa i komunikacji na Uniwersytecie w Oksfordzie oraz w Worcester College. W swoich badaniach skoncentrowała się na badaniu słownika mentalnego, zmian językowych oraz języka mediów. Wykładała na wielu uniwersytetach w USA, Europie i Indiach. Jest autorką takich książek, jak: *Linguistics, The Articulate Mammal: An Introduction to Psycholinguistics, Language Change: Progress or Decay?* i *Introducing Language and Mind*.

³³ Drzewo Porfiriusza to forma przedstawiania zależności rodzajów i gatunków w postaci drzewa. Stanowi ono podstawę współczesnych taksonomii.

Struktury hierarchiczne nie są jednorodne i można je uznać za szczególny przypadek sieci, w której występują tylko zależności hierarchiczne. Istotą podejścia sieciowego jest znaczący udział w organizacji struktur językowych relacji kojarzeniowych, które charakteryzują się symetrycznością. Szersze wykorzystanie relacji kojarzeniowych może zapewnić elastyczność języka i większe dostosowanie do indywidualnych potrzeb użytkowników. Przejawem tzw. sieciowości są różnego rodzaju gniazda pojęciowe, semantyczne, leksykalne. Rozszerzenie istniejącego modelu semantycznego na model sieciowy pozwoli w większym zakresie uwzględniać w budowie i wyszukiwaniu informacji inne rodzaje relacji niż relacje hierarchiczne.

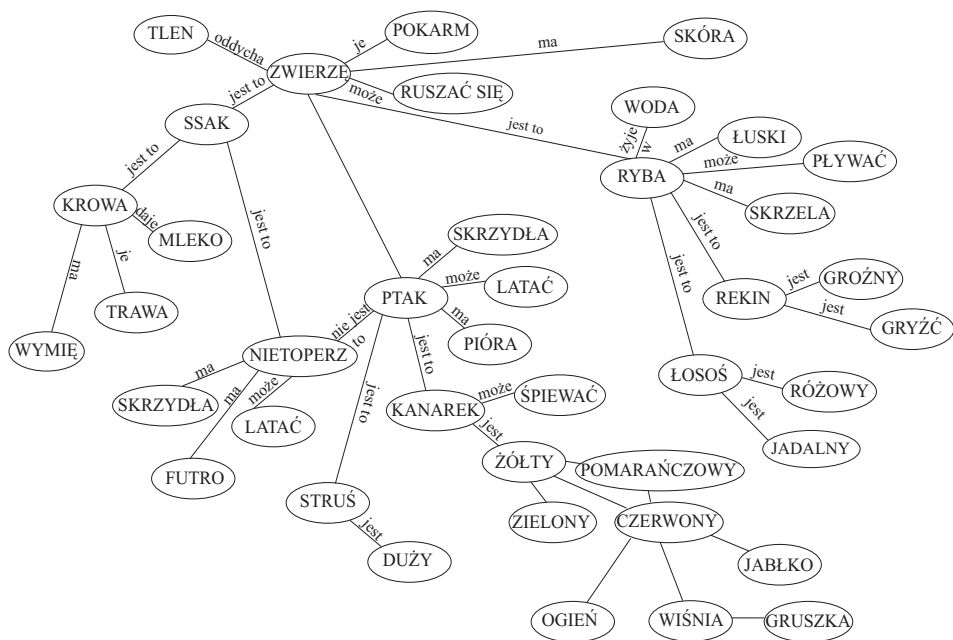
Modele semantyczne różnią się liczbą uwzględnianych relacji. Większa liczba relacji w modelu świadczy o bardziej leksykalnym niż pojęciowym ujmowaniu relacji, co owocuje orientacją bardziej leksykalną (postkoordynacja) lub bardziej pojęciową (prekoordynacja)³⁴. Liczba zastosowanych relacji może także wynikać ze szczegółowego podejścia do struktury pola semantycznego danego języka.

Szczególnie interesujące i przydatne z punktu widzenia moich rozważań są teorie pamięci semantycznej, które opisują sposób, w jaki zostały uporządkowane reprezentacje pojęciowe w systemie poznawczym człowieka. Są to systemy organizacji pojęć niehierarchiczne. Teorie te przyjmują, że pojęcia są przechowywane w umyśle ludzkim w postaci struktury przypominającej sieć złożoną z węzłów oraz połączeń między nimi, czyli tzw. ścieżek. W każdym z węzłów sieci jest zakodowana reprezentacja pojęciowa. Ścieżkom łączącym poszczególne węzły odpowiadają relacje semantyczne zachodzące między pojęciami reprezentowanymi przez te węzły. Sieć jest zbudowana tak, aby zapewnić ekonomię poznawczą. Niewątpliwie części węzłów odpowiadają pojęcia wyrażane w formie słów kluczowych. Słowa odpowiadające pojęciom są zatem reprezentowane w pamięci w postaci tzw. punktów węzłowych (ang. *nodes*), połączonych z sobą relacjami semantycznymi. W ten sposób powstaje sieć połączeń. Jeżeli pojęcia są z sobą blisko związane, to charakteryzuje je duża liczba wzajemnych połączeń (linków) odpowiadających ich wspólnym właściwościom. Istotność (siła, moc) relacji zachodzących pomiędzy poszczególnymi pojęciami jest odzwierciedlana za pomocą wagi ścieżki. Im relacja jest silniejsza, tym większą wagę ma połączenie między nimi (ziarnistość informacji³⁵). Połączenie pojęć ścieżką o dużej wadze sprawia, że skraca się czas udzielania odpowiedzi. Gdy pojęcia są powiązane słabiej, wówczas waga ich połączenia (waga ścieżki) jest mała. Dotyczy to również znacznego dystansu semantycznego pomiędzy pojęciami, co ułatwia podejmowanie decyzji negatywnych.

³⁴ „Prekoordynacja – w opozycji do postkoordynacji – cecha systemu informacyjno-wyszukiwawczego polegająca na tym, że funkcję wyszukiwawczą pełni samodzielnie tylko niektóre wyrażenia charakterystyki wyszukiwawczej dokumentu (najczęściej stojące na pierwszym miejscu), pozostałe zaś umożliwiają dostęp do zbioru wyszukiwawczego jedynie pośrednio, w sposób zależny od rezultatów wyszukiwania, za pomocą wyrażen samodzielnych wyszukiwawczo. Prekoordynacja jest charakterystyczna dla zbiorów informacyjnych »tradycyjnego« (konwencjonalnego) typu, na przykład kartkowe katalogi biblioteczne, drukowane indeksy itp.” (Bojar 2002, oprac., s. 206). Nie chodzi tu więc o budowanie wyrażen złożonych, bo każda charakterystyka wyszukiwawcza jest wyrażeniem złożonym, a te można tworzyć w każdym języku (poza językami unitermowymi, bo tu jest tyle oddzielnych charakterystyk, ile użytych unitermów).

³⁵ Ziarnistość zawartych w nich metainformacji pozwala określić charakterystyki systemowe w języku słów kluczowych.

Przetwarzanie konkretnego pojęcia uaktywnia odpowiadający mu węzeł sieci i sprawia, że jego pobudzenie (aktywacja) stopniowo rozprzestrzenia się poprzez ścieżki łączące go z węzłami reprezentującymi pojęcia związane z nim semantycznie. Dociera do węzłów bezpośrednio sąsiadujących z węzłem źródłowym, a także do innych – dalszych, ale znaczeniowo związanych z bodźcem, który uruchomił (pobudził) dany obszar sieci i leżących wzdłuż wywołanego przez niego łańcucha skojarzeń. Poziom (stopień) aktywacji konkretnego węzła sieci jest proporcjonalny do wagi ścieżki, wzdłuż której biegnie, a także do liczby połączeń wychodzących od węzła źródłowego. Pobudzenie wzdłuż ścieżek łączących sąsiadujące węzły sieci sprawia, że pojawienie się bodźca poprzedzającego ułatwia przetwarzanie słowa i odpowiadającego mu pojęcia w dalszej kolejności, jeżeli słowo to pozostaje w relacji semantycznej z bodźcem poprzedzającym. W ten sposób dochodzi do usprawnienia przejawiającego się w przyspieszeniu podejmowania decyzji semantycznych angażujących jego znaczenie (ang. *lexical decision task*). Węzeł źródłowy wykazuje aktywność tak długo, jak długo jest przetwarzane odpowiadające mu pojęcie. Pobudzenie (aktywacja) semantyczne zanika wraz z upływem czasu, a także wraz z wydłużaniem się pokonywanego dystansu. Ulega osłabieniu również w sytuacji, gdy system przetwarzania zapoczątkowuje aktywność konkurencyjną (Collins, Loftus 1975). W tym kontekście na uwagę zasługuje model sieciowy Allana M. Collinsa i Elizabeth F. Loftus (1975), dalej rozwijany przez Johna R. Andersona (1983). Model ten ilustruje rys. 4.



Rys. 4. Sieć pamięci semantycznej według modelu A.M. Collinsa i E.F. Loftusa (1975).
Za: Nęcka, Orzechowski, Szymura 2007, s. 156

Proces, o którym mowa, zachodzi automatycznie, tzn. bez udziału uwagi i świadomości. Osoba przetwarzająca pojęcia także nie ma bezpośredniego wpływu na zakres aktywacji, tzn. że wskutek aktywacji mogą zostać uruchomione również pojęcia położone w bardzo odległych rejonach sieci.

Obniżenie progu przetwarzania bodźca docelowego następuje także w wyniku działania mechanizmów w większym stopniu kontrolowanych przez osobę, czyli gdy świadomie kieruje ona swoją uwagą w stronę konkretnej reprezentacji pojęciowej odpowiadającej konkretnemu pojęciu (w stronę konkretnego obszaru pamięci semantycznej). Świadome ukierunkowanie uwagi na wybrane reprezentacje pojęciowe, a tym samym przeniesienie uwagi w rejon sieci, z którego pochodzi oczekiwane słowo docelowe, powoduje przyspieszenie decyzji. Przyspieszenie decyzji wskutek działania opisanego mechanizmu dotyczy jednak tylko pojęć pochodzących z kategorii oczekiwanej przez daną osobę. Mechanizm uaktywniania węzłów sieci stanowiących reprezentacje odrębnych pojęć może być automatyczny. Istnieje też możliwość kontrolowanego (ze świadomym zaangażowaniem uwagi) procesu uaktywniania sieci semantycznej i podejmowania decyzji leksykalnych. Proces ten jest badany za pomocą procedury poprzedzania semantycznego. Warto zwrócić uwagę, że mózg traktowany jest jak sieć neuronowa. Zagadnienie siły i zasięgu działania mechanizmów aktywacyjnych pozostaje jednak nadal otwarte.

Interesujący kierunek badań nad organizacją zapisu pamięciowego człowieka, użyteczny i stanowiący inspirację dla kognitywnego modelowania organizacji informacji i wiedzy w systemach dokumentacyjnych za pomocą języków słów kluczowych, stanowią badania związków między wyodrębnionymi w pamięci jednostkami zapisu (pojęciami). Problematyką tą szczegółowo zajmowała się B. Sosińska-Kalata (Sosińska-Kalata 1999). Poniżej przestawię teorię sieci semantycznych, która znajduje zastosowanie w modelowaniu języków słów kluczowych. Inne teorie, jak teoria cech semantycznych, zostały szczegółowo omówione przez Sosińską-Kalatę. Nie są przydatne w dalszych rozważaniach. Zarówno badania psychologiczne, jak i lingwistyczne do tej pory nie dostarczyły przekonujących przesłanek, aby sądzić, że, operując pojęciami lub wyrażeniami językowymi, ludzie najpierw wyodrębniają, a następnie łączą z sobą ich istotne cechy dystynktywne. Przedstawione modele organizacji pamięci semantycznej mają odniesienie do słownika mentalnego, zwanego też leksykonem umysłowym.

Model związków między pojęciami – o którym wspomniałem – powstał na podstawie teorii procesów poznawczych Joya Paula Guilforda, co miało zapewnić zgodność identyfikacji relacji między najważniejszymi pojęciami w semantycznej strukturze indeksowanych dokumentów ze sposobem, w jaki użytkownicy wyodrębniają związki międzypojęciowe w strukturze wiedzy. Zdaniem Guilforda (Guilford 1978) umysł ludzki realizuje określoną liczbę operacji poznawczych, wykorzystując dwa podstawowe mechanizmy przetwarzania bodźców: mechanizm asocjacji i mechanizm dyskryminacji (odróżniania). Każdy z nich operuje na trzech poziomach przetwarzania informacji. Mechanizm asocjacji zmierza do wzrostu integracji asocjacyjnej na trzech poziomach: poziomie uświadomienia (ang. *cognition*), poziomie asymilacji pamięciowej (ang. *memory*) i poziomie oceny (ang. *evaluation*). Uświadomo-

mienie polega na prostej identyfikacji pojęcia, pamięć pozwala na uchwycenie związków pojęć w czasie, ocena zaś umożliwia powstanie stałych powiązań na podstawie wcześniejszej wiedzy i poznawczego doświadczenia. Mechanizm dyskryminacji bodźców pozwala na dokładność percepcji na trzech etapach: rozpoznanie bodźca, myślenie konwergencyjne i myślenie dywergencyjne. Rozpoznanie umożliwia stwierdzenie tożsamości elementów na mocy nierozróżnialności ich cech indywidualnych. Myślenie konwergencyjne pozwala na identyfikację równoważności percypowanych elementów, przy równoczesnej świadomości ich cech indywidualnych. Myślenie dywergencyjne umożliwia identyfikację wyraźnych i dobrze rozpoznanych różnic między percypowanymi bodźcami.

Niezależnie od badań nad strukturą pamięci semantycznej prowadzono szereg eksperymentów nad pamięcią semantyczną w perspektywie teorii kategoryzacji, w których można dostrzec pewne podobieństwa do przedstawionych wcześniej badań nad pamięcią. Jedną z teoretycznych implikacji tych badań jest teza, że słowa należące do danej kategorii są kodowane w pamięci blisko siebie i tworzą poznawczą jednostkę wyższego rzędu. Wydobycie z pamięci jednego elementu takiej jednostki-kategorii aktywizuje jednocześnie wszystkie jej pozostałe elementy. Sugeruje to możliwość badania w świetle współczesnych teorii kategoryzacji. Wskazuje także na przestrzenną organizację zapamiętanych informacji, która została uwzględniona w omówionych w tym rozdziale sieciowych modelach pamięci. Problematyka ta już była przedmiotem opracowania Jadwigi Woźniak (Woźniak 2000), dlatego pozostawiam ją poza zakresem moich rozważań. Warto jednak zasygnalizować istnienie stosunkowo młodej, ale dynamicznie rozwijającej się dziedziny nauki, zwanej neuronauką (ang. *neuroscience*), która zajmuje się badaniem ludzkiego systemu nerwowego, w tym badaniem mózgu. Poszukuje biologicznych uwarunkowań percepcji informacji oraz pamięci. Mózg traktuje jako samoorganizujący się system.

W teorii i praktyce języków informacyjno-wyszukiwawczych sieci semantyczne były intuicyjnie wykorzystywane od dawna w językach słów kluczowych opartych na słowniku jawnym w postaci tradycyjnej lub w nowoczesnej formie bazy leksykalnej online. Takie słowniki elektroniczne znajdują już zastosowanie w wielu dziedzinach. W przypadku słowników niejawnych bazę leksykalną stanowi słownik mentalny człowieka i/lub zbiór terminów użytych w indeksowaniu, zaś w przypadku słowników jawnych bazą leksykalną jest słownik drukowany (także w wersji komputerowej) lub słownik online.

Sieci semantyczne, bazując na rekonstruowaniu procesów umysłowych człowieka, w tym procesów przetwarzania języka, są pomocne w tworzeniu języków słów kluczowych. Wyniki badań psychologicznych, psycholingwistycznych i językoznawczych stanowią podstawę przyjęcia podstawowych założeń paradygmatu sieciowego w systemach wyszukiwania informacji, w tym w językach słów kluczowych, które są jednym z elementów takich systemów. W dalszej części rozważań zajmę się szczegółowo słownikiem mentalnym.

1.3.2. Leksykon umysłowy (słownik mentalny) jako wzorzec/model organizacji systemu leksykalnego i słownikowego języków słów kluczowych

Postawienie w teorii języków informacyjno-wyszukiwawczych, w tym w teorii języków słów kluczowych, pytania o naturę funkcjonowania językowego zainspirowało badania nad słownikiem czy leksykonem umysłowym, pojmowanym jako magazyn i miejsce procesów pamięciowych oraz procesów produkcji i odbioru tekstów języka informacyjno-wyszukiwawczego. Ze względu na związek języka z posiadaną i nabywaną wiedzą zagadnienie dotyczące leksykonu umysłowego można uznać za kluczowe dla funkcjonowania języków swobodnych słów kluczowych. Badania te mają z konieczności interdyscyplinarny charakter. Wykorzystując teorie, terminy, metody i rezultaty badań psychologii i językoznawstwa, pozwalają na wyciąganie wniosków i opracowanie na potrzeby teorii języków informacyjno-wyszukiwawczych odpowiedniego modelu języków słów kluczowych, a więc wykraczającego poza te szczególne dziedziny.

Tworzenie teorii języków słów kluczowych wymaga zdobycia danych o strukturze tych języków oraz o ich elementach, poziomach i powiązaniach między nimi, co będzie przedmiotem części drugiej opracowania. Jednym z głównych czynników, wyznaczających makro- i mikrostrukturę słowników tych języków, jest makro- i mikrostruktura słownika umysłowego, a więc jego uporządkowanie semantyczne, w tym także uporządkowanie hierarchiczne, ale nie tylko. Rekonstrukcja struktury semantycznej leksykonu umysłowego pozwala odsłonić specyfikę i mechanizmy ludzkiej pamięci i myślenia oraz procesów przetwarzania językowego, a także właściwości słownictwa języków słów kluczowych.

Poddając analizie struktury organizacji informacji leksykalnej w umyśle człowieka, trzeba odwołać się do wiedzy na ten temat, jaką udało się uzyskać w badaniach psychologicznych i w innych naukach poznawczych. Umysł posiada umiejętność odzwierciedlania tego, co istotne w zjawiskach świata poznawanego przez człowieka. Najbliższe poglądom współczesnej psychologii poznawczej jest stanowisko, że człowiek aktywnie uczestniczy w konstruowaniu obrazu rzeczywistości, posługując się danymi, których dostarcza mu doświadczenie. To, co człowiek ma w umyśle, stanowi zapis informacyjny zdeterminowany jednak przez reguły rządzące ludzkim systemem poznawczym. Zapis ten jest odpowiednio zorganizowany i utrwalony w pamięci. Rodzi się więc pytanie, w jaki sposób taki zapis powstaje i jaki jest sposób jego organizacji.

Współczesne badania nad tymi problemami koncentrują się wokół dwóch nurtów, nazywanych paradygmatem neuropsychologicznym i paradygmatem teorioinformacyjnym. Oba te nurty kształtują współczesne konstrukcje modeli umysłu, stanowiąc dwa uzupełniające się aspekty opisu: pierwszy podkreśla funkcjonalny aspekt procesów pamięciowych, drugi – aspekt strukturalny, skupiając się na modelach logicznego lub asocjacyjnego zapisu wiedzy.

W polskim piśmiennictwie z zakresu języków informacyjno-wyszukiwawczych rozróżnia się wiedzę językową i wiedzę pozajęzykową (wiedza o świecie). Takie rozważania były i są podejmowane przede wszystkim przez badaczy języka pojmowanego jako jedna ze struktur poznawczych. Rozróżnienie to wywodzi się z teorii lingwistycznych, jeszcze z czasów Ferdinanda de Saussure'a, który akcentował dwu- aspektowość badań językoznawczych: aspekt systemowy systemu znakowego oraz aspekt wykonawczy (akty mowy) (Saussure 1961). Kategorię języka de Saussure'a zastępuje się dzisiaj, wprowadzoną przez Noama Chomsky'ego, kategorią kompetencji językowej (ang. *competence*), zaś kategorię mówienie – kategorią wykonania (ang. *performance*), przy czym oba aspekty języka są traktowane jako indywidualnie zróżnicowane własności poszczególnych użytkowników języka (Lyons 1998).

Chomsky w swojej gramatyce generatywno-transformacyjnej przyjmuje, że reprezentacja wiedzy językowej w umyśle ma postać ustrukturalizowanego zbioru reguł generowania i transformacji struktur językowych – czyli pewnego słownika umysłowego – którymi reguły te operują oraz reguł przypisywania tym mentalnym strukturom językowym odpowiednich interpretacji semantycznych, które pozwalają odnosić te wyrażenia do elementów opisywanego przez nie świata.

Model kompetencji językowej Chomsky'ego ma już bogatą literaturę przedmiotu, także w języku polskim. Jego szczegółowa analiza nie należy jednak do zakresu moich rozważań, dlatego charakterystykę tej koncepcji ograniczam, wskazując tylko na te elementy, do których będę odwoływać się w dalszych rozdziałach książki.

Teoria Chomsky'ego i wprowadzone w niej pojęcia struktury powierzchniowej, struktury głębokiej oraz słownika umysłowego odegrały ważną rolę w badaniach psycholingwistycznych. Problematykę semantyczną najpełniej uwzględnia teoria standardowa. Ogólną koncepcję wiedzy językowej w tym ujęciu można opisać jako rodzaj wiedzy, który pozwala użytkownikowi języka łączyć określone znaczenia (pojęcia) z formą fizyczną, przyporządkowaną mu w określonym języku naturalnym. Powstałe w ten sposób struktury głębokie i struktury powierzchniowe są mentalnymi strukturami językowymi. Wiedza wyrażona za pomocą języka przyjmuje postać struktur frazowych: struktura głęboka reprezentuje znaczenie w umyśle (plan treści), natomiast struktura powierzchniowa wyrażenia (plan wyrażania).

Próby wyznaczenia granicy między współdziałającymi systemami wiedzy językowej i pozajęzykowej podejmowało niewielu badaczy. Interesujący model systemu poznawczego człowieka zaprezentowała Ida Kurcz (Kurcz 1987). Opiera się on na rozróżnieniu umiejętności posługiwania się językiem i wiedzy konceptualnej. Nie podejmując się szczegółowego omówienia tego modelu, zwracam uwagę tylko na zawartą w nim koncepcję słownika mentalnego, zwanego przez Kurcz słownikiem umysłowym.

Słownik umysłowy jest układem poznawczym, którego zadaniem jest identyfikowanie znaczenia percypowanych bodźców werbalnych. Obiektami informacji zawartej w słowniku są leksemy, stanowiące fragment wiedzy językowej dotyczącej leksyki i reguł jej funkcjonowania. Zadaniem słownika umysłowego jest ustalanie struktury semicznej leksemu, przez którą rozumie się charakterystyczny dla niego zespół dystynktywnych cech semantycznych.

Dla moich rozważań szczególnie ważna jest semantyczna organizacja leksykonu umysłowego. Ze względu na słownik umysłowy, na którym bazują języki swobodnych słów kluczowych, warto zwrócić uwagę na formy jego organizacji. Ida Kurcz twierdzi (za wieloma psychologami), że istnieje co najmniej kilka równoległych struktur organizacji informacji leksykalnej, z którymi wiążą się różne sposoby dostępu do niej. Model I. Kurcz odwzorowuje strukturę intensjonalnego znaczenia leksemów w formie sieci semantycznej, której łuki wyznaczają podstawowe relacje paradygmatyczne: synonimię, antonimię oraz hiponimię/hiperonimię. Podstawę organizacji słownika umysłowego stanowią elementarne cechy semantyczne, a więc zapis struktury znaczenia intensjonalnego leksemów. Badaczka traktuje znaczenie, podobnie jak w semantyce komponentalnej, jako zestaw dystynktywnych cech semantycznych tworzących strukturę semiczną leksemu. W słowniku umysłowym zakodowana jest struktura semantyczna, tj. semantyczne związki językowe. Semantyczna wiedza językowa w słowniku umysłowym została wyznaczona przez relacje paradygmatyczne między leksemami, determinujące zakres ich wymienialności w określonych kontekstach, zachodzące niezależnie od wiedzy o opisywanej rzeczywistości. Na wiedzę językową, zinternalizowaną w umyśle każdego człowieka, składa się zatem zarówno kompetencja językowa, rozumiana jako abstrakcyjna zdolność posługiwania się językiem, jak i kompetencja komunikacyjna, będąca umiejętnością użycia języka odpowiednio do zamierzeń nadawcy oraz wiedzy o pozostałych elementach układu komunikacyjnego. Za te procesy odpowiedzialne są odpowiednie analizatory wiedzy językowej i wiedzy pozajęzykowej.

Wiedza językowa obejmuje informacje o wyrażeniach i regułach posługiwania się nimi w opisie świata. Jest przede wszystkim odwzorowaniem umiejętności posługiwania się językiem. Wiedza pozajęzykowa to konceptualizacja świata wyrażona za pomocą odpowiedniego systemu kategorii pojęciowych i zachodzących między nimi relacji. Pomiedzy konceptualnym i werbalnym odwzorowaniem rzeczywistości nie ma jednak izomorfizmu. Nieizomorficzność organizacji leksykalnej wiedzy językowej i semantycznej wiedzy o systemie pojęć może powodować trudności komunikacyjne związane z identyfikacją właściwego znaku językowego dla pojęcia, którym posługuje się użytkownik. Na potrzeby systemów wyszukiwania informacji i kształtowania odpowiednich systemów leksykalnych języków słów kluczowych niezbędna jest więc identyfikacja i wyeliminowanie nie tylko luk informacyjnych w wiedzy o rzeczywistości, zinternalizowanej w umyśle użytkownika systemu, ale również identyfikacja luk werbalizacyjnych związanych z brakiem odpowiedniego fragmentu wiedzy językowej, której „wypełnienie” stanowi podstawę selekcji dokumentów zajmujących poszukiwane fragmenty wiedzy o rzeczywistości.

W ustalaniu systemowej reprezentacji dokumentu, a więc w opisie treści dokumentu w języku systemu, można wyróżnić (za duńską szkołą strukturalizmu) trzy wzajemnie zależne płaszczyzny odwzorowywania rzeczywistości, nazywane planami: treści, werbalnym i notacyjnym. Plan treści odpowiada płaszczyźnie odwzorowania konceptualnego, plan werbalny stanowi płaszczyznę odwzorowywania rzeczywistości w języku naturalnym użytkownika (zawarta jest tu m.in. wiedza tworząca słownik umysłowy), plan notacyjny to z kolei płaszczyzna odwzorowania rzeczywi-

stości w językach, w których informacje o niej są reprezentowane w systemie, a więc m.in. w językach słów kluczowych.

W kontekście interakcji użytkownika z systemem dokumentacyjnym wyodrębnia się dwa rodzaje wiedzy językowej:

- wiedza o języku (najczęściej naturalnym), w którym użytkownik werbalizuje swoje problemy wyszukiwawcze;
- wiedza o języku, w którym musi być ona reprezentowana w komunikacji z systemem informacyjno-wyszukiwawczym.

Pomiędzy wiedzą użytkownika o świecie, jego wiedzą językową dotyczącą języka naturalnego i wiedzą dotyczącą języka systemu zachodzą pewne sprzężenia, w ramach których użytkownik dostosowuje zakres identyfikacji wiedzy do własnych możliwości jej werbalizacji w języku naturalnym oraz do możliwości formalnego języka systemu. Prowadzić to może do nieuprawnionego, nadmiernego uogólniania.

Według przedstawionej interpretacji reprezentacji wiedzy w umyśle, problem jej organizacji można sprowadzić do dwóch podstawowych zagadnień:

- generowanie i definiowanie pojęć;
- strukturalizacja systemu pojęciowego i sytuacyjnego obrazu rzeczywistości.

Za szczególnie ważny aspekt modelowania języków słów kluczowych na potrzeby systemów dokumentacyjnych uważam więc odwzorowywanie w nich zarówno wiedzy o języku naturalnym i językach specjalistycznych, która umożliwi odwzorowanie wiedzy prywatnej użytkownika w pytaniach kierowanych do systemu, jak i wiedzy o języku systemu, która umożliwi adekwatną transformację tego odwzorowania.

Rozważania nad psycholingwistycznymi determinantami języka słów kluczowych, poza identyfikacją determinant, które pozwalają wyjaśnić pewne zjawiska językowe, pomogły określić kierunek dalszego „rozwoju” tych języków. Wspomniane czynniki, jakkolwiek ważne, nie są jednak wystarczające. Niezbędne jest uwzględnienie również czynników wewnątrzjęzykowych. Za najważniejsze uważam: terminologię, pole semantyczne, kompetencje językowe oraz relacje kojarzeniowe. Ich rolę w odniesieniu do języka słów kluczowych omówię po kolei.

1.4. Językowe determinanty słów kluczowych

Słowa kluczowe używane w wyszukiwaniu informacji, to elementy języka, najczęściej terminologii, dlatego też w dalszej części rozważań zajmę się podobieństwami i różnicami między słowem kluczowym a terminem, polem semantycznym języka słów kluczowych i języka naturalnego, oraz kompetencją językową użytkowników słów kluczowych. Pozwoli to określić językowe uwarunkowania tożsamości języka słów kluczowych.

1.4.1. Termin jako jednostka leksykalna języka naturalnego i języka słów kluczowych

Słowa kluczowymi są zwykle terminy, czyli wyrażenia oznaczające pojęcia fachowe. Czasem używane są wyrazy języka ogólnego, ale w znaczeniu specjalistycznym (na przykład „stopa” – w naukach technicznych: „wystająca część maszyny, służąca do jej przymocowania do podstawy”). Niektóre terminy (na przykład w matematyce) są ściśle, inne (zwłaszcza w naukach humanistycznych) – nieostre, a czasami nawet wieloznaczne.

Terminologia stanowi więc – jak już sygnalizowałem w rozdziale 1.1.2 – istotny element systemu leksykalnego języka słów kluczowych. Ustalanie precyzyjnej terminologii dyscyplin naukowych wiąże się ze zmierzaniem w kierunku kontroli słownictwa. W systemach gromadzących informacje o piśmiennictwie naukowym operowanie terminologią naukową jest konieczne i naturalne. Nie jest ono jednak niezbędne w systemach informacji specjalistycznej, ale niekoniecznie naukowej, na przykład w bazach Krajowej Izby Gospodarczej, tzw. INFOR-u, w systemach ogólnej informacji społecznej albo w bazach informacji prasowej.

Słowa kluczowe to najczęściej terminy³⁶. A termin – według Hanny Jadackiej – to „jednowyrazowy lub wielowyrazowy odpowiednik pojęcia z określonej dziedziny nauki lub techniki, mający znaczenie wyraźne, używany przez specjalistów w tekstach fachowych” (Jadacka 1978)³⁷. Są to więc wyrazy i związki wyrazowe pretendujące do określeń jednoznacznych. Wymogi stawiane terminom to przede wszystkim dokładne określenie znaczenia i stałość ich użycia. W zasadzie terminy nie powinny mieć synonimów, jednak, w wyniku oddziaływania różnych czynników zewnętrznych, ten sam przedmiot poznania często określany jest kilkoma terminami, na przykład „książka” i „wydawnictwo zwarte”. W ten sposób synonimia staje się zjawiskiem obecnym nie tylko w języku potocznym, lecz także w specjalistycznej terminologii naukowo-technicznej. Jednoznaczność, wysuwana często jako swoista cecha terminu, jest postulatem słusznym, ale tylko postulatem.

Słowo kluczowe to rodzaj metainformacji, która pełni następujące funkcje: informacyjną, identyfikacyjną, interpretacyjną, wyszukiwawczą, koordynacyjną i organi-

³⁶ W tym kontekście warto zwrócić uwagę na zamieszczoną we *Wstępie do Słownika słów kluczowych językoznawstwa slawistycznego* definicję słowa kluczowego: „Słowa kluczowe to elementarne wyrażenia języka słów kluczowych, równokształtne z terminami, ważnymi i charakterystycznymi dla danej dziedziny wiedzy oraz mogącymi reprezentować treść dokumentu” (http://www.ispan.waw.pl/cnis/slownik_slow_kluczowych.htm). Widać tu wyraźne ograniczenie do terminologii, z nadaniem jej specjalnej funkcji, którą jest reprezentowanie treści dokumentu. Słowniki terminologiczne zawierają większość terminów, które odpowiadają wymaganiom stawianym słowom kluczowym (często wraz z odpowiednimi odesłaniami). W wielu wypadkach terminy, a zwłaszcza terminy wielowyrazowe, wymagają jednak zmian ze względu na szyk wyrazów czasami odbiegający od naturalnego lub możliwość zastąpienia ich słowami kluczowymi odpowiadającymi ich elementom składowym.

³⁷ Podobnie termin jest określony w *Słowniku terminologicznym informacji naukowej*: „Termin – wyraz (w znaczeniu 3) lub połączenie wyrazowe o ściśle ustalonym znaczeniu w danej dziedzinie nauki lub techniki” (s. 115).

zacyjną. Słowa kluczowe tworzą więc zasób metainformacyjny samoistny, wydzielony (jak w przypadku języka słów kluczowych) albo niewydzielony (jak w przypadku języka naturalnego).

Termin jest elementarną jednostką języka nauki i techniki, zaś słowo kluczowe to wyrażenie języka słów kluczowych. Termin jest szczególnym przypadkiem nazwy, natomiast słowo kluczowe jest szczególnym przypadkiem terminu w funkcji metainformacyjnej. Słowo kluczowe stanowi więc typową nazwę obiektów znajdujących się w polu zainteresowań danej dyscypliny i wyodrębnionych w sposób powszechnie akceptowany przez uprawiających ją badaczy.

Słowa kluczowe to podzbiór słownictwa języka naturalnego, któremu zleca się pełnienie określonych specjalnych funkcji. Ten podzbiór istnieje nadal w języku naturalnym, a więc nie jest czymś wydzielonym, zamkniętym i mającym swoją własną egzystencję. Jest to analogiczna sytuacja jak z terminologią. W przypadku nauki ten podzbiór w znacznym stopniu pokrywa się z terminologią. Słowa kluczowe to elementy tego podzbioru, ale nie ograniczają się do niego.

Status słów kluczowych opiera się na mierze dyskretnej. Użytkownik, posługując się tymi samymi wyrażeniami w języku naturalnym, nakłada swoją siatkę znaczeń, która w swej istocie nie jest dyskretna. To system dokonuje modyfikacji nałożonej przez użytkownika na rzeczywistość dokumentacyjną siatki pojęciowej, odpowiednio do wymogów stosowanego języka słów kluczowych, którego jednostki leksykalne mają charakter naturalny, aczkolwiek w sferze denotacji dyskretnej.

Słowa kluczowe różni od innych wyrażen języka naturalnego w funkcji metainformacyjnej lub wyszukiwawczej ich odniesienie do rzeczywistości dokumentacyjnej. Supozycja informacyjna oznacza tu wykorzystanie wyrażen języka naturalnego do oznaczania nazw przedmiotów/tematów dokumentów, a nie nazw elementów rzeczywistości pozatekstowej. Słowa kluczowe, opisujące dany dokument, w sumie tworzą/wyznaczają pewną pod/przestrzeń informacyjną w danym zbiorze wyszukiwawczym.

1.4.1.1. Status terminu w systemach leksykalnych języków słów kluczowych

O statusie terminów jako jednostek leksykalnych języka informacyjno-wyszukiwawczego pisałem m.in. w artykule zamieszczonym w „Zagadnieniach Informacji Naukowej” (Babik 1999b). Poczynione tam ustalenia zamierzam teraz odnieść do języków słów kluczowych.

System leksykalny w językach paranaturalnych jest generowany na podstawie systemu leksykalnego języka naturalnego, najczęściej jego części, terminologii. Systemy terminologiczne odwzorowują z kolei systemy pojęciowe porządkujące wiedzę w danej dziedzinie. Odwzorowanie to nie jest jednak izomorficzne, tzn. system pojęć i system terminologiczny nie są paralelne. Systemy pojęciowe ciągle ewoluują i są bogatsze od systemów terminologicznych, zmierzających do uściślenia i ujednoczenia interpretacji pojęć na określonym etapie rozwoju wiedzy danej dziedziny. Z kolei system leksykalny języka informacyjno-wyszukiwawczego odwzorowuje jedynie te pojęcia, które wyznaczają pola tematyczne istotne dla organizacji i wyszukiwania informacji o zbiorze dokumentów z danej dziedziny.

W szczególnej sytuacji znajdują się systemy dziedzinowe, gdyż słownictwo specjalistyczne odgrywa w nich zasadniczą rolę w analizie treści dokumentów oraz wyborze haseł w celu indeksowania dokumentów. Język naturalny nie wyczerpuje w pełni terminologii konkretnej dziedziny, dlatego – przede wszystkim na potrzeby systemów dziedzinowych – tworzy się własne języki, oparte na terminologii danej dziedziny.

Język specjalistyczny oparty na terminologii pełni funkcję klasyfikowania elementów rzeczywistości i nadawania nazw wydzielonym klasom. W zależności od konkretnej potrzeby elementy rzeczywistości są grupowane w różne klasy. Każdy język dzieli elementy rzeczywistości według swoich własnych kryteriów.

Nadając terminowi lub wyrazowi status słowa kluczowego, trzeba zdać sobie sprawę z cech wewnętrznych i zewnętrznych terminu bądź wyrazu. Mogą bowiem pojawić się rozbieżności pomiędzy znaczeniem terminu a treścią pojęcia, wynikające z nieznamości istoty rzeczy czy unikania wysiłku umysłowego, a nawet braku kultury słowa. Terminy naukowe oraz słowa kluczowe powinny nasuwać prawidłowe skojarzenia z pojęciami powstałymi w ludzkim umyśle pod wpływem znaku językowego. Niestety, w piśmiennictwie fachowym występują terminy utworzone wprawdzie zgodnie z zasadami słowotwórstwa, ale ubogie w treść lub nasuwające błędne skojarzenia myślowe. Przyczyną tego może być:

- nieznamość terminologii danej dziedziny nauki;
- pośpieszne tworzenie neologizmów językowych w sposób sprzeczny z zasadami semantyki i słowotwórstwa, co powoduje zaśmiecanie terminologii;
- wprowadzenie do tekstu sztucznych i napsuszonych terminów naukowych w celu nadania pracy naukowej charakteru prekursorskiego (odkrywczego);
- błędy w tłumaczeniu z jednego języka na drugi.

Przy tworzeniu systemów leksykalnych języków słów kluczowych jednym ze źródeł słownictwa jest terminologia zgromadzona w słownikach terminologicznych oraz używana w opublikowanych tekstach konkretnych dokumentów. Słowniki terminologiczne, ze względu na często zbyt długi proces opracowywania i publikacji, nie odnotowują wielu nowych terminów, rejestrują je natomiast stale aktualizowane banki danych, zwane terminologicznymi bankami danych. Banki te – podobnie jak kartoteki wzorcowe³⁸ – stanowią narzędzia normalizacji leksyki.

Banki terminologii (banki terminów, terminologiczne banki danych) są produktem dwóch współczesnych zjawisk: ogromnego wzrostu liczby terminów oraz rozwoju informatyki (techniki komputerowej). Stanowią one współczesną formę podręcznych kartotek, będących od dawna w użyciu wśród pracowników nauki, tłumaczy, leksykografów i dokumentalistów. Dane banku terminologicznego odwzorowują system terminologiczny danej dziedziny, dyscypliny wiedzy lub techniki. Nie należy jednak utożsamiać banków terminów z bankami leksyki, przewyższającymi zasoby banków terminologicznych, które nie stanowią przedmiotu naszych zainteresowań.

³⁸ Kartoteki wzorcowe gromadzą różne oboczności nazw charakteryzujących dokumenty i pozwalają wprowadzać np. nazwy osobowe, geograficzne, korporatywne w postaci znormalizowanej, tworząc pewien ład leksykalny w świecie chaosu wieloznaczności, charakterystycznej dla języków naturalnych.

Powszechny dostęp do coraz „inteligentniejszych” komputerów, umożliwiających dialog użytkownika z bankiem terminów, stwarza nowe możliwości wykorzystywania terminologii w językach słów kluczowych. Jest to o tyle istotne, że znajomość terminologii przez specjalistów w zakresie informacji i dokumentalistów jest niezbędna przy identyfikowaniu treści dokumentów i opracowywaniu języków. Źródła leksyki języków informacyjno-wyszukiwawczych i źródła danych banków terminów są więc wspólne.

1.4.1.2. Znaczenie leksykalne terminów jako jednostek leksykalnych języka słów kluczowych

Cechą charakterystyczną języków słów kluczowych jest wykorzystywanie nazw (terminów)³⁹ języka naturalnego jako elementarnych jednostek leksykalnych. Punktem wyjścia przy ustalaniu znaczenia elementarnych jednostek leksykalnych języka informacyjno-wyszukiwawczego jest rzeczywistość dokumentacyjna (Hutchins 1978). Elementarna jednostka leksykalna służy do identyfikacji i reprezentacji fragmentów rzeczywistości dokumentacyjnej, to jest treści i/lub formy dokumentów. Istota reprezentacji polega na odwzorowywaniu i oznaczaniu. Odwzorowywanie wiąże się z systemem pojęć (obiekt – działalność poznawcza – termin), oznaczanie z funkcją identyfikacyjną znaczenia leksykalnego.

W językach informacyjno-wyszukiwawczych jako językach sztucznych, miejsce to jest w jakiś sposób explicite wskazywane tam, gdzie jest określona paradygmatyka. W języku słów kluczowych nie ma możliwości wyrażania informacji o tym, w którym miejscu struktury paradygmatycznej znajduje się dana jednostka leksykalna. Brak wskazania relacji semantycznych, w które wchodzi ona z innymi jednostkami leksykalnymi danego języka informacyjno-wyszukiwawczego, decyduje o tym, że jednostkę tę należy traktować jako niezależną (na gruncie danego języka) od innych jednostek leksykalnych tego języka.

Termin to jednostka leksykalna użyta w specjalnej funkcji: nazywania obiektów badań określonej dziedziny wiedzy lub działalności ludzkiej. Słowo kluczowe to termin lub wyraz języka ogólnego użyty w funkcji nazywania treści zawartych w dokumentach i/lub pytaniach informacyjnych dotyczących określonej dziedziny wiedzy lub dyscypliny naukowej (w funkcji metainformacyjnej). Zatem różnice pomiędzy terminem a jednostką leksykalną języka słów kluczowych tkwią w ich funkcjach i planie treści. Systemy terminologiczne funkcjonują według reguł języka naturalnego. Systemy leksykalne języków słów kluczowych funkcjonują według reguł języka naturalnego (język swobodnych słów kluczowych) lub praw języków sztucznych (ję-

³⁹ Tylko niewielka część terminologii znajduje się w podstawowych normach terminologicznych danego obszaru normalizacyjnego, zawartych m.in. w słownikach terminologicznych. Są to terminy znormalizowane, czerpane z Polskich Norm w postaci niezmienionej, których zarówno forma, jak i znaczenie określone są tymi normami. Wykorzystanie znormalizowanej terminologii do tworzenia systemów leksykalnych języków informacyjno-wyszukiwawczych stawia dodatkowe wymagania dotyczące prac analitycznych nad poprawnością terminologii stosowanej w danej dziedzinie i zawartej w słownikach terminologicznych oraz innych dokumentach normalizacyjnych.

zyk kontrolowanych słów kluczowych). Sprawia to, że ich plany treści i plany wyrażania posiadają pewną specyfikę. Słowami kluczowymi jest tylko część terminów oraz niektóre wyrażenia leksyki ogólnej języka naturalnego.

Każdy język naturalny opisuje rzeczywistość na sobie właściwy sposób. Dotyczy to również terminologii naukowej, która /stanowi podstawowy zasób leksykalny języków słów kluczowych. Mimo zbieżności form, w różnych językach jest ona nośnikiem różniących się treści. Język składa się z planu treści i wyrażania, a plan treści tworzy konotacja jego wyrażen. Konotacja zaś to cała wiązka cech semantycznych i nie tylko tych koniecznych i wystarczających (konstytutywnych) do tego, aby zgodnie z regułami semantyki danego języka można je przyporządkować danemu elementowi rzeczywistości, ale także innych przysługujących mu cech (przypadłości), skupionych przez dane wyrażenie w jednej strukturze. W tej sytuacji warunkiem porozumienia staje się identyczność denotacji.

Terminologia ułatwia nie tylko dostęp do wiedzy, ale też jest źródłem słów kluczowych umożliwiających odwzorowanie treści dokumentów dla potrzeb systemów informacyjno-wyszukiwawczych, co jest szczególnie ważne w systemach internetowych. Systemy te bowiem reprezentują treść zgromadzonych w nich informacji (dokumentów) przez odwzorowujące je słowa kluczowe w funkcji metainformacyjnej. W omawianych językach jest niezbędna znajomość systemów terminologicznych poszczególnych dyscyplin naukowych i ustaleń terminologii jako nauki o tych systemach.

Istnieje (jest ustalona) pewna konwencja etykietowania tekstu przy użyciu słów kluczowych. Konwencjonalność języków słów kluczowych polega na tym, że słowa kluczowe to przede wszystkim rzeczowniki⁴⁰. Specyfika słów kluczowych polega na tym, że w sposób konwencjonalny określają one elementy treści dokumentu. Konwencja wielu słów kluczowych dotyczy samej ich formy. Zatem słowo kluczowe jest szczególnym przypadkiem nazwy. Jest to nazwa mająca konwencjonalnie ustalone przez użytkownika języka znaczenie w funkcji metainformacyjnej.

Przedmiotem rozważań były tu tylko niektóre, wybrane aspekty terminu⁴¹. Nie wyczerpuje to całokształtu problematyki znaczenia i statusu terminu w systemie leksykalnym języka informacyjno-wyszukiwawczego. Dla pełności obrazu konieczne jest uwzględnienie przy analizie znaczenia terminu stosunku terminu do innych terminów w systemie terminologicznym, w całym zasobie terminów i poza nim oraz do całego systemu językowego. Rozważania nad miejscem terminologii w systemach leksykalnych języków informacyjno-wyszukiwawczych są elementem poznawczej orientacji badawczej, nazywanej organizacją wiedzy, stanowiącej po-

⁴⁰ Kategorie rzeczownikowe (rzeczowniki, frazy nominalne) jako etykiety denotują od razu całą wiązkę cech a nie pojedynczą cechę. Poza tym wiążą się silniej z wyobraźnią, umożliwiając tworzenie odpowiednich obrazów mentalnych.

⁴¹ Terminem jest nazywana również elementarna jednostka leksykalna języka informacyjno-wyszukiwawczego. Niektórzy autorzy prac z zakresu systemów wyszukiwania informacji traktują „termin” bardzo szeroko, jako synonim jednostki leksykalnej języków paranaturalnych. Poglądy takie reprezentuje m.in. Ewa Chmielewska-Gorczyca. Należy jednak dostrzec wyraźne różnice w rozumieniu tego wyrażenia. W systemie leksykalnym języka informacyjno-wyszukiwawczego zakres terminu nie pokrywa się z zakresem terminu w terminologii (systemie terminologicznym). Plan wyrażania jest ten sam, ale plany treści różnią się, gdyż wyrażenia te pełnią różne funkcje i w związku z tym mają różne denotacje.

moc w rozwiązywaniu teoretycznych i praktycznych problemów teorii i praktyki języków informacyjno-wyszukiwawczych.

1.4.2. Pole semantyczne języka słów kluczowych a języka naturalnego

Kolejnym elementem definicyjnym tożsamości języka słów kluczowych jest pole semantyczne. Celem jego opisu jest określenie podobieństw i różnic pomiędzy polem semantycznym wymienionych rodzajów języków, a także wyjaśnienie ich źródeł oraz konsekwencji. W dalszej części rozważań zamierzam wykazać, że istnieje pewna odpowiedniość między polem semantycznym języka naturalnego a polem semantycznym języka słów kluczowych.

1.4.2.1. Charakterystyka pola semantycznego języka naturalnego

W polskim piśmiennictwie na temat pola semantycznego języka informacyjno-wyszukiwawczego i języka naturalnego pisały m.in. Elżbieta Artowicz, Barbara Sosińska-Kalata oraz Bożenna Bojar. Zajęcie się tym problemem uważam za ważne, gdyż stanowi jeden z elementów określających tożsamość języka. Definicja pola semantycznego języka określa pole semantyczne danego języka naturalnego jako ustrukturalizowany zbiór cech obiektów realnej rzeczywistości (uniwersum), którego podzbiorem (wiązkom) można przyporządkować, zgodnie z regułami semantyki danego języka naturalnego, wyrażenia tego języka. Cechy te są porządkowane według różnych kryteriów, tworząc wielopoziomowe struktury wyrażane w systemie leksykalnym języka naturalnego za pomocą różnych typów relacji paradygmatycznych i syntagmatycznych. Język naturalny odwzorowuje całą rzeczywistość, stosownie do pełnionych przezeń funkcji. Każdy język naturalny odwzorowuje rzeczywistość w sposób wielo-wieloznaczny. Nie ma więc tu wzajemnie jednoznacznego przyporządkowania pomiędzy elementami języka i elementami odwzorowywanej przezeń rzeczywistości. Znajduje to wyraz w systemie paradygmatycznym danego języka naturalnego.

Język naturalny pełni funkcję klasyfikowania elementów rzeczywistości. Każdy język dzieli tę rzeczywistość według swoich własnych kryteriów i dlatego języki naturalne są właściwie nieporównywalne (na przykład różny podział wyrazu „śnieg” w języku fińskim i w języku polskim). Służy on również do wydzielania tych elementów, dzielenia ich na klasy i nadawania tym klasom nazw. W samej rzeczywistości najczęściej brak wyraźnych granic między tymi klasami. Ich ustalenie jest każdorazowo wynikiem arbitralnej decyzji klasyfikującego, tj. użytkownika języka, stąd elementy rzeczywistości są grupowane w bardzo różne klasy, zależnie od konkretnej potrzeby. Użytkownik danego języka etnicznego wybiera ze skończonej liczby cech obiektu tylko te cechy istotne lub relewantne tego obiektu, których znajomość wystarczy mu do użycia danego wyrażenia języka naturalnego (Bojar 2005).

System słownikowy każdego języka naturalnego ma budowę luźną. Oznacza to, że wyrażenia języka naturalnego nie dają się ułożyć w jedno drzewo klasyfikacyjne.

Klasyfikacja ta nie spełnia więc warunków stawianych poprawnej klasyfikacji. Nie przeprowadza bowiem konsekwentnie podziału rozłącznego. W każdym języku naturalnym najwięcej jest wyrażen nazywających konkretne przedmioty. Język naturalny rozbudowuje tę klasyfikację stosunkowo rzadko, gdyż rzadko istnieje potrzeba posługiwania się wyrażeniami o wysokim stopniu abstrakcji, a więc niosącymi mało informacji. W razie potrzeby klasyfikacja ta może być rozbudowana poprzez łączenie lub dzielenie zakresów, co powoduje tworzenie się rozgałęzionych wielopoziomowych drzew klasyfikacyjnych. Tak jest w języku nauki. Przykładem takiej klasyfikacji może być systematyka roślin i zwierząt. Należy jednak zauważyć, że zakres wyrazów jest w języku naturalnym w pewnej mierze przypadkowy. Wyrazy te, włączone w drzewo klasyfikacyjne języka nauki, nabierają często innej treści, na przykład w języku potocznym zakres nazwy „zwierzę” obejmuje tylko ssaki i to z wyjątkiem ludzi (ptaki, ryby, gady, płazy to nie zwierzęta). W języku naukowym natomiast wyraz „zwierzę” jest jednostką na wyższym poziomie abstrakcji. Przykład ten ilustruje względność zasady poprawnej klasyfikacji, która wymaga, aby klasyfikacja była naturalna, tzn. aby przedmioty należące do zakresu jednego członu podziału były do siebie bardziej podobne niż przedmioty należące do różnych członów (Bojar 2005).

Jednostki języka naturalnego bardzo często nie wyznaczają zakresu odwzorowywanych przez nie zbiorów w sposób ostry i jednoznaczny. Przykładem mogą tu być wyrazy: „rzeka”, „potok”, „strumień” – którym trudno przyporządkować jednoznaczne zbiory desygnatów. Takie sytuacje są dla języka naturalnego typowe. Dokładne sprecyzowanie granic zakresu poszczególnych wyrażen jest możliwe w językach sztucznych.

Istotny postęp w zakresie tej problematyki przyniosło sformułowanie na bazie strukturalizmu teorii pola znaczeniowego, która opiera się na hipotezie, że znaczenia poszczególnych wyrazów są strukturalizowane poprzez wzajemne powiązania wewnątrz odrębnych sfer, zwanych polami znaczeniowymi, pokrywającymi w całości obszar doświadczenia poznawczego człowieka.

Pole semantyczne języka naturalnego może też być rozpatrywane na różnych poziomach. Jeśli chodzi o jego wewnętrzną strukturę, to nie ma ona mozaikowego charakteru – jak uważali niektórzy teoretycy pola semantycznego języka naturalnego – polegającego na ścisłym, jednoznacznym odgraniczaniu zakresu jego pól, lecz zachodzą one na siebie wzajemnie, więc ich granice są zmienne. Pole semantyczne języka naturalnego składa się więc z przenikających się wzajemnie i zespalających się pól elementarnych. W wypadku języka naturalnego trzeba też przyjąć równoczesność rozwoju planu treści i planu wyrażania, gdyż język naturalny stale się rozwija i w miarę powstawania nowych obiektów tworzy dla nich nowe nazwy.

1.4.2.2. Charakterystyka pola semantycznego języków słów kluczowych

Języki słów kluczowych, odpowiednio do funkcji i zakresu tych języków, odwzorowują pewien fragment rzeczywistości, zwany rzeczywistością dokumentacyjną. Wspecjalizowaną funkcją tych języków jest opis dokumentów, czyli utrwalonych

informacji oraz formułowanie pytań użytkowników, co umożliwi wyszukiwanie potrzebnych użytkownikom informacji.

Pole semantyczne języka słów kluczowych możemy określić jako zbiór cech rzeczywistości dokumentacyjnej relewantnych dla (potencjalnego) użytkownika systemu informacyjno-wyszukiwawczego, którego elementom (dokumentom) można przyporządkować, zgodnie z regułami semantyki danego języka, słowa kluczowe. Kryteria doboru tych cech są zależne od konkretnego systemu. Wyznaczają je własności treści dokumentów rejestrowanych w danym systemie informacyjno-wyszukiwawczym, czyli atrybuty zawartej w nich informacji oraz potrzeby informacyjne użytkowników.

Pole semantyczne języka słów kluczowych zależy od rodzaju/odmiany tego języka: w języku swobodnych słów kluczowych rozwija się spontanicznie, jak w języku naturalnym, natomiast w sztucznym języku słów kluczowych jest projektowane i budowane zgodnie z przyjętymi regułami, a następnie uzupełniane w sposób zorganizowany, obejmując swoim zakresem ściśle określony obszar. Cel i przeznaczenie języka słów kluczowych określają więc zakres jego pola semantycznego, które jest po prostu sumą zakresów jego wyrażenń wyznaczonych na podstawie ich konotacji. Zakres ten jest mniejszy od zakresu pola semantycznego języka naturalnego – tym mniejszy, im bardziej wyspecjalizowany jest język informacyjny.

Cechą odróżniającą wyrażenia języka słów kluczowych od wyrażenń języka naturalnego jest to, że te ostatnie denotują bezpośrednio elementy rzeczywistości (uniwersum), natomiast jednostki leksykalne języka słów kluczowych denotują zbiory dokumentów, zapisanych informacji o elementach uniwersum, dzielone na odpowiednie podzbiory w procesie wyszukiwania informacji, dlatego relacje semantyczne w tym języku jedynie nawiązują do relacji między zakresami nazw w języku naturalnym. W rzeczywistości są to relacje zachodzące między zbiorami dokumentów zindeksowanych za pomocą słów kluczowych.

Biorąc pod uwagę relacje, jakie zachodzą pomiędzy słowami kluczowymi, trzeba zaznaczyć, że są one najczęściej przeniesieniem relacji paradygmatycznych językowych i relacji pozajęzykowych w inną rzeczywistość, a mianowicie w rzeczywistość dokumentacyjną. Zagadnienia te doczekały się precyzyjniejszych rozstrzygnięć na gruncie teorii języków informacyjno-wyszukiwawczych. Akceptowane przez wszystkich podejście do zagadnienia relacji w językach informacyjno-wyszukiwawczych zaproponowała B. Bojar. Wyszła od stwierdzenia, że języki informacyjno-wyszukiwawcze są przeznaczone do zupełnie innego celu aniżeli języki naturalne. Służą do indeksowania dokumentów i pytań informacyjnych, a desygnatami ich jednostek leksykalnych nie są określone wycinki rzeczywistości, lecz teksty dokumentów. Z tego wynika, że na przykład rozpatrując relacje generyczne, należy interpretować je nie dosłownie jako relacje między zakresami nazw, lecz jako relacje między zbiorami dokumentów zaindeksowanych za pomocą tych nazw. Tak więc relacje między zakresami nazw są wykorzystywane do oznaczania relacji między zbiorami dokumentów. Ta dokumentacyjna interpretacja relacji w językach informacyjno-wyszukiwawczych jest szczególnie przydatna przy rozpatrywaniu relacji między podstawowymi przedmiotami zainteresowań poszczególnych dziedzin a tymi dzie-

dzinami (Bojar 2005). Na wspomnianą dokumentacyjną interpretację relacji w językach informacyjno-wyszukiwawczych powołują się w swoich pracach m.in. Olgierd Ungurian i Eugeniusz Ścibor.

Jak widać, semantyka języka słów kluczowych może się różnić od semantyki języka naturalnego, na którego słownictwie może być oparty słownik słów kluczowych. Nawet w przypadku identyczności zasobów leksykalnych mamy do czynienia z dwoma różnymi językami, gdyż elementom leksykalnym języka słów kluczowych przyporządkowuje się klasy dokumentów, w których charakterystykach wyszukiwawczych wyrazy te występują. Pomiedzy jednostkami języka słów kluczowych a jednostkami języka naturalnego zachodzą też różnice w planie wyrażania:

1. słowa kluczowe mają tylko jedną formę – jeśli ma postać rzeczownika, to jest podana w mianowniku liczby pojedynczej lub mnogiej i nie jest odmieniana przez przypadki i liczby, co występuje na przykład w wyrazach języka polskiego będących rzeczownikami;
2. wielowyrazowe słowa kluczowe stanowią nierozdzieloną całość, nie podlegają podziałowi na wyrazy składowe (na przykład słowo kluczowe „biblioteka główna” nie może być rozbite na wyrazy „biblioteka” i „główna”) i z punktu widzenia reguł gramatycznych języka słów kluczowych są traktowane tak samo jak każde inne słowo kluczowe. Zasób leksykalny języków słów kluczowych jest znacznie uboższy od zasobu leksykalnego języka naturalnego, z którego dany język czerpie swoje słownictwo, na przykład z języka polskiego; tylko bowiem niektóre wyrażenia występujące w języku naturalnym są potrzebne do odtwarzania treści i formy dokumentów w danym systemie wyszukiwawczym nawet wtedy, gdy system ten ma zakres uniwersalny, tj. obejmuje wszystkie dziedziny nauki i działalności praktycznej. Słowa kluczowe ograniczają się w zasadzie do informacji jednego rodzaju, a mianowicie metainformacji o treści dokumentu.

Aby język słów kluczowych mógł efektywnie pełnić swoje wyspecjalizowane funkcje, powinien odznaczać się określonymi cechami. W szczególności należy pozbyć go synonimii i homonimii (wieloznaczności), charakterystycznych dla języka naturalnego, a przynajmniej wyeliminować negatywne skutki tych zjawisk.

W języku naturalnym, bazującym na kompetencji semantycznej człowieka, często brak formalnych wykładników wskazujących na rodzaje powiązań pomiędzy poszczególnymi elementami tego języka. Prowadzi to czasami do wieloznaczności w rodzaju wyrażenia: „starzy mężczyźni i kobiety”, mającego dwie różne możliwości interpretacyjne. Innym rodzajem wieloznaczności możliwej do wyeliminowania w językach słów kluczowych (na przykład poprzez wprowadzenie nawiasów) jest wieloznaczność wynikająca z braku ustalenia preferencji łączenia: na przykład wyrażenie języka naturalnego „sód i potas lub cynk” może mieć w języku słów kluczowych przypisane dwie różne struktury. W języku naturalnym o poprawnym znaczeniu takich struktur wieloznacznych mówi kontekst, akcent lub wiedza pozajęzykowa – komputer (maszyna) takiej kompetencji nie ma (nie zawsze zresztą tę kompetencję ma człowiek). W systemie informacyjno-wyszukiwawczym nie wyszukujemy według znaczeń wyrazów, lecz według ich cech formalnych, tzn. kolejnych znaków tworzących daną jednostkę leksykalną.

Pole semantyczne języków naturalnych nie ma charakteru statycznego, bo tworzy się i rozwija w sposób żywiołowy (naturalny), natomiast pole semantyczne języka słów kluczowych może być sztuczne, tzn. budowane i modyfikowane (projektowane) w sposób zorganizowany, a jakiegokolwiek zmiany w tym języku mogą być wprowadzone tylko przez kompetentny organ. I jeszcze jedna istotna różnica: pole semantyczne języka słów kluczowych jest prezentowane tylko za pomocą rzeczowników i fraz rzeczownikowych, podczas gdy pole semantyczne języka naturalnego jest odwzorowywane przez wyrażenia o różnych kategoriach gramatycznych.

Chociaż różnice pomiędzy polami semantycznymi języka słów kluczowych i języka naturalnego mają swoje konsekwencje w systemach paradygmatycznym i syntagmatycznym, to jednak pole semantyczne języka słów kluczowych powinno być modelowane na polu semantycznym języka naturalnego w tym sensie, że jego struktura powinna wykorzystywać w możliwie szerokim zakresie strukturę pola semantycznego języka naturalnego. Pozwoliłoby to użytkownikom systemu informacyjno-wyszukiwawczego, stosującym język słów kluczowych w procesie wyszukiwania, w pewnym stopniu wykorzystywać znajomość systemu języka naturalnego, co dzięki możliwości wyszukiwania asocjacyjnego uczyniłoby te systemy bardziej przyjaznymi dla użytkownika.

1.4.3. Modelowanie języka słów kluczowych na bazie języka naturalnego

1.4.3.1. Relacje kojarzeniowe w języku naturalnym i języku słów kluczowych

Intensywne badania nad strukturami kojarzeniowymi człowieka są prowadzone od początku XX wieku. Sprowadzają się one do wykrywania różnego rodzaju związków, które skutkują powstawaniem pól kojarzeniowych, dających możliwość badania struktury leksykonu umysłowego człowieka. Centralną część takiego pola kojarzeniowego stanowią stałe związki między słowami. Pozostałe, przypadkowe związki wyrażają reakcje indywidualne i sytuacyjne.

Już w początkowej fazie rozwoju systemów wyszukiwania informacji podejmowano próby przeniesienia całej złożoności struktur kojarzeniowych języków naturalnych do języków informacyjno-wyszukiwawczych. Jednakże kwestionowanie przez niektórych językoznawców uniwersalności mechanizmu kojarzenia sprawiło, że struktury te coraz bardziej upraszczano, co wpłynęło na dotychczasowe rozumienie relacji kojarzeniowych. Przyjmuje się, że relacjami kojarzeniowymi są te wszystkie relacje paradygmatyczne, które nie mieszczą się w ramach przyjętego układu, a więc nie mieszczą się w grupie logicznych relacji równoznaczności lub podrzędności i nadrzędności zakresowej. Poza problemami terminologicznymi pojawiły się istotne różnice w interpretacji tych relacji.

W Polsce o relacjach kojarzeniowych pisała B. Bojar. W kontekście budowy tezaursów zajmowała się nimi Barbara Wereszczyńska-Cisło, a relacje syntagmatycz-

ne badała Ewa Chmielewska-Gorczyca. W 1989 roku w „Zagadnieniach Informatyki Naukowej” ukazał się też tekst Leonida V. Sacharnego, na który się powołuję, dotyczący struktur kojarzeniowych w języku naturalnym i języku informacyjno-wyszukiwawczym, ale rozpatrywanych tylko w kontekście języka haseł przedmiotowych (Sacharnyj 1989). W odniesieniu do języków słów kluczowych nikt się tym tematem do tej pory nie zajmował, bo w klasycznych językach słów kluczowych nie ma struktur kojarzeniowych, gdyż są to języki bez własnej paradygmatyki. Jeżeli jednak języki słów kluczowych mają być przyjazne dla użytkownika, to niezbędne jest przyjęcie w nich tej samej podstawy procesów kojarzeniowych co w językach naturalnych. To bardzo ważne dla tej klasy języków.

W badaniach nad charakterem procesów kojarzeniowych eksponowany jest zazwyczaj semantyczny aspekt relacji między bodźcami słownymi a przyporządkowanymi im asocjacjami. Niewątpliwie czynnikiem ukierunkowującym przebieg tych procesów są związki znaczeniowe i to one w głównej mierze decydują o rodzaju powstałego w pamięci semantycznej skojarzenia. Ostateczny wybór dokonuje się jednak przy wykorzystaniu informacji o zróżnicowaniu wyrazów pod względem ich częstości. Spośród słów, które w sensie semantycznym kojarzą się z danym bodźcem, najczęściej podawane są te charakteryzujące się największym współczynnikiem frekwencji.

Biorąc pod uwagę treść wyrażenia stanowiącą podstawę skojarzeń, wyróżnić można:

- skojarzenia paradygmatyczne, w znaczeniu używanym przez psychologów i psycholingwistów – jako synonim terminu „skojarzenia logiczne”, w których wyrażenia są powiązane relacją synonimii lub antonimii, relacją rodzaj–gatunek lub gatunek–rodzaj;
- skojarzenia syntagmatyczne, w których wyrażenia są związane relacją współwystępowania w formach słownikowych lub tekstowych;
- skojarzenia tematyczne, w których wyrażenia występują jako słowa kluczowe charakteryzujące jakąś sytuację.

Relacje asocjacyjne w języku są więc relacjami zachodzącymi między wyrażeniami podobnymi pod względem formy lub treści albo między wyrażeniami występującymi często razem w jednym kontekście. Niezależnie od różnych interpretacji samej natury relacji kojarzeniowych przyjmuje się, że są to związki między wyrażeniami językowymi. Związki te mogą mieć charakter bardziej lub mniej bezpośredni, jawny lub niejawny, skondensowany w znaczeniu wyrażenia.

Dotychczasowe badania eksperymentalne w znacznym stopniu odsłoniły sam mechanizm kojarzenia. Wykryte związki odpowiadają wyróżnionym rodzajom skojarzeń. To, w jaki sposób związki te odpowiadają konkretnym sytuacjom, szczegółowo opisał Leonid V. Sacharnyj (Sacharnyj 1989). Jego uwagi wykorzystam w dalszych rozważaniach. Związki te stanowią podstawę budowy różnego rodzaju odesłań.

Wykryte związki odpowiadają trzem grupom sytuacji:

- wyrażenie wydaje się odbiorcy jako ogólne oznaczenie jakiejś sytuacji (lub typu sytuacji) jednym wyrażeniem. Faktycznie jest to wyrażenie/zdanie/tekst. Pytany, w odpowiedzi na wyrażenie S, używa słowa R, będącego synonimem S, na przykład: budynek – dom – kamienica. W rzeczywistości są to relacje paradygmatyczne, wyrażające logiczne zależności synonimii, antonimii, rodzaj–gatunek,

gatunek–rodzaj. Eksperymenty wykazały, że zdania zbudowane z takich wyrażen połączonych relacjami paradygmatycznymi w 2/3 przypadków wyrażają struktury, które nie są spotykane w innych typach skojarzeń. Są to następujące struktury: poszczególne wyrazy paradygmatycznej pary w zdaniu stanowią oddzielne zdania/teksty, a właściwie metateksty, opisujące inną, bardziej ogólną sytuację niż ta, która jest opisana za pomocą oddzielnych wyrazów, przy czym wyrazy te mogą być traktowane jako synonimy sytuacyjne lub przeciwstawienia (antonimia) sytuacyjne, na przykład: osiedle – mieszkanie. Takie przeciwstawienia mogą zachodzić zarówno w poziomie (stosunki koordynacji, współzależność), jak i w pionie (stosunki nadrzędności i podrzędności).

- wyrażenie wydaje się odbiorcy częścią znaczenia frazy użytej na określenie jednej sytuacji (lub typu sytuacji). Wtedy wskazuje on na drugi wyraz, który jakby dopełnia go i razem w sensie formalnym stanowią frazę językową (syntagmę), na przykład: woda – woda pitna. Stąd nazwa tego typu skojarzeń: skojarzenia syntagmatyczne (składniowe). W zdaniach frazy językowe nazywają jakiś fragment opisywanej sytuacji. Często są to stałe związki frazeologiczne, ale też mogą być połączenia doraźne. Tego typu struktury kojarzeniowe dotyczą nie tylko pojedynczych wyrazów, ale i większych struktur składniowych (fraz, zdań). Ze względu na formę analityczną wyrażają one dokładniej zróżnicowanie rodzajowe sytuacji oznaczanej danym wyrazem. Warto zauważyć, że na takich skojarzeniach opiera się większość wielowyzrazowych słów kluczowych lub słów kluczowych mających postać fraz językowych, a także budowa hasel przedmiotowych wielowyzrazowych i rozwiniętych.
- wyrażenie jest częścią analitycznie podzielonego wielowyzrazowego oznaczenia sytuacji (lub typu sytuacji). Dany wyraz jest słowem kluczowym dla danej sytuacji najważniejszym, najbardziej trafnym. Często w celu opisu jakiejś sytuacji niezbędne jest użycie jako minimalnego zestawu słów kluczowych nie jednego wyrażenia, lecz kilku, na przykład: Kraków – Wawel. Tego typu skojarzenia nazywane są skojarzeniami tematycznymi. W rzeczywistości odzwierciedlają relacje między danym słowem kluczowym a zestawem słów kluczowych, tj. pozostałymi słowami kluczowymi.

Strefa utożsamiania lub przeciwstawiania wyrażen realizuje się w zależności od oceny użytkownika, tak jak w relacjach paradygmatycznych lub w relacjach syntagmatycznych, bądź jak w relacjach tematycznych. Na przykład, jeżeli użytkownik korzysta z systemu odsyłaczy, to tym samym uznaje w tej konkretnej sytuacji wyrazy za quasi-synonimy, tj. wyrazy, których znaczenie jest tożsame. W czasie wyszukiwania może natomiast wystąpić sytuacja, że te same wyrazy raz mogą być traktowane jako identyczne, a innym razem jako przeciwstawne. Takie podejście do struktur kojarzeniowych opiera się na przyjęciu tezy, że w językach informacyjno-wyszukiwawczych charakter kojarzeniowy mają nie tylko te relacje, które nazywa się *ex definitione* kojarzeniowymi, lecz także te, które nazywane są relacjami paradygmatycznymi i w ogóle wszelkie relacje, w tym syntagmatyczne. Oznacza to, że wszelkie relacje językowe mają naturę kojarzeniową.

Pole kojarzeniowe to nie prosty zbiór/zestaw wyrazów, gdyż relacje kojarzeniowe mają charakter zarówno systemowy, jak i tekstowy. W zależności od typu tekstu powstają różne typy struktur kojarzeniowych. Struktury kojarzeniowe aktualizują się z różną częstotliwością/siłą. Dlatego w polu kojarzeniowym można wydzielić część centralną (jądro), w której występują najczęstsze struktury (stereotypowe), i część peryferyjną, w której występują struktury doraźne, sytuacyjne.

Niezależnie od relacji leksykalnych wiele skojarzeń typowych można połączyć w pewne grupy według kryterium tematycznego. Inne skojarzenia można grupować według stopnia ich związku z desygnatem nazwy: niektóre nawiązują bezpośrednio do desygnatu i jego cech (na przykład: uniwersytet – studenci), w przypadku innych związek z desygnatem jest luźniejszy (na przykład: smar – samochód).

Dla rozważań dotyczących języków słów kluczowych przydatne są również następujące uwagi L.V. Sacharnego dotyczące pola kojarzeniowego języków naturalnych:

- złożoność analizy struktury pola kojarzeniowego języka naturalnego jest związana z częstością występowania struktur. Od centrum do peryferii często się zmniejsza, ale też nie jest możliwe dokonanie ostrych jej rozgraniczeń. Trudno jest określić granice peryferii;
- ranga struktur kojarzeniowych może zmieniać się w zależności od grupy użytkowników;
- charakter relacji paradygmatycznych między składnikami pola kojarzeniowego w językach naturalnych nie jest ostry i stały, lecz zmienny i zależny od sytuacji.

Zdaniem Bojar: „Relacje paradygmatyczne należą do szeroko rozumianych relacji asocjacyjnych, czyli relacji skojarzeniowych. W umyśle użytkownika języka to właśnie te wyrażenia najczęściej się z sobą kojarzą, struktura paradygmatyczna języka jest bowiem ważną częścią kompetencji językowej, a do orzeczenia o tym, jakiego typu relacje paradygmatyczne łączą dane wyrażenia, wystarcza tylko znajomość systemu językowego” (Bojar 2005, s. 182).

W systemie paradygmatycznym języka słów kluczowych „pion” został zredukowany niejako do zera, stąd brak w tych językach relacji nadrzędności i podrzędności. W niektórych językach słów kluczowych dopuszcza się występowanie wymienionych typów relacji, ale wówczas mamy do czynienia z nieklasycznymi językami słów kluczowych.

Uważam, że poszczególne rodzaje języków słów kluczowych opierają się na różnych typach skojarzeń: od prostych skojarzeń swobodnych (języki swobodnych słów kluczowych), poprzez skojarzenia kierowane (języki słów kluczowych oparte na słownikach wzorcowych słów kluczowych), po skojarzenia ciągłe (języki słów kluczowych oparte na słownikach kontrolowanych, zwłaszcza wzorcowych słownikach słów kluczowych).

Relacje kojarzeniowe to podstawowe elementy struktury zapamiętywania, przechowywania i przetwarzania informacji językowej/leksykalnej. Wynika to z badań kognitywnych ludzkiego mózgu dotyczących zapamiętywania i przetwarzania informacji. Już Ferdinand de Saussure pisał o skojarzeniach jako podstawie relacji językowych. Relacje hierarchiczne uważam za szczególnie przypadek relacji koja-

rzeniowych, gdyż opierają się na pewnej klasie skojarzeń. Relacje kojarzeniowe są podstawą sieci, z wyróżnianiem węzłów i linków, a system to szczególnie przypadek sieci, która odwzorowuje to, co dzieje się w mózgu. Sam system oddaje tylko część tego procesu. Jest to podejście kognitywne, a zarazem sieciowe w teorii języków informacyjno-wyszukiwawczych.

1.4.3.2. Kompetencja językowa użytkowników języka słów kluczowych i jej determinanty

W dalszym ciągu zajmę się kompetencją językową dotyczącą języka naturalnego i języka słów kluczowych. Według definicji słownikowej kompetencja językowa jest to „opanowanie przez użytkownika języka systemu języka, umożliwiające mu tworzenie i rozpoznawanie tekstów tego języka” (Bojar 2002, oprac., s. 140).

Termin ten pojawił się w piśmiennictwie lingwistycznym na przełomie lat pięćdziesiątych i sześćdziesiątych XX wieku i został wprowadzony przez Noama Chomsky’ego, dla którego język to kompetencja niezbędna do posługiwania się językiem. Strukturaliści traktują język jako strukturę, która istnieje niezależnie od użytkownika i którą da się opisać. Językoznawcy kognitywni dodatkowo zwracają uwagę na zależność ludzkiego poznania od doświadczenia oraz na fakt, że w procesie posługiwania się językiem, oprócz wiedzy językowej, ważną rolę odgrywa wiedza o świecie, z której korzystają zarówno nadawca, jak i odbiorca komunikatu językowego, a która stanowi podstawę ich wzajemnego zrozumienia.

Kompetencja językowa jest umiejętnością stopniowalną. W przypadku języka naturalnego pozwala ona w większym lub mniejszym stopniu, w zależności od stopnia jego opanowania, budować nowe struktury (zdania/teksty). Również w odniesieniu do języków sztucznych, jakimi są języki informacyjno-wyszukiwawcze, kompetencja językowa może być zerowa lub maksymalna. Co więc kształtuje kompetencję językową użytkownika języka słów kluczowych? Aby odpowiedzieć na to pytanie, trzeba dostrzec związek kompetencji językowej w zakresie języka słów kluczowych z kompetencją w zakresie języka naturalnego, wiedzą pozajęzykową (wiedza o rzeczywistości i wiedza dokumentacyjna), cechami języka, w tym wiedzą o sposobach prezentacji języka, a także jego użyciem⁴².

Kompetencja językowa różnych kategorii użytkowników, w zależności od tego, czy ma się do czynienia z użytkownikiem-twórcą, czy użytkownikiem końcowym, jest różna. W odniesieniu do języków słów kluczowych można mówić wyłącznie o kompetencji gramatycznej i kompetencji semantycznej. Kompetencja gramatyczna dotyczy stopnia opanowania zasad budowy struktur złożonych, głównie tworzenia zdań i tekstów oraz ich interpretacji. Kompetencja semantyczna dotyczy stopnia opanowania znaczenia poszczególnych struktur językowych, w tym znaczenia jednostek leksykalnych, całych zdań oraz tekstów w danym języku.

⁴² O związkach między językiem naturalnym a językami informacyjno-wyszukiwawczymi w kontekście kompetencji językowej wielokrotnie pisała B. Bojar.

Do rozpoznawania znaczeń wyrażeń języka informacyjno-wyszukiawczego potrzebny jest kontekst paradygmatyczny, który określa/definiują relacje, w jakich występuje dane wyrażenie wobec innych wyrażeń w strukturze języka (w słowniku). Relacje te są specyfikowane w strukturze artykułu słownikowego, a więc na poziomie mikrostruktury słownika. Dla użytkownika istotnym czynnikiem ułatwiającym opanowanie języka informacyjno-wyszukiawczego są reguły analogii, dzięki którym język ten staje się bardziej przejrzysty i przewidywalny dla użytkownika.

Na stopień opanowania języka ma wpływ sama struktura języka, wielkość jego leksyki (słownik), reguły gramatyczne (budowa wyrażeń złożonych), analogie rozwiązań stosowanych w języku. Elementy te z jednej strony powinny zapewnić możliwość odzwierciedlenia w danym języku różnych elementów treści dokumentu, zaś z drugiej – powinny być dostosowane do możliwości percepcyjnych użytkownika.

Istotnym elementem kompetencji językowej jest wiedza o świecie, jej uporządkowanie i kategoryzacja, która umożliwia poznanie znaczenia konstrukcji językowych. W przypadku języków informacyjno-wyszukiawczych (JIW) wiedza pozajęzykowa jest wiedzą zawartą w treści dokumentów. Jeżeli użytkownik ma odpowiednią wiedzę z jakiejś dziedziny, to łatwiej mu interpretować znaczenie jednostek leksykalnych oraz odtworzyć siatkę znaczeń/pól semantycznych tworzących zakres słownictwa danej dziedziny wiedzy i jej poszczególnych zagadnień. „Uczenie się JIW – jak pisze Jadwiga Sadowska – polega więc m.in. na porównywaniu wzorca wiedzy o rzeczywistości pozajęzykowej z wzorcem jej uporządkowania za pomocą jakiegoś JIW. Im bardziej wypełniony jest wzorzec wiedzy o rzeczywistości w umyśle użytkownika, tym łatwiej i szybciej zdobywa on kompetencję w zakresie JIW” (Sadowska 2010).

Kompetencja w zakresie języka informacyjno-wyszukiawczego ma charakter nabyty. Wpływają na to następujące czynniki:

- wiedza o semantyce języka naturalnego;
- wiedza o strukturze i zasobie leksykalnym języka informacyjno-wyszukiawczego;
- wiedza o rzeczywistości pozajęzykowej (o świecie, danej dziedzinie).

Kompetencja językowa w zakresie języka słów kluczowych jest przede wszystkim pochodną kompetencji językowej w zakresie języków naturalnych, w tym języków specjalistycznych (terminologii) oraz pochodną wiedzy o rzeczywistości językowej (o świecie, o danej dziedzinie). Czynnikiem zwiększającym kompetencję jest dodatkowo doświadczenie związane z posługiwaniem się tym językiem.

W przypadku języków słów kluczowych zarówno ich twórcy, jak i użytkownicy korzystają ze znajomości języka naturalnego, wykorzystując swoją wiedzę o języku naturalnym – zwłaszcza dotyczącą znaczenia wyrażeń – do poznawania języków informacyjno-wyszukiawczych. Klasyczne języki słów kluczowych stanowią krańcowy przypadek, w którym słownik języka informacyjno-wyszukiawczego nie zawiera objaśnień znaczeń jednostek leksykalnych. Domyślnie zakłada się wykorzystanie przez użytkownika części, bo ograniczonej do odpowiednich kategorii gramatycznych, leksyki języka naturalnego, z wyjątkiem tzw. stop-listy. To założenie o wykorzystaniu znajomości znaczeń leksyki języka naturalnego w języku informacyjno-wyszukiawczym nie zawsze przekłada się na konkretnie przyjęte rozwiązania.

Języki słów kluczowych są najprostszą formą paranaturalnych języków informacyjno-wyszukiwawczych. Słowa kluczowe są równokształtne z wyrażeniami języka naturalnego. Znaczenie słów kluczowych nie jest podawane *explicite*. Zakłada się, że użytkownik tego języka:

- sam je odkryje na podstawie swojej kompetencji językowej z zakresu odpowiedniego języka naturalnego;
- sam odkryje zasadę, że słowa kluczowe odwzorowują/są etykietami⁴³ elementów treści dokumentów, a nie rzeczywistości pozajęzykowej/pozatekstowej, a więc że ich znaczenie odpowiada równokształtnym z nimi wyrażeniom języka naturalnego użytym w funkcji metainformacyjnej.

Język słów kluczowych może być prezentowany w postaci jawnego słownika/listy słów kluczowych nieswobodnych, znormalizowanych, ale tylko w zakresie formy wyrażen. Znaczenie (treść) słów kluczowych jest zdeterminowane kompetencją językową jego użytkowników w zakresie semantyki odpowiedniego języka naturalnego. System leksykalny języka słów kluczowych może też być prezentowany za pomocą słownika negatywnego (tzw. stop-listy), czyli spisu wyrażen danego języka naturalnego, które nie stanowią wyrażen języka słów kluczowych. Jeszcze innym rodzajem prezentacji języka jest zestaw kryteriów, jakie powinny spełniać wyrażenia danego języka naturalnego, które mogą być użyte w funkcji słów kluczowych. Zestaw tych kryteriów nazywa się słownikiem zadaniem instrukcją (słownikiem indeksowania). Ten sposób prezentacji słownictwa jest właściwy dla języków swobodnych słów kluczowych. W tym przypadku język słów kluczowych w ogóle nie jest prezentowany, lecz stanowi część kompetencji językowej w zakresie języka naturalnego poszczególnych użytkowników języka słów kluczowych, tj. ich kompetencji leksykalnej i semantycznej, w tym kompetencji dotyczącej paradygmatyki systemu semantycznego.

Kompetencja użytkownika języka swobodnych słów kluczowych w zakresie tego języka kształtuje się w trakcie używania (język indeksowania) i/lub metodą „prób i błędów” w procesach wyszukiwania (język wyszukiwawczy). Różnice w kompetencji językowej poszczególnych użytkowników języka swobodnych słów kluczowych mogą być duże, szczególnie kiedy bazą słownikową tego języka nie jest język ojczysty, jak w przypadku poszukiwań internetowych, lub gdy stosowany w danym systemie język słów kluczowych jest inny niż język reprezentowanych przez te słowa kluczowe dokumentów.

Wprawdzie od pięćdziesięciu lat trwają prace nad zaopatrzeniem komputera w kompetencję językową i możliwość komputerowego jej modelowania, co znacznie ułatwiłoby kontakt człowieka z systemem informacyjno-wyszukiwawczym za pomocą języka słów kluczowych, ale jak do tej pory ten nurt badań w zakresie prze-

⁴³ Słowa kluczowe to swoistego rodzaju etykiety fragmentów rzeczywistości opisywanej w dokumentach. Etykieta jest wyrażeniem pełniącym funkcję adresu względem zestawu danych (danej treści) umożliwiającej jego identyfikację (Bojar 2002, oprac.). Używając słów jako etykiet kategorii tematycznych, ryzykuje się, że różni użytkownicy mogą je różnie rozumieć, tzn. przypisywać im różną konotację i denotację.

tworzenia języka naturalnego – poza próbą wyposażenia tzw. agentów wyszukiwawczych w kompetencję językową – stanowi jedno z wielkich wyzwań przyszłości⁴⁴.

1.4.3.3. Struktury języka słów kluczowych jako efekt modelowania struktur języka naturalnego

Ten, kto włada językiem, włada też związkami paradygmatycznymi, które występują między elementami pola kojarzeniowego języka naturalnego, przede wszystkim jego centralnej części. Elementy te mogą być strukturami niepodzielnymi (wyrazami), jak i podzielnymi (frazami językowymi lub zestawami słów kluczowych), dlatego w trakcie wyszukiwania informacji użytkownik może aktualizować odpowiednie pole kojarzeniowe (zdanie, tekst), przede wszystkim jego centralną część i identyfikować każdy lub tylko poszczególne składniki pola kojarzeniowego, które mogą podlegać peryfrazie w ramach relacji paradygmatycznych tego pola. Oznacza to, że jednostki leksykalne języka słów kluczowych lub ich kombinacje, o ile występują w języku, są wykorzystywane w procesie wyszukiwania informacji jako składnik odpowiednich pól kojarzeniowych.

Znajomość standardowych struktur pola kojarzeniowego z możliwością zastosowania peryfrazy to znajomość języka większa niż znajomość tylko zasobu leksykalnego. W komunikowaniu się za pomocą języka naturalnego związki paradygmatyczne w polu kojarzeniowym (semantycznym) mają charakter dynamiczny i w zależności od sytuacji mogą realizować się jako tzw. quasi-synonimy lub quasi-antonimy.

Aby mógł zająć proces wyszukiwania informacji, niezbędna jest zbieżność dialogu użytkownik – indeksator. Mogą tu zaistnieć niebezpieczeństwa w postaci:

- odmienności w rozpoznawaniu i wybraniu elementów treści i ich cech;
- odmienności w nazywaniu poszczególnych elementów treści (w sposób adekwatny lub uogólniony);
- odmienności w wyrażaniu treści.

Użytkownik formułuje swoje pytanie, wychodząc od najbardziej „naturalnych”, znanych mu sposobów nazywania interesujących go obiektów. Są to zwykle sposoby zgodne z normami/zwyczajami danego języka etnicznego. Struktura tekstu, jakim jest charakterystyka wyszukiwawcza, w przypadku języka słów kluczowych zawiera jednostki leksykalne, które zwykle spotyka się w tekście dokumentu. Użytkownik ma zawsze wiele możliwości formułowania pytania.

Niezależnie od tego, jak bardzo byłby rozbudowany aparat odsyłaczowy, ma on ograniczone możliwości. Wynika to z elastyczności sformułowań użytkownika odwołującego się do języka naturalnego i schematyczności języka sztucznego, operującego ograniczonym zasobem jednostek leksykalnych i powiązań między nimi.

Formułowanie struktur kojarzeniowych w języku swobodnych słów kluczowych przebiega pod wpływem pól kojarzeniowych języka naturalnego, a w przypadku ję-

⁴⁴ W Polsce problematyką tą w ramach sztucznej inteligencji zajmuje się m.in. Zygmunt Vetulani, który w 2004 roku opublikował wyniki badań w tym zakresie. Godne uwagi są również badania zespołu pod kierunkiem Wiesława Lubaszewskiego w zakresie lingwistyki komputerowej, których rezultatem jest m.in. *Słownik fleksyjny języka polskiego* (Kraków 2001).

zyka słów kluczowych o słownictwie kontrolowanym pod wpływem gotowego słownika słów kluczowych. Realizuje się to w procesie indeksowania przez indeksatora i w czasie wyszukiwania informacji przez użytkownika. W podejściu strukturalnym tekst jest traktowany jako struktura statyczna; w podejściu funkcjonalnym w owej strukturze próbujemy dostrzec i zbadać oraz zrekonstruować dynamikę procesu komunikowania. Z punktu widzenia funkcji teksty języka słów kluczowych różnią się w istotny sposób. Charakterystyki wyszukiwawcze mają charakter statyczny, zaś instrukcje wyszukiwawcze jako struktury językowe są dynamiczne, a więc zmienne⁴⁵. Tu należy się jednak pewne wyjaśnienie. W wyszukiwaniu informacji posługujemy się jeszcze tzw. językiem wyszukiwawczym, zwanym też językiem kwerend. Jest to najczęściej język komend wyszukiwawczych oparty na etykietach struktury bazy danych, na przykład SQL, komplementarny w stosunku do języka słów kluczowych⁴⁶. Słownictwo tego języka tworzą nazwy operacji wykonywanych przez komputer w celu wyszukania informacji w zbiorze wyszukiwawczym. Język ten znajduje się jednak poza zakresem moich rozważań.

W kontekście dotychczasowych analiz niewątpliwie trafna jest następująca definicja „słowa kluczowego”, zamieszczona w *Glossary of Terms Relating to Thesauri and Other Form of Structured Vocabulary for Information Retrieval*: „Słowo kluczowe to wyrażenie lub fraza językowa występująca w języku naturalnym dokumentu, uznane za znaczące dla wyszukiwania informacji” (Will 2005)⁴⁷. W komentarzu autorskim do tej definicji Leonard Will sygnalizuje dodatkowe znaczenia tego terminu, oprócz preferowanego. Informuje, że wyrażenie to jest używane dowolnie (ang. *loosely*), przeważnie w dwóch innych możliwych znaczeniach, które są często mieszane. Używania terminu „słowa kluczowe” w tych znaczeniach należy unikać. Chodzi tu o utożsamianie słowa kluczowego z terminem ze słownika kontrolowanego, przypisanego dokumentowi oraz jako identyfikatora. Warto też zwrócić uwagę na konsultantów definicji tego terminu. Byli nimi: Stella Dextre Clarke, Alan Gilchrist i Ron Davies, angielscy specjaliści od tezaursów i innych słowników budowanych na potrzeby wyszukiwania informacji. Podana tu definicja słowa kluczowego jest spójna z definicją zamieszczoną w najnowszej brytyjskiej normie budowy tezaursów BS 8723: 2005. „Structured vocabularies for information retrieval” oraz definicją w normie amerykańskiej ANSI/NISO Z39.19 (NISO 2004). Dokładna analiza i porównanie tych definicji pozwala zauważyć, że w znaczeniu słów kluczowych następuje

⁴⁵ Niewątpliwie jest to pewnego rodzaju uproszczenie, gdyż w rzeczywistości w różnego typu językach informacyjno-wyszukiwawczych jest różnie.

⁴⁶ W poszczególnych systemach informacyjno-wyszukiwawczych mogą być stosowane różne zestawy komend. W skali światowej istnieje jeden standard określony normą ISO 8777: 1993 Commands for interactive text searching (CCL – Common Command Language). Odpowiada jej norma PN – ISO 8777: 1999. Informacja i dokumentacja. Komendy w interaktywnym wyszukiwaniu tekstowym.

⁴⁷ “Keyword – a word or phrase occurring in the natural language of a document that is considered significant for retrieval” (Will 2005).

jakby przesunięcie akcentu w kierunku ich „naturalności”, tj. grupy słów znaczących i użytecznych w procesach wyszukiwania informacji, i zmiana ich statusu z jednostek leksykalnych języka sztucznego na wyrażenia języka naturalnego wykorzystywane w specjalnej funkcji (tj. bez zmiany ich statusu). Następuje rezygnacja z operowania pojęciem statusu jednostek leksykalnych.

Celem tej części rozważań była charakterystyka słów kluczowych jako problemu badawczego w nauce o informacji, wynikającego z wielokontekstowości ujmowania i różnorodności ich odniesień i funkcji. Krytyczna analiza piśmiennictwa na ten temat pozwoliła ujawnić duże zróżnicowanie rozumienia wyrażenia „słowo kluczowe”, począwszy od języka naturalnego (literaturoznawstwo), poprzez informatykę, a skończywszy na teorii i praktyce wyszukiwania informacji. Dotyczy to także dziedzin stosowania, w tym Internetu. Są więc różne znaczenia i różne konteksty używania słów kluczowych, stanowiąc elementy teorii słowa kluczowego. Słowa kluczowe różnią się tym od innych wyrażeń języka naturalnego, że są to wyrażenia języka naturalnego, które zostały użyte w specjalnej funkcji, związanej z organizacją informacji w celu jej wyszukiwania, co powoduje zmianę rodzaju denotacji i może wywołać zmianę zakresu tych wyrażeń oraz daje możliwość ich interpretacji dokumentacyjnej. Dane wyrażenie, zwane słowem kluczowym, jest bowiem dodatkowo włączane w system językowy języka słów kluczowych lub użyte w ramach tego systemu, co nadaje mu specjalny status (konotację bezpośrednią) i nową funkcję. W sensie konotacji pośredniej (odniesienia do rzeczywistości realnej) nadal jednak pozostaje wyrażeniem języka naturalnego.

Istnieją różne odmiany słów kluczowych i pełnią one różne funkcje, zależnie od dziedziny i konkretnych potrzeb. Nawet w działalności informacyjnej, w tym w systemach informacyjno-wyszukiwawczych, ich natura nie jest jednorodna, gdyż są to zarówno wyrażenia języka naturalnego w funkcji metainformacyjnej (język swobodnych słów kluczowych), jak i wyrażenia sztucznego języka kontrolowanych słów kluczowych, to jest zaprojektowanego inną metodą niż indukcyjna. W tej sytuacji ustalenie tożsamości języka słów kluczowych nie jest zadaniem łatwym i prostym.

W nowoczesnych systemach wyszukiwawczych słowa kluczowe często pełnią funkcję łączy hipertekstowych w treści dokumentu lub jego fragmentach. Są one niezwykle proste, jeśli chodzi o ich wybór, i stanowią podstawę do wielu skojarzeń i szybkich przeskoków w sieci. Ta łatwość tworzenia/wyboru jest zarazem zaletą i wadą stosowanego indeksowania swobodnego, które jest w istocie etykietowaniem poszczególnych elementów treści dokumentu lub potrzeby informacyjnej. Ze względu na stosunkową łatwość wyboru mamy do czynienia z ich mnogością; nie są one systematycznie planowane. W wyniku indywidualnych i subiektywnych skojarzeń indeksatora stanowią element łączący poszczególne fragmenty tekstu, poszczególne dokumenty, a nawet w tzw. *semantic grid* – usługi sieciowe. Dowolność myślenia powoduje, że dane słowo kluczowe jako etykieta dokumentu nie zawsze oznacza to, co indeksator – a potem użytkownik – kojarzy z daną nazwą. Słowo kluczowe jako etykieta staje się bardziej zrozumiałe dzięki otaczającemu go kontekstowi.

Współczesny poziom rozwoju technologii komputerowej daje możliwości sieciowe, które pozwalają na wyrafinowane przedstawianie zależności leksykalnych,

dlatego należałoby zmienić paradygmat teorii tych języków z systemowego na sieciowy. Organizacja sieciowa leksyki to układ związków między jednostkami leksykalnymi języka charakteryzujący się głównie powiązaniem poziomymi. W praktyce pomiędzy elementami takiej struktury może występować bardzo duże zróżnicowanie powiązań: od „luźnych” relacji kojarzeniowych, po związki o charakterze hybrydowym, a nawet zależności hierarchiczne. Formy powiązań elementów sieci mogą przybierać formę wertykalną (pionową) i formę horyzontalną (poziomą). Z teoretycznego punktu widzenia sieć stanowić może zakwestionowanie hierarchii jako sposobu regulowania zdarzeń. Dążenie do obniżenia znaczenia hierarchii powoduje zastąpienie komunikacji pionowej stosunkami poziomymi. Dzięki osłabieniu zależności hierarchicznych wzrasta poziom elastyczności i adaptacyjności. Powiązania między elementami składowymi struktury sieciowej umożliwiają wielostronność i wzajemność przekazu informacji, co stanowi podstawę istnienia sieci. Tendencja do eliminacji zależności hierarchicznych nie oznacza ich negacji i zupełnego usunięcia. W sieci istnieje pełna autonomia jej elementów, chociaż może być zróżnicowana. Oprócz wymienionych cech sieci do zasadniczych walorów tego rozwiązania strukturalnego należy zaliczyć ciągłą zmienność struktury, przejawiającą się w tymczasowości, co zapewnia jej dynamikę. Sprzyja to kompletności sieci, a minimum specyfikacji – uzyskaniu zdolności do samoorganizacji. Natura relacji między częściami obejmuje znacznie więcej niż jednokierunkowe zależności. Zależności symetryczne powodują wzajemną komplementarność i elastyczność działania wskutek eliminacji zależności hierarchicznych lub ich znacznego osłabienia.

W istocie znaczenie słowa kluczowego określa wiązka różnych relacji (kojarzeniowych), w jakie wchodzi słowo kluczowe w danym języku. Relacje te generują odpowiednią, leksykalną strukturę sieciową. Pójście w kierunku sieci sprowadza się do uwzględniania i wykorzystywania w większym stopniu w językach słów kluczowych konotacji, a nie tylko denotacji słów kluczowych. Helen Brownson, przemawiając na konferencji w Dorking, poświęconej zagadnieniom klasyfikacji, powiedziała 14 maja 1957 roku: „Problem [wyszukiwania informacji], przynajmniej jak niektórzy badacze go widzą, polega na przetwarzaniu pojęć i zachodzących między nimi relacji w postaci wyrażonej w języku naturalnym dokumentów na język nieco bardziej sformalizowany, z kontrolą synonimów i uproszczoną składnią. Obecnie uzasadnione jest przypuszczenie, że im dalej posuniemy się w metodach formalizacji i mechanizacji tłumaczenia języka naturalnego na język sformalizowany i na kod dla celów manipulowania maszynowego, tym bardziej prawdopodobne będzie osiągnięcie ekonomicznie możliwego do zrealizowania wyszukiwania maszynowego na dużą skalę”⁴⁸.

⁴⁸ “The problem, as some investigators see it at least, is to transform concepts and their relationship as expressed in the language of documents, into a somewhat more regularized language, with synonyms controlled and syntactic structures simplified. Now it is reasonable to think that the further we can go in routinizing and mechanizing the techniques of translating ordinary language into a regularized language and of coding for machine manipulation, the more will we be likely to achieve economically feasible machine searching on a large scale”. H. Brownson: *Proceedings of the International Study Conference on Classification for Information Retrieval*. ASLIB 1957, s. 99–100). Cyt. za: B.C. Vickery: *Thesaurus – a new word in documentation*. “Journal of Documentation” 1960, no 16, s. 181–189. Tłumaczenie pochodzi z: Robowski 1974, s. 65.

Celem tej części rozprawy było też zwrócenie uwagi na to, co wnosi do języka słów kluczowych sam użytkownik i co jest niezbędne, aby języki te „odpowiadały” użytkownikowi, w tym pokazanie, że organizacja ich słownictwa ma swoje podłoże w strukturze pamięci semantycznej, która to struktura powinna być odwzorowywana w leksyce języków słów kluczowych. Sieci relacji w mózgu powinna odpowiadać sieć relacji w języku słów kluczowych, odwzorowywana m.in. w słowniku jawnym i niejawnym (mentalnym) tego języka. Tu niezbędne okazało się odniesienie do tych właściwości języka słów kluczowych, które są wypełniane/uzupełniane przez użytkownika w postaci jego wiedzy o języku i rzeczywistości językowej, a więc kompetencji językowych.

Powtórzę więc, bo to ważne, że w działalności informacyjnej termin „słowo kluczowe” jest używany w dwóch znaczeniach, często niezbyt wyraźnie uświadamianych. W pierwszym, szerokim znaczeniu, **słowem kluczowym nazywa się każde wyrażenie (wyraz) języka naturalnego wybrane z tytułu, z tekstu dokumentu lub spoza dokumentu, wykorzystywane do etykietowania treści dokumentu lub dowolne słowa z pytania informacyjnego charakteryzujące jego treść.** W drugim znaczeniu **słowo kluczowe jest rozumiane jako jednostka leksykalna języka słów kluczowych.** Słowom kluczowym stawia się warunek, aby z informacyjnego punktu widzenia były słowami znaczącymi, to jest relewantnymi, a więc mającymi wartość wyszukiwawczą w danym systemie informacyjno-wyszukiwawczym. Fakt nieuświadomienia sobie odmienności tych dwóch znaczeń, i w konsekwencji nierozróżnianie ich, powoduje, że wyszukujący po prostu posługują się słowami kluczowymi jako dla nich znaczącymi słowami w sensie informacyjnym, nie traktując ich jako elementów języka słów kluczowych.

Z natury rzeczy słowa kluczowe są wyrazami (wyrażeniami) języka naturalnego, użytymi w funkcjach specjalnych, zależnie od potrzeb danej dziedziny ich użycia. Słowa kluczowe pełnią różne funkcje. Akcentowanie tzw. nastawienia sytuacji informacyjnej na zawartość treściową sprawia, że metainformacyjna funkcja języka naturalnego jest realizowana w różny sposób, co znajduje wyraz w różnych dziedzinach, w których używa się pojęcia „słowo kluczowe”. Szczególną sytuację dla słowa kluczowego stwarza czerpanie z zasobów naturalnego języka ogólnego, co owocuje w postaci języków informacyjno-wyszukiwawczych o notacji paranaturalnej, jakimi są języki słów kluczowych.

Odpowiedź na pytanie o naturę słów kluczowych, używanych w omówionych w tej części rozważań „kontekstach”, sprowadza się w istocie do następującej konstatacji: słowa kluczowe raz są jednostkami leksykalnymi języka naturalnego w jego funkcji metainformacyjnej, tak jak w Internecie, innym razem są jednostkami leksykalnymi sztucznego języka słów kluczowych. W obu przypadkach są jednak genetycznie związane z językiem naturalnym. **W procesach informacyjnych słowa kluczowe to wyrazy (wyrażenia) języka naturalnego, użyte do opisu tekstów dokumentów w celu późniejszego ich wyszukiwania.** Oznacza to, że udało się nam już uzyskać część odpowiedzi na zasadniczą kwestię pracy, czyli pytanie o tożsamość języków słów kluczowych. W tym kontekście warto postawić następujące pytania:

- z jakiego typu wyrażen (rejestru) języka naturalnego korzysta język słów kluczowych?
- jak język słów kluczowych postępuje z wyrażeniami, które włącza do swego słownika?
- jakie możliwości budowania zdań i tekstów ma język słów kluczowych?
- gdzie leży granica między językiem słów kluczowych a językiem naturalnym?

Poszukiwanie pełnej odpowiedzi będzie kontynuowane w kolejnych częściach pracy.

Do tej pory nie rozpatrywałem języków słów kluczowych z punktu widzenia ich struktury i funkcji ani zróżnicowania typologicznego w systemie informacyjno-wyszukiwawczym. Zajmę się tym w drugiej części rozprawy.

Część druga

Słowo kluczowe jako jednostka leksykalna języka słów kluczowych

W tej części zamierzam zająć się słowami kluczowymi jako jednostkami leksykalnymi języka słów kluczowych, przede wszystkim w jego „wzorcowej”, klasycznej wersji i porównać je ze słownictwem innych języków paranaturalnych⁴⁹. Ma to na celu określenie, na tle pozostałych paranaturalnych języków informacyjno-wyszukiwawczych, specyfiki odwzorowywania rzeczywistości dokumentacyjnej w tym języku jako jednego z elementów definiujących jego tożsamość.

Ważny krąg problemów związanych ze sposobem tego odwzorowywania stanowią: struktura pola semantycznego, system leksykalny i słownik języka słów kluczowych. Elementy te nie doczekały się do tej pory w nauce o informacji pogłębionej refleksji. Zazwyczaj patrzono na nie przez pryzmat systemów kategorii leksykalnych języka słów kluczowych, co wprawdzie pozwoliło dostrzec wiele istotnych problemów, ale nie wszystkie. Omawiając proces tworzenia języków słów kluczowych, w tym tworzenie kategorii leksykalnych, nie ograniczam się do samej analizy produktu. Opieram się również na ustaleniach językoznawców dotyczących teorii pól semantycznych.

Z analizy wzorcowej wersji języka, literatury przedmiotu oraz konkretnych języków słów kluczowych wynika, że proces tworzenia tego języka ma w dużej mierze charakter intuicyjny. Postuluję więc konieczność świadomej refleksji teoretycznej w tym zakresie, podpowiadając jednocześnie, z jakich cech klasycznego języka słów kluczowych warto czerpać inspiracje.

⁴⁹ Do zdefiniowania języka słów kluczowych posłużyły mi słowniki terminologiczne z zakresu informacji naukowej, językoznawstwa oraz języków informacyjno-wyszukiwawczych. Do opracowania zasad konstrukcji języka słów kluczowych, jego leksyki i gramatyki wykorzystałem także książki i artykuły dotyczące języków informacyjno-wyszukiwawczych oraz informacji naukowej. Wiele informacji odnoszących się do stosowania języka słów kluczowych w narzędziach wyszukiwawczych służących do przeszukiwania Internetu zaczerpnąłem właśnie stamtąd.

2.1. Geneza języka słów kluczowych

2.1.1. Prehistoria języka słów kluczowych

Język słów kluczowych, podobnie jak język haseł przedmiotowych, wywodzi się z okresu średniowiecza. Wtedy powstawały pierwsze katalogi i inwentarze biblioteczne, w których stosowano układy przedmiotowe, a jako hasło przyjmowano główny wyraz tytułu. Katalogi alfabetyczne, budowane według głównego słowa tytułu, nie były jeszcze katalogami przedmiotowymi, lecz po prostu pierwszymi formami zapewnienia dostępu do treści książek poprzez alfabetyczne wyliczenie słów wskazujących na ich tematy. Równoprawne jest więc twierdzenie, że użyto wówczas po raz pierwszy słów kluczowych, jak i twierdzenie, że były to pierwsze formy haseł przedmiotowych.

Prototypem języków słów kluczowych są też indeksy słów kluczowych⁵⁰, których genezy można doszukiwać się w wykazach wyrazów głównych tytułu, stosowanych chętnie w bibliotekarstwie, a także w idei konkordancji, czyli alfabetycznych zestawieniach słów albo zwrotów wyrazowych przy badaniach tekstów dzieł literackich. Główny wyraz tytułu to wyraz najważniejszy w tytule, wskazujący istotne elementy treści danego dokumentu opatrzonego tym tytułem, etykieta tego dokumentu⁵¹. Zauważmy, że tak właśnie definiuje się słowo kluczowe w wielu słownikach terminologicznych. Pierwsze konkordancje wyrazowe i rzeczowe (tematyczne) sporządzane były do Biblii już w XIII wieku. Z późniejszych najbardziej znana jest konkordancja do dzieła *Summa Theologica* św. Tomasza z Akwinu. Przeniesienie koncepcji wyizolowania z kontekstu słów kluczowych do dokumentacji wiąże się z pracami Hansa Petera Luhna⁵² dotyczącymi indeksu typu KWIC, stąd przyjmuje się, że języki słów kluczowych były modelowane bezpośrednio na hasłach przedmiotowych, a pośrednio na średniowiecznych konkordancjach. Dla obu tych języków podstawą ewolucyj-

⁵⁰ Istnieją różne teorie na temat powstania języka słów kluczowych, które tworzą jego prehistorię. Oto niektóre z nich: 1. Język słów kluczowych był modelowany na języku haseł przedmiotowych z rozluźnieniem gramatyki pozycyjnej i rezygnacji z podziału jednostek leksykalnych w języku haseł przedmiotowych na tematy i okreśniki, co stwarzało różnorodny status tych jednostek (Por. Ćwiekowa 1988). 2. Język haseł przedmiotowych był modelowany na języku słów kluczowych (w XVIII wieku we Francji można było wybrać tylko słowo z tytułu dokumentu). Potem nastąpiło rozluźnienie. Adam Łysakowski podaje (*Katalog przedmiotowy: podręcznik*, Warszawa, PZWS 1946) przykłady, w których istotnie są to wyrażenia podobne do słów kluczowych, co mogłoby oznaczać, że teoria języka haseł przedmiotowych kształtowała się na bazie słów kluczowych (por. też: D.M. Norris: *A History of Cataloguing and Cataloguing Methods 1100–1850, with an introductory survey of ancient times*, London, Grafton and Co. 1939).

⁵¹ Słowo kluczowe to rodzaj etykiety „przyczepianej” do danego tekstu, mającej informować o jego treści w zależności od konkretnych lub potencjalnych potrzeb informacyjnych użytkownika języka słów kluczowych.

⁵² Hans Peter Luhn (1896–1964) – pracownik firmy IBM, który w 1958 roku zaprezentował nową formę indeksu słów kluczowych utworzonego za pomocą komputera, tzw. Keyword-in-Context (KWIC) index.

na były katalogi przedmiotowe⁵³, w których w „klasyczny” sposób odzwierciedlano analizę tekstu z punktu widzenia treści i przedmiotu dzieła (Ćwiekowa 1988).

Genezy języków słów kluczowych powszechnie upatruje się też w związku z powstaniem i rozwojem – rewolucyjnego jak na owe czasy – indeksowania współrzędnego (koordynowanego). Za jego twórcę uważa się amerykańskiego logika i matematyka Mortimera Taubego⁵⁴, z którego nazwiskiem związane jest zastosowanie indeksowania współrzędnego w praktyce informacyjnej⁵⁵. W swoich pracach pisał on o współrzędnych dokumentu w znaczeniu, w jakim używa się tego terminu w matematyce. Tak jak w matematyce współrzędne jednoznacznie określają położenie punktu na płaszczyźnie lub w przestrzeni, tak wyrażenia języka informacyjno-wyszukiwawczego, pełniące funkcje współrzędnych, powinny wskazać, gdzie dany dokument (który można przedstawić jako punkt) znajduje się w przestrzeni informacyjnej (zbiorze informacyjnym) i tym samym umożliwić wyszukanie tego dokumentu w odpowiedzi na pytanie informacyjne. Jego zdaniem indeksowanie współrzędne, zwane też indeksowaniem kojarzeniowym, można stosować w systemach językowych, w których do koordynacji jednostek leksykalnych można używać operacji logicznych. Dla Taubego źródłem tego pomysłu była możliwość koordynacji liczb w maszynach cyfrowych, a także algebra Boole’a i logika klasyczna (schematy logiczne). Zasadę tę nazwał indeksowaniem koordynowanym (ang. *co-ordinate indexing*).

W indeksowaniu współrzędnym słowa kluczowe występujące w charakterystykach wyszukiwawczych dokumentu nie są powiązane żadnymi relacjami paradygmatycznymi w momencie ich wprowadzania do systemu informacyjno-wyszukiwawczego. Każde słowo kluczowe jako jednostka leksykalna języka słów kluczowych grupuje opisy dokumentów, w których charakterystykach wystąpiło. To grupowanie następuje na podstawie identycznej formy wyrażen języka naturalnego w supozycji materialnej. Wyszukiwanie dokumentu odbywa się dzięki koordynacji słów kluczowych charakterystyki wyszukiwawczej dokumentu i instrukcji wyszukiwawczej pytania informacyjnego sformułowanej w tym samym języku informacyjno-wyszukiwawczym, a więc na zasadzie identyczności formy słów kluczowych.

Niewątpliwą zasługą Taubego jest nie tylko „przeniesienie” samego pomysłu na wyszukiwanie informacji, lecz przede wszystkim stworzenie odpowiednich narzędzi, które realizowały ten pomysł. Były to: język unitermów jako lingwistyczne narzędzie oraz tradycyjny system UNITERM, oparty na inwersyjnym modelu organizacji zbioru, w którym niezbędne okazały się dwie kartoteki: kartoteka kart unitermowych oraz kartoteka opisów bibliograficznych dokumentów ułożonych według numerów. W ten sposób informacja zakodowana za pomocą danej jednostki leksykalnej nie znajduje

⁵³ Początek katalogom przedmiotowym w Europie dał Real Katalog, stworzony przez Martina Schrettingera w pierwszej połowie XIX wieku.

⁵⁴ Mortimer Taube (1910–1965) urodził się w Jersey City (New Jersey). Studiował filozofię na uniwersytecie w Chicago, a następnie na University of California w Berkeley, gdzie w 1935 roku doktoryzował się z tego zakresu. W 1936 roku skończył studia biblioteczarskie. W latach 1953–1959 redagował serię *Studies in Coordinate Indexing*.

⁵⁵ Po raz pierwszy indeksowanie (wyszukiwanie) współrzędne zastosowano w systemie kart obrzeźnie perforowanych w USA w połowie lat trzydziestych XX wieku (Kilgour 1997).

się w jednym miejscu, i nie jest gotowa jak w tradycyjnym katalogu przedmiotowym, gdzie opisy dotyczące danego tematu są gromadzone w określonym miejscu. Jest rozproszona, a może być uzyskana dopiero jako rezultat procesu wyszukiwania informacji i wykonania odpowiednich operacji logicznych. Jest to niewątpliwie ważny krok w kierunku „inteligentnego” wyszukiwania, w rozumieniu posługiwania się logiką. Współczesne systemy wyszukiwania informacji działają według tego schematu. Stworzony przez Taubego język unitermowy był pierwszym językiem informacyjno-wyszukiwawczym opartym na indeksowaniu współrzędnym.

Unitermem (ang. *unit term* – termin prosty) był termin przejęty z języka naturalnego, nierozkładalny, a więc jednowyrazowy, który mógł być niekiedy zaopatrzonej w uwagi dotyczące jego znaczenia. Charakterystyka wyszukiwawcza dokumentu składała się z kilku takich jednowyrazowych terminów, zapisanych w dowolnej kolejności. Taube dopuszczał jednak kontrolę słownictwa i podział na kategorie. Rozróżniał unitermy wolne, równokształtne z wyrazami o jednym tylko rdzeniu i unitermy związane – równokształtne z wyrazami złożonymi motywowanymi lub połączeniami wyrazowymi, obwarowując jednak tego typu konstrukcje licznymi ograniczeniami. Unitermom w planie treści miała odpowiadać klasa pojęć jednostkowych, tj. nierozkładalnych już na pojęcia prostsze, bardziej elementarne. Można więc uważać, że słowa kluczowe kontynuują ideę unitermów w zakresie gramatyki, a tylko częściowo słownictwa, gdyż są najczęściej słowami prostymi – ale nie muszą nimi być (występują również jednostki wielowyrazowe) – i nie podlegają też ścisłej kontroli.

Taube, badając użycie wyrazów języka naturalnego do indeksowania współrzędnego, zaproponował system UNITERM, który polegał na tworzeniu języka przez rozbijanie standardowych haseł przedmiotowych i klasyfikacji na zestawy pojedynczych terminów (Taube, Gull, Wachtel 1952). Nowy sposób indeksowania pozwalał na indeksowanie dokumentów o wielu tematach, rozpatrywanych z różnych punktów widzenia. Należy zwrócić uwagę, że Taube był logikiem, a nie językoznawcą, i pomysł ten zaczerpnął z logiki. Problemem tym zajmował się także Calvin Northrup Mooers, uważany za twórcę języka deskryptorowego (Mooers 1963). Podstawowa różnica między deskryptorami Mooersa⁵⁶, stosowanymi w wyszukiwaniu informacji w „Zato-Coding System”, stworzonym przez niego w 1950 roku, a unitermami Taubego polega na tym, że słownik deskryptorów powstawał niejako z góry, a priori, był gotowy, zanim przystąpiono do indeksowania dokumentów, natomiast unitermy tworzyły słownictwo tego języka w miarę indeksowania dokumentów. Można tu dostrzec wyraźną zbieżność z obecnie stosowanymi dwiema odmiennymi metodami budowy języków informacyjno-wyszukiwawczych: metodą dedukcyjną i metodą indukcyjną. Język unitermowy był budowany metodą indukcyjną, zaś język deskryptorowy metodą dedukcyjną. Terminy wprowadzane do słownika czerpano z tytułów, tekstu dokumentów i przez rozbijanie na poszczególne wyrazy dawnych haseł przedmiotowych.

⁵⁶ De facto przy deskryptorach Calvina N. Mooersa umieszczane były uwagi dotyczące łączenia danego deskryptora z innymi deskryptorami. Uwagi te ograniczały łączliwość deskryptorów. Można je więc uznać za rodzaj środka gramatycznego.

J.C. Costello uważa, że „unitermy są to hasła, które wyrażają pojedyncze pojęcia (*unit concepts*). Wszystkie unitermy posiadają tę samą rangę hierarchiczną. W sensie logicznym żaden z unitermów nie zajmuje wyższej lub niższej pozycji w stosunku do innych unitermów, jak to ma miejsce w klasyfikacjach hierarchicznych. Pojedynczych unitermów nie zestawiało się też z innymi według wstępnie ustalonej kolejności, jak w przypadku haseł przedmiotowych” (Costello 1961). Jako unitermów można było używać także nazw własnych, nazw geograficznych, nazw firm.

W celu uniknięcia szumu informacyjnego wywoływanego przez niekontrolowane słownictwo, z którego nie usunięto wieloznaczności powodowanej przez synonimie, quasi-synonimie czy homonimie, indeksatorzy stosowali rejestrowanie niektórych relacji zachodzących między unitermami na odwrocie karty unitermowej lub w słowniku unitermów⁵⁷. Powstawały więc słowniki unitermów zawierające odsyłacze „zob.” i „zob. też”. Środki te miały służyć do kontroli słownictwa. Charakterystyczną cechą pierwszych słowników unitermów i deskryptorów⁵⁸ było to, że opracowywano je przede wszystkim dla jednego ośrodka. Opublikowane słowniki miały służyć jedynie jako pomoc i wzór dla innych ośrodków, natomiast nie zalecano ich bezpośredniego wykorzystania.

Istotny wpływ na pomysł indeksowania współrzędnego i koncepcję języka unitermowego M. Taubego wywarła praca Vannevara Busha *As we may think*, w której dowodzi, że myślenie ludzkie nie odbywa się według schematu klasyfikacji, lecz „it operates by association” i może ono posłużyć jako model zapamiętywania (przechowywania) informacji i jej wyszukiwania (Bush 1945). Taube wyobrażał sobie, że dostarcza indeksującemu pomoc w postaci „manual dictionary of associations”, który nie wyjaśnia słowa, jak leksykon, lecz podporządkowuje wszystkie pokrewne słowa jednemu słowu na potrzeby systemu informacyjno-wyszukiwawczego określonej dziedziny wiedzy (Taub 1955). Wykorzystanie takiego słownika zawierającego skojarzenia pojęć okazało się bardzo przydatne w procesach wyszukiwania informacji.

Język unitermowy stanowi odmianę języka słów kluczowych. Najważniejszą jego cechą jest możliwość tworzenia dużej liczby zdań języka ze stosunkowo małej liczby elementarnych jednostek leksykalnych (unitermów). Wadą języków unitermowych jest większe niż w innych językach informacyjnych prawdopodobieństwo występowania fałszywych koordynacji unitermów, będących przyczyną szumu informacyjnego (Costello 1961). Prawdopodobnie była to zasadnicza przyczyna rezygnacji ze stosowania tych języków, tak bardzo popularnych w latach pięćdziesiątych XX wieku⁵⁹. Wyparły je stopniowo języki słów kluczowych i języki deskryptorowe,

⁵⁷ Na oznaczenie relacji zachodzących między jednostkami leksykalnymi języków unitermowych używano najczęściej tylko odsyłaczy typu „zob.” i „zob. też”. Dla niektórych terminów rejestrowano nawet zależności hierarchiczne. Konsekwencją takiego podejścia było stopniowe zanikanie w USA różnic pomiędzy terminami „deskryptor”, „uniterm”, „słowo kluczowe”, stąd często są one stosowane zamiennie. Najczęściej w piśmiennictwie amerykańskim na określenie deskryptora lub unitermu obecnie używa się nazwy „keyword” = słowo kluczowe lub „term” (termin).

⁵⁸ Deskryptor – w literaturze zachodniej jest stosowany na oznaczenie jednostek leksykalnych paraturalnych języków informacyjno-wyszukiwawczych, w przeciwieństwie do języków o słownictwie sztucznym (symbole).

⁵⁹ Obecnie języki unitermowe są używane bardzo rzadko. Były one jednak wielkim wydarzeniem w historii języków informacyjno-wyszukiwawczych. Obok deskryptorów Mooersa, które pojawiły się

w których zmalało prawdopodobieństwo występowania fałszywych koordynacji, dzięki wprowadzeniu wyrażen wielowyrazowych jako elementarnych jednostek leksykalnych tych języków oraz takich środków gramatycznych jak wskaźniki więzi i wskaźniki roli.

Widać wyraźnie, że język unitermowy podlegał rozwojowi w kierunku języka słów kluczowych. Narodziny języka słów kluczowych wiążą się z odejściem już na początku lat pięćdziesiątych XX wieku od systemu UNITERM, związanym z zaakceptowaniem możliwości prekoordynacji słów kluczowych.

2.1.2. Historia języka słów kluczowych

Przy założeniu, że języki unitermowe są wczesną odmianą języków słów kluczowych, za datę narodzin języków słów kluczowych można przyjąć początek lat pięćdziesiątych XX wieku, kiedy to opracowano język dla systemu UNITERM. Warto zwrócić uwagę na cechy „pierwotnej” odmiany języka słów kluczowych: prosta struktura zarówno na osi paradygmatycznej, jak i syntagmatycznej, przeważnie jednowyrazowe, uporządkowane alfabetycznie wyrażenia zaczerpnięte z języka naturalnego, brak specyfikacji relacji hierarchicznych i kojarzeniowych w słowniku, brak kontroli słownictwa. W istocie jest to „czysty”, „klasyczny” język swobodnych słów kluczowych. To właśnie od takiej postaci języka słów kluczowych wszystko się zaczęło.

Charakterystyka treści dokumentu – identycznie jak w języku unitermów – ma postać ciągu już nie unitermów, lecz słów kluczowych wymienionych w dowolnej kolejności, których znaczenie w sumie odzwierciedla treść dokumentu. Na wzór unitermów języki słów kluczowych zachowały jako formę dla elementarnych jednostek leksykalnych wyrażenia równokształtne z wyrażeniami języka naturalnego. Również w tym kierunku zmierzały języki deskryptorowe, w tym powstałe w 1950 roku deskryptory Mooersa. Słowa kluczowe stały się tu tworzywem deskryptorów – terminów ustalonych jako obowiązujące dla oznaczania przedmiotów dokumentów (kodowania) w ich charakterystyce wyszukiwawczej i w procesie wyszukiwania informacji o tych dokumentach. Język słów kluczowych rozwinął się więc z języka unitermowego, który jest uważany za pierwowzór języka słów kluczowych (Chu 2003).

Tak więc języki słów kluczowych:

- mają ten sam rodowód co języki haseł przedmiotowych/hasła przedmiotowe;
- były modelowane na językach haseł przedmiotowych (Ćwiekowa 1988);

w 1950 roku, dały początek indeksowaniu współrzędnemu. Były one stosowane przeważnie w systemach opartych na tzw. małej mechanizacji. Podstawę ich realizacji stanowiły kartoteki unitermowe. To zaważyło na fakcie, że języki te kojarzy się i omawia przy okazji prezentacji kartoteki unitermowej. Zwykle nie zauważa się, że języki unitermowe mogą być, i są, stosowane także w systemach zautomatyzowanych. Przykładem jest wykorzystanie tego języka w systemie informacyjno-wyszukiwawczym patentów amerykańskich Uniterm System to US Patents.

– są odrębne od języków deskryptorowych i języków haseł przedmiotowych, bo różnią się od nich strukturą⁶⁰.

W literaturze przedmiotu sposób narodzin języka słów kluczowych nie jest tak oczywisty. Do końca nie wiadomo, czy unitermy stanowiły etap w rozwoju tego języka, czy były swoistego rodzaju „odpryskiem” w jego rozwoju. Niewątpliwie początek językom słów kluczowych dały swobodne słowa kluczowe.

2.2. Język słów kluczowych – problemy definicyjne

Najbardziej ogólną, modelową charakterystykę języka słów kluczowych⁶¹ zawiera definicja tego terminu zamieszczona w *Słowniku encyklopedycznym informacji, języków i systemów informacyjno-wyszukiwawczych* (Bojar 2002, oprac., s. 111). Oto ona: „**JĘZYK SŁÓW KLUCZOWYCH** – *język informacyjno-wyszukiwawczy o *notacji paranaturalnej, bez określonej explicite w *słowniku (2) *paradygmatyki, w którym jednostki *słownika (1) nazywane są *słowami kluczowymi (2), a na *gramatykę (1) składają się reguły *indeksowania współrzędnego”. Definicja ta stanowi przykład tzw. wąskiej definicji języka słów kluczowych, gdyż obejmuje swoim zakresem tylko klasyczny język słów kluczowych⁶². Nie mieszczą się w niej języki słów kluczowych niespełniające wymagań tej definicji. Nazywam je nieklasycznymi językami słów kluczowych. W bibliotekarstwie i nauce o informacji uważa się jednak te ostatnie również za języki słów kluczowych⁶³.

W zgodzie z aktualnym rozwojem języka słów kluczowych, za niezbędne uznałem rozszerzenie zakresu tego terminu, tak aby objąć wszystkie istniejące języki słów kluczowych, w tym również nieklasyczne języki słów kluczowych, gdyż w praktyce rozumienie języków słów kluczowych uległo znacznemu poszerzeniu właśnie w tym kierunku. Biorąc pod uwagę tzw. uzus, przyjmuję na potrzeby dalszych rozważań – częściowo zbieżną z definicją zaproponowaną przez Bożennę Bojar – następującą szerszą definicję języka słów kluczowych: „**język słów kluczowych jest językiem informacyjno-wyszukiwawczym o notacji paranaturalnej, którego słownictwo stanowią wyrażenia zwane słowami kluczowymi, a gramatykę stanowią reguły indeksowania współrzędnego**”. Definicja ta jest szersza dzięki nieuwzględnieniu w niej ograniczenia dotyczącego paradygmatyki języka. Przedstawieniu „pojemno-

⁶⁰ Eugeniusz Ścibor (1982) twierdzi odwrotnie. Uważa, że języki słów kluczowych są tylko pozornie odrębne od języków deskryptorowych i języków haseł przedmiotowych, gdyż nie jest to umotywowane ich własnościami strukturalnymi. Jednak nie ma on tu racji.

⁶¹ Joanna Tomasiak-Beck w artykule pt. *Nowe terminy w informacji naukowej* (1997) odnotowała brak terminu „język słów kluczowych” w *Słowniku terminologicznym informacji naukowej* (STIN), opublikowanym w 1979 roku. Język ten był jednak już znany w momencie opracowywania tego słownika, lecz – jej zdaniem – „było mało publikacji na ten temat i przez autorów słownika termin ten nie został uznany za godny odnotowania” (s. 192).

⁶² Warto zwrócić uwagę, że w tym słowniku termin „słowo kluczowe” posiada dwa znaczenia, zaś we wcześniejszym *Słowniku terminologicznym informacji naukowej* – tylko jedno (por. rozdział 1.1.2).

⁶³ To aposterioryczne/indukcyjne podejście do sposobu definiowania języka słów kluczowych różni się od dominującego do tej pory apriorycznego/dedukcyjnego podejścia w definiowaniu tego języka.

ści” tej rozszerzonej definicji języka słów kluczowych będzie poświęcona przede wszystkim trzecia część rozprawy. W dalszych rozważaniach charakterystyka języka słów kluczowych zostanie wzbogacona o omówienie systemu językowego tego języka w jego wersji klasycznej dla języka polskiego.

2.2.1. Język słów kluczowych jako element systemu informacyjno-wyszukiwawczego

Język słów kluczowych jest elementem systemu informacyjno-wyszukiwawczego. W odniesieniu do dokumentów pełni funkcje charakterystyczne dla języków informacyjno-wyszukiwawczych, które z definicji są językami sztucznymi, specjalnie konstruowanymi dla systemów informacyjno-wyszukiwawczych. Jednocześnie zbiór jego jednostek leksykalnych reprezentuje zasób terminologiczny określonej dyscypliny i – zależnie od jego kompletności, szczegółowości i sposobu organizacji – może służyć do reprezentowania problemów, tematów i zagadnień, które mieszczą się w polu zainteresowania danej dyscypliny. Aktualizacja słownictwa, odpowiednio do etapu i poziomu rozwoju dziedziny nauki i działalności praktycznej, jest jednym z ważniejszych problemów teorii języków informacyjno-wyszukiwawczych.

Język słów kluczowych w systemie wyszukiwania informacji ma charakter pragmatyczny. Słownik tego języka umożliwia przejście z języka naturalnego, w którym napisane są teksty dokumentów i teksty pytań użytkowników, na teksty w języku słów kluczowych. Język ten pełni więc funkcję metainformacyjną – służy do opisu informacji o informacji zawartej w dokumentach, oraz funkcję wyszukiwawczą, bo umożliwia wyszukiwanie informacji.

Możliwości i ograniczenia systemu wyszukiwania informacji, w którym stosowany jest dany język, stanowią pryzmat, przez który patrzy się na różne języki informacyjno-wyszukiwawcze w celu ich porównania, oceny i ewentualnie wyboru. Umożliwiają je typowe i indywidualne elementy systemu. Do typowych należą zbiory dokumentów i środki techniczne, a indywidualnymi nie będziemy się zajmować. Język stanowi tu „naczynie połączone” z repertuarem technik wyszukiwawczych, jakim dysponuje dany system (wagi, maskowanie, operatory Boole’a, operatory kontekstowe). Na formy prezentacji oraz sposoby realizacji języka duży wpływ mają uwarunkowania techniczne.

2.2.2. Jeden czy wiele języków słów kluczowych, czyli o naturze języka słów kluczowych

W polskim piśmiennictwie z definicji uważa się języki słów kluczowych – podobnie jak inne języki informacyjno-wyszukiwawcze – za języki sztuczne⁶⁴. Natury języków

⁶⁴ Cecha sztuczności jest cechą definicyjną języków informacyjno-wyszukiwawczych do dziś. Świadczy o tym następująca definicja tych języków w najnowszym polskim słowniku terminologicznym z informacji naukowej, która podaje, że jest to „język sztuczny, którego wyspecjalizowaną funkcją jest funkcja

słów kluczowych nie da się zdefiniować prosto. Są to języki paranaturalne, tj. oparte na słownictwie języka naturalnego, ale podobnie jak inne języki informacyjno-wyszukiwawcze są zredukowane pod względem funkcjonalno-strukturalnym, co oznacza, że w stosunku do języka naturalnego mają ograniczoną liczbę pełnionych funkcji, nie wszystkie elementy struktury języka naturalnego, posługują się wyłącznie kodem pisanym. Z definicji są niekontrolowane lub kontrolowane w bardzo ograniczonym stopniu. W klasycznej postaci języka swobodnych słów kluczowych słowa kluczowe są wyrażeniami języka naturalnego użytymi w funkcji metainformacyjnej. Wyrażenia te są ujednocione i uproszczone do potrzeb wyszukiwania informacji w systemach dokumentacyjnych. Rodzaj i stopień ujednoczenia oraz uproszczenia wyrażeń języka naturalnego zależy od aktualnych potrzeb i możliwości technologii wyszukiwawczych. Określa to odpowiednia instrukcja indeksowania swobodnego, a w przypadku jej braku – pragmatyka ogólna języka słów kluczowych.

Język słów kluczowych w odmianie języka swobodnych słów kluczowych to język, którego elementarnymi jednostkami leksykalnymi są słowa kluczowe derywowane z indeksowanego tekstu dokumentu lub wybrane z zasobu czynnego słownika (słownika mentalnego) indeksatora, służące do indeksowania swobodnego. W procesie indeksowania swobodnego za pomocą tego języka „mamy do czynienia z identycznościową transformacją tekstu charakterystyki słownej dokumentu i nieidentycznościową transformacją jego funkcji, polegającą na tym, że oprócz funkcji metainformacyjnej tekst ten zaczyna pełnić funkcję wyszukiwawczą” (Bojar 2002, oprac., s. 88). Stąd słusznie uważa się język swobodnych słów kluczowych za język naturalny. Z czasem pojawiła się nowa odmiana języka słów kluczowych, charakteryzująca się m.in. słownictwem kontrolowanym, która nie mieści się w definicji klasycznego języka słów kluczowych.

Języki słów kluczowych, podobnie jak inne języki informacyjno-wyszukiwawcze, mogą być projektowane w sposób celowy, tj. „tworzone wtedy, gdy trzeba w sposób jednoznaczny, precyzyjny i ekonomiczny przekazywać relewantne informacje o rzeczywistości o dobrze określonych granicach i relacjach zachodzących w niej między dobrze określonymi obiektami” (Bojar 2006, s. 7). Nie zapewnia tego język swobodnych słów kluczowych. W przypadku projektowania języka słów kluczowych konstrukcja poprzedza jego użycie. Taki język tworzy się metodą dedukcyjną, aczkolwiek wyrażenia języka naturalnego – podobnie jak w dotychczas omawianej odmianie – stanowią podstawę takiego języka słów kluczowych. Stąd taki język jest rezultatem swoistego rodzaju ingerencji projektanta w sferę leksyki, semantyki i składni języka naturalnego w celu lepszego przystosowania tych elementów do pełnienia konstytutywnych dla tego języka funkcji: metainformacyjnej i wyszukiwawczej. Biorąc pod uwagę sposób jego powstawania, taki język słów

metainformacyjna polegająca na odwzorowywaniu cech informacji i/lub cech nośników informacji, na których są one utrwalone, przez tworzenie ich reprezentacji zwanej charakterystyką wyszukiwawczą dokumentu, oraz funkcja wyszukiwawcza polegająca na wyrażaniu zapytań informacyjnych i umożliwieniu odnalezienia w zbiorze wyszukiwawczym tych informacji, które spełniają warunki wyrażone w zapytaniu” (Bojar 2002, oprac., s. 104–105; por. także: Zacharow 2002 oraz Beling, Port, Strohl-Goebel, ed. 2006, s. 63). Ale już nie wszyscy traktują tę cechę jako cechę definicyjną tych języków). (por. Bertram 2005).

kluczowych trzeba uznać za twór sztuczny. Za tym może iść struktura tego języka, w tym umowność denotacji i konotacji jego wyrażen (słów kluczowych) tworzących słownik słów kluczowych. W systemie informacyjno-wyszukiwawczym istotna jest denotacja (w wyszukiwaniu informacji), zaś w indeksowaniu – konotacja, czerpana z języka naturalnego. Sztuczność języka słów kluczowych jest pochodną sposobu budowy języka, a więc użytej metody jego tworzenia, a nie charakteru jego słownictwa. Zatem sztucznymi są te języki słów kluczowych, których struktura, a przede wszystkim system leksykalny, została wcześniej zaprojektowana, co sprowadza się do użycia innej metody niż metoda indukcyjna. W piśmiennictwie sztuczność bywa rozumiana różnie. Jedni odnoszą ją do języków, które nie służą do komunikowania się ludzi z sobą, choć przecież języki słów kluczowych nie służą do tego, lecz do korzystania z systemu informacyjno-wyszukiwawczego, a za jego pośrednictwem – z dokumentów (jako opozycja do języków naturalnych). Drudzy łączą ją ze sztucznym słownictwem języka. W rozprawie sztuczność odnoszę do języka w rozumieniu przyjętym w językoznawstwie, to jest wiąże ją ze sposobem powstawania/tworzenia języka.

O tym, czy dany język słów kluczowych jest naturalny czy sztuczny, przesądza więc metoda jego budowy. W przypadku użycia metody indukcyjnej mamy do czynienia z językiem naturalnym, zaś w przypadku innej metody niż indukcyjna – z językiem sztucznym. W tym wypadku zarówno system leksykalny, jak i relacje mogą nawet znacznie odbiegać od odpowiadających im w języku naturalnym. Niezależnie od odmiany język ten opiera się na indeksowaniu współrzędnym i ma sztuczną składnię, gdyż bazuje ona na obcej językowi naturalnemu konkatenacji, dołączaniu w dowolnej kolejności słów kluczowych w tekstach tego języka (charakterystyce wyszukiwawczej dokumentu i instrukcji wyszukiwawczej zapytania).

2.3. Elementy strukturalno-funkcjonalne określające tożsamość języka słów kluczowych

System języka słów kluczowych może być prezentowany w postaci następującej formuły:

$$J = \langle V, G, S, P \rangle$$

gdzie:

V – słownik/alfabet/słownictwo danego języka,

G – gramatyka,

S – semantyka,

P – pragmatyka.

Składa się ze słownika i gramatyki, które pozwalają na generowanie wyrażen złożonych (zdan i tekstów), reguły semantyki, pozwalające na przyporządkowywanie

wyrażeń tego języka elementom odwzorowywanej rzeczywistości pozajęzykowej oraz reguły pragmatyki ustalające sposób poprawnego używania języka w tworzeniu komunikatów adekwatnych do konkretnych potrzeb.

2.3.1. Funkcje języka słów kluczowych

Języki słów kluczowych pełnią wyspecjalizowane funkcje, do których należą:

1. grupa funkcji podstawowych, zwanych funkcjami pierwotnymi, konstytutywnymi; są to:
 - funkcja metainformacyjna, polegająca na odwzorowywaniu cech dokumentów,
 - funkcja wyszukiwawcza, zwana również heurystyczną, polegająca na sterowaniu procedurami wyszukiwania informacji o dokumentach, ich treści, wskazanej przez użytkownika w skierowanym do systemu pytaniu informacyjnym. Należy zauważyć, że funkcja wyszukiwawcza niejako zawiera funkcję metainformacyjną. Funkcja metainformacyjna antycypuje funkcję wyszukiwawczą, gdyż jeżeli nie ma czegoś opisanego (w charakterystyce dokumentu i pytaniu informacyjnym), to nie sposób tego wyszukać. Obie te funkcje są ściśle związane; jeśli jest dokument indeksowany, to po to, aby go można było wyszukiwać. Sama funkcja wyszukiwawcza jest możliwa tylko w wyszukiwaniu pełnotekstowym;
2. dodatkowe funkcje, zwane funkcjami wtórnymi lub pochodnymi.

Funkcje podstawowe przekładają się na dwie odmiany funkcjonalne języka słów kluczowych: język indeksowania i język wyszukiwawczy. Ewa Chmielewska-Gorczyca pisze wręcz o dwóch podsystemach języka informacyjno-wyszukiwawczego (Chmielewska-Gorczyca 1991).

W tradycyjnych systemach informacyjno-wyszukiwawczych nie ma potrzeby oddzielania języka wyszukiwawczego od języka indeksowania. Język wyszukiwawczy stara się upodobnić i odzwierciedlać język indeksowania, co idealnie sprawdza się w przypadku jednego języka pełniącego obie funkcje. Wraz z rozwojem nowoczesnych technologii informacyjnych i wyszukiwawczych następuje stopniowe oddzielanie się obu języków⁶⁵. Spowodowane jest to przede wszystkim automatyzacją procesów wyszukiwania informacji i wynikającym z tego „zapleczem” systemowym. Występują tu różnice w zasobach leksykalnych, na przykład w języku indeksowania wykorzystuje się słownik jawny, zaś w języku wyszukiwawczym stosuje się indeksowanie swobodne (niekontrolowane słowa kluczowe lub pełne teksty dokumentów). Wtedy zasoby leksykalne i ich struktura są bardzo różne. Jeżeli jednak ma być możliwy proces porównywania ich elementów na podstawie formalnych cech jednostek leksykalnych, a nie cech semantycznych, to słowniki obu odmian funkcjonalnych wprawdzie nie muszą być identyczne, tj. całkowicie pokrywać się, ale pewne ich podzbiory powinny być identyczne. W takim przypadku, w celu usprawnienia tłu-

⁶⁵ Zjawisko to zauważył już w 1978 roku A.I. Czerny, pisząc: „W systemie informacyjno-wyszukiwawczym stosuje się często nie jeden, ale dwa języki informacyjno-wyszukiwawcze: jeden – do indeksowania dokumentów, drugi – do indeksowania zapytań informacyjnych” (Czerny 1978, s. 108).

maczenia z jednego języka na drugi, staje się niezbędny jakiś język pośrednik, który odwzorowywałby struktury pola semantycznego obu języków.

Takie podejście spowodowało, że obecnie – ze względu na kryterium funkcjonalne – wyraźnie wyróżnia się:

- (pod)język indeksowania (ang. *indexing language*), zwany też językiem informacyjnym, pełniący funkcję opisu treści i/lub formy dokumentów. Językiem tym posługują się bibliotekarze i dokumentaliści oraz osoby opisujące zawartość i formę dokumentów;
- (pod)język wyszukiwawczy (ang. *retrieval language*⁶⁶), pełniący funkcję opisu potrzeb informacyjnych użytkownika systemu w postaci pytania użytkownika, a zarazem sterujący procesem wyszukiwania informacji.

Nie są to dwa różne języki, lecz jeden język, który może pełnić dwie różne funkcje.

Oprócz omówionych funkcji języki słów kluczowych – podobnie jak inne typy języków informacyjno-wyszukiwawczych – mogą pełnić dodatkowe funkcje, zwane funkcjami wtórnymi lub pochodnymi. Są to:

- funkcja organizująca (porządkująca), polegająca zwykle na alfabetycznym porządkowaniu i strukturalizacji zbiorów informacyjnych i dokumentów;
- funkcja określania zakresów tematycznych zbiorów, dokumentów i informacji.

Funkcje te nie są warunkiem koniecznym ani wystarczającym do uznania danego języka za język informacyjno-wyszukiwawczy, ale mogą stanowić istotny element w konstruowaniu funkcjonalnej definicji języków słów kluczowych. Taka definicja jest rodzajem opisu tych języków w kategoriach pełnionych przez nie funkcji w systemie informacyjno-wyszukiwawczym. Pełnione przez ten język funkcje umożliwiają wyznaczenie różnego rodzaju pól: pola informacyjnego, pola dokumentacyjnego, pola tematycznego oraz pola semantycznego.

2.3.2. Struktura języka słów kluczowych

Strukturę omawianych języków rozpatrzę dwojako: w wymiarze wertykalnym, tj. poziomów tych języków, w którym jest to struktura wielopoziomowa, oraz w wymiarze horyzontalnym, czyli elementów tych języków, w którym jest to struktura wieloelementowa.

2.3.2.1. Wymiar wertykalny – poziomy języka słów kluczowych

W językach paranaturalnych, do których zaliczany jest język słów kluczowych, nie ma sensu wyróżnianie – jak to się czyni w przypadku języków o słownictwie sztucznym – poziomu znaków graficznych, gdyż słowa kluczowe nie są w tych językach tworem budowanym z liter, lecz są przejmowane w całości z języków naturalnych. W istocie pierwszy poziom tego języka jest poziomem elementarnych jednostek leksykalnych, czyli słów kluczowych. Wyróżnianie poziomu elementarnych znaków graficznych ma sens jedynie w odniesieniu do słowa kluczowego, gdyż rzeczywiście

⁶⁶ Termin „information retrieval” wprowadził w 1950 roku C.N. Mooers.

składa się ono z elementarnych znaków graficznych, którymi są w przeważającej części litery (poszczególnych alfabetów), cyfry i znaki specjalne. Słowa kluczowe stanowią – powtarzam – pierwszy poziom tego języka, poziom semantycznie pełnych jednostek leksykalnych. W związku z tym, że są to jednostki leksykalne równokształtne z wyrazami lub połączeniami wyrazowymi języka naturalnego, noszą miano jednostek paranaturalnych, a słownictwo języka słów kluczowych jest słownictwem paranaturalnym.

Przepisy formułujące zasady tworzenia tego poziomu języka słów kluczowych dla języka polskiego podał w swojej publikacji Olgierd Ungurian, określając części mowy mogące być słowami kluczowymi (Ungurian 1976). Zaleca on, aby nazwy występujące jako słowa kluczowe były jednowyrazowe, przestrzegając równocześnie przed pochopnym rozbijaniem terminów wielowyrazowych, gdyż może to przyczynić się do powstawiania nieodpowiednich koordynacji, generujących szum informacyjny. Zwraca także uwagę na wysuwanie w terminach wielowyrazowych na pierwszą pozycję nazw cech charakterystycznych.

Obowiązującym sposobem postępowania przy indeksowaniu prac przez ich autorów jest tworzenie swobodnych słów kluczowych, zwanych też autorskimi słowami kluczowymi. Często, ale nie zawsze, czynność ta odbywa się stosownie do dostarczonej wcześniej bardzo ogólnej instrukcji.

Słowa kluczowe, które są wybierane w analizie treści dokumentu, zwykle otrzymują ujednoliconą formę gramatyczną (na przykład gdy mają być umieszczone w indeksie KWOC, WADEX) lub pozostają w niezmienionej formie gramatycznej (na przykład w indeksie KWIC, KWIT)⁶⁷. Według większości takich instrukcji słowa kluczowe, które podlegają opracowaniu leksykograficznemu, mogą być rzeczownikami, przymiotnikami lub liczebnikami. Rzeczowniki występujące w tekście, a uznane przez indeksatora za odpowiednie, stają się słowami kluczowymi. Inne części mowy, jak na przykład przymiotniki lub czasowniki, przekształca się w rzeczowniki o możliwie równorzędnym znaczeniu. Zasadniczo słowa kluczowe są podawane w pierwszym przypadku liczby pojedynczej. Liczba mnoga jest dopuszczalna dla rzeczowników, dla których nie jest używana liczba pojedyncza. Dotyczy to tylko tych języków, w których istnieją zróżnicowane kategorie gramatyczne (przypadek, liczba, rodzaj), tak jak w języku polskim. W liczbie mnogiej zaleca się podawanie:

- rzeczowników o charakterze *pluralia tantum*, na przykład: nożyczki, drzewa;
- nazw rodów, na przykład: Połanieccy, Radziwiłłowie, Jagiellonowie;
- nazw grup związków chemicznych i minerałów, na przykład: tłuszcze, białka, boksyty;
- nazwy grup osób, na przykład: dzieci, kombatanzi, liberałowie (ale nazwy grup zawodowych podaje się w liczbie pojedynczej, na przykład: górnik, lekarz, nauczyciel);
- nazw wyższych grup zwierząt, na przykład: ssaki, koty (jako nazwa rodziny, ale kot domowy jako nazwa gatunku w liczbie pojedynczej);
- wyrażen powszechnie stosowanych w liczbie mnogiej, na przykład: wody przybrzeżne, badania naukowe.

⁶⁷ Więcej na ten temat w rozdziale 3.1.3.1.6.

Przymiotniki stają się samodzielnymi jednostkami leksykalnymi w charakterystykach wyszukiwawczych tylko wtedy, gdy nie ma możliwości zastąpienia ich rzeczownikami, na przykład: społeczny, międzynarodowy, lub gdy ich znaczenie odnosi się do wielu obiektów. Pełnią wówczas funkcję modyfikatorów, które mogą być łączone z wieloma rzeczownikami, na przykład: automatyczny, regulacyjny. Przymiotników nie stosuje się do słów kluczowych przy określaniu nazw własnych państw, w przypadku określania narodowości, w opisach zdarzeń zachodzących na terenie innych państw. Przymiotniki będące samodzielnymi słowami kluczowymi są sprowadzane do pierwszego przypadku liczby pojedynczej rodzaju męskiego.

Słowem kluczowym może być pojedynczy wyraz, może je tworzyć również wielowyrazowa fraza. Wyrazy, między którymi istnieje związek frazeologiczny, tzn. połączenie wyrazów, którego znaczenia nie można bezpośrednio wywnioskować ze znaczenia jego składników, na przykład: biały kruk, paż królowej, boża krówka, są też słowami kluczowymi.

W językach słów kluczowych używa się również złożonych z wielu wyrazów słów kluczowych, ale tylko w następujących przypadkach:

- gdy znaczenie terminu jako całości nie jest sumą znaczeń poszczególnych składników, lecz stanowi związek o nowym innym znaczeniu, na przykład: złoty środek, świnka morska;
- gdy stanowią nazwy własne, na przykład: Kazimierz Wielki, Kaplica Zygmuntowska, Czerwone Wierchy;
- gdy wyrazy składniowe zachowują swe podstawowe znaczenie, jednak stopień ich zespolenia jest znaczny, na przykład: skok w dal, silnik spalinowy, maszyna rolnicza.

Słowa kluczowe mogą być wielowyrazowe i składać się z połączenia rzeczownika z innym rzeczownikiem, z przymiotnikiem, imiesłowem przymiotnikowym lub wyrażeniem przyimkowym⁶⁸.

Wyrażenia, których można używać jako słów kluczowych, zwykle tworzą wykaz słów kluczowych użytych w opisach dokumentów. Taki wykaz stale się aktualizuje. Jest ogólną zasadą, że jeżeli znaczenie wyrażenia wielowyrazowego stanowi sumę znaczenia prostych wyrażen wchodzących w jego skład, to właśnie te wyrażenia proste powinny stać się słowami kluczowymi. Jeżeli znaczenia takiego wyrażenia nie można otrzymać z połączenia znaczenia jego elementów lub też jeżeli te elementy nie są przyjęte w języku słów kluczowych jako samodzielne słowa, to wyrażenie to może zostać uznane za słowo kluczowe, na przykład: edukacja medialna.

W literaturze przedmiotu można spotkać również katalogi wyjątków od reguł podanych przez O. Unguriana. Generalnie sprowadzają się one do nierozrywania struktur językowych tak zespolonych, że ich otrzymanie z elementów prostszych nie jest możliwe. Dotyczy to m.in. wspomnianych już *pluralia tantum*, fraz wielowyrazowych, nazw własnych i wyrażen wielowyrazowych. Instrukcje przeważnie nie ograniczają liczby słów w słowie kluczowym. Na podstawie przeanalizowanego materiału można stwierdzić, że nawet tak proste instrukcje nie są przestrzegane.

⁶⁸ Na temat zasad gramatycznych formułowania słów kluczowych pisze również E. Chmielewska-Gorczyca w pracy *Indeksy permutacyjne* (Warszawa 1977, Materiały Szkoleniowe CINTE nr 1). Rozdział 8 poświęcony jest opracowaniu leksykograficznemu zbioru słów kluczowych.

Słowa kluczowe służą do budowy kolejnego poziomu, czyli poziomu zdań języka słów kluczowych, które z kolei tworzą teksty, zwane charakterystykami/instrukcjami wyszukiwawczymi⁶⁹, stanowiące trzeci, najwyższy poziom języka słów kluczowych. Poziomami języka słów kluczowych zajmę się jeszcze przy omawianiu pragmatyki języka słów kluczowych (rozdział 3.2.3).

2.3.2.2. Wymiar horyzontalny – elementy języka słów kluczowych

Elementami języka słów kluczowych są: słownik⁷⁰ (słownictwo), reguły interpretacji (semantyka), reguły składniowe (gramatyka), a także reguły pragmatyki.

2.3.2.2.1. Słownictwo języka słów kluczowych

Języki słów kluczowych posługują się słownictwem paranaturalnym⁷¹, na ogół niekontrolowanym, derywowanym ze słownictwa języka naturalnego. Z formalnego punktu widzenia słownictwo tych języków tworzą słowa kluczowe, które najczęściej przyjmują postać prostych (słowa kluczowe) lub rozwiniętych fraz nominalnych (frazy kluczowe), na przykład naturalne nazwy klas: medycyna, zwierzęta, samochody. Wszystkie słowa kluczowe mają identyczną charakterystykę składniową; są samodzielne składniowo i wyszukiwawczo.

2.3.2.2.1.1. System leksykalny języka słów kluczowych

System leksykalny języka słów kluczowych jest tylko w niewielkim stopniu zorganizowany wewnętrznie (stopień bywa różny w zależności od rodzaju języka słów kluczowych), otwarty (łatwość przejmowania z języka naturalnego nowych jednostek językowych), niestabilny (częste zmiany relacji pomiędzy elementami słownika). Dotyczy to tylko swobodnych słów kluczowych, w innych językach słów kluczowych jest wprost przeciwnie. W klasycznym języku słów kluczowych nie ma żadnych powiązań pomiędzy słowami kluczowymi poza ich współwystępowaniem. Właśnie tym różnią się te języki od innych, w których funkcja każdej jednostki leksykalnej jest wyznaczana przez sieć relacji funkcjonalnych, w jakie ona wchodzi⁷². Zmiana tej sieci pociąga za sobą zmianę funkcji poszczególnych jednostek leksy-

⁶⁹ Tekst w języku słów kluczowych można porównać z koncepcją tekstu idealnego Rolanda Barthesa, „który (...) nie ma początku, ma charakter odwracalny, można doń się dostać przez rozliczne wejścia, z których żadne nie powinno pochopnie zostać uznane za główne” (Barthes 1999, s. 40).

⁷⁰ W przeszłości często nie zauważano, że słownik słów kluczowych jest tylko jednym z elementów języka słów kluczowych. W skrajnych przypadkach nawet utożsamiano go z językiem. Nie uświadamiano sobie faktu, że słownikowi musi towarzyszyć cały aparat językowy w postaci gramatyki i reguł semantycznych.

⁷¹ Ze względu na paranaturalny charakter słownictwa w piśmiennictwie anglojęzycznym słowa kluczowe, podobnie jak w innych językach paranaturalnych, często nazywa się deskryptorami, zaś ich słownik – tezauresem.

⁷² Zdaniem F. de Saussure’a znaczenie słowa staje się możliwe do uchwycenia tylko w zderzeniu tego słowa z innymi współistniejącymi z nim wyrażeniami (por. Saussure 2004).

kalnych. System leksykalny języka słów kluczowych dysponuje własnym mechanizmem adaptacji jednostek leksykalnych.

Dotychczasowa praktyka pokazuje, że systemy leksykalne języka słów kluczowych i innych języków paranaturalnych zmierzają do odwzorowywania struktury terminologii poszczególnych dziedzin nauki lub struktury terminologii zawartej w pewnym zbiorze dokumentów oraz terminologii odwzorowującej własności utrwalania i prezentacji informacji zawartej w dokumentach.

Klasyczny język słów kluczowych z definicji nie ma systemu leksykalnego zorganizowanego w oparciu o relacje hierarchiczne. Odpowiednia segmentacja pola semantycznego jest możliwa tylko na bazie struktury fasetowej słownictwa. Użytkownik języka czerpie każdorazowo z tego magazynu słownictwa, zarówno produkując, jak i interpretując teksty⁷³. System leksykalny tego języka powinien mieć odpowiedni zakres oraz odpowiednią wartość selekcyjną jednostek leksykalnych, czyli ich szczegółowość.

Jeżeli system leksykalny danego języka słów kluczowych opiera się na derywacji wyrażzeń z tytułów dokumentów, wówczas istnieje możliwość stosunkowo łatwej automatyzacji procesu wyboru słów kluczowych, głównie na podstawie tzw. stop-listy. Dzięki temu leksyka takiego języka opiera się na aktualnie używanej terminologii i wzajemnych związkach słów kluczowych (terminów).

Wartość informacyjna słów kluczowych w języku swobodnych słów kluczowych jest zależna od stopnia wpływu kontekstu towarzyszącego słowu kluczowemu na selekcyjną faktyczną „kluczowość słowa”. Wynika ona z obliczenia kluczowości poszczególnych wyrazów. Wartość kluczowości jest obliczana poprzez porównanie statystycznej częstości występowania danego wyrażenia w badanym zbiorze tekstów z przeciętną statystyczną częstością jego występowania w dużym korpusie języka ogólnego. Dzięki temu istnieje możliwość identyfikacji wyrazów, których częstość występowania jest nieproporcjonalnie duża w porównaniu do przeciętnej częstości ich występowania w języku ogólnym. Słowa kluczowe to zatem nie te wyrażenia, które występują najczęściej, lecz te, które występują w badanych tekstach wyraźnie częściej niż ma to miejsce w przypadku języka ogólnego. Zazwyczaj należą do jednej z trzech kategorii: nazwy własne, podstawowe terminy z danej dziedziny oraz powtarzające się słowa funkcjonalne. Dwie pierwsze kategorie są przydatne do identyfikacji słownictwa specjalistycznego danej dziedziny, natomiast trzecia nie. Jej eliminacja jest możliwa za pomocą tzw. stop-listy (listy wykluczeń), zawierającej rejestr wyrażzeń, które powinny być pomijane w procesie przetwarzania. Stosując tę procedurę, można uzyskać listę słownictwa obejmującą centralną terminologię używaną w danej dziedzinie, zawierającą najbardziej kluczowe słownictwo danej dziedziny. Wartość kluczowości jest również przydatna w algorytmach automatycznego wiązania ekwiwalentów.

System leksykalny bardziej rozwiniętych języków słów kluczowych jest bardzo rozwarstwiony. Obok leksyki, stanowiącej centrum systemu leksykalnego danego języka słów kluczowych, występują różnego typu podsystemy leksykalne (słowniko-

⁷³ Może też istnieć system leksykalny mieszany oraz system leksykalny rozproszony.

we), na przykład podsystemy terminologiczne danej dziedziny lub poszczególnych dziedzin. Każdy z tych podsystemów dysponuje z zasady pewnym zespołem swoich cech strukturalnych.

2.3.2.2.1.2. System słownikowy języka słów kluczowych

Słowa kluczowe są zwykle porządkowane i prezentowane w układzie alfabetycznym w postaci spisu słów kluczowych używanych w indeksowaniu i wyszukiwaniu dokumentów. W ten sposób powstaje wykaz słów kluczowych, który w wypadku posługiwania się językiem słów kluczowych stanowi słownik tego języka⁷⁴. W odróżnieniu od słowników innych języków informacyjno-wyszukiwawczych wykazy słów kluczowych nie mają charakteru słowników zamkniętych. Są to więc słowniki otwarte (por. koncepcja słowników otwartych w: Babik 1994). E. Ścibor pisze, że „W odróżnieniu od słowników innych języków informacyjnych wykaz słów kluczowych nigdy nie jest słownikiem zamkniętym, a osoba charakteryzująca dokumenty nie jest skrepowana istnieniem lub nieistnieniem danego słowa kluczowego w wykazie. Jedynym wymaganiem jest sprawdzenie, czy dane słowo występuje w wykazie – jeżeli nie, to należy je do niego wpisać. Wykaz słów kluczowych jest więc słownikiem »rosnącym«, a każdemu słowu kluczowemu występującemu w wykazie odpowiada w zbiorze informacyjnym przynajmniej jeden dokument scharakteryzowany za pomocą tego słowa” (Ścibor 1982, s. 21). Z tego stwierdzenia wynika, że słowniki słów kluczowych mogą być budowane tylko metodą indukcyjną oraz, że cechą różniącą języki słów kluczowych od innych jest ich otwarty charakter. Uważam jednak, że Ścibor nie ma tu racji, gdyż to zależy od przyjętych w danym systemie zasad i nie jest cechą właściwą wszystkich słowników słów kluczowych.

Makrostruktura ani mikrostruktura⁷⁵ słownika słów kluczowych do tej pory nie zostały określone odpowiednimi normami międzynarodowymi lub krajowymi. Zadaniem słownika jest prezentacja języka, a w szczególności eksplikacja znaczeń słów kluczowych. Pozycją słownika słów kluczowych jest artykuł słowa kluczowego, składający się ze słowa kluczowego lub słów kluczowych związanych różnymi relacjami, które są połączone za pomocą odpowiednich odsyłaczy⁷⁶. Jest to sposób definiowania słów kluczowych umożliwiający także wprowadzanie definicji projektujących lub definicji regulujących, a więc zmieniających denotację i konotację wyrażań języka naturalnego w funkcji słów kluczowych. Jest on elementem tzw. mikrostruktury słownika. Nie dotyczy to klasycznego języka słów kluczowych.

Słowniki słów kluczowych stanowią zbiory słów kluczowych służących jako środek leksykalny do reprezentowania w formie metainformacji treści dokumentów

⁷⁴ W tradycji anglo-amerykańskiej wszelkie słowniki języków paranaturalnych nazywa się tezaurusami. W polskiej terminologii języków informacyjno-wyszukiwawczych jego zakres został ograniczony tylko do słowników języków deskryptorowych. Praktyka ta utrzymuje się do tej pory w nazewnictwie na Zachodzie, gdzie słowniki tego typu, niezależnie od języka, nadal nazywane są tezaurusami.

⁷⁵ Makrostrukturę słownika słów kluczowych stanowi przyjęta w danym słowniku siatka haseł słownikowych. Mikrostruktura dotyczy struktury wewnętrznej poszczególnych haseł słownikowych.

⁷⁶ Widać tu wyraźnie wpływ języka haseł przedmiotowych.

oraz potrzeb informacyjnych. W słownikach tych są eksplikowane tylko niektóre związki semantyczne zachodzące pomiędzy słowami kluczowymi jako elementami słownika, mające ułatwiać proces wyszukiwania informacji na różnych poziomach szczegółowości.

Słowniki języków słów kluczowych⁷⁷, podobnie jak innych języków informacyjno-wyszukiwawczych, są słownikami dwujęzycznymi i przekładowymi⁷⁸, pozwalającymi na tłumaczenie wyrażonych w języku naturalnym tekstów charakterystyk słownych dokumentów i pytań informacyjnych użytkowników na odpowiednie teksty języka słów kluczowych⁷⁹. W przypadku języków słów kluczowych i innych języków paranaturalnych język naturalny pojawia się tylko w określonych przypadkach, tzn. wtedy, gdy wyrażenie języka naturalnego potencjalnie mogłoby stać się słowem kluczowym. Jednak w danym języku słów kluczowych jest reprezentowane przez inne wyrażenie, często synonimiczne z nim w języku naturalnym lub zakresowo szersze albo węższe.

⁷⁷ Pojęcie „słownik” staje się obecnie pojęciem centralnym, zwłaszcza w kontekście rozwoju teoretycznych koncepcji języków informacyjno-wyszukiwawczych oraz projektowania systemów wyszukiwania informacji w związku z bibliotekami cyfrowymi. Znaczące wzbogacenie tego pojęcia zauważył m.in. Michael Buckland ze Szkoły Zarządzania Informacją i Systemami Uniwersytetu Kalifornijskiego (USA). Postrzega on słownik jako najważniejszy podstawowy typ struktury danych opisujących zasoby dokumentów i informacji. Nazwą tą obejmuje wszelkiego rodzaju indeksy i kartoteki wzorcowe (Buckland 1999).

⁷⁸ Słowniki słów kluczowych są słownikami przekładowymi z języka naturalnego na sztuczny język informacyjno-wyszukiwawczy. Trudno się zgodzić z autokorektą E. Chmielewskiej-Gorczyca, gdy twierdzi, że „tezauryusy można (...) traktować jako słowniki przekładowe, ale nie z języka naturalnego na sztuczny JIW, a z quasi-naturalnego języka indeksatora i użytkownika (języka swobodnych słów kluczowych) na język deskryptorowy” (Chmielewska-Gorczyca 1995, s. 5). Słowa kluczowe jako elementy charakterystyki słownej dokumentu niosą metainformacje odwzorowujące elementy treści dokumentów, mimo iż przyznaje im się status wyrażen języka naturalnego. Konsekwencją uznania słownika słów kluczowych za słownik przejścia z języka naturalnego na sztuczny język słów kluczowych i odwrotnie jest uznanie go za słownik przekładowy, służący do przekładu pewnych specjalnych tekstów języka naturalnego na język informacyjny i odwrotnie. Tymi tekstami są charakterystyki słowne dokumentów i pytań informacyjnych, a rezultatem tłumaczenia – charakterystyki wyszukiwawcze dokumentów i instrukcje wyszukiwawcze zapytań. Rozumiany w ten sposób słownik słów kluczowych jest zawsze słownikiem wielojęzycznym, przynajmniej dwujęzycznym.

⁷⁹ Jednym z pierwszych, który potraktował słownik języka informacyjno-wyszukiwawczego jako słownik przekładowy był Jean-Claude Gardin, uważany za ojca leksykografii dokumentacyjnej (Gardin 1966). W publikacji pt. *Éléments d'un modèle pour la description des lexiques documentaires* wyróżnił leksykografię dokumentacyjną, która zajmuje się badaniem i sposobami tworzenia słowników dokumentacyjnych oraz leksykografię naturalną, zajmującą się badaniem i budową słowników językowych języków naturalnych. Leksykografia dokumentacyjna – jego zdaniem – opiera się na dwóch podejściach: podejściu charakteryzującym się tym, że słowniki nie podają w jawny sposób (explicite) odpowiedniości między jednostkami leksykalnymi języka informacyjno-wyszukiwawczego/terminami a wyrażeniami języka naturalnego reprezentowanymi w języku informacyjno-wyszukiwawczym przez dane jednostki leksykalne (słowniki implicite); oraz podejściu charakteryzującym się tym, że słowniki języka informacyjno-wyszukiwawczego w jawny sposób podają listę tych odpowiedników poprzez zestawienie pewnego rodzaju słownika dwu- lub wielojęzycznego, w którym z jednej strony znajdują się wyrażenia języka lub języków naturalnych, z drugiej zaś przyjęte ekwiwalenty w słowniku dokumentacyjnym. Gardin proponował nazwanie tych słowników tezaursami, co było zgodne z tendencjami w ówczesnej praktycznej działalności informacyjnej na Zachodzie. Nie wnikając w szczegóły, J.C. Gardin nazywa – podobnie jak autorzy z krajów zachodnich – terminami jednostki leksykalne języków o słownictwie paranaturalnym, niezależnie od typu języka.

W językach słów kluczowych występują dwa typy słowników, w których jednej jednostce leksykalnej języka słów kluczowych odpowiada:

- tylko jedno wyrażenie (wyraz lub połączenie wyrazowe) języka naturalnego, zawsze równoznaczne z jednostką leksykalną języka słów kluczowych;
- jedno lub kilka wyrażen języka naturalnego, w tym przynajmniej jedno równoznaczne z jednostką leksykalną języka (tak jak w języku deskryptorowym).

W języku informacyjno-wyszukiwawczym słownik jest tablicą przejść z języka naturalnego dokumentów na język informacyjny, jednak w słownikach słów kluczowych „wskazanie jednostek języka naturalnego powiązanych relacją odpowiedniości jest całkowicie lub częściowo niejawne, dlatego że w słownikach języka słów kluczowych nie podaje się w sposób jawny jednostek języka naturalnego, które są równoznaczne z daną, wymienioną w słowniku jednostką leksykalną języka słów kluczowych” (Ścibor 1982, s. 120).

Dla precyzji opisu treści dokumentu w językach słów kluczowych i wykluczenia szumu informacyjnego przydatne są ustalenia relacji między słowami kluczowymi. Struktura słowników słów kluczowych podlega stałemu rozwojowi. Początkowo słownik słów kluczowych posiadał wyłącznie alfabetyczny zestaw słów kluczowych. Z biegiem czasu praktyka wykazała potrzebę wzbogacenia metod tworzenia słowników słów kluczowych o nowe środki, które już nie mieszczą się w definicji klasycznego języka słów kluczowych, o czym szerzej będzie mowa w trzeciej części rozprawy.

Metodyka budowy słowników słów kluczowych przewiduje gromadzenie słów kluczowych, opracowanie słownictwa, wprowadzenie relacji, opracowanie ich struktury. Sposób opracowania słownika słów kluczowych zależy od kategorii użytkowników systemu wyszukiwania informacji, od stanu danej dyscypliny naukowej/dziedziny działalności praktycznej w okresie opracowywania; niemały wpływ mają tu również przyjęte, często subiektywne, konwencje zespołu. Siatka słów kluczowych może więc posiadać luki w słownictwie, tj. braki słów kluczowych ze względu na brak publikacji dotyczących określonych zagadnień lub podwójne słowa kluczowe dla przedmiotów, treści istotnych historycznie obok współczesnych, ząębających się zakresem. Może też wykazywać nierównomierne wyszczególnianie przedmiotów (nazw) jednej dyscypliny lub danego obszaru wiedzy w stosunku do innych, zgodnie jednak z ustalonym zakresem danego języka słów kluczowych.

Język słów kluczowych w odmianie języka swobodnych słów kluczowych, bazującego na słowniku mentalnym, jest znacznie uboższy od języka kontrolowanych słów kluczowych, opartego na jawnym słowniku, gdyż nie zawiera odsyłaczy. Poza tym słowa kluczowe charakteryzuje przypadkowość, dlatego często zachodzą na siebie zakresami. Nie ma bowiem struktury wiążącej poszczególne swobodne słowa kluczowe w jakiś system, oczywiście poza relacją współwystępowania, co pozwala na swobodne stosowanie i tworzenie słów kluczowych⁸⁰.

Oprócz tradycyjnej, klasycznej formy słownika słów kluczowych można wyróżnić słownik hybrydowy, który łączy do swojego zasobu w całości lub częściowo inne gotowe zbiory słownictwa. Za nowsze formy słowników słów kluczowych uwa-

⁸⁰ Niewątpliwie na wybór słów kluczowych wpływa dyscyplina naukowa/rodzaj działalności praktycznej, typ sytuacji zadaniowej oraz typ intelektu.

za się kartotekę/listę wzorcową słów kluczowych⁸¹ oraz kartotekę/listę wzorcowych słów kluczowych⁸².

Niewątpliwą zaletą kartotek/list wzorcowych słów kluczowych jest wzbogacenie informacji o znaczeniu słów kluczowych oraz ich stosowaniu w opisie rzeczowym dokumentów. Pozwala to zapobiegać utożsamianiu przez użytkowników denotacji równoznacznych wyrażen dwóch różnych języków (języka naturalnego i języka słów kluczowych), chociaż mających taką samą postać, do czego często skłania kompetencja w zakresie języka naturalnego. Żeby jednak z tych metajęzykowych informacji skorzystać, użytkownik musi zdawać sobie sprawę z możliwości występowania takich różnic, powodujących wyszukanie innych informacji. Zwykle jest jednak intuicyjnie skłonny polegać na własnej kompetencji językowej z zakresu języka naturalnego.

Obecnie w zautomatyzowanych systemach wyszukiwania informacji słowniki słów kluczowych są już przeważnie zintegrowane z systemem, co wnosi szereg usprawnień do procesu indeksowania i wyszukiwania informacji. Jako narzędzie wyszukiwawcze słownik słów kluczowych zyskuje nową jakość. Jego zawartość można podglądać na każdym etapie procesu wyszukiwania informacji oraz modyfikować wyjściową instrukcję wyszukiwawczą w miarę uzyskiwania kolejnych informacji. Informacje ze słownika słów kluczowych wzbogacają indeks zbioru wyszukiwawczego, a zbiór wyszukiwawczy wzbogaca słownik na przykład o liczbę wystąpień wyrażen języka mówiącą, ile dokumentów jest zaindeksowanych danym słowem kluczowym, co ułatwia m.in. przyjęcie odpowiedniej strategii wyszukiwawczej. Dotyczy to w szczególności słowników w formie kartotek haseł wzorcowych, list wzorcowych słów kluczowych. Są to słowniki online zintegrowane z systemem wyszukiwania informacji.

Przejawem takiej integracji są również katalogi OPAC, w których kartoteka wzorcowa słów kluczowych jest przechowywana w postaci rekordów o odpowiednio zdefiniowanej strukturze i zawartości informacyjnej, zintegrowana ze zbiorem głównym systemu. Ten typ kartoteki stanowi w zasadzie zbiór kartotek zawierających leksykę różnych kategorii: hasła osobowe, hasła korporatywne, tytuły ujednolicone, hasła rzeczowe. Wymienione rodzaje haseł bibliograficznych oraz hasła charakterystyki rzeczowej są ściśle z sobą związane. Jest to możliwe dzięki nazwaniu i odwzorowaniu w kartotece wzorcowej, jako jednym zbiorze jednostek leksykalnych, relacji

⁸¹ Wykazami słów kluczowych są tzw. kartoteki/listy wzorcowe słów kluczowych. Określić je można jako skończony, uporządkowany alfabetycznie zbiór słów kluczowych typowych dla piśmiennictwa z danej dziedziny. Lista wzorcowa słów kluczowych może być pomocna przy budowie słownika języka słów kluczowych, lecz nie może go zastąpić ze względu na obejmowanie słownictwa występującego w ściśle określonym pod względem tematycznym i ilościowym zbiorze dokumentów.

⁸² Kartoteki/listy wzorcowych słów kluczowych stanowią jedną z form organizacji i kontroli leksyki w językach informacyjno-wyszukiwawczych, pełnią funkcje tablic klasyfikacji, słowników alfabetyczno-systematycznych i instrukcji indeksowania. Ta forma organizacji i kontroli słownictwa różni się od innych form stosowanych w zautomatyzowanych systemach katalogowych. Kartoteka wzorcowych słów kluczowych zawiera słownictwo (zasób leksykalny) przydatne dla użytkownika tworzącego instrukcję wyszukiwawczą. Jest to więc nieskończony zbiór słów kluczowych. Oprócz leksyki kartoteka taka zawiera szereg metainformacji o znaczeniu wyrażen i sposobach ich wykorzystywania w procesie wyszukiwania informacji.

zachodzących pomiędzy różnymi hasłami. Są to następujące relacje paradygmatyczne: relacja ekwiwalencji (odpowiedniości) wyszukiwawczej, relacje hierarchiczne (generyczna i mereologiczna)⁸³ oraz wybrane relacje kojarzeniowe⁸⁴.

Klasyczny język słów kluczowych relację ekwiwalencji opiera wyłącznie na kompetencji użytkownika w zakresie posługiwania się językiem naturalnym. Kompetencja ta bywa wspomagana przez system językowy języka słów kluczowych, a więc element sztuczny. Jest to rodzaj ingerencji „zewnętrznej” w stosunku do owej „pierwotnej” struktury i wówczas mamy do czynienia już nie z „klasycznymi” językami słów kluczowych, lecz z „nieklasycznymi” językami słów kluczowych. Tu granica jest płynna, stanowiąc swoistego rodzaju kontinuum, zaś faktyczne granice są ustalane w sposób arbitralny przez dany system informacyjno-wyszukiwawczy.

Człony relacji odpowiedniości, a właściwie elementy klasy równoważności, oznaczone w zbiorze wyrażeń języka naturalnego i reprezentowane przez jedno wyrażenie nie muszą być ani synonimami, ani wyrażeniami bliskoznacznymi. Stanowią istotną różnicę między przekładalnością słownika języka informacyjno-wyszukiwawczego i słownika języka naturalnego. J. Woźniak-Kasperek określa to w ten sposób: „Błędne jest interpretowanie relacji ekwiwalencji jako związku zachodzącego między wyrażeniami języka słów kluczowych i każdym z elementów przyporządkowanej mu klasy wyrażeń języka naturalnego. Przeciwnie: relacja ekwiwalencji zachodzi między denotacją i konotacją wyrażeń języka informacyjno-wyszukiwawczego a sumą denotacji i konotacji wyrażeń odrzuconych wyrażenia przyjętego. Inaczej mówiąc, gdyby wyrażenie języka naturalnego tworząc klasę ekwiwalencji reprezentowaną jako całość przez jedno wyrażenie języka słów kluczowych należało do leksyki tego języka, to tu ich znaczenie nie byłoby różnicowane, a więc wskazywałoby same zbiory dokumentów relewantnych” (Woźniak 1999, s. 45). Jeżeli zatem w danym słowniku słów kluczowych znajdujemy słowo kluczowe x i termin odrzucony y , to znaczy, że w tym słowie kluczowym znaczenie słowa kluczowego x jest sumą logiczną znaczeń x i y . Można to również odnieść do tekstów w języku słów kluczowych.

Cechy związane z różnymi sposobami przejścia z języka naturalnego na język informacyjno-wyszukiwawczy i odwrotnie, obok innych cech związanych ze słownictwem, prezentacją tego słownictwa oraz gramatyką, posłużyły E. Ściborowi do skonstruowania typologii strukturalnej języków informacyjnych (Ścibor 1982). Ze względu na sposób przejścia z języka naturalnego na język informacyjno-wyszukiwawczy i odwrotnie, wyróżnił on najpierw trzy typy słowników tych języków, a więc odpowiednio słowniki, w których przejście to jest niejawne, częściowo jawne i całkowicie jawne. Zwracam uwagę na możliwość wykorzystania tej typologii w odniesieniu do słowników języka słów kluczowych. W słownikach pierwszego typu w sposób jawny podaje się tylko elementarne jednostki leksykalne danego języka słów kluczowych, które są zawsze paranaturalne. Jednostki te jako paranaturalne są równokształtne z jednym i tylko jednym wyrażeniem języka naturalnego. Takich wyrażeń języka naturalnego nie umieszcza się w słowniku słów kluczowych, gdyż

⁸³ Relacje hierarchiczne są systemotwórcze.

⁸⁴ Relacje kojarzeniowe są sieciotwórcze.

w słownikach tych nie podaje się jednostek języka naturalnego równokształtnych z jednostkami języka słów kluczowych, powiązanych z nimi relacją odpowiedniości wyszukiwawczej. Umieszczenie w słowniku słów kluczowych danego słowa kluczowego jest dyrektywą lub może być wzorcem charakteryzowania tą jednostką leksykalną wszystkich dokumentów, w których tekście występuje dane słowo i jest ono istotne dla wyrażenia treści. Dane słowo kluczowe jest również używane do charakteryzowania tych dokumentów, w których w sposób jawny lub ukryty występują synonimy danego słowa. Oczywiście następuje to tylko wtedy, gdy w słowniku danego języka nie została umieszczona elementarna jednostka leksykalna będąca synonimem danego słowa kluczowego. Zależy to od konkretnego języka. W słownikach charakteryzujących się całkowicie niejawnym przejściem z języka naturalnego na język słów kluczowych, i odwrotnie, nie wprowadza się ani żadnych oznaczeń relacji odpowiedniości wyszukiwawczej między wyrażeniami języka naturalnego a słowami kluczowymi jako elementarnymi jednostkami leksykalnymi języka słów kluczowych, ani żadnych uwag na temat synonimii zachodzącej pomiędzy poszczególnymi wyrażeniami języka naturalnego. Nie ma więc żadnych wskazówek dotyczących stosowania danego słowa kluczowego. Ustalenie odpowiedniości wyszukiwawczej między słowem kluczowym a nierównokształtnym z nim wyrażeniem języka naturalnego najczęściej opiera się na wiedzy lub intuicji osoby charakteryzującej dokument.

Relacja ekwiwalencji ma duży wpływ na efektywność systemu informacyjno-wyszukiwawczego i skuteczność poszukiwań prowadzonych przez użytkownika. Jeżeli w słowniku brakuje jawnych zapisów ustanawiających odpowiedniość pomiędzy wyrażeniami języka naturalnego i słowami kluczowymi, może zdarzyć się, że użytkownik systemu posługujący się tym językiem dokona, opierając się na swojej osobniczej wiedzy i kompetencji, innego przyporządkowania i będzie szukał dokumentu pod innym słowem kluczowym niż to, które przydzielił dokumentowi indeksator. Ba, nawet jawnie określona relacja, ale nierelevantna z punktu widzenia potrzeb użytkowników i charakteru opisywanych dokumentów, może sprawić, że na poziomie opisu rzeczowego dokumenty stają się nierozróżnialne. Ciężar selekcji i oceny relewancji leksykalnej spada więc na użytkownika, który zwykle nie jest świadomy sytuacji, w jakiej się znalazł, ani nie dysponuje odpowiednią wiedzą. W języku słów kluczowych ten mechanizm przejścia realizowany jest w różny sposób: od słownika jawnego, po słownik niejawny (por. rozdział 1.4.3.2).

Słowniki charakteryzujące się całkowicie niejawnym przejściem z języka słów kluczowych na język naturalny, i odwrotnie, są stosowane obecnie bardzo rzadko. Wyrażenia języka naturalnego powiązane relacją odpowiedniości wyszukiwawczej ze słowem kluczowym są podawane z reguły w wielu słownikach słów kluczowych, a także w słownikach haseł przedmiotowych.

2.3.2.2.2. System semantyczny języka słów kluczowych

Słowa kluczowe mają swoją autonomię semantyczną, tzn. w tekstach języka słów kluczowych każde słowo wyznacza odpowiednią klasę dokumentów i poza relacją współwystępowania nie jest powiązane semantycznie z innymi występującymi w tek-

ście jednostkami leksykalnymi, a więc znaczenie poszczególnych słów kluczowych w tekście nie zależy od pozostałych. Tak samo jest w słowniku słów kluczowych, gdzie poszczególne słowa kluczowe są autonomiczne semantycznie i nie wskazują w sposób systemowy na określone, typowe fragmenty rzeczywistości dokumentacyjnej. Paradygmatyka języków słów kluczowych bywa jawna lub niejawna. W tym drugim przypadku bazuje na kompetencjach i możliwościach odkrycia jej przez użytkowników tych języków.

Funkcje słów kluczowych łączą się nierozdzielnie z procesami denotacji i konotacji w językach informacyjno-wyszukiwawczych. To zagadnienie dualnego charakteru znaku języka informacyjno-wyszukiwawczego w polskiej literaturze przedmiotu wnikliwie przeanalizowała i opisała B. Sosińska-Kalata, wyróżniając:

- dwa typy denotacji, czyli dwa rodzaje obiektów, które desygnują wyrażenia tych języków: desygnaty bezpośrednie – dokumenty opisywane przez język informacyjno-wyszukiwawczy, oraz desygnaty pośrednie, czyli obiekty rzeczywistości pozadokumentacyjnej opisywane w dokumentach;
- dwa typy konotacji: konotację bezpośrednią, którą stanowią zbiory cech desygnatów bezpośrednich, czyli zbiory własności dokumentów, oraz konotację pośrednią: własności obiektów rzeczywistości pozadokumentacyjnej (Sosińska 1989).

Istnieje ścisły związek między tymi sferami interpretacji znaczenia znaków owych języków. W systemach dokumentacyjnych, żeby uzyskać informację o obiekcie rzeczywistości pozajęzykowej, trzeba odwołać się do rzeczywistości tekstowej. Z kolei możliwość informowania przez system o przedmiotach rzeczywistości pozadokumentacyjnej jest uzależniona od uprzedniego odwzorowania w jego zbiorach (dokumentach) pewnej wiedzy. Stąd słowa kluczowe to wyrażenia języka naturalnego użyte metainformacyjnie, tj. odniesione do informacji zawartych w dokumentach, a nie bezpośrednio do obiektów rzeczywistości.

Reguły semantyczne języków informacyjno-wyszukiwawczych, oparte na teorii mnogościowych działaniach sumy i iloczynu zbiorów desygnatów bezpośrednich, odnoszą się do denotacji bezpośredniej znaków tych języków. Do denotacji pośredniej sięga się przy określaniu w systemie leksykalnym relacji mereologicznych oraz części relacji kojarzeniowych, które nie polegają na współwystępowaniu wyrażen w tekstach, a także przy konstruowaniu wyrażen o strukturze semantycznie umotywowanej. O relacjach, w jakie wchodzi wyrażenie, tj. o miejscu wyrażenia w strukturze języka informacyjnego, decyduje konotacja bezpośrednia i pośrednia, czyli zespół cech semantycznych wynikających z postulatów znaczeniowych danego języka, niezbędnych i wystarczających (tzw. konotacja charakterystyczna) do identyfikacji bezpośrednich i pośrednich desygnatów wyrażen.

W organizacji systemu leksykalnego języka słów kluczowych ważną rolę odgrywa pole semantyczne, czyli strefa pojęciowa, której znaczenia składowe wyrażane są przez słowa kluczowe. Zarówno podział słownictwa na pola semantyczne, jak i wewnętrzna organizacja poszczególnych pól są swoiste dla każdego z języków słów kluczowych. Znaczenie słów kluczowych rozumiane jest jako relacja wewnątrzsystemowa, zachodząca pomiędzy jednostkami tego systemu, jakkolwiek bazująca na kompetencji użytkownika w zakresie języka naturalnego. W języku swobodnych

słów kluczowych są wyrażenia języka naturalnego w funkcji metainformacyjnej, stąd swoje znaczenie czerpią w całości z języka naturalnego. Słownik i słowa kluczowe podlegają zatem takim samym procesom zmian jak wyrazy języka naturalnego. Można więc mówić o rejestrach słów kluczowych, czego przykładem są dziedzinowe słowniki słów kluczowych. Zbiory słów kluczowych mają charakter względny, stąd na przykład naukowiec będzie szukał słów kluczowych w terminologii danej dziedziny, handlowiec w nazewnictwie/nomenklaturze. W innych odmianach języka słów kluczowych wprowadzie znaczenie ich wyrażeń również jest czerpane z języka naturalnego, ale w części jest modyfikowane stosownie do potrzeb systemu, tworząc mikropola semantyczne odbiegające od takich samych w języku naturalnym, czyniąc przez to – według zasady *pars pro toto* – ten język językiem sztucznym, chociaż można by się upierać nie bez racji, że nadal jest to język naturalny. Szczególnego typu stosunkami funkcjonalnymi istniejącymi wewnątrz pola semantycznego danego języka słów kluczowych są relacje antonimii i relacje synonimii.

Istotą języka słów kluczowych jest jego płaska⁸⁵, lub prawie płaska⁸⁶, struktura paradygmaticzna. Słowa kluczowe są nośnikami informacji o treści dokumentu. Opisy systemowe są zatem formą organizacji informacji za pomocą języka słów kluczowych, opartą na płaskiej strukturze tego języka, która niczego nowego nie wnosi poza informacją o współwystępowaniu takich a nie innych informacji. W związku z tym informacja jest bardzo uboga, bo odwołuje się tylko do listy elementów treści danego dokumentu lub pytania informacyjnego. Jediną relacją paradygmaticzną tego języka jest relacja współrzędności, a więc równoważności poszczególnych wyrazów w funkcji kluczy wyszukiwawczych. Ten rodzaj relacji jest oznaczany w słownikach słów kluczowych za pomocą odsyłacza „zob.”. Dotyczy to wyłącznie klasycznych języków słów kluczowych. Niektórzy wprowadzają w „paradygmatyce” tego języka również odsyłacz „zob. też” na oznaczenie innych rodzajów relacji, ale czyni się to tylko w nieklasycznych odmianach języka, i/lub relacje syntagmatyczne, których wykładnikami językowymi są wskaźniki więzi, roli i wagi.

W relacyjnej organizacji paradygmatyki nieklasycznego języka słów kluczowych najczęściej wyróżnia się trzy typy relacji: relację synonimii, sporadycznie relację hierarchiczną⁸⁷ i tzw. relację kojarzeniową⁸⁸, włączając do tej ostatniej relację tematyczną. Charakterystyka tych związków jest przedstawiona w wielu publikacjach, dlatego nie zamierzam omawiać ich szczegółowo.

Języki słów kluczowych od samego początku nie wykorzystywały organizujących i porządkujących walorów relacji hierarchicznych, przyjmując płaską strukturę pola semantycznego i charakterystyczną dla sieci relację współwystępowania. To sie-

⁸⁵ Struktura płaska to struktura, której elementy nie są powiązane żadną relacją hierarchiczną. Występuje ona w językach słów kluczowych oraz wczesnych typach języków deskryptorowych, to jest w językach, w których słownikach oznaczono jedynie relację odpowiedniości wyszukiwawczej.

⁸⁶ Struktura prawie płaska oznacza strukturę płaską zakłóconą tylko lokalnie, a nie globalnie.

⁸⁷ Relacje hierarchiczne dają uporządkowanie sekwencyjne typu 1:1.

⁸⁸ Relacje kojarzeniowe wprowadzają inny rodzaj uporządkowania informacji, a mianowicie za pomocą węzłów i linków. Są to uporządkowania typu: 1:n, n:1 oraz n:n. Relacje kojarzeniowe umożliwiają globalne uporządkowanie, a nie tylko lokalne, jak relacje hierarchiczne.

ciowe podejście w tej klasie języków znajduje swoją podstawę/uzasadnienie w sieciowej strukturze procesów zapamiętywania i przetwarzania informacji przez człowieka, związanej z procesami kojarzeniowymi. Dzięki takiemu podejściu ważniejszy stał się fakt połączeń poszczególnych elementów tworzących sieć niż rodzaj tego połączenia, który jest charakterystyczny i istotny dla systemu, czyniąc tę strukturę „wyzisolowaną” z otoczenia. Charakterystyczną cechą sieci jest jej otwartość, zmienność i modyfikowalność. Pełna realizacja tego podejścia stała się możliwa dopiero w epoce komputerowej, stanowiąc podstawę sieciowego paradygmatu w językach informacyjno-wyszukiwawczych. To, w początkowej fazie nie do końca uświadamiane przez twórców języków słów kluczowych, oparcie na poznawczej sferze ludzkich skojarzeń, będącej wyznacznikiem sieciowości, stopniowo jest przejmowane również przez języki tradycyjne. Szczegółowego opisu formalnego takiego systemu leksykalnego języka dokonał ostatnio Alain Polguère z University of Montreal w Kanadzie (Polguère 2008).

Rezygnacja w języku słów kluczowych z relacji (powiązań) hierarchicznych rodzi jednak pewne problemy. Trzeba wcześniej przewidzieć w „płaskim” opisie nakładanie się zakresów słów kluczowych odnoszących się w specyficzny sposób do informacji z poziomu nadrzędnego. W wielu przypadkach taki „płaski” opis staje się skrajnie skomplikowany, a nawet trudno czytelny dla użytkownika tego języka z powodu zbyt dużej liczby powiązań. Brak relacji hierarchicznych nie pozwala na stosowanie indeksowania hierarchicznego, które może być narzędziem zwiększania kompletności odpowiedzi systemu. Jest to ważne, gdyż w języku słów kluczowych nie istnieje możliwość zwiększania dokładności odpowiedzi, jeśli konkretny język słów kluczowych nie dysponuje odpowiednio szczegółowymi wyrażeniami, możliwymi do użycia zarówno w budowie charakterystyk wyszukiwawczych dokumentów, jak i instrukcji wyszukiwawczych.

Ważnym składnikiem organizacji systemu leksykalnego języka słów kluczowych jest też podział tego systemu na klasy funkcjonalne, na przykład poprzez kategoryzację. Podziały takie stanowią szczególnie typ ingerencji w wewnętrzną organizację systemu leksykalnego języka.

Słowa kluczowe zapewniają łatwe i szybkie odnajdowanie informacji za pomocą znanych użytkownikowi nazw interesujących go obiektów, na przykład nazwisk, nazw towarów, tytułów dokumentów, ale wyszczególniając informację o jednostkowych obiektach, nie zapewniają generalizacji.

Dla języków słów kluczowych i innych języków paranaturalnych charakterystycznym zjawiskiem językowym jest synonimia wyszukiwawcza. Jest to relacja między wyrażeniami języka informacyjno-wyszukiwawczego mającymi identyczną denotację bezpośrednią i konotację bezpośrednią oraz identyczną denotację pośrednią i konotację pośrednią. Występuje na poziomie elementarnych jednostek leksykalnych, a także często na poziomie syntaktycznym. Może to być synonimia między wyrażeniem elementarnym a wyrażeniem złożonym, na przykład „choroby układu krążenia” (wyrażenie elementarne) i „Układ krążenia. Choroby” (wyrażenie złożone). Jej przyczyną są błędy popełniane przez projektantów tych języków. Synonimia wyszukiwawcza w niektórych językach informacyjno-wyszukiwawczych ma cha-

rakter systemowy, na przykład w UKD zasada odwracalności symboli połączonych znakiem łączącym symetrycznej zależności semantycznej (dwukropkiem). Jeżeli w wyszukiwaniu uwzględnimy zjawisko synonimii wyszukiwawczej, to strat informacji w procesach wyszukiwania informacji i obniżenia współczynnika kompletności wyszukiwania nie ma (Bojar 2002, oprac.).

Przypomnę tylko, że synonimia jest językową relacją równoważności, a więc relacją symetryczną, zwrotną i przechodnią, która zachodzi pomiędzy werbalnymi reprezentacjami tego samego pojęcia. Synonimia występuje wyłącznie w tych językach informacyjno-wyszukiwawczych, w których nie wyznacza się form preferowanych, a wyszukiwanie uwzględnia możliwość reprezentowania określonego pojęcia przez wiele semantycznie i/lub wyszukiwawczo ekwiwalentnych wyrażen. Do niedawna relacja ta występowała w strukturze organizacji systemu leksykalnego tylko niektórych języków słów kluczowych i dlatego w literaturze przedmiotu traktowana była marginalnie. Więcej uwagi poświęcano relacji ekwiwalencji wyszukiwawczej. Obecnie jej znaczenie wzrasta w związku z coraz częstszym wprowadzaniem jej do tezaurusów wyszukiwawczych dostępnych online.

Synonimia w językach słów kluczowych jest traktowana szeroko, szerzej niż w języku naturalnym, na przykład: Antyklerykalizm – Klerykalizm – w języku słów kluczowych są synonimami; w języku naturalnym nie. Dwa wyrażenia są w języku słów kluczowych (na poziomie jednostek leksykalnych) synonimami wyszukiwawczymi tylko wtedy, gdy jest to wyraźnie zaznaczone w słowniku danego języka, na przykład: Językoznawstwo – Lingwistyka. Synonimami w języku słów kluczowych są również wyrażenia, które w języku naturalnym nie są synonimami, na przykład: Watykan – Papiestwo, Atmosfera – Powietrze, Religia – Ateizm (por. Bojar 2005).

Języki słów kluczowych mają własne swoiste zasady prezentacji struktury języka, formalizacji i prezentacji systemów językowych. Języki swobodnych słów kluczowych (w klasycznej postaci swobodnych słów kluczowych) z założenia nie prezentują relacji paradygmatacznych. Jedynym sposobem prezentacji systemów semantycznych klasycznych języków słów kluczowych jest odwołanie się – *explicite* lub *implicite* – do znaczenia skorelowanych ze słowami kluczowymi wyrażen języka naturalnego. Opiera się to na założeniu, że użytkownik języka sam odkryje znaczenie słowa kluczowego na podstawie znajomości znaczenia przyporządkowanego mu lub z nim równoznacznego wyrażenia języka naturalnego. Nieklasyczne języki słów kluczowych prezentują tylko wybrane relacje paradygmataczne za pomocą systemów odsyłaczy. Prezentacja *explicite* relacji paradygmatacznych tych języków jest jednym ze sposobów częściowej prezentacji ich systemu semantycznego, ograniczonej do wybranych opozycji semantycznych, nie zawsze wyraźnie określonych, stąd niebezpieczeństwo interferencji językowej. Częściowa prezentacja paradygmataki staje się bardziej jawna za pomocą systemu odsyłaczy. Komplementarnym, a w języku swobodnych słów kluczowych jedynym, sposobem prezentacji systemu semantycznego jest odwoływanie się – *explicite* lub *implicite* – do znaczenia wyrażen języka naturalnego, a szczególnie jego podzbioru leksyki zwanego terminologią, skorelowanych z równoznacznymi słowami kluczowymi. Znaczenie swobodnych słów kluczowych jest wyznaczone przez paradygmatakę języka naturalnego z zastrzeżeniem, że języki

słów kluczowych i języki naturalne odwzorowują dwie różne rzeczywistości. Języki słów kluczowych odwzorowują rzeczywistość dokumentacyjną, gdyż wyrażeniom tych języków przyporządkowane są klasy dokumentów zawierających odpowiednie informacje. Języki naturalne odwzorowują rzeczywistość obiektywną, gdyż ich wyrażenia są przyporządkowane klasom realnych obiektów (relacji, sytuacji).

Ze względu na sposób organizacji leksyki języka słów kluczowych można zastanawiać się nad zasadnością określenia go jako system leksykalny. Trafniejszym terminem byłaby, być może, „sieć leksykalna” języka słów kluczowych, co sygnalizowałoby inne niż dotychczas spojrzenie na leksykę tych języków.

Języki słów kluczowych dają bardzo uproszczony obraz pola semantycznego, a zarazem odwzorowywanej rzeczywistości, ze względu na małą siłę relacji semantycznych. Małą siłą organizującą mają relacje kojarzeniowe. Tak samo relacje tematyczne jako relacje organizujące są bardzo słabe semantycznie. Wprawdzie języki te mają mniejszą siłę semantyczną niż inne, ale za to oferują większą elastyczność i adaptacyjność oraz znacznie, „asocjacyjnie”, poszerzają użytkownikowi pole widzenia.

2.3.2.2.3. System gramatyczny języka słów kluczowych

Elementem określającym tożsamość języka słów kluczowych jest także rodzaj gramatyki. Gramatyka języka słów kluczowych w porównaniu z gramatyką języka naturalnego czy innych języków informacyjno-wyszukiwawczych jest prostsza. Jest to tzw. gramatyka zerowa⁸⁹, w której kolejność elementarnych jednostek leksykalnych jest dowolna, a miejsce każdej jednostki leksykalnej może być zmienione, nie powodując zmiany znaczenia zdania jako całości. Wszystkie słowa kluczowe tworzące zdanie są na tym samym poziomie, czyli są od siebie niezależne, a więc nie istnieje między nimi podporządkowanie. W charakterystyce wyszukiwawczej dokumentu, jako tekście w języku słów kluczowych, nie stosuje się żadnych znaków umownych, które oddzielałyby lub za pomocą wykładników relacji syntagmatycznych⁹⁰ łączyły z sobą elementarne jednostki leksykalne. Właśnie znaczny stopień (być może maksymalny) uproszczenia i redukcji składni jest jedną z charakterystycznych cech języka słów kluczowych. Języki słów kluczowych dzięki indeksowaniu współrzędnemu mają najprostszą składnię, sprowadzającą się do konkatenacji słów kluczowych. Ten

⁸⁹ W tym typie gramatyki jedyną regułą gramatyczną jest reguła dołączania (konkatenacji), zaś jedyną relacją syntagmatyczną jest relacja współwystępowania, czyli kookurencji elementarnych jednostek leksykalnych w zdaniu w dowolnej kolejności. Interpretacja znaczeniowa każdej takiej jednostki jest niezależna od kontekstu. Każda tego typu jednostka leksykalna stanowi samodzielny klucz wyszukiwawczy. Istnieje też interpretacja, że kontekst poszczególnym słowom kluczowym w tekście nadają pozostałe współwystępujące w nim słowa kluczowe. Ma to miejsce tylko w wyszukiwaniu manualnym, i to jedynie na podstawie wiedzy wyszukiwającego o rzeczywistości. Gramatykę tę umownie nazywa się gramatyką zerową, co nie oznacza, że oprócz reguły dołączania nie ma innych reguł, na przykład istnieje reguła, że słowo kluczowe może w wyrażeniu złożonym (zdaniu) występować tylko jeden raz, co uważa się za domyślne.

⁹⁰ Relacja syntagmatyczna to relacja między jednostkami leksykalnymi prostymi w strukturze wyrażenia złożonego. W językach informacyjno-wyszukiwawczych wyróżnia się syntagmatyczne relacje współwystępowania (kookurencji) oraz syntagmatyczne relacje zależności semantycznej, specyfikowanej lub niespecyfikowanej.

rodzaj gramatyki jest właściwy dla wszystkich języków informacyjno-wyszukiwawczych opartych na indeksowaniu współrzędnym.

W języku słów kluczowych nie tworzy się wyrażeń złożonych, chyba że przez wyrażenia złożone rozumie się zdania, a więc sekwencje słów kluczowych derywowanych z dokumentu. Wyrażenia złożone w sensie wielowyrazowych słów kluczowych nie są wyrażeniami złożonymi języka słów kluczowych. Słowami kluczowymi mogą być frazy nominalne języka naturalnego. Inne struktury języka naturalnego nie występują w języku słów kluczowych.

Gramatyka języków słów kluczowych jest prezentowana niekonsekwentnie. Najczęściej podawane są tylko niektóre reguły składniowe. Przeważnie zakłada się, że użytkownik języka sam je wyabstrahuje z podanych przykładów wyrażeń złożonych albo zachowa się zdroworozsądkowo, na przykład nie będzie powtarzał tych samych słów kluczowych w wyrażeniu złożonym, choć nigdzie nie podaje się *explicite*, że takie powtórzenia są niedopuszczalne.

2.4. Inżynieria języka naturalnego na potrzeby języka słów kluczowych

2.4.1. Język słów kluczowych a język naturalny

Nawiązując do swoich wcześniejszych publikacji (Babik 1996b, 2007) i rozważań w rozdziale 1.4.3.3, gdzie stwierdziłem, że języki słów kluczowych stanowią przykład wykorzystania słownictwa języków naturalnych do wyszukiwania informacji⁹¹, uważam, że nie oznacza to pełnego wykorzystania języków naturalnych jako języków informacyjno-wyszukiwawczych. Ta na pozór błaha różnica między tymi dwoma sformułowaniami, wynikająca z „przeznaczeniowej” definicji języka informacyjno-wyszukiwawczego, uwzględniającej ich funkcje w systemie informacyjno-wyszukiwawczym, do tej pory często jest przyczyną nieporozumień oraz niepowodzeń w procesie organizacji słownictwa tych języków. Z definicji „przeznaczeniowej” wynika, że w sensie funkcjonalnym język ten służy do przedstawiania istotnych elementów treści dokumentu w celu umożliwienia ich powtórnej identyfikacji w zbiorze informacyjnym w wyszukiwaniu informacji tylko tych opisów dokumentów, które odpowiadają na pytanie użytkownika, tzn. których charakterystyka wyszukiwawcza pokrywa się z instrukcją wyszukiwawczą.

W celu automatyzacji pewnych reguł użycia języka, musimy przyjąć, że jego użycie polega na operowaniu kształtami wyrażeń, stąd niezbędna jest ich formaliza-

⁹¹ W piśmiennictwie dotyczącym języków informacyjno-wyszukiwawczych zwykle opisuje się słownictwo tych języków tylko z perspektywy synchronicznej, jak gdyby nie dostrzegając perspektywy diachronicznej. Trzeba mieć jednak świadomość, że znaczenia wyrazów uznawanych za słowa kluczowe i same słowa kluczowe zmieniają się zarówno w czasie, jak i w przestrzeni. Dotyczy to w równej mierze ich zakresów, jak i treści.

cja, ponieważ w systemie informacyjno-wyszukiwawczym operacje są wykonywane na tekstach języka naturalnego, w którym tworzone są dokumenty oraz używanego w systemie języka informacyjno-wyszukiwawczego. Automatyzacja dotyczy stosowania reguł słownikowych, formowania i jednoznacznego przekształcania wyrażeń bez konieczności rozumienia treści samych wyrażeń. Może ona dokonywać się tylko wtedy, gdy reguły użycia języka są operacjami na kształtach wyrażeń. Dzięki temu takie operacje nie stwarzają trudności interpretacyjnych. W zautomatyzowanym systemie informacyjno-wyszukiwawczym następuje więc rejestracja kształtu wyrażenia, a nie utrwalanie oznaczeń i porcji wiedzy. Komputer nie rozumie wyrażeń. Jest zdolny tylko rozróżniać ich kształty (ciągi znaków) i operować nimi. Logicy uważają, że języki naturalne, a zwłaszcza język potoczny, nie nadają się do formalizacji, gdyż nie można ściśle opisać ich składni bez odwoływania się do treści, co oznacza, że w tym przypadku nie jest możliwa ścisła odpowiedniość między operacjami na kształtach wyrażeń a operacjami na znaczeniach.

Autorzy języków słów kluczowych nakładają na język naturalny szereg ograniczeń w celu niezbędnej formalizacji tego języka. Tworzywo, którym jest słownictwo języka naturalnego, daje w języku słów kluczowych wiele możliwości wyrażania treści dokumentów (Bojar 2006). Tak też można postrzegać fakt funkcjonowania w Polsce i na świecie wielu języków słów kluczowych. W obrębie jednego języka słów kluczowych zasady doboru leksyki oraz formułowania słów kluczowych winny stanowić spójny system, natomiast nie musi zachodzić pełna, prosta i łatwa przekładalność między słowami kluczowymi różnych języków słów kluczowych, jeżeli takiej kompatybilności z góry nie założono⁹². Najistotniejsza różnica między językiem naturalnym a językiem słów kluczowych tkwi w regułach semantycznych, tj. w semantyce tych języków, a więc w ich planie treści. Zgodnie z regułami semantycznymi dowolnego języka naturalnego słowa oznaczają przypisane im wycinki rzeczywistości. Słowa kluczowe zaś oznaczają zawsze tylko dokumenty omawiające dany wycinek rzeczywistości. Różnice występują również w planie wyrażania.

Języki słów kluczowych w semantycznej strukturze nie odwzorowują bezpośrednio rzeczywistości pozajęzykowej, opisanej w tekstach dokumentów, jak to jest w językach naturalnych, lecz czynią to pośrednio w swoich tekstach (charakterystykach i instrukcjach wyszukiwawczych). Tym samym języki te odwzorowując teksty języków naturalnych, odwzorowują określony podzbiór ich leksyki relewantnej dla użytkowników informacji zawartej w tych tekstach. „Tak jak każdy język naturalny nakłada inną siatkę semantyczną na rzeczywistość pozatekstową, tak każdy język informacyjno-wyszukiwawczy nakłada na całe uniwersum dokumentacyjne swoją własną strukturę, zależną przede wszystkim od przewidywanych potrzeb informacyjnych użytkowników” (Bojar 2006, s. 7). Mimo wspomnianej redukcji pozostaje jednak podobieństwo języka słów kluczowych do języka naturalnego, dlatego słusznie te języki są zaliczane do klasy języków paranaturalnych.

⁹² Z faktu, że języki słów kluczowych bazują na językach etnicznych wynika dla tych języków wiele specyficznych problemów, których źródłem są poszczególne języki etniczne.

Język informacyjno-wyszukiwawczy opisuje nie bezpośrednio obiekty (elementy uniwersum), lecz dokumenty traktujące o tych obiektach, stąd przy projektowaniu słownika tego języka i wprowadzaniu zmian bierze się pod uwagę nie tylko samo istnienie danych terminów, ale także ich występowanie w dokumentach (ang. *literary warrant*). Terminy mogą istnieć w języku naukowym, technicznym, potocznym, ale mogą nie być używane w dokumentach, gdyż są nazwami obiektów, o których nie pisze się dostatecznie często.

Języki słów kluczowych bazują na języku naturalnym w jego różnych odmianach etnicznych, ale też na językach sztucznych, jak na przykład języku chemii czy matematyki, co stwarza dodatkowe problemy generowane przez języki bazowe, m.in. dotyczące odpowiedniej dla danego języka naturalnego (na przykład języka angielskiego) kategoryzacji rzeczywistości. Każdy z języków naturalnych przeprowadza odmienną kategoryzację rzeczywistości.

Dane wyrażenie staje się jednostką leksykalną języka słów kluczowych, zwaną słowem kluczowym, wtedy, gdy jest wyjęte z tekstu dokumentu i zaczyna reprezentować cały tekst, a nie tylko jakiś element rzeczywistości w tekście. Koncepcja izolowania z kontekstu słów kluczowych pochodzi od H.P. Luhna, o czym już wspominałem w rozdziale 2.1.1.

W języku słów kluczowych unika się używania jako słów kluczowych na oznaczenie elementów treści dokumentów wyrażen typu „badanie”, „zastosowanie” czy „nauczanie”. Są to wyrażenia, które w innych typach języków paranaturalnych pełnią funkcje wyrażen aspektowych (pomocniczych), jak na przykład w Języku Haseł Przedmiotowych Biblioteki Narodowej. Próby traktowania tego typu wyrażen jako słowa kluczowe zwykle generują niepotrzebny szum informacyjny, na przykład interpretacja zdania w języku słów kluczowych „Finanse. Badania” ujawnia niejednoznaczność tej konstrukcji, gdyż nie wiadomo, czy chodzi o badania nad finansami, czy o wpływ finansów na badania.

Języki słów kluczowych operują słowami kluczowymi o różnym stopniu szczególności. Ma to swoje konsekwencje w postaci mnożenia jednostek leksykalnych w zdaniach tego języka, na przykład ciąg słów kluczowych stanowiący tekst w języku słów kluczowych opisujący dokument x , można zastąpić istniejącym w tym języku słowem kluczowym y . Szczegółowość słów kluczowych w języku słów kluczowych powoduje zwiększenie ich liczby w zdaniu/tekście. Chociaż w niektórych nieklasycznych językach słów kluczowych wprowadza się w roli słów kluczowych wyrażenia opisowe, na przykład: „Białka w żywieniu człowieka”, to w klasycznym języku słów kluczowych słowami kluczowymi mogą być tylko wyrażenia proste typu: „Białko”, „Białko spożywcze”.

Granica między językiem słów kluczowych a językiem naturalnym ujawnia się w momencie tworzenia metainformacyjnych reprezentacji dokumentów, czyli w procesie indeksowania. Jest to ciąg następujących transformacji informacyjnych: analiza, selekcja i tłumaczenie. Analizowane są dokumenty, a przede wszystkim zawarte w nich informacje. Rezultatem tego procesu jest metainformacja, która stanowi reprezentację informacji zawartej w tekście dokumentu, wyrażona w języku naturalnym, choć nie zawsze w tym samym języku co dokument oraz informacja o samym

dokumencie. W ostatnim kroku indeksowania, którym jest przekład na język danego systemu, wyselekcjonowane metainformacje są wyrażane za pomocą środków leksykalnych i gramatycznych, którymi dysponuje ten język. W naszym przypadku mają postać słów kluczowych.

Zbiór wyrażeń języka naturalnego, który stanowi tworzywo dla języka słów kluczowych, jest dla potrzeb tych języków poddawany odpowiednim zabiegom językowym, co powoduje, że języki słów kluczowych istnieją w wielu odmianach. Bazę wyjściową dla wszystkich tych języków stanowi język naturalny w odpowiedniej wersji etnicznej, na przykład język polski w jego współczesnej postaci. Języki słów kluczowych stanowią przykład obróbki tworzywa, jakim jest język naturalny i jego słownictwo, na potrzeby języka systemów wyszukiwania informacji o dokumentach.

2.4.2. Koncepcja wielofunkcyjnych słowników otwartych

Jak już sygnalizowałem w pierwszej części rozprawy, w języku słów kluczowych w znacznym stopniu – chociaż nie aż w takim, jak w języku deskryptorowym – występuje zbliżenie systemu leksykalnego tego języka do systemu terminologicznego. Następuje transformacja tego systemu w system/sieć leksykalną języka słów kluczowych.

W przypadku systemów specjalistycznych ważne okazuje się słownictwo języka naturalnego w jego odmianie specjalistycznej, w formie terminologii danej dziedziny wiedzy, nauki lub techniki. Jest faktem, że specjaliści z danej dziedziny, formułując pytania kierowane do systemu, posługują się przede wszystkim terminologią funkcjonującą w ich dziedzinie. Praktyka ta – zresztą zauważona już wcześniej – znalazła wyraz m.in. w nazwie elementarnych jednostek leksykalnych, na których oznaczenie używa się również nazw „termin”, „termin wyszukiwawczy”.

Skoro użytkownicy specjalistycznych systemów informacyjno-wyszukiwawczych posługują się przy formułowaniu pytań informacyjnych, bardziej lub mniej świadomie, terminologią danej dziedziny zawartą w tekstach dokumentów oraz słownikach terminologicznych stanowiących odwzorowanie systemów terminologicznych, powstaje pytanie, czy nie można wykorzystać słowników terminologicznych w wyszukiwaniu informacji lub budować słowników terminologicznych, które pełniłyby jednocześnie funkcje słowników języka informacyjno-wyszukiwawczego. Takie wielofunkcyjne narzędzia wyeliminowałyby dodatkowe sięganie podczas wyszukiwania informacji do tradycyjnych słowników terminologicznych.

Ogólną koncepcję budowy takich słowników przedstawiłem w jednej z moich publikacji (Babik 1995). Opiera się ona na wzbogaceniu zakresu wykorzystania tradycyjnych słowników terminologicznych języka naturalnego o funkcje pełnione przez słowniki języków informacyjno-wyszukiwawczych. Jako element języka i składnik systemu takie słowniki pełniłyby następujące funkcje:

- informacyjną (normatywną, wyjaśniającą), związaną ze wspomaganiem procesu tworzenia tekstów dokumentów źródłowych w języku naturalnym;
- przekładową, związaną z wyrażaniem potrzeb informacyjnych w odpowiednim specjalistycznym języku;

- metainformacyjną, związaną z tworzeniem opisów dokumentów dla systemów informacji (CHWD);
- wyszukiwawczą, związaną z funkcjonowaniem systemu wyszukiwania informacji, tworzeniem instrukcji wyszukiwawczych (IWZ).

Budowa takich wielofunkcyjnych słowników wymagałaby rozwiązania szeregu problemów teoretycznych i praktycznych (metodycznych), związanych przede wszystkim z niezbędnymi dla rozszerzonego zestawu funkcji takich słowników środkami i metodami definiowania terminów.

Wynikająca z różnych potrzeb użytkowników koncepcja pełnienia przez jeden słownik funkcji informacyjnej słowników terminologicznych języka naturalnego oraz funkcji metainformacyjnej i funkcji wyszukiwawczej języka słów kluczowych wymaga odpowiednich środków do niezbędnej formalizacji definiowania terminów, co w konsekwencji prowadzi do postawienia pytania o przedmiot i formę definicji słownikowej, wystarczającej do realizacji funkcji tych słowników.

Zaletą wielofunkcyjnych słowników terminologicznych z punktu widzenia potrzeb języka słów kluczowych byłoby zdjęcie z użytkownika obowiązku szczegółowej znajomości systemu, jego zawartości informacyjnej i organizacji. Byłoby to zbieżne z najnowszymi tendencjami w metodologii projektowania systemów informacyjno-wyszukiwawczych oraz z możliwością oddzielnego projektowania poszczególnych modułów systemów informacji. Słowniki te odwzorowywałyby systemy terminologiczne poszczególnych dziedzin, a więc podawałyby gotowe terminy odpowiednio powiązane z pozostałymi, stanowiąc spójny system terminologiczny. Stanowiłyby jednocześnie system leksykalny języka słów kluczowych.

Ten typ wielofunkcyjnych słowników stanowi nową, dość precyzyjnie określoną formę organizacji i prezentacji słownictwa języka słów kluczowych, a więc terminologii w jej funkcji metainformacyjnej (reprezentacji treści dokumentów), informacyjnej i wyszukiwawczej, a jednocześnie pomoc w indeksowaniu i wyszukiwaniu informacji.

Wyspecjalizowane funkcje słowników języka słów kluczowych (metainformacyjna i wyszukiwawcza) wymuszają wypracowanie bardziej sformalizowanego, niż w tradycyjnych słownikach terminologicznych, aparatu definicyjnego jednostek leksykalnych (terminów) tych języków, dostosowanego także do komputerowego przetwarzania informacji. Formalizacja i stosowane środki normalizacji słów kluczowych nie ograniczają przydatności tych słowników do pełnienia funkcji informacyjnej, właściwej tradycyjnym słownikom terminologicznym. Słownik terminologiczny mógłby pełnić funkcje słownika słów kluczowych przy założeniu, że zebrane w nim słownictwo stanowić będzie klasę otwartą, to znaczy, że może być ono ciągle uzupełniane o nowe terminy.

Pomysł budowy wielofunkcyjnych słowników terminologicznych, to jest słowników pełniących jednocześnie funkcje: informacyjną, metainformacyjną i wyszukiwawczą, wymaga oparcia aparatu definicyjnego tych słowników na konotacyjnej interpretacji znaczenia i wykorzystania najnowszych technologii leksykograficznych.

2.5. Zróżnicowanie języka słów kluczowych

2.5.1. Kryteria podziału i odmiany języka słów kluczowych

Problem ten sygnalizowałem już w rozdziale 1.3. Tutaj omówię go szczegółowo. W dawnej literaturze przedmiotu nie zawsze odróżniano dwie odmiany języka słów kluczowych: język słów kluczowych jako język naturalny i język słów kluczowych jako język sztuczny. Przyczyną tego było utożsamianie ich planów treści, w których to planach tkwią istotne różnice między nimi.

Słowa kluczowe jako elementarne jednostki leksykalne języków słów kluczowych są równokształtne z wyrażeniami języka naturalnego, który stanowi bazę do generowania słów kluczowych. Znaczenia słów kluczowych nie są podawane *explicit*. Zakłada się, że użytkownik tego języka sam domyśli się, na podstawie swojej kompetencji językowej w zakresie odpowiedniego języka naturalnego, oraz odkryje zasadę, iż słowa kluczowe odwzorowują elementy treści dokumentów, a nie elementy rzeczywistości pozatekstowej. Znaczenie słów kluczowych odpowiada więc równokształtnym z nimi wyrażeniom języka naturalnego użytym w funkcji metainformacyjnej. Można powiedzieć, że w wyszukiwaniu informacji wyrażenia języka naturalnego jako słowa kluczowe są używane w supozycji materialnej (a więc w sytuacji, w której chodzi wyłącznie o formę, a nie treść), a nie w supozycji normalnej, zwanej też zwykłą, gdzie chodzi wyłącznie o treść, a nie formę. W tym drugim przypadku używa się tych wyrażen w indeksowaniu.

W tekście dokumentu nie wszystkie wyrażenia w języku naturalnym lub języku specjalistycznym (terminologia) są jednakowo ważne. Tylko te najważniejsze (przeważnie nie na najwyższym poziomie abstrakcji) to słowa kluczowe, co oznacza, że są przydatne do opisu i wyszukiwania danego dokumentu.

W działalności informacyjnej realizowane są różne rodzaje języków słów kluczowych. Możemy wyróżnić systemy otwarte oraz autorskie języki słów kluczowych, tj. języki, które mają autorskie systemy słownikowe (leksykalne). W przypadku systemów otwartych (jak w języku swobodnych słów kluczowych), każdy wybiera dowolne słowa kluczowe, przez co nie jest zapewniona taka spójność pojęciowa, jak w przypadku języków słów kluczowych opartych na słowniku jawnym, karcie wzorcowej lub na słowniku wzorcowych słów kluczowych. Systemy autorskie to systemy opracowane przez jednego autora lub grupę autorów współpracujących z sobą. Charakteryzują się tym, że tworzy się je z zachowaniem spójności pojęciowej (terminologicznej) danej dziedziny.

Języki słów kluczowych już wcześniej podzieliłem na klasyczne języki słów kluczowych i nieklasyczne języki słów kluczowych. Definicję klasycznego języka słów kluczowych podała B. Bojar (por. rozdz. 2.2). Nie określiła jednak możliwych odstępstw od rozwiązań modelowych. Używanie tu przymiotnika „klasyczny” uważam za uzasadnione. W swej istocie dotyczy on przede wszystkim różnic w naturze i paradygmatykach tych języków. Klasycznym językiem słów kluczowych jest język swobodnych słów kluczowych. Nie posiada on systemu odsyłaczy. Bazuje na słow-

niku mentalnym człowieka, tj. słowniku niejawnym. Wspomaganie się systemem odsyłaczy (poza odsyłaczem „zob.”) wprowadza w system leksykalny system preferencji, co sprawia, że język ten traci atrybut klasycznego, stając się nieklasycznym językiem słów kluczowych.

Stopień kontroli słownictwa stanowi kryterium wyróżnienia następujących odmian języka słów kluczowych:

- język słów kluczowych o słownictwie niekontrolowanym słownikiem (języki swobodnych słów kluczowych)⁹³;
- język słów kluczowych o słownictwie częściowo kontrolowanym słownikiem, na przykład słowniki słów kluczowych wydawane przez Pracownię OIN PAN w Krakowie w swej istocie miały pełnić funkcje wykazów wzorcowych słów kluczowych;
- język słów kluczowych o słownictwie kontrolowanym słownikiem (słownik-wykaz słów kluczowych, kartoteka wzorcowa słów kluczowych jako leksykograficzna baza danych).

E. Ścibor wyróżnił w ramach typologii strukturalnej języków informacyjno-wyszukiwawczych następujące rodzaje języka słów kluczowych (Ścibor 1982):

- najprostszy język słów kluczowych (bez usunięcia synonimii, bez odsyłaczy i bez hierarchii);
- język słów kluczowych bez hierarchii, z odsyłaczami „zob.” lub najprostszy język deskryptorowy, o słowniku w postaci teaurusu alfabetycznego prostego (odsyłacze „U” – „używaj”);
- język słów kluczowych bez hierarchii, z odsyłaczami „zob. też”.

Przytoczone tu rozróżnienie wymaga komentarza. Jako kryterium Ścibor przyjął charakter paradygmatyki języka. Kryterium to okazało się jednak mało dyskretne, gdyż nie pozwala na wytyczenie klarownej granicy między językami słów kluczowych a językami deskryptorowymi.

2.5.2. Zalety i wady języka słów kluczowych

U podstaw popularności tych języków – moim zdaniem – leżą ich następujące cechy:

- słownictwo naturalne lub paranaturalne, tzn. słownictwo języka naturalnego w funkcji metainformacyjnej lub równokształtne z wyrażeniami języka naturalnego, ułatwiające posługiwanie się tym językiem;
- możliwość swobodnego indeksowania współrzędnego;
- elastyczność pod względem absorpcji nowych tematów i obszarów wiedzy wchodzących w zakres danego systemu, tj. łatwość wprowadzania nowych terminów, co ułatwia szybką asymilację i efektywne odwzorowywanie nowych zakresów badanych, nowych dziedzin i zagadnień – w przypadku środowiska internetowego,

⁹³ Indeksowanie swobodne – indeksowanie w języku informacyjno-wyszukiwawczym, którego zasób leksykalny jest klasą otwartą i może być zawsze powiększony o każdy nowy element posiadający cechę (zbiór cech) określoną w tzw. instrukcji indeksowania swobodnego. Termin „indeksowanie swobodne” jest używany w opozycji do terminu „indeksowanie słownikowe (normatywne)”.

które szybko i żywiołowo reaguje na nowe kierunki w badaniach naukowych, jest to bardzo ważne;

- ahierarchiczność słów kluczowych, co daje możliwość wskazywania przynależności danego dokumentu/informacji do wielu klas, a jednocześnie powoduje niekiedy konieczność wielokrotnego przeszukiwania zbiorów dla wyszukania kompletnej informacji;
- otwartość.

Ten rodzaj opracowania rzeczowego zakłada maksymalne zbliżenie do języka naturalnego i związaną z tym giętkość języka opisu, oznaczającą m.in. stałą aktualizację zasobu słów kluczowych stosownie do zmian obserwowanych w terminologii zawartej w opisywanych dokumentach. Stąd język ten doskonale nadaje się do wykorzystania jako uniwersalny język informacyjno-wyszukiwawczy dla sieci ośrodków informacji o zróżnicowanej tematyce opracowywanych zbiorów, na przykład w systemie SYNABA.

Za wady języków słów kluczowych uważam:

- duże ryzyko subiektywizmu, związane z intuicyjnością wyboru terminów, które może prowadzić do braku spójności;
- uzależnienie od języków etnicznych, co uniemożliwia powszechność ich stosowania;
- duża, a w zasadzie permanentna w przypadku języków słów kluczowych o słownictwie niekontrolowanym, niestabilność systemu językowego;
- brak możliwości semantycznej strukturalizacji informacji;
- powstawanie szumu informacyjnego na skutek fałszywych skojarzeń słów kluczowych;
- niekontrolowane słownictwo, z którego nie usunięto wieloznaczności wynikającej z synonimii i homonimii.

Niewątpliwą wadą języków słów kluczowych jest więc większe niż w innych językach prawdopodobieństwo wystąpienia fałszywych koordynacji jednostek leksykalnych będących przyczyną szumu informacyjnego, tj. wydawania użytkownikom opisów dokumentów nierelevantnych w stosunku do pytania informacyjnego. Na przykład połączenie jednostek leksykalnych „Energia” i „Produkcja” może powodować wydanie informacji o dokumentach zarówno dotyczących produkcji energii, jak i energii dla celów produkcji. Ograniczenie fałszywych koordynacji w językach słów kluczowych uzyskuje się przez wielowyzrazowe słowa kluczowe. Języki słów kluczowych pozwalają na dużą swobodę wyboru terminów oraz zapewniają możliwość wielowariantowego i wieloaspektowego wyrażania treści dokumentów i wpływających do systemu pytań informacyjnych.

Niezależnie od poczynionych tu krytycznych uwag języki słów kluczowych uchodzą za najbardziej zbliżone do języka naturalnego spośród paranaturalnych języków informacyjno-wyszukiwawczych lub – jak w przypadku swobodnych słów kluczowych – są językiem naturalnym. Nie mają one trudności w wyrażaniu terminów najnowszych.

2.6. Klasyczny język słów kluczowych a inne języki paranaturalne

Spróbuję teraz określić miejsce języka słów kluczowych wśród innych języków zaliczanych do klasy paranaturalnych języków informacyjno-wyszukiwawczych. Nie jest to zadanie łatwe, gdyż dla poszczególnych typów języków informacyjno-wyszukiwawczych historycznie ukształtowały się odmienne terminologie. Stąd te same obiekty lingwistyczne są różnie nazywane w różnych językach lub ich typach czy odmianach, często w zależności od różnych tradycji ich opisu. Utrudnia to ocenę poprawności poszczególnych typów języków, ocenę ich możliwości, a także ich porównywanie. Tę różnorodność, powodującą czasami chaos terminologiczny, próbuje się rozwiązać, adaptując wybrane terminy językoznawcze, ale propozycje te z różnych względów, a często z powodu ich nieznamomości, nie zawsze są stosowane. Te same funkcje jednostek leksykalnych, odmienne nazywane w różnych typach języków informacyjno-wyszukiwawczych, powodują, że ich nazwy są stosowane często zamiennie, jako synonimy. Kolejne podrozdziały stanowią kontynuację problematyki zasygnalizowanej w rozdziale 1.2.

2.6.1. Język słów kluczowych a język deskryptorowy

To połączenie dwóch typów języków informacyjno-wyszukiwawczych jest symptomatyczne i odzwierciedla wspólny rodowód tych języków, które jednak różnią się między sobą pod względem struktury i sposobu użytkowania. Języki deskryptorowe ze słownikami w formie tezaurusów oraz języki słów kluczowych oparte na słowniku jawnym znajdują szerokie zastosowanie przede wszystkim w związku z automatyzacją procesów informacyjnych. Za moment narodzin tego kierunku w naszym kraju można uznać rok 1965, w którym odbyła się w Berlinie konferencja poświęcona budowie tezaurusów, a następnie symposium na temat mechanizacji i automatyzacji wyszukiwania informacji w resorcie przemysłu ciężkiego, zorganizowane w 1966 roku w Szklarskiej Porębie. Jednym z postulatów wysuniętych na tym symposium było tworzenie „słowników pojęć”, które wkrótce zaczęto nazywać „tezaurusami”. Kontakt uczestników tych konferencji okazał się na tyle twórczy, że już w 1969 roku opublikowano pierwszy polski tezaurus dokumentacyjny, dotyczący zagadnień transportu bliskiego. Natomiast słowniki słów kluczowych zaczęto wydawać dopiero pod koniec lat osiemdziesiątych XX wieku na potrzeby konkretnych systemów informacyjno-wyszukiwawczych. W Polsce największe zasługi ma tu Ośrodek Informacji Naukowej Polskiej Akademii Nauk (OIN PAN), a zwłaszcza zamiejscowa Pracownia Informacji Naukowej tego Ośrodka w Krakowie⁹⁴. Słowniki te stanowią cenne źródło

⁹⁴ W drugiej połowie lat osiemdziesiątych XX wieku OIN PAN rozpoczął opracowywanie i publikowanie słowników słów kluczowych z różnych dziedzin, głównie z zakresu nauk społecznych. W latach 1988–1993 ukazały się słowniki dotyczące następujących dziedzin: ekonomia polityczna, filozofia,

słownictwa i dobry punkt wyjścia do opracowania tezaurusów dla dziedzin, których dotyczą. Wcześniej słowniki słów kluczowych były stosowane, ale nie były publikowane. Przedtem języki oparte na indeksowaniu współrzędnym, tj. języki deskryptorowe i języki słów kluczowych, nie były stosowane w bibliotekarstwie, przynajmniej w bibliotekarstwie polskim, gdyż nie były przydatne w katalogu kartkowym lub drukowanym do szeregowania opisów według kryteriów semantycznych.

W sposobie korzystania ze słowników porównywanym do języków istnieją poważne różnice. Posługując się tezaurem, można w indeksowaniu używać tylko deskryptorów znajdujących się w tezaurucie jako ich normatywnym wykazie. Stosując język słów kluczowych, używamy słów kluczowych ze słownika słów kluczowych lub tworzymy je w sposób niekontrolowany i swobodny. Gramatyka języka słów kluczowych jest identyczna z gramatyką większości języków deskryptorowych.

Nowym zjawiskiem zauważalnym w najnowszych tezaurusach jest niewykluczenie możliwości operowania synonimami w wyszukiwaniu informacji. Zamiast wskaźników U/UF (U/NU) stosowany jest wskaźnik SY (synonim). Oznacza to tendencję do rezygnacji z różnicowania statusu jednostek leksykalnych na deskryptory i askryptory, a więc na wyrażenia języka deskryptorowego i wyrażenia języka naturalnego. Zjawisko to można uznać za symptomatyczne także dla innych języków paraturalnych. Oprócz tego w wielu nowoczesnych tezaurusach wprowadza się szczegółową specyfikację powiązań kojarzeniowych. Przykładem takiego tezaurusa jest tezaurus Unified Medical Language System (UMLS); zdefiniowano tam ponad czterdzieści rodzajów relacji, z których większość dotyczy skojarzeń zachodzących między pojęciami współwystępującymi w różnego typu kontekstach. Tego typu rozwiązania zbliżają tezaury do słowników języka naturalnego. Wspomniane tu zjawisko można odnieść również do języków słów kluczowych. Sprawia to, że problem synonimii leksykalnej jak gdyby rozwiązuje się sam.

Zarówno język słów kluczowych, jak i język deskryptorowy bazują na języku naturalnym. Leksyka jest prezentowana w układzie alfabetycznym. W najprostszej formie słownik słów kluczowych stanowi alfabetycznie uporządkowaną listę jednostek leksykalnych bez specyfikacji związków pojęciowych pomiędzy tymi jednostkami. Jednak często takie listy zawierają specyfikację relacji międzypojęciowych, co stwarza trudność w ich odróżnieniu, zwłaszcza od tezaurusów. Przykładem może być Medical Subject Headings (MeSH), który wprawdzie stanowi listę słów kluczowych, ale w piśmiennictwie jest określany jako tezaurus.

językoznawstwo, marksizm-leninizm, nauka o polityce, pedagogika, psychologia, socjologia, botanika, historia Polski, literaturoznawstwo, religioznawstwo, a także została opracowana w języku angielskim neuropsychofarmakologia. Oprócz tego w ośrodku poznańskim został przygotowany *Słownik słów kluczowych z metodologii nauk społecznych*. Należy podkreślić wyjątkowy wkład organizacyjny w tym zakresie Aliny Fitowej, ówczesnej kierowniczki Pracowni w Krakowie. Koncepcja opracowywania i publikowania słowników słów kluczowych powstała w 1985 roku na ruinach koncepcji budowy tezaurusa nadrzędnego dla nauk społecznych. Wobec niemożliwości wykonania tego zadania nowy zespół Pracowni zdecydował się na języki słów kluczowych. Nie wykorzystano jednak zebranej do tej pory leksyki na potrzeby budowy tezaurusów. W późniejszym czasie, a zwłaszcza w momencie likwidacji Ośrodka w 1994 roku, te cenne zbiory uległy zniszczeniu.

Granice między językiem słów kluczowych a językiem deskryptorowym są płynne. Istnienie wykazu słów kluczowych wyprzedza zwykle powstanie tezauryśa wówczas, gdy ten ostatni budowany jest metodą indukcyjną, tj. na podstawie tekstów dokumentów znajdujących się rzeczywiście w danym zbiorze informacyjnym. Przeważnie elementem słownictwa takich list towarzyszą synonimy i terminy pokrewne. Forma ta jednak stopniowo zanika. Za moment, w którym język słów kluczowych przekształca się (zmienia swoją jakość) w język deskryptorowy, uważa się wprowadzenie do wykazu słów kluczowych wyrażeń języka naturalnego nieprzyjętych do charakteryzowania dokumentów w danym języku i połączonych odsyłaczami „U” (używaj) lub „zob.” z jednostkami leksykalnymi stosowanymi do charakteryzowania dokumentów oraz wyszukiwania informacji. Takie elementarne jednostki leksykalne otrzymują status deskryptorów, tj. jednostek języka deskryptorowego, zaś wyrażenia języka naturalnego nieprzyjęte w danym języku deskryptorowym nazywa się askryptorami. Wykaz słów kluczowych, z chwilą wprowadzenia do niego terminów nieprzyjętych w danym języku (tj. z innego języka), przekształca się w tezaurus alfabetyczny prosty (Bielicka, Ścibor 1982).

Obecnie słowa kluczowe występują prawie wyłącznie jako wyrażenia samoistne. Cechą charakterystyczną (i jednocześnie nietypową dla tezauryśa) są łańcuchowe związki nazw i odsyłacze. Dzięki temu jest możliwa realizacja w szerokim zakresie zasady prekoordynacji. Właśnie w tym – odmiennie niż E. Ścibor – Jutta Bertram z Instytutu Informacji i Dokumentacji Uniwersytetu w Poczdamie (Niemcy), a obecnie z Fachhochschule w Burgerland (Austria), upatruje istotną różnicę między listą słów kluczowych a tezaurem (Bertram 2005, s. 131).

W przypadku porównywanych tu dwóch typów języków informacyjno-wyszukiwawczych bardzo często myli się stosowane do ich opisu odmienne terminologie, na przykład Grażyna Pierzchalska we wstępie do *Słownika naukowego słów kluczowych transportu morskiego* (Pierzchalska 1986) utożsamia słowo kluczowe z deskryptorem. Interesujący jest fakt, że Józef Robowski w swojej pracy poświęconej analizie języków deskryptorowych (Robowski 1974), porównując te języki z innymi językami informacyjnymi opartymi na indeksowaniu współrzędnym, nie zauważył istnienia słów kluczowych, prawdopodobnie utożsamiając je z unitermami lub deskryptorami.

Uważam, że trafego porównania elementarnych jednostek leksykalnych tych języków, to jest deskryptorów ze słowami kluczowymi dokonał Josef Koblitz (Koblitz 1968). Istotne różnice pomiędzy tymi jednostkami leksykalnymi przedstawiono na rysunku 5.

Łatwo zauważyć, że zarówno język słów kluczowych, jak i język deskryptorowy mają dwie wspólne cechy: korzystają ze słownictwa języków naturalnych oraz wykorzystują reguły indeksowania współrzędnego. Różnią się rodzajem uporządkowania i stopniem kontroli słownictwa. Języki deskryptorowe mają słownictwo w pełni kontrolowane i uporządkowane. Jednostki leksykalne języków słów kluczowych w zasadzie są swobodne, tj. kontrolowane tylko do pewnego stopnia.

	Kryterium	Deskryptor	Słowo kluczowe
1.	Sposób prezentacji	Pojedyncze słowa lub wyrażenia złożone	Tak samo
2.	Stopień normalizacji	Deskryptor powinien być terminem, tj. jednostką znormalizowaną w zakresie językowym i pojęciowym, słowa wieloznaczne otrzymują dookreślenie lub odsyłacz	Słowa wieloznaczne otrzymują dookreślenie
3.	Funkcje	Deskryptor pełni funkcje metainformacyjne, utożsamiane z funkcją semantyczną	Słowo kluczowe pełni funkcje metainformacyjne, utożsamiane z funkcją semantyczną.
4.	Wykorzystanie	Organizacja informacji	Tak samo
5.	Relacje generyczne i kojarzeniowe	Są uwzględniane	Nie są uwzględniane
6.	Słownik	Tezaurus	Lista słów kluczowych
7.	Normalizacja	Łatwa	Trudna

Rys. 5. Porównanie słowa kluczowego z deskryptorem (źródło: Koblitz 1968)

2.6.2. Język słów kluczowych a język haseł przedmiotowych

Języki słów kluczowych wykazują wiele podobieństw do języków haseł przedmiotowych. Również ich słownictwo składa się z paranaturalnych elementarnych jednostek leksykalnych, a w wykazach (słownikach) słów kluczowych przeważa uporządkowanie alfabetyczne. Języki te jednak różnią się od języków haseł przedmiotowych przede wszystkim statusem jednostek leksykalnych. Języki słów kluczowych mają tylko jednostki autosyntaktyczne⁹⁵, zaś języki haseł przedmiotowych mają jednostki leksykalne o zróżnicowanym statusie: tematy jako autosyntaktyczne jednostki leksykalne oraz określniki jako synsyntaktyczne jednostki leksykalne. Zróżnicowany status jednostek leksykalnych związany jest z gramatyką pozycyjną tych języków.

W miarę rozwoju języków słów kluczowych coraz bardziej widoczne są analogie między niektórymi elementami ich struktury. Na przykład istnieją podobieństwa pomiędzy relacjami w języku słów kluczowych a odsyłaczami w języku haseł przedmiotowych, między tematem a słowem kluczowym, między słownikiem języka haseł przedmiotowych a słownikiem słów kluczowych⁹⁶. Języki słów kluczowych mają strukturę

⁹⁵ Autosyntaktyczne jednostki leksykalne w opozycji do synsyntaktycznych jednostek leksykalnych to takie wyrażenia, które samodzielnie mogą tworzyć zdania.

⁹⁶ Jaskrawym przykładem mylenia tych dwóch języków informacyjno-wyszukiwawczych jest charakterystyka wyszukiwawcza zamieszczona w opisie artykułu Magdaleny Olszewskiej i Joanny Siudy pt.

prostszą od języków haseł przedmiotowych i języków deskryptorowych. Istotnymi cechami języków słów kluczowych jest współrzędność elementarnych jednostek języka – słów kluczowych w procesie tworzenia zdań i tekstów w tych językach, oraz koordynacja słów kluczowych dopiero w procesie wyszukiwania informacji, a więc odwrotnie niż w językach haseł przedmiotowych. Języki słów kluczowych różni od języków haseł przedmiotowych przede wszystkim inna „filozofia”⁹⁷ – języki haseł przedmiotowych odwzorowują tematy dokumentów, ich ważność w dokumencie i sposób ujęcia, a języki słów kluczowych nie mają tych funkcji. Języki słów kluczowych odwzorowują elementy treści dokumentów bez ich kategoryzacji i wartościowania.

W odróżnieniu od języka deskryptorowego, języki słów kluczowych z definicji nie prezentują leksyki w postaci pełnego (normatywnego) słownika w odróżnieniu od języków haseł przedmiotowych – nie stosują kwalifikacji jednostek leksykalnych na samodzielne i niesamodzielne, bo nie ma w nich takiego rozróżnienia – wszystkie są autosyntaktyczne. Na tle zasygnalizowanych tu odmienności języki słów kluczowych jawią się jako najmniej sformalizowane.

Celem tej części rozprawy było przedstawienie modelu klasycznego języka słów kluczowych, jego leksyki, gramatyki oraz metod przejmowania słownictwa z języka naturalnego. Starałem się wyodrębnić zbiór cech dystynktywnych tego języka, określających w sumie jego tożsamość. Studium piśmiennictwa dotyczącego języków słów kluczowych pozwoliło mi zauważyć, że w teorii języków informacyjno-wyszukiwawczych w wielu przypadkach nie ma jednoznacznych rozstrzygnięć, akceptowanych przez wszystkich specjalistów. Świadczy to o tym, że zarówno te języki, jak i ich teoria rozwijają się i stale poszukują nowych rozwiązań. Postęp w tej dziedzinie polega nie tyle na osiągnięciu niekwestionowanych rozstrzygnięć, ile na dostrzeganiu coraz to nowych uwarunkowań i powiązań, czyli na uświadamianiu sobie coraz bardziej bogactwa i złożoności badanych języków. Właśnie propozycje różnych rozwiązań tych problemów pomagają dokładniej zdać sobie sprawę z funkcji, jakie pełnią słowa kluczowe w procesach indeksowania i wyszukiwania informacji. Wyodrębniony zbiór tzw. cech konstytutywnych języka słów kluczowych, chociaż z natury rzeczy niekompletny, pozwala jednak na stworzenie spójnego obrazu tego języka, definiując w ten sposób tożsamość i „granice” klasycznego języka słów kluczowych.

Wykorzystanie jhp KABA w katalogu komputerowym Biblioteki Papieskiej Akademii Teologicznej w Krakowie (http://www.ebib.info publikacje/matkonf/kaba/Olszewska_Siuda.php), w którym zamiast deklarowanych słów kluczowych w języku polskim i języku angielskim faktycznie podano hasła przedmiotowe.

⁹⁷ J. Woźniak-Kasperek (1999) porównała to metaforycznie z rodzajem filtru pomiędzy światem ludzkiej myśli a potencjalną operacyjnością systemu informacyjno-wyszukiwawczego. W wyniku zastosowania tego filtru, którym jest dany język informacyjno-wyszukiwawczy, w naszym wypadku język słów kluczowych, otrzymujemy w większym lub mniejszym stopniu różniący się odmienny obraz rzeczywistości dokumentacyjnej. To właśnie za pośrednictwem języka dokonuje się odwzorowywanie rzeczywistości dokumentacyjnej w rzeczywistość metainformacyjną. Języki słów kluczowych muszą dysponować odpowiednimi środkami i mechanizmami umożliwiającymi realizację tego zadania.

Warto zauważyć, że w klasycznych językach słów kluczowych priorytetowo traktuje się zasób leksykalny, który służy przede wszystkim do identyfikacji elementów treści dokumentów/pytań informacyjnych, z jednoczesnym pomniejszeniem i/lub nieuwzględnieniem jego roli w odwzorowywaniu relacji semantycznych pomiędzy obiektami odwzorowywanej rzeczywistości dokumentacyjnej. Języki słów kluczowych w odmianie języków swobodnych słów kluczowych stanowią więc zredukowaną do potrzeb optymalizacji procesów informacyjnych odmianę języka naturalnego. Model struktury wiedzy stosowany w klasycznych językach słów kluczowych odwołuje się do systemu poznawczego człowieka. W języku tym w ogóle nie definiuje się znaczenia słów kluczowych.

Naturalnie w języku kontrolowanych słów kluczowych istnieje pewnego rodzaju umowność znaczeniowa słów kluczowych. Jest ona jednak względna, ale sprawia wrażenie ścisłego podobieństwa słów kluczowych w warstwie semantycznej do wyrażen języków naturalnych. Niewątpliwie słowa kluczowe jako jednostki leksykalne klasycznego języka słów kluczowych należy uznać za jednostki leksykalne języka naturalnego. Języki słów kluczowych mogą być też tworem sztucznym ze wszystkimi płynącymi z tego konsekwencjami. Niemniej nawet jako języki sztuczne mają jednostki leksykalne, które są równokształtne z wyrażeniami języka naturalnego. Specyfika tego języka polega więc na tym, że wprawdzie odwołuje się on – chociaż w różny sposób – do znajomości słów w języku naturalnym, lecz – jak żaden język naturalny – łączy je w teksty tylko na zasadzie konkatenacji. Stąd teksty w języku słów kluczowych odbiegają od wzorców tekstów w języku naturalnym.

Podsumowując rozważania nad tożsamością klasycznego języka słów kluczowych, przyjmuję, że „język słów kluczowych to język informacyjno-wyszukiwawczy o notacji paranaturalnej, bez wyraźnie określonej w słowniku paradygmatyki, w których jednostki słownika nazywane są słowami kluczowymi, a na gramatykę składają się reguły indeksowania współrzędnego” (Bojar 2002, oprac., s. 111). Najszerzej ujmuje się język słów kluczowych, określając go jako język informacyjny o słownictwie paranaturalnym i tzw. gramatyce zerowej, równoznacznej z indeksowaniem współrzędnym, polegającym na charakteryzowaniu złożonej treści dokumentu przez koordynację współrzędnych, dających się swobodnie zestawiać elementarnych jednostek leksykalnych. Słownikiem języka kontrolowanych słów kluczowych jest uporządkowany alfabetycznie wykaz słów kluczowych wybranych z tekstów dokumentów, który nigdy nie jest zamknięty i w zasadzie nie są w nim wyrażane relacje między słowami kluczowymi, które stanowią elementarne jednostki leksykalne języka słów kluczowych. Języki słów kluczowych często wykorzystują słowniki negatywne, tzw. stop-listy, tj. słowniki wskazujące wyrażenia języka naturalnego, których nie wolno używać przy indeksowaniu, podczas gdy pozostałe wyrazy występujące w indeksowanych tekstach są akceptowane jako słowa kluczowe.

Zapoczątkowane w latach pięćdziesiątych XX wieku wykorzystanie technologii komputerowej do gromadzenia i przetwarzania informacji tekstowej z jednej strony doprowadziło do modyfikacji tradycyjnych typów języków informacyjno-wyszukiwawczych (języków haseł przedmiotowych i klasyfikacji biblioteczno-bibliograficznych), z drugiej zaś strony do zbudowania nowych modeli tych języków, czego

przykładem są języki słów kluczowych oraz języki deskryptorowe. Do tej pory nurty te wzajemnie się przenikają, co uwidacznia się w ewolucji m.in. słowników tych języków. Rozwój języków słów kluczowych, głównie ich słowników, stosowanie elementów hierarchii, klasyfikacji i kategorii oraz fasetowość spowodowały, że trudno jest nadal opowiadać się za jednolitością tych języków.

W rozdziale tym zajmowałem się językiem słów kluczowych w ogóle, a nie konkretnymi językami słów kluczowych. O tym będzie mowa w następnej, ostatniej części rozprawy. Zajmę się w niej językami słów kluczowych konkretnych systemów informacyjno-wyszukiwawczych.

Część trzecia

Słowa kluczowe w systemach wyszukiwania informacji

W tej części pracy zamierzam opisać, jak wyglądają praktyczne realizacje (implementacje) języka słów kluczowych, w tym jakie są różnice między rzeczywiście istniejącymi językami słów kluczowych stosowanymi we współczesnych systemach informacyjno-wyszukiwawczych a omówionym w poprzedniej części modelem klasycznego języka słów kluczowych. Ma to na celu pokazanie, jak w omawianych językach słów kluczowych, zwłaszcza w ich słownikach, zdefiniowano ich tożsamość, w tym jak określano (i poszerzano) granice języka, czerpiąc wzory z języka naturalnego oraz z innych języków paranaturalnych. Uczyniłem też szereg odniesień do problemów pragmatycznych w dotychczasowych i nowych obszarach zastosowań języków słów kluczowych, jakimi są systemy WWW.

3.1. Analiza porównawcza języków słów kluczowych opartych na słowniku jawnym

Celem tego rozdziału jest prezentacja wyników analizy struktury wybranych słowników słów kluczowych jako jednego z elementów języka słów kluczowych. Miała ona na celu wyodrębnienie cech konstytutywnych języka słów kluczowych na bazie słownictwa tych języków. Przy założeniu otwartości tego typu narzędzi lingwistycznych, stosowanych w systemach wyszukiwania informacji, chodzi mi o pokazanie ich różnorodności oraz odmienności przyjętych rozwiązań w stosunku do modelu klasycznego języka słów kluczowych jako czynników determinujących ich strukturę i funkcje. Uważam, że pozwoli to ustalić elementy konstytuujące tożsamość i pokazać zróżnicowanie języków słów kluczowych.

3.1.1. Charakterystyka materiału badawczego

Materiał badawczy stanowiły słowniki słów kluczowych oraz ich opisy w literaturze przedmiotu, wstępy, a także recenzje. Obecnie brak ewidencji takich słowników. Wprawdzie od 1968 roku działał dom clearingowy, powstały pod auspicjami UNESCO przy Instytucie Informacji Naukowej, Technicznej i Ekonomicznej w Warszawie, który wydawał co roku „Bibliographic Bulletin of the Clearinghouse at IINTE”, zawierający opisy bibliograficzne tworzonych na świecie systemów klasyfikacji, tezaurusów, słowników terminologicznych, list haseł przedmiotowych oraz słowników słów kluczowych opracowanych w innych językach niż język angielski, a od 1977 roku – także opublikowanych w języku angielskim, ale w 1989 roku zaprzestał dalszej działalności. Biuletyn, o którym mowa, nie zawierał jednak informacji o wszystkich publikacjach z tego zakresu, gdyż nie istniał obowiązek informowania o nich. Pewne informacje o tej klasie języków zawiera opracowanie *Języki informacyjne opracowane, adaptowane i wykorzystywane w Polsce (stan na 31.12.1994 r.)*, stanowiące katalog języków informacyjnych, będący odpowiednikiem zawartości powstałej w 1993 roku w IINTE bazy danych dotyczącej języków informacyjnych opracowanych, adaptowanych i wykorzystywanych w Polsce. Dane do tej bazy zostały zebrane za pomocą ankiety opracowanej w IINTE, którą wysłano do 306 placówek informacji. W związku z tym, że odpowiedzi otrzymano jedynie od 166 instytucji, dane zawarte w bazie i we wspomnianym katalogu są niepełne, gdyż odnoszą się wyłącznie do języków stosowanych w placówkach, które odpowiedziały na ankietę. Katalog ten zawiera informacje tylko o pięćdziesięciu stosowanych w Polsce językach słów kluczowych. Opis każdego języka, a w naszym przypadku języka słów kluczowych, składa się z następujących elementów:

- nazwa języka;
- tytuł słownika;
- charakterystyka ilościowa zasobu leksykalnego. W tej rubryce podawano liczbę słów kluczowych wymienionych jawnie w słowniku. Oddzielnie podano jednostki leksykalne preferowane, czyli stosowane w indeksowaniu dokumentów i pytań informacyjnych, a oddzielnie jednostki niepreferowane, czyli wyrażenia synonimiczne lub quasi-synonimiczne, podane w słowniku, ale których stosowanie nie jest dozwolone;
- typ stosowanej gramatyki;
- zakres tematyczny języka;
- wzorzec.

Tej grupie danych towarzyszy zestaw informacji dotyczących systemu informacyjnego, w którym jest stosowany język słów kluczowych. Dane te okazały się jednak niekompletne i w wielu przypadkach są niewystarczające do przeprowadzenia dogłębnej analizy cech definiujących tożsamość (konstrytuwanych i indywidualnych o charakterze przypadłościowym) opisanych w tym informatorze języków słów kluczowych. Z punktu widzenia potrzeb moich badań jednym z największych mankamentów jest brak informacji o zastosowanej metodzie budowy języka.

Znaczna część prezentowanych języków słów kluczowych zawiera jednostki leksykalne bez stosowania preferencji, ale część z nich zawiera jednostki preferowane i niepreferowane, czyli stosuje preferencję. Zamieszczone opisy języków słów kluczowych świadczą o stosowaniu w nich tzw. gramatyki zerowej, częściowo pozycyjnej i w jednym przypadku gramatyki pozycyjnej. Przyjęte kryteria kwalifikowania języków do grupy języków słów kluczowych pozwalają stwierdzić, że kryterium preferencji nie stanowiło cechy dystynktywnej tych języków odróżniającej je od języków deskryptorowych. Również „dowolnie” traktowano rodzaj gramatyki.

Przytoczone dane pochodzą w warunkach polskich z „epoki przedinternetowej”, toteż ich aktualność jest dyskusyjna. Wskazują one na możliwości wykorzystania przynajmniej niektórych języków słów kluczowych, w tym do organizacji zasobów internetowych. Niezależnie od tego słuszne jest stwierdzenie, które narzuca się samo przez się, że nigdy nie istniała, i nadal nie istnieje, kompletna informacja o stosowanych w Polsce i za granicą językach informacyjno-wyszukiwawczych, w tym językach słów kluczowych, dlatego materiał badawczy mogły stanowić tylko wybrane języki słów kluczowych, a wyniki badań mogą być uogólnione tylko w ograniczonym stopniu na wszystkie języki słów kluczowych.

W zdecydowanej większości przypadków podstawę charakterystyki stosowanych w językach słów kluczowych słowników stanowią zebrane przeze mnie słowniki. Ich opisy bibliograficzne zamieściłem w *Aneksie*.

3.1.2. Cele badawcze

W ramach moich badań nad tożsamością języków słów kluczowych postawiłem sobie następujące szczegółowe pytania:

- w jaki sposób w poddanych badaniom językach słów kluczowych została określona tożsamość tych języków;
- jakie inne elementy, w tym innych języków informacyjno-wyszukiwawczych, zawierają te języki w stosunku do ich wersji klasycznych;
- w jakim kierunku zmiernają te języki.

Wyniki analizy makro- i mikrostruktur wybranych słowników słów kluczowych ilustrują omawiany w rozprawie problem tożsamości języków słów kluczowych. Nie stanowią one recenzji tych słowników. Rezultaty moich badań przedstawię tak, aby pokazać wzrastający stopień komplikacji systemów słownikowych (rozbudowy systemu paradygmatycznego) tych języków. W opisie poszczególnych języków starałem się uwzględnić (w miarę możliwości) następujące dane: tytuł słownika, nazwę systemu, w którym jest stosowany słownik, rodzaj wydawnictwa, zakres słownika, zawartość słownika, przeznaczenie słownika, uzasadnienie opracowania, jedno- lub wielojęzyczność słownika, metodę opracowania słownika, problemy występujące przy opracowywaniu słownika, kompletność słownictwa, stopień szczegółowości elementarnych jednostek leksykalnych, gramatykę języka słów kluczowych, uwzględnione relacje, sposób/metodę kontroli słownictwa, metodę wyboru słów-

nictwa, metodę opracowania słownictwa/słownika, źródła słownictwa, autorstwo słownika, literaturę dotyczącą danego języka.

3.1.3. Wyniki badań własnych

3.1.3.1. Słowa kluczowe systemów tradycyjnych

3.1.3.1.1. Klasyczne słowniki słów kluczowych

Najprostszego typu słownika używają klasyczne języki słów kluczowych. Są to słowniki jednojęzyczne, to jest bazujące na jednym języku naturalnym. Następuje w nich „naturalizacja” wyrażen języka naturalnego użytych w funkcji metainformacyjnej. Przykładem systemów wykorzystujących klasyczne języki słów kluczowych są: SYNABA, CYTBIN, system BOINTE w Gliwicach, GOSPODARKA czy BAZTECH. Słowniki tych systemów stanowią najprostsze alfabetyczne listy słów kluczowych bez jakichkolwiek informacji dodatkowych.

System SYNABA to Centralny Specjalistyczny System Informacji o Pracach Naukowo-Badawczych, Rozwojowych i Ekspertyzach Naukowych. Baza danych tego systemu zawiera informacje na temat badań naukowych prowadzonych w Polsce. Składa się z dwóch podbaz: SYNABA1 i SYNABA2. SYNABA1 zawiera 69 586 opisów prac naukowo-badawczych wprowadzonych w latach 1990–1998 do systemu. Ma ona charakter statyczny, tzn. nie jest na bieżąco uzupełniana o nowe rekordy. SYNABA2 to stale rosnący zbiór opisów systematycznie wprowadzanych do systemu od 1999 roku (obecnie ponad 30 000 opisów).

SYNABA posługuje się językiem swobodnych słów kluczowych w odmianie autorskich słów kluczowych. Twórcami opisów systemowych prac naukowo-badawczych są ich autorzy, którzy wypełniając odpowiednie karty systemu SYNABA, opisują je za pomocą słów kluczowych równokształtnych ze znaczącymi terminami danej pracy, nie kierując się żadną szczegółową instrukcją indeksowania, lecz wyłącznie ogólnymi zasadami indeksowania w piśmiennictwie z tego zakresu stanowiącymi tzw. pragmatykę ogólną języka swobodnych słów kluczowych.

Autorskie (albo odautorskie) słowa kluczowe stanowią odmianę języka swobodnych słów kluczowych wyróżnioną ze względu na osobę indeksującą dokument, którą jest autor. W tym wypadku źródłem słów kluczowych jest tytuł pracy, jej abstrakt oraz sam tekst dokumentu, a słownik ma charakter niejawni i opiera się na słowniku mentalnym autora. Jeżeli słowa kluczowe są pobierane z wymienionych źródeł, to sposób indeksowania pokrywa się z indeksowaniem derywacyjnym.

Opisy systemowe dokumentów spływają do systemu SYNABA, w którym są poddawane przez zespół systemu SYNABA korekcie i odpowiedniej obróbce leksykograficznej, polegającej na eliminowaniu słów kluczowych/terminów o znaczeniu ogólnym, tj. występujących we wszystkich dziedzinach, jak na przykład wyrażenie „badanie”, które nie ma charakteru selektywnego, uzupełnianiu charakterystyki wyszukiwawczej o słowa kluczowe, które zdaniem tego zespołu powinny w niej się

znaleźć, ujednoczeniu przypadków i liczby, kolejności słów w wielowyrazowych słowach kluczowych oraz frazach kluczowych, co sprawia, że słownictwo tego języka jest częściowo kontrolowane⁹⁸. Narzędziem pomocniczym w tego rodzaju kontroli leksykograficznej jest jawny słownik słów kluczowych systemu SYNABA zawierający 69 642 słowa kluczowe (stan na 28 października 2008 roku)⁹⁹. Został on zbudowany metodą indukcyjną poprzez kumulowanie słów kluczowych występujących w indeksowanych dokumentach, tj. rzeczywiście występujących w opisach systemowych prac naukowo-badawczych. Słownik ten jest słownikiem tylko języka indeksowania jako jednej z odmian funkcjonalnych języka słów kluczowych. Ma charakter otwarty, tzn. zasób jego słownictwa powiększa się w miarę pojawiania się w opisach nowych słów kluczowych. Pełni jednocześnie funkcję kontrolną w tym sensie, że jest leksykograficznym narzędziem ujednoczania słów kluczowych, co oznacza, że ewentualne wątpliwości w przypadku różnicy formy słowa kluczowego rozstrzygane są na korzyść formy słowa kluczowego już umieszczonego w tym słowniku.

Omawiany słownik nie jest publicznie dostępny dla użytkowników systemu ani na etapie wyszukiwania informacji, ani na etapie opisu prac naukowo-badawczych. Ma więc charakter wewnętrzny (do użytku służbowego) i pomocniczy (nieobowiązkowy) jako narzędzie częściowej kontroli słownictwa języka autorskich słów kluczowych. Jest elementem pragmatyki szczegółowej języka słów kluczowych stosowanego w systemie SYNABA. W odróżnieniu od słowników teoretycznych jest to słownik praktyczny, gdyż został opracowany dla konkretnego systemu. Omawiany słownik pełni więc funkcję kontrolną jako narzędzie korygujące informacje w opisach systemowych prac naukowo-badawczych dostarczanych do systemu. Na tym etapie zespół systemu SYNABA korzysta z przygotowanej na potrzeby tego systemu *Instrukcji indeksowania prac badawczych za pomocą swobodnych słów kluczowych*. Instrukcja ta została opracowana przez Lucynę Annę Bielicką przy współpracy zespołu SYNABA i opublikowana w 1984 roku przez Ośrodek Specjalistycznych Systemów Informacyjnych CIINTE wraz z *Instrukcją klasyfikowania materiałów informacyjnych według Polskiej Klasyfikacji Tematycznej i Macierzowym systemem ochrony informacji*. Stopka wydawnicza tej publikacji zawiera m.in. informację o nakładzie 1505 egzemplarzy formatu B5, objętości ark. druk. 1,5. Instrukcja ta została osobno wydrukowana przez powstały w roku 1990 w Warszawie Ośrodek Przetwarzania Informacji, nadzorowany przez Komitet Badań Naukowych (obecnie przez Ministerstwo Nauki i Szkolnictwa Wyższego). Wydanie to nie zawiera żadnych informacji o dacie druku, nakładzie i autorstwie. Porównanie jego tekstu z tekstem

⁹⁸ Informacje dotyczące stosowania tej procedury w systemie SYNABA otrzymałem w 2008 roku od Tadeusza Pietrzyka, długoletniego pracownika i członka tego zespołu zarówno w CINTE, jak i w Ośrodku Przetwarzania Informacji w Warszawie.

⁹⁹ Bardzo cenne informacje merytoryczne dotyczące tego słownika uzyskałem w 2007 roku od Jolanty Szumowskiej, członka zespołu SYNABA II. Z uwagi na fakt dużej liczby słów kluczowych (około 70 000 = 258 stron tradycyjnego wydruku), funkcjonowania tego słownika w ważnym dla nauki i badań naukowych w Polsce systemie oraz niewielką liczbę takich słowników w Polsce uznałem za niezbędne jego zasygnalizowanie w rozprawie. W lutym 2007 roku słownik ten liczył około 80 000 słów kluczowych. Z uwagi na małą operatywność aż tak dużego zasobu leksykalnego zawartość słownika poddano „melioracji”, której efektem była redukcja liczby słów kluczowych do około 70 000, a więc o około 10%.

instrukcji z 1984 roku pozwoliło mi ustalić, że dokonano tylko drobnych przeróbek formalnych niemających charakteru merytorycznego, a polegających tylko na zmianie nazwy „Centrum Informacji Naukowej, Technicznej i Ekonomicznej” na „Ośrodek Przetwarzania Informacji przy Komitecie Badań Naukowych” na okładce i w tekście. Instrukcja ta – podobnie jak poprzednia – ma charakter wewnętrzny i nie jest dostępna dla twórców opisu prac ani dla korzystających z systemu SYNABA. Instrukcją tą posługują się pracownicy SYNABY. Z uwagi na to, że zawiera ona większość zasad tzw. pragmatyki ogólnej języków słów kluczowych, uważam za stosowne jej analizę i szczegółowe omówienie, zdając sobie sprawę, że problematyka ta koresponduje wyraźnie z przedmiotem rozważań w rozdziale 3.2.

Instrukcja ta – jak już wspominałem – została opracowana w 1984 roku przez Lucynę Annę Bielicką i dotyczy indeksowania prac badawczych za pomocą swobodnych słów kluczowych w systemie SYNABA. Jest więc elementem pragmatyki szczegółowej języka tego systemu. Stanowi ona formę słownika języka słów kluczowych zadaną instrukcją indeksowania. Tekst instrukcji składa się z sześciu punktów.

Punkt 1 stanowi wprowadzenie i zawiera podstawowe informacje dotyczące języka, reguł indeksowania swobodnego, wyboru słowa kluczowego oraz tworzenia tekstów w języku słów kluczowych. Intencją twórców instrukcji było to, aby stanowiła ona pomoc metodyczną w indeksowaniu prac naukowo-badawczych przekazywanych do systemu SYNABA dla osób „sporządzających analizy dokumentacyjne”. Widać ewidentnie, że dotyczy ona jedynie języka słów kluczowych w jego funkcji języka indeksowania. Ponieważ analizy dokumentacyjne sporządzają autorzy prac naukowo-badawczych, można przypuszczać, że w pierwotnym zamierzeniu twórców miała być ona powszechnie dostępna. Fakt wyczerpania się nakładu instrukcji i nieumieszczenie jej tekstu w Internecie sprawia, że posługują się nią tylko członkowie zespołu SYNABA, który w tym wypadku pełni rolę komórki wiodącej w zakresie stosowania i formułowania zasad pragmatyki szczegółowej tego języka.

W punkcie 2 określono etapy indeksowania swobodnego. Są nimi:

- analiza treści/formy pracy naukowo-badawczej;
- wybór słów kluczowych;
- nadawanie odpowiedniej formy słowom kluczowym.

Dalsza część instrukcji zawiera szczegółowe omówienie wymienionych etapów indeksowania.

Punkt 3 dotyczy analizy treści prac naukowo-badawczych, której celem jest ustalenie istotnych elementów treści indeksowanych dokumentów. Określono w nim podstawy analizy stanowiące jednocześnie źródła słów kluczowych, za które uznano: tytuł pracy, tytuły rozdziałów i podrozdziałów, teksty analiz dokumentacyjnych i streszczeń autorskich. W analizie treści za bardzo pomocne uznano również fragmenty prac wyróżnione graficznie, wykresy, tablice, wstęp i zakończenie. Instrukcja zaleca przeprowadzenie tej analizy w celu określenia:

- dyscypliny naukowej lub gałęzi gospodarki narodowej, w ramach których znajduje się przedmiot badań;
- przedmiotu badań;
- materiału badawczego;

- metody badawczej;
- właściwości przedmiotu badań, tj. przyjętego w pracy punktu widzenia, ujęcia przedmiotu badań;
- umiejscowienia w czasie i przestrzeni przedmiotu badań.

Wyszczególnione dane, poza przedmiotem badań, podaje się w postaci słów kluczowych, ale tylko wtedy, gdy mają istotne znaczenie w omawianej pracy.

W punkcie 4 zamieszczono kryteria oceny zestawu wybranych słów kluczowych, które mają stanowić charakterystykę wyszukiwawczą. Są to kompletność i szczegółowość indeksowania. Wątpliwości, które wyrażenia powinny stać się słowami kluczowymi, instrukcja słusznie rozstrzyga zarówno na korzyść nazw ogólnych, jak i szczegółowych.

Punkt 5 jest najobszerniejszy. Zawiera kryteria określania i leksykograficznego ujednociania formy słów kluczowych. W wyborze słów kluczowych instrukcja preferuje „ściśle trzymanie się tekstu pracy. Słowa kluczowe spoza tekstu są wprowadzane wtedy, gdy w pracy występują sformułowania niejasne i niezręczne” (s. 3). Kryteria dotyczą także wyboru terminów spośród synonimicznych i bliskoznacznych, rozstrzygnięć w kwestii jedno- lub wielowyrazowości słów kluczowych, ujednociania deklinacyjnego i szyku wyrazów w wielowyrazowych słowach kluczowych. Z instrukcji wynika, że słowami kluczowymi powinny być tylko takie jednostki leksykalne, których znaczenia nie da się rozdzielić, w sensie semantycznym, na pojedyncze składniki. Instrukcja dopuszcza jako słowa kluczowe skróty, w tym akronimy, nazwiska, nazwy instytucji, nazwy geograficzne, określenia czasu, ale tylko wtedy, gdy są konieczne i ważne dla treści danej pracy.

W punkcie 6 zilustrowano za pomocą ośmiu przykładów sposób stosowania zasad formalnych, zawartych w punktach 1–5. Poszczególne przykłady zawierają opis bibliograficzny pracy naukowo-badawczej, abstrakt/streszczenie autorskie oraz charakterystykę wyszukiwawczą w postaci zestawu słów kluczowych. Stanowią one ilustrację stosowania zasad zarówno pragmatyki ogólnej¹⁰⁰, jak i pragmatyki szczegółowej systemu SYNABA. Jako wymagania pragmatyki szczegółowej podano:

- wielkość charakterystyki wyszukiwawczej (nie może ona liczyć więcej niż 300 znaków);
- maksymalną długość jednego słowa kluczowego (łącznie z wewnętrznymi spacjami nie może przekroczyć 30 znaków);
- sposób pisania i oddzielania od siebie poszczególnych słów kluczowych (słowa kluczowe mają być pisane dużymi literami i należy je oddzielać przecinkiem).

¹⁰⁰ W *Słowniku encyklopedycznym informacji, języków i systemów informacyjno-wyszukiwawczych* (Bojar 2002, oprac.) pragmatykę języków informacyjno-wyszukiwawczych dzieli się na pragmatykę ogólną, czyli metodykę stosowania danej klasy języków, która zawiera reguły ogólne, oraz pragmatykę szczegółową, opisującą reguły właściwe tylko dla konkretnych jednostkowych języków informacyjno-wyszukiwawczych.

Oto wybrany fragment słownika słów kluczowych systemu SYNABA:

(...)	Abramowski Edward
5-HIAA (kwas 5-hydroksyindolooctowy)	abrazja
5-metylocytozyna	abrazja powietrzna
5-metylotetrahydrofolian	abrazja żwirów rzecznych
5-pentafluoropropenyouracyl	ABS
5-trifluorowinyouracyl	absencja chorobowa
6 Program Ramowy	absencja wyborcza
6-merkaptopuryna	absolut
6-OH-dopamina	absolutne retrakty otoczeniowe
7 Program Ramowy	absolutność
7 psu	absolutyzm
8-hydroksy-2-deoksyguanozyna (8-OHdG)	absolwenci uczelni
8-oksyguanina	absolwent
99mTc-MIBI	absolwent szkoły wyższej
9-akrydony	absolwent uczelni technicznej
9-aminoakrydyny	absorbancja
9-metylotryptycen	absorbpcja związków chemicznych
AAA+ Boks VII	absorber energii
ABA	absorber nieliniowy
abciximab	absorber półprzewodnikowy nieliniowy
„a,b-dehydro-a-aminokwasy”	absorpcja
abelowe różności	absorpcja amoniaku
aberracja chromosomowa	absorpcja atomowa
aberracja chromosomowa liczbowa	absorpcja dielektryczna
aberracja optyczna	absorpcja dipolowa
aberracja sferyczna	absorpcja dwutlenku siarki
aberracja subtelomerowa	absorpcja dwutlenku węgla
aberracje chromosomowe	absorpcja energii
ablacja	absorpcja energii zderzenia
ablacja igłowa przezcewkowa	absorpcja (ESA)
ablacja laserowa	absorpcja gazów
ablacja migotania przedsionków	absorpcja mikrofal
ablacja prądem o częstotliwości radiowej (ablacja RF)	absorpcja mikrofalowa
ablacja przeznaczyńowa	absorpcja neutronów
Ablewicz Jerzy biskup	absorpcja nieliniowa
abolicjonizm	absorpcja optyczna
abonament	absorpcja podczerwieni
aborcja	absorpcja promieniowania
aborcja nielegalna	absorpcja przejściowa
Aborygen	absorpcja rentgenowska
ABR	absorpcja SO ₂
Abraham	absorpcja światła
Abraham Władysław	absorpcja wodoru
	abstrakcyjność językowa

abstynencja	adaptacja kulturowa
abstynencja tytoniowa	adaptacja literatury
(...)	adaptacja organizacji
acyltransferaza lecytynowo-cholesterolowa	adaptacja polityczna
(LCAD)	adaptacja postnatalna
adamantan	adaptacja rośliny
Adampol	adaptacja sceniczna prozy
Adams Quincy John	adaptacja selektywna
Adamski Stanisław biskup	adaptacja społeczna
adamsyt	adaptacja systemu
adaptabilność	adaptacja twórcza
adaptacja	adaptacja wysiłkowa
adaptacja Biblii	adaptacja zawodowa
adaptacja budynku	adaptacyjna MES
adaptacja do konfliktu	adaptacyjne witryny internetowe
adaptacja do stresu	adaptacyjność rodziny
adaptacja filmowa	(...)
adaptacja fizjologiczna	

Słownik słów kluczowych systemu SYNABA podaje słowa kluczowe uporządkowane alfabetycznie i nie zawiera żadnych odsyłaczy. Jest to słownik jawny klasycznego języka słów kluczowych w wydaniu autorskim. Słowami kluczowymi są wyrażenia języka naturalnego ujednolicone stosownie do wymagań języka polskiego. Słownik zawiera przeważnie jednowyrazowe słowa kluczowe, ale też wielowyrazowe słowa kluczowe oraz frazy kluczowe. Słowami kluczowymi są też nazwy własne, w tym osobowe oraz akronimy. Zakres słownika, podobnie jak zakres systemu SYNABA, obejmuje wszystkie dziedziny nauki, z których wpłynęły do systemu opisy prac naukowo-badawczych. Słownictwo tego języka jest paranaturalne, ale zawiera też jednostki leksykalne równokształtne z terminami innych sztucznych języków specjalistycznych, na przykład chemii.

Podobnie wyglądają *Alphabetical List of the Key-Words for the Clearinghouse File at IINTE* oraz słownik słów kluczowych bazy danych CYTBIN (http://ibin.us.edu.pl/cbn/lista_slow_kluczowych.doc). Baza ta jest tworzona przez Instytut Bibliotekoznawstwa i Informatyki Uniwersytetu Śląskiego i dotyczy cytowań prac z zakresu informacji naukowej i bibliotekoznawstwa. Oto fragment tego słownika:

(...)

Egzemplarz obowiązkowy

Ekologia informacji

Ekonomia

Ekonomia on-line

Elektroniczne dostarczanie dokumentów

Emigracja

Estetyka

Estonia

ETSI
Etyka zawodowa
Europa
Ewicz Kazimierz
Ewidencja zbiorów bibliotecznych
Excel
Filmoznawstwo
Filologia polska
Filozofia
Finansowanie bibliotek
Format wprowadzania danych
Francja
Fundacja Bertelsmanna
(...)

Tego typu jawnym słownikiem języka słów kluczowych posługuje się również Branżowy Ośrodek Informacji Naukowej, Technicznej i Ekonomicznej Instytutu Metali Nieżelaznych w Gliwicach, wiodący w Polsce ośrodek informacji w zakresie dokumentacji i informacji o przerobie rud, metalurgii metali nieżelaznych oraz dziedzin pokrewnych (<http://www.imn.gliwice.pl/Bointe/sownik.htm>).

Klasycznym językiem słów kluczowych jest też język słów kluczowych „Bazy danych o zawartości polskich czasopism technicznych” BAZTECH. Jest to bibliograficzno-abstraktowa baza danych, rejestrująca opisy artykułów z polskich czasopism z zakresu nauk technicznych oraz wybranych czasopism z zakresu nauk ścisłych i ochrony środowiska. Baza ta od 1998 roku rejestruje opisy artykułów z około 500 tytułów czasopism, publikacje z wydawnictw własnych uczelni technicznych i instytutów naukowo-badawczych, opublikowane materiały z konferencji naukowych. Projekt ten obecnie realizują 22 instytucje oraz 55 instytucji współpracujących. Od 1999 roku baza jest udostępniana w sieci Internet na serwerze ICM (Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW) pod adresem <http://baztech.icm.edu.pl> i razem z innymi tworzy zasoby Biblioteki Wirtualnej Nauki (BWN). Właścicielem i twórcą bazy jest Konsorcjum BazTech. Baza zawiera autorskie słowa kluczowe lub nadawane przez indeksującego w języku polskim i języku angielskim. Lista słów kluczowych polskich jest dostępna pod adresem: http://baztech.icm.edu.pl/cgi-bin/btgetdes_pl.cgi. Od twórców tej bazy uzyskałem informacje, że w systemie tym generuje się (dokonuje wyboru) słowa kluczowe także z języka haseł przedmiotowych KABA, przy czym określniki mogą stawać się słowami kluczowymi, stąd charakterystyczną cechą tego języka są zapożyczenia leksykalne z KABY. Polega to na redukcji hasła przedmiotowego do postaci tematu, któremu rzadko towarzyszy grupa określników. Tym samym zdania języka haseł przedmiotowych przyjmują postać wykazu tematów reprezentujących ich przedmiot i treść. Również niektóre określniki stają się autosyntaktycznymi jednostkami leksykalnymi, przyjmując postać i status słów kluczowych.

Przykładem zagranicznego słownika omawianego typu może być francuski słownik słów kluczowych zdrowia publicznego pt. *Index des mots-clés de la base de*

données de l'ENSP au 8 février 1990. Warto zwrócić uwagę, że tego typu proste struktury organizacji leksykalnej języka słów kluczowych pojawiają się również w tezaurusach, czego przykładem mogą być „czyste” alfabetyczne wykazy deskryptorów w różnych językach naturalnych, zamieszczone na przykład w *Multilingual European Thesaurus on Health Promotion (Deutsch-English-Français-Nederlands)*, będącym publikacją Komisji Europejskiej. Różnica jest jednak istotna. Wykazy te są zamknięte, jakkolwiek w swoim zewnętrznym kształcie są identyczne z wykazami słów kluczowych, zawierają wyłącznie słownictwo w pełni kontrolowane.

W zaprezentowanych słownikach klasycznych języków słów kluczowych z założenia nie jest w ogóle prezentowana struktura paradygmatyczna języka słów kluczowych. Znaczenie słów kluczowych rozumiane jest jako relacja wewnątrzsystemowa języka naturalnego, zachodząca pomiędzy jednostkami tego systemu językowego, bazująca na kompetencji jego użytkowników, dlatego przy korzystaniu z tych słowników trzeba posługiwać się wyłącznie informacją o znaczeniu słów kluczowych w ich funkcji metainformacyjnej, wynikającą z kompetencji językowej użytkownika konkretnego języka naturalnego, gdyż ta nie jest tu systemowo modyfikowana. Języki tego typu odwołują się więc do wiedzy językowej i pozajęzykowej oraz kompetencji użytkownika w zakresie języka naturalnego, którego leksykę reprezentują, co stanowi ważny element tożsamości tych języków. Wybór słów kluczowych jest zgodny z ogólnymi zasadami wyboru i opracowywania słów kluczowych w języku polskim, podanymi m.in. w omawianej *Instrukcji indeksowania prac badawczych za pomocą słów kluczowych*.

Słowniki klasycznych języków słów kluczowych charakteryzują się niejawnym przejściem z języka naturalnego na język słów kluczowych, gdyż implicite występują w nich wyrażenia języka naturalnego w funkcji metainformacyjnej, powiązane ze słowami kluczowymi relacją odpowiedniości wyszukiwawczej, co jest charakterystyczne dla tej odmiany języków słów kluczowych. Słowniki tych języków faktycznie stanowią podzbiór wyrażeń języka naturalnego. Identycznym formom są przypisane znaczenia skorelowane ze znaczeniami odpowiednich wyrażeń w języku naturalnym.

Omawiane w dalszej części pracy słowniki należą do nieklasycznych języków słów kluczowych.

3.1.3.1.2. Słowniki słów kluczowych z systemami odesłań

Nieco więcej informacji o strukturze semantycznej systemu języka podają nowsze prezentacje języków słów kluczowych, w których listom słów kluczowych towarzyszą różnego rodzaju odsyłacze, będące wykładnikami „podstawowych” relacji słownikowych, do których zalicza się relację synonimii wyszukiwawczej oraz relację pokrewieństwa¹⁰¹. Przykład stanowi większość słowników zbudowanych przez Ośrodek Informacji Naukowej PAN. Do tej grupy zaliczam również: *Słownik hasel z zakresu zdrojownictwa i dziedzin pokrewnych* (1977), *Słownik naukowy słów*

¹⁰¹ Por. przypis 12.

kluczowych transportu morskiego (Pierchalska 1986), *Słownik słów kluczowych z metodologii nauk społecznych* (Sprys 1990), *Indeks alfabetyczny słów kluczowych (zdrowia publicznego)* (1993), *Słownik słów kluczowych w logopedii* (Surowaniec 1993), *Słownik języka informacyjno-wyszukiwawczego Biblioteki Społeczno-Ekonomicznej* (Pietraszkiewicz, Stankiewicz, Żmijewska 1997), *Słownik słów kluczowych językoznawstwa sławistycznego* (Rudnik-Karwatowa, Karpińska 1999) oraz słowniki jednojęzyczne w innych językach niż język polski: *Schlagwortverzeichnis Pädagogik* (Fischbach 1970), *Spisok ključovych slov po farmakologii, himičeskiye sredstva* (Spisok 1984), *Deskriptornyj slovar po informatike* (1986), *Spisok ključovych slov po farmakologii* (1987), *Spisok normalizovannoj leksiki po filozofii i sociologii* (1989).

Jak pokazuje praktyka leksykograficzna języków słów kluczowych, słowniki słów kluczowych często zaopatruje się w odsyłacze. Przyczyną tego jest niewystarczalność struktury słowników klasycznych języków słów kluczowych do odwzorowania rzeczywistości dokumentacyjnej w systemie informacyjno-wyszukiwawczym. Przykład wzbogacenia list słów kluczowych o system odsyłaczy, który explicite pokazuje elementy struktury paradygmatycznej języka, stanowią właśnie języki słów kluczowych zbudowane przez OIN PAN. Jednak i w tym przypadku paradygmatyka języka implicite tkwi w kompetencji jego użytkownika w zakresie języka polskiego, a więc interpretacja znaczenia poszczególnych słów kluczowych zależy przede wszystkim od wiedzy użytkownika w zakresie języka naturalnego (w jego wersji specjalistycznej, czyli terminologii) i jego wiedzy pozajęzykowej. Mimo że słowniki te nie funkcjonują w realnie działających systemach informacyjno-wyszukiwawczych, potraktuje je jako źródło materiału, za którego pomocą zilustruję ogólne prawidłowości rozwoju języków słów kluczowych.

Budowanie języków słów kluczowych dla nauk społecznych jest wyjątkowo trudne. Wiąże się m.in. z interdyscyplinarnością wielu problemów badawczych, brakiem precyzyjnej terminologii, zróżnicowaniem metod badawczych. Charakterystyczną cechą tej grupy nauk są różne teorie posługujące się własnymi systemami terminologicznymi, różnego rodzaju nazwami własnymi, jak na przykład nazwy osobowe, nazwy koncepcji, nazwy teorii, nazwy organizacji. Ma to istotny wpływ na funkcjonowanie systemów wyszukiwania informacji, gdyż płynność granic i interdyscyplinarny charakter wielu problemów badawczych nauk społecznych sprawia, że na dane pytanie informacyjne możemy w odpowiedzi otrzymać opisy dokumentów relewantne dla kilku dyscyplin naukowych.

Tak trudnego zadania podjęła się w latach 1986–1993 krakowska Pracownia Informacji Naukowej OIN PAN. Opracowała słowniki słów kluczowych dla nauk społecznych, które miały się stać elementem języka słów kluczowych systemu informacji nauk społecznych w Polsce¹⁰². Z tego zakresu opublikowano łącznie jedenaście słowników¹⁰³. Słowniki te mają charakter słowników teoretycznych, bo nie były tworzone dla żadnego konkretnego systemu, zresztą w czasie ich publikowania

¹⁰² Por. przypis 94.

¹⁰³ Oprócz wymienionych opracowano i opublikowano słownik słów kluczowych z zakresu neuropsychofarmakologii oraz słownik słów kluczowych z botaniki.

takie systemy w Polsce jeszcze nie istniały. Słowniki zostały przygotowane według jednolitej metodologii, stworzonej i opracowanej przez tę Pracownię¹⁰⁴. Została ona sformalizowana w postaci odpowiedniej instrukcji wyboru słów kluczowych w naukach społecznych, która określała cel prac, definiowała słowa kluczowe, podając ich przykłady, wskazywała źródła słownictwa, a także sposoby jego wyboru i kontroli kompletności. Zalecenia metodyczne zamieszczono w notach *Od wydawcy*.

Przyjęta metodyka opracowywania słownictwa opierała się na wykorzystaniu relacji ekwiwalencji rozszerzonej o następujące relacje: relację polisemii (synonimii wyszukiwawczej), oznaczanej za pomocą odsyłacza „zob.”, relację pokrewieństwa oznaczaną za pomocą odsyłacza „zob. też.”, co znalazło odbicie w mikrostrukturze tych słowników. Relacje pokrewieństwa obejmowały przede wszystkim relacje kojarzeniowe oraz relacje antonimii, charakterystyczne dla innych języków informacyjno-wyszukiwawczych. Makrostruktura tych słowników dopuszczała układ fasetowy słownictwa. Między innymi to uczyniło budowane słowniki elementem nieklasycznych języków słów kluczowych.

Prace nad słownikami słów kluczowych miały charakter zespołowy. Polegały na współpracy dwóch zespołów specjalistów, a mianowicie średniej kadry naukowej z poszczególnych dyscyplin naukowych, która dokonywała wyboru słów kluczowych¹⁰⁵, oraz specjalistów w zakresie informacji, którzy opracowywali te zasoby pod względem metodologicznym, głównie leksykograficznym.

Słowniki zostały opracowane metodą dedukcyjną, co jest charakterystyczne dla słowników teoretycznych. Języki słów kluczowych oparte na tych słownikach są językami sztucznymi. Słowniki te zostały wyposażone we wstęp od wydawcy zawierający informacje o istocie i przeznaczeniu słowników, wstęp metodologiczny autorów, alfabetyczną listę słów kluczowych. Niektóre z nich zawierają dodatkowo wykaz najważniejszych prac, z których czerpano słownictwo oraz wykazy nazw osobowych, nazw ciał zbiorowych, a także ujednolicone tytuły czasopism.

Słowniki słów kluczowych OIN PAN nie mają charakteru normalizacyjnego. Miały służyć jako pomoc przy indeksowaniu i wyszukiwaniu dokumentów, stąd ich pragmatyczny charakter oraz otwartość na nowe słownictwo i podatność na ciągłą aktualizację. Były budowane z myślą o wykorzystaniu ich w obu odmianach funkcjonalnych języka słów kluczowych, tzn. zarówno w języku indeksowania, jak i w języku wyszukiwawczym.

Słownictwo tych języków ma charakter paranaturalny, zaś pod względem stopnia kontroli jest częściowo kontrolowane. Same zaś słowniki mogą pełnić zarówno funkcję kontrolną (kontrola pełna, całkowita), jak i funkcję narzędzia częściowej kontroli w postaci słownika wzorcowych słów kluczowych.

¹⁰⁴ Prace nad zasobami leksykalnymi poszczególnych słowników doprowadziły do powstania zespołu odpowiedzialnego za opracowanie i posługiwanie się swoistą metodyką budowy tych słowników, którą można by nazwać krakowską szkołą budowy słowników języków informacyjno-wyszukiwawczych. Metodykę tworzenia omawianych słowników słów kluczowych opisali W. Babik i A. Fitowa.

¹⁰⁵ Dla opracowujących wykazy słów kluczowych zorganizowano seminarium naukowe z udziałem specjalistów w zakresie języków informacyjno-wyszukiwawczych.

Z założenia gramatyka tych języków miała być gramatyką opartą wyłącznie na indeksowaniu współrzędnym (język indeksowania i język wyszukiwawczy). Słowniki nie zawierają wykazów modyfikatorów. Jednym wyjątkiem, który dopuszcza możliwość stosowania gramatyki częściowo pozycyjnej, jest *Słownik słów kluczowych z historii Polski (z elementami historii powszechnej)*.

Większość słowników słów kluczowych, które ukazały się drukiem, to słowniki opracowane przez zespoły mieszane, składające się ze specjalistów danej dziedziny wiedzy oraz specjalistów z zakresu języków informacyjno-wyszukiwawczych.

Wspomniany system informacji w zakresie nauk społecznych miał być wspomagany przez opublikowany przez poznański Ośrodek Informacji Naukowej PAN *Słownik słów kluczowych z metodologii nauk społecznych* (Sprys 1990). Jest on efektem indywidualnej pracy jednej osoby, a mianowicie Zofii J. Sprys. Ten wykaz słów kluczowych obejmuje terminologię z zakresu metodologii, filozofii wraz z estetyką, logiki formalnej, metodologii ogólnej, antropologii, historii, psychologii, socjologii, teologii religii, a także częściowo językoznawstwa, lingwistyki, literaturoznawstwa, pedagogiki, nauk ekonomicznych, politycznych i prawnych, nauk o kulturze i sztuce oraz teorii komunikacji, zachowania, gier i decyzji (Wstęp s. 5–6). Jego opracowanie oparto na tej samej metodologii, którą zastosowano w krakowskiej Pracowni OIN PAN.

Bogaty repertuar uwag zastosowano w opracowanym przez krakowską Pracownię OIN PAN *Słowniku słów kluczowych z botaniki*. Oto fragment tego słownika (s. 17):

(...)

Dylatacja

Dyluwium

zob. **Plejstocen**

Dymorfizm (płciowy)

Dynamika

Dynamika populacji

zob. też **Demografia**

Liczebność

Przeżywalność

Reprodukcja

Rozrodczość

Śmiertelność

Zagęszczenie

Dysjunkcja

Dystrofia

Dystroficzne jeziora

zob. **Jeziora dystroficzne**

Dywergencja

(...)

Tu informacji o znaczeniu wyrażen jest więcej niż w klasycznych językach słów kluczowych. Dzięki dodatkowym uwagom ma ona zresztą inny charakter niż we wspomnianych językach. Wprowadzony system odsyłaczy, niezależnie od kompe-

tencji użytkownika w zakresie języka naturalnego i jego wiedzy pozajęzykowej, ustanawia pewne relacje paradygmatyczne, które często modyfikują znaczenie konkretnych wyrażen języka naturalnego w ich funkcji metainformacyjnej. W przypadku odsyłacza „zob. też” mamy do czynienia z niespecyfikowanymi relacjami paradygmatycznymi, w tym relacjami asocjacyjnymi, zwanymi też relacjami pokrewieństwa, których wykorzystanie w procesach wyszukiwania informacji w Internecie umożliwia tworzenie sieci. Słowo kluczowe odwzorowujące kategorię tematyczną pozwala na prezentację relewantnych CHWD i wyświetlenie skojarzonych z nim wyrażen pochodzących z takiego słownika/indeksu. Przykładem słownika słów kluczowych, w którym „wyspecyfikowano” tylko relacje kojarzeniowe, jest *PEDAGOG. Słownik słów kluczowych bibliotecznego systemu informacyjno-wyszukiwawczego* (Matwiejczuk, Okoń 2002) Biblioteki Uniwersyteckiej w Opolu.

W omawianej grupie słowników zastosowano sieć relacji asocjacyjnych odwzorowujących powiązania znaczeniowe pomiędzy słowami kluczowymi. Wzbogacenie struktury paradygmatycznej języka słów kluczowych o tego typu zależności semantyczne było rzadko stosowane. W omawianych słownikach rodzaje relacji kojarzeniowych¹⁰⁶ nie są specyfikowane: słowom kluczowym przyporządkowane zostały tylko skojarzone terminy. Stosownie do pragmatyki ogólnej języków słów kluczowych układ jest alfabetyczny, słowa kluczowe będące rzeczownikami występują w formie mianownika liczby pojedynczej, zaś w liczbie mnogiej tylko tzw. *pluralia tantum*. W wielowyrazowych słowach kluczowych zachowano obowiązujący w języku polskim szyk wyrazów¹⁰⁷. Przy wyborze słów kluczowych kierowano się powszechnie przyjętymi kryteriami, jak bieżące i częste stosowanie, poprawność terminologiczna, wyrazistość strukturalna, rodzimość, zwięzłość.

Do tej klasy słowników zaliczam również *Słownik słów kluczowych w logopedii* (Surowaniec 1993), opracowany przez Józefa Surowańca z Uniwersytetu Pedagogicznego w Krakowie. Słownik powstał na bazie *Słownika terminów logopedycznych*¹⁰⁸, który stanowił główne źródło leksyki dla budowanego słownika słów kluczowych, ale i dostarczył paradygmatyki dla języka. Co do możliwości wykorzystania zasobu leksykalnego tego typu słowników, to należy mieć na uwadze, że słownik słów kluczowych nie zawiera wszystkich istniejących w danej dziedzinie terminów. Jest to zadanie słownika terminologicznego, a nie słownika języka informacyjno-wyszukiwawczego. Słownik takiego języka zawiera tylko te terminy, które zostały uznane za elementarne jednostki leksykalne tego języka przydatne w procesie in-

¹⁰⁶ Zależności kojarzeniowe zwane są słabymi, gdyż istnieje jedynie pewne prawdopodobieństwo, że dokumenty wyszukane za pomocą terminów połączonych relacją skojarzeniową z terminem reprezentującym słowo kluczowe będą relewantne dla użytkownika. Za silne uważa się zależności hierarchiczne, gdyż użytkownik może być pewny, że przynajmniej część z dokumentów, którym przypisano dany termin, będzie dla niego relewantna.

¹⁰⁷ Stosowanie w obrębie fraz kluczowych metody pozycyjnej w języku polskim sprawia, że najpierw jest wyraz określający, a potem określający.

¹⁰⁸ Por.: J. Surowaniec: *Słownik terminów logopedycznych*. Kraków 1992; J. Surowaniec: *Podręczny słownik logopedyczny*. Kraków 1993.

deksowania. Słownik słów kluczowych może stanowić jednak dobrą podstawę do dalszych prac nad porządkowaniem i normalizacją terminologii w danej dziedzinie.

Innym przykładem omawianej grupy słowników słów kluczowych jest *Słownik słów kluczowych językoznawstwa sławistycznego*¹⁰⁹ (Rudnik-Karwatowa, Karpińska 1999), który powstał na potrzeby opisu dokumentów w bazie danych z zakresu językoznawstwa sławistycznego. Oto fragment tego słownika z 1999 roku (s. 32):

(...)

derywacja

derywacja afiksalna zob. afiksacja

derywacja alternacyjna

derywacja dezintegralna

derywacja fleksyjna zob. derywacja paradygmatyczna

derywacja morfologiczna

derywacja paradygmatyczna

derywacja prefiksalna zob. prefiksacja

derywacja prefiksально-sufiksalna

derywacja przedrostkowa zob. prefiksacja

derywacja przyrostkowa zob. sufiksacja

derywacja semantyczna

derywacja słowotwórcza

derywacja sufiksalna zob. sufiksacja

derywacja syntaktyczna

derywacja ujemna zob. derywacja wsteczna

derywacja wymienna

(...)

oraz nowszego wydania tego słownika, opublikowanego w 2006 roku (rys. 6).

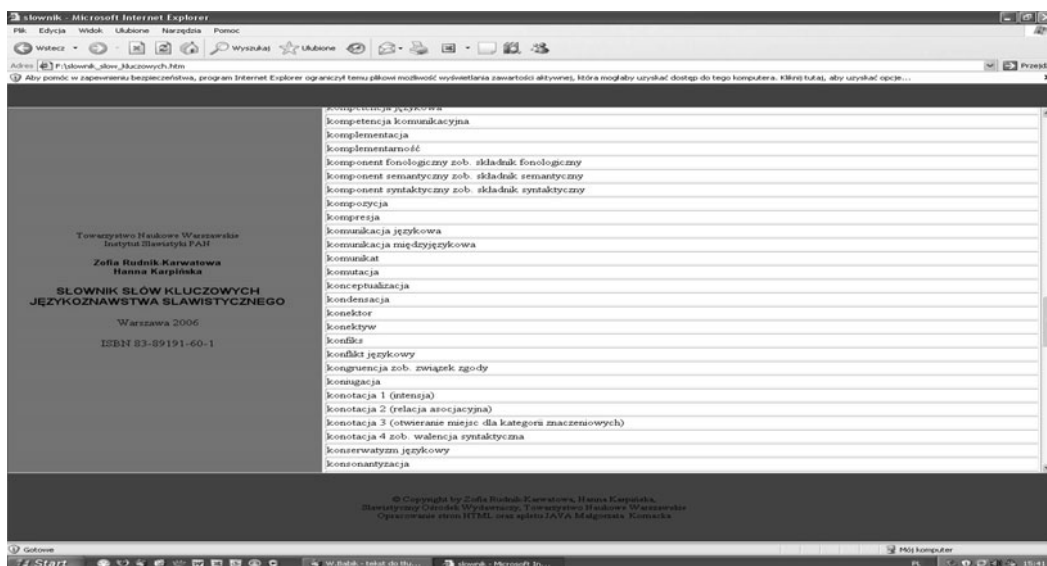
Podane tu informacje odnoszą się do najnowszego wydania tego słownika. Prezentuje on słowa kluczowe językoznawstwa sławistycznego. Jest to zatem dziedzinowy słownik specjalistyczny. We wstępie podano informacje, które przytoczę poniżej. Leksyka odwzorowuje pole semantyczne dziedziny oraz jej produkcję wydawniczą. Słownik zawiera ponad 3000 słów kluczowych¹¹⁰. Jego zadaniem jest pomoc w indeksowaniu i wyszukiwaniu dokumentów, przede wszystkim w tworzonej bazie danych iSybisław, która stanowi elektroniczną edycję międzynarodowej bibliografii językoznawstwa sławistycznego¹¹¹. W stosunku do wcześniejszej wersji, słownik uzupełniono na podstawie reprezentatywnych tekstów bieżącego piśmiennictwa i jego opisów dokumentacyjnych oraz na podstawie źródeł leksykograficznych. Przy gromadzeniu zasobu leksykalnego zastosowano więc metody korpusowe, natomiast sam słownik opracowano metodą indukcyjno-dedukcyjną. Jak pisał autorki we *Wstępie*, słownik

¹⁰⁹ Nowsza wersja tego słownika została opublikowana na CD-ROM-ie w 2006 roku.

¹¹⁰ Wersja słownika z 1999 roku zawierała około 2500 słów kluczowych.

¹¹¹ Bibliografia ta zawiera obecnie około 50 000 opisów dokumentów. Materiał dotyczący lat 1992–1997 został opublikowany w postaci książkowej.

ma być traktowany jako kontrolowana lista słów kluczowych. Jest jednojęzyczny, a jego układ alfabetyczny. Charakterystyczną cechą języka słów kluczowych w tym słowniku jest oparcie systemu leksykalnego wyłącznie na systemie terminologicznym z zachowaniem zasady *literary warrant*¹¹². Do słownika nie wprowadzono nazw języków niesłowiańskich i nazw własnych (poza nazwami szkół językoznawczych). Jediną relacją syntagmatyczną zachodzącą między słowami kluczowymi jest relacja współwystępowania. Struktura słownika jest płaska, co oznacza, że nie uwzględniono w nim żadnych relacji hierarchicznych. W słowniku zastosowano wskaźnik relacji „zob.”, który odsyła od terminów języka naturalnego do słowa kluczowego będącego terminem preferowanym języka naturalnego. Słowa kluczowe podano czcionką prostą. Terminy niebędące słowami kluczowymi wyróżniono kursywą. Terminy polisemiczne zostały opatrzone cyframi arabskimi oraz uzupełnione (w nawiasie) dopowiedzeniem precyzującym zakres znaczeniowy (cyfry stanowią integralną część słowa kluczowego). Gramatyka tego języka to reguły indeksowania współrzędnego. Sposób prezentacji słownictwa jest typowy dla języków słów kluczowych.



Rys. 6. Słownik słów kluczowych językoznawstwa sławistycznego (wersja online)
(Źródło: http://www.ispan.waw.pl/cnis/slownik_slow_kluczowych.htm)

¹¹² Metoda *literary warrant* została opracowana na początku ubiegłego wieku przez Edwarda Wyn-dhama Hulme’a i polega na analizie kolekcji pod kątem tematów odwzorowywanych w procesie indeksowania. Uzależnia wybór danego terminu od częstości jego występowania w literaturze. Na podstawie analizy treściowej dokumentów kolekcji bierze się pod uwagę wyłącznie te tematy, które posiadają co najmniej jednego reprezentanta w postaci CHWD. Niedopuszczalne jest tworzenie tzw. tematów pustych. W ten sposób generuje się w języku słów kluczowych wykaz wyłącznie aut syntaktycznych jednostek leksykalnych stosowanych do reprezentacji treści dokumentów zbioru informacyjnego. Niektórzy autorzy wyróżniają również tzw. *user warrant* oraz *concept warrant*. Możliwy jest też model hybrydowy.

W słownikach nieklasycznych języków słów kluczowych, których przykładem jest omówiony ostatnio słownik, z założenia prezentuje się kompletną lub tylko częściową strukturę paradygmatyczną systemu semantycznego języka słów kluczowych, stosując jako narzędzie pomocnicze odpowiedni system odsyłaczy. Znaczenie słów kluczowych rozumiane jest tu jako relacja wewnątrzsystemowa, zachodząca pomiędzy jednostkami tego systemu, bazująca w znacznej części, ale nie zupełnie, na kompetencji użytkownika, dlatego przy korzystaniu z języków słów kluczowych trzeba posługiwać się nie tylko informacją o znaczeniu, wynikającą z kompetencji semantycznej użytkownika konkretnego języka naturalnego, gdyż ta często jest systemowo modyfikowana. Tego typu języki odwołują się więc do wiedzy językowej i pozajęzykowej oraz do kompetencji językowych użytkownika w zakresie języka, którego leksykę reprezentują. Problem staje się bardziej skomplikowany, gdy jesteśmy użytkownikami języka słów kluczowych opartego na innym języku etnicznym, który nie jest naszym językiem ojczystym. Pierwotna kompetencja językowa ukształtowana pod wpływem języka ojczystego inaczej odwzorowuje rzeczywistość niż inny język etniczny, na którym „oparty” jest język słów kluczowych, na przykład język słów kluczowych bazujący na języku angielskim. Ilustrację tego stanowi *Słownik słów kluczowych z neuropsychofarmakologii* w języku angielskim (s. 20):

(...)

D-145

see **Memantine**

D-amphetamine

see **Amphetamine**

D-cyclazocine

see **Cyclazocine**

D-methamphetamine

see **Methamphetamine**

D-tubocurarine

see **Tubocurarine**

DADLE

DAGO

Damage

Dark

DAVA

DBH

Derisoquine

Decortication

Defense

see also **Avoidance**

Degeneration

Degradation

see also **Metabolism**

Dementia

(...)

Zaprezentowany fragment słownika pozwala zauważyć, że słowniki tego rodzaju charakteryzują się częściowo jawnym przejściem z języka naturalnego na język słów kluczowych. W sposób jawny podawane są paranaturalne słowa kluczowe oraz powiązane relacją odpowiedniości wyszukiwawczej, ale nierównokształtne z nimi wyrażenia języka naturalnego, którymi są synonimy, quasi-synonimy danego słowa kluczowego, dopowiedzenia, homonimy. Wyrażenia nierównokształtne ze słowami kluczowymi, z którymi są one powiązane relacją odpowiedniości wyszukiwawczej, są przyporządkowane tylko niektórym słowom kluczowym umieszczonym w słowniku. Na relację odpowiedniości wyszukiwawczej wskazują tu odpowiednie odsyłacze.

W omawianej grupie języków słów kluczowych słowa kluczowe są wyrażeniami paranaturalnymi, tzn. równokształtnymi z odpowiednimi wyrażeniami języka naturalnego, ale są jednocześnie jednostkami leksykalnymi sztucznych języków słów kluczowych. W prostych spisach słów kluczowych, jak w przykładowo przytoczonych powyżej fragmentach słowników, dodatkowe informacje o znaczeniu słowa kluczowego podawane są tylko wtedy, gdy odpowiednie wyrażenie języka naturalnego jest homonimem lub wyrażeniem wieloznacznym. Wtedy słowa kluczowe opatrywane są dodatkową informacją pozwalającą na ustalenie jednoznacznego znaczenia. Zwykle tego typu informacje nazywane są dopowiedzeniem i zawierają wskazówki pozwalające na ustalenie znaczenia słowa kluczowego, informujące na przykład o dziedzinie wiedzy, w jakiej to wyrażenie jest używane, typie kontekstów, w jakich występuje, nazwie kategorii semantycznej, do jakiej należy. Tego rodzaju „wtręty” wprowadzone do klasycznego wykazu słów kluczowych czynią język oparty na takim słowniku językiem sztucznym.

W przypadku zaprezentowanych języków słów kluczowych do ustalenia paradygmatyki danego języka użytkownikowi języka słów kluczowych w funkcji języka wyszukiwawczego nie wystarczy kompetencja w zakresie języka naturalnego, niezbędna jest jeszcze odpowiednia wiedza językowa dotycząca danego języka słów kluczowych i wiedza o rzeczywistości (wiedza pozajęzykowa).

Użytkownicy słów kluczowych są przekonani, że indeks/słownik słów kluczowych zaopatrzonego w system odsyłaczy jest jeszcze słownikiem słów kluczowych i że mamy jeszcze do czynienia z językiem słów kluczowych, choć faktycznie takie „struktury” można traktować już jako proste tezaury. Granice są tu dość płynne. Języki te mieszczą się w przyjętej w rozprawie szerokiej definicji języka słów kluczowych.

3.1.3.1.3. Słowniki słów kluczowych z elementami fasetyzacji

Do tej grupy słowników należą te, które oprócz alfabetycznego zrzębu głównego zawierają dodatkowo wykazy słów kluczowych różnych kategorii. Prezentując problem tożsamości języka słów kluczowych w tej grupie słowników, omówię *Słownik słów kluczowych z językoznawstwa* oraz *Słownik słów kluczowych z historii Polski*. Oto fragment opracowanego przez OIN PAN *Słownika słów kluczowych z językoznawstwa* (s. 23–24):

(...)

Struktura

Struktura fonologiczna

Struktura frazowa (gramatyka generatywna)

Struktura głęboka (gramatyka generatywna)

Struktura gramatyczna

Struktura hierarchiczna

Struktura językowa

Struktura leksykalna

Struktura logiczna

Struktura pośrednia (gramatyka generatywna)

Struktura powierzchniowa (gramatyka generatywna)

Struktura prozodyczna

Struktura semantyczna

Struktura syntaktyczna

Struktura wyjściowa (gramatyka generatywna)

Strukturalizm

Stwardnienie

zob. **Dyspalatalizacja**

Odmiękczenie

Styl

Stylistyka

Stylistyka lingwistyczna

Stylistyka literacka

Subiekt

Subkategoryzacja (gramatyka generatywna)

Subkod

Substantywizacja

Substrat

Substytucja

Substytut

Sufiks

zob. **Przyrostek**

Sufiks produktywny

Sufiks strukturalny

(...)

Towarzyszy mu *Wykaz języków, dialektów i gwar* (s. 31–36). Oto wybrany fragment:

(...)

Języki ugryjskie

Język chantyjski

Język mansyjski

Język węgierski

DIALEKTY POLSKIE

Dialekty polskie

Dialekt kaszubski

Dialekt małopolski

Dialekt mazowiecki

Dialekt śląski

Dialekt wielkopolski

Nowe dialekty mieszane

GWARY

Gwara chwalimska

Gwara jabłonkowska

Gwara lasowska

Gwara łowicka

(...)

Strukturę fasetową ma również opublikowany przez OIN PAN *Słownik słów kluczowych z historii Polski (z elementami historii powszechnej)*. Zawiera on następujące fasy: „Pojęcia”, „Fakty i wydarzenia”, „Nazwy własne”, „Nazwy geograficzno-historyczne. Modyfikatory”. Umożliwia to dostęp do zasobów leksyki za pomocą odpowiedniego wykazu. Wykazy te są uporządkowane alfabetycznie. Poszczególne jednostki takich wykazów stanowić mogą nazwy węzłów dla powiązań hipertekstowych. Fragment fasy „Pojęcia” tego słownika (s. 84):

Właściciel

Włodarz

Włościanin

zob. **Chłop**

Włość

Włóczęstwo

Włókna

zob. **Łan**

Włókiennictwo

Wodzostwo

Wojewoda

zob. też. **Palatyn**

Województwo

Wojna

Wojna agresywna

zob. też **Wojna niesprawiedliwa**

Wojna celna

Wojna chłopska

Wojna domowa

Wojna elekcyjna

Wojna graniczna

Wojna lokalna

Wojna narodowa

Wojna narodowo-wyzwoleńcza

Wojna niesprawiedliwa

zob. też. **Wojna agresywna**

Wojna obronna

Wojna o niepodległość

Wojna plemienna

Wojna pozycyjna

Wojna przewencyjna

Wojna psychologiczna

Wojna religijna

zob. też. **Krucjata**

Wojna rewolucyjna

Wojna ruchoma

Wojna sprawiedliwa

Wojna sukcesyjna

Wojna światowa

Wojna totalna

„Wojny gwiazdne”

Wojska lądowe

zob. **Potęga lądowa**

Wojska lotnicze

Wojsko

Wojsko kwarciane

zob. też **Kwarta**

Wojsko najemne

Wojna elekcyjna
Wolna elekcja „viritim”
Wolne miasto
Wolni
Wolnizna
Wolnomularstwo
 zob. Masoneria
 zob. też **Łoża wolnomularska**
Wolnomysłicielstwo
 zob. też **Ruch wolnomysłicielski**
Wolności obywatelskie
Wolności szlacheckie
Wolność
Wolność gospodarcza
 zob. też. **Leseferyzm**
Wolność jednostki
Wolność polityczna
Wolność religijna
Wolność szlachecka
Woluntaryzm ekonomiczny
 zob. Woluntaryzm gospodarczy

Woluntaryzm gospodarczy
 zob. **Woluntaryzm ekonomiczny**
Woluntaryzm polityczny
Wódz
Wójt
Wpływ gospodarczy
Wpływ kulturalny
Wpływ polityczny
Wspomnienia
 zob. też. **Pamiętnik**
Wspólnota
Wspólnota narodowa
Wspólnota pierwotna
 zob. też **Organizacja rodowa**
Wspólnota terytorialna
Współczesność
 zob. **Teraźniejszość**
Współpraca międzynarodowa
Współpraca międzypartyjna
Współzawodnictwo międzysystemowe

Przykładem słownika w układzie fasetowym – oprócz prezentowanego wcześniej *Słownika słów kluczowych z historii Polski* – może być też lista słów kluczowych pt. *Keywords for indexing*, zamieszczona w „Materials Research Bulletin” (vol. 37, no 10, 2002, s. IV–V). Jest to lista kontrolowanych słów kluczowych, gdyż tylko słowa z tej listy mogą być użyte do indeksowania prac naukowych. Słownictwo zostało podzielone na cztery następujące kategorie: A. Type of material, B. Preparation and Processing, C. Techniques, D. Properties and Phenomena. W ich ramach słowa kluczowe zostały uporządkowane alfabetycznie. Lista ta nie zawiera żadnych odsyłaaczy. Przy niektórych słowach kluczowych w nawiasach półokrągłych podano ich akronimy, na przykład „nuclear magnetic resonance (NMR)”. Poniżej zamieszczam fragment tej listy z uwagi na fakt odniesienia się do niej przy omawianiu pragmatyki języka słów kluczowych:

A. TYPE OF MATERIAL

alloys
 amorphous materiale
 carbides
 ceramics
 chalcogenides
 composites
 electronic materials
 elements
 fluorides
 (...)

B. PREPARATION AND PROCESSING

chemical synthesis

crystal growth
epitaxial growth
intercalation reactions
laser annealing
(...)

C. TECHNIQUES

atomic force microscopy
differential scanning calorimetry (DSC)
electrochemical measurements
electron diffraction
electron energy loss spectroscopy (EELS)
electron microscopy
high pressure
impedance spectroscopy
(...)

Dodatkowe wykazy słów kluczowych stanowią leksykograficzne udogodnienie w wieloaspektowym indeksowaniu dokumentów w językach słów kluczowych.

Do tej grupy słowników słów kluczowych zaliczam również: *Układ gniazdowy i słownik słów kluczowych wybranych kategorii kultury: etos, obrzędy, demonologia, magia* oraz *Układ gniazdowy i słownik słów kluczowych wybranych kategorii kultury. Medycyna ludowa*, gdyż mają dodatkowe wykazy słownictwa. Ze względu na przyjęte rozwiązania strukturalne tworzą nową grupę słowników słów kluczowych, różniącą się od poprzednich, którą omówię w następnym podrozdziale. Obecnie słowniki słów kluczowych mają przeważnie strukturę fasetową, gdyż fasetowość stała się jedną z istotnych cech większości współczesnych języków informacyjno-wyszukiwawczych, w tym języków słów kluczowych. Ten rodzaj kategoryzacji semantycznej na poziomie metajęzykowym ułatwia odnalezienie w słowniku systemu języka adekwatnych do potrzeb słów kluczowych.

3.1.3.1.4. Słowniki słów kluczowych w układzie gniazdowym

Nowym rozwiązaniem niedawno zastosowanym w słownikach słów kluczowych, jest układ gniazdowy słów kluczowych wspomagany systemem terminologicznym. Ten sposób prezentacji języka słów kluczowych dostarcza jeszcze więcej informacji o jego strukturze semantycznej. Zastosowano go w języku słów kluczowych etnologii. Język ten powstał, a właściwie powstaje, od roku 1993 w ramach prac zespołu kierowanego przez prof. Czesława Robotyckiego z Instytutu Etnologii i Antropologii Kulturowej Uniwersytetu Jagiellońskiego nad porządkiem terminologii tej dyscypliny. Uznano, że terminologia nie tylko ułatwia dostęp do wiedzy, ale jest też źródłem słów kluczowych, umożliwiających odwzorowanie treści dokumentów (źródła etnograficzne) na potrzeby systemu informacyjno-wyszukiwawczego, dlatego zdecydowano się połączyć te prace z pracami nad słownikiem, a właściwie słownikami słów kluczowych dla poszczególnych kategorii kultury, które to słowniki mają być wykorzystane w budowanym przez ten Instytut systemie informacji o źródłach

archiwalnych dotyczących polskich Karpát o nazwie PROKES. Wcześniej zespół ten opublikował w 1995 roku rezultaty swoich prac jako *Układ słów kluczowych do bazy danych o źródłach etnograficznych. Kultura ludowa Karpát Polskich* (Pod red. Cz. Robotyckiego. „Zeszyty Naukowe UJ”, seria Varia, t. CCCXLIII, Kraków 1995). Dalsze prace zespołu, rozszerzonego o autora tej rozprawy, doprowadziły do opublikowania pod wspólnym tytułem dwóch tomów *Układu gniazdowego terminów i słownika słów kluczowych wybranych kategorii kultury*, zawierających opracowane według jednolitej metodologii kolejne części słownika (słowników) słów kluczowych z tej dziedziny. Pierwszy tom został poświęcony następującym kategoriom kultury: etos, obrzędy, demonologia, magia, zaś drugi medycynie ludowej. Jak łatwo zauważyć, słownik ten jako całość z założenia ma mieć strukturę fasetową.

Wydawnictwo składa się z dwóch elementów: układu gniazdowego terminów, stanowiącego spis terminów nauki o kulturze w formie systematycznego wykazu terminów, lokalizującego je w odpowiednich polach wyznaczających ich znaczenie oraz użycie, a także alfabetycznego wykazu słów kluczowych (nazwanego słownikiem słów kluczowych), służącego do odwzorowywania i wyszukiwania dokumentów, który jest pochodną gniazdowego układu terminów, uboższą o wskazanie explicite łączących je relacji. Integracja obu tych elementów była podyktowana szczególną sytuacją terminologii w tej dziedzinie. Dzięki takiemu rozwiązaniu użytkownik może przyporządkowywać treściom w strukturze wiedzy (układ gniazdowy) wiedzę prezentowaną fragmentarycznie w dokumentach.

Listom słów kluczowych towarzyszą spisy terminów w układzie gniazdowym. Pokazują one explicite kategoryzację semantyczną jednostek leksykalnych, a także elementy struktury paradygmatycznej tego języka. Są to w zasadzie pola asocjacyjne łączące terminy podstawowe w etnologii, nazywające główne przedmioty badań z wyrażeniami kojarzącymi się z nimi na zasadzie bliżej nieokreślonych i niespecyfikowanych w słowniku relacji paradygmatycznych. Tego typu układy są jednak bardzo przydatne dla użytkownika. Do *Układu gniazdowego...* dołączono fasety w postaci wykazów świąt, świętych oraz kościołów (wyznań) i roślin leczniczych. Tak prezentowany język nie jest już klasycznym językiem słów kluczowych, którego charakterystyczną cechą stanowi z założenia płaska struktura leksyki, a jego paradygmatyka tkwi implicite w kompetencji jego użytkownika w zakresie odpowiedniego języka naturalnego. Ma to niewątpliwy wpływ na proces indeksowania. Nieukazana explicite paradygmatyka nie może być wykorzystana w wyszukiwaniu.

Drugi tom słownika został przygotowany według tych samych zasad¹¹³. Terminologia została zaprezentowana w układzie gniazdowym, co pozwala lokalizować poszczególne terminy w odpowiednim polu wyznaczającym ich znaczenie i użycie. Do wyrażania związków między terminami użyto kropek i wcięć, które z zasady nie odwzorowują relacji hierarchicznych, lecz, zależnie od typu gniazda terminów, pokazują jakiś rodzaj zależności. Nazwy szerokich kategorii oznaczono wytłuszczonym drukiem, na przykład **WIEK DZIECIĘCY**. W tej części wprowadzono też wyrażenia zaznaczone kursywą, które nie są terminami, a których zadaniem jest jedynie

¹¹³ Por. *Wstęp do Układ gniazdowy...*, s. 7–14.

porządkowanie terminologii, na przykład *CHOROBY KOBIECE*. Obok terminów hasłowych, po przecinku, podano kursywą ich synonimy, na przykład *BÓL BRZUCHA*, Morzysko, Rzezanie, Rznawka. W nawiasach kwadratowych umieszczono objaśnienia, na przykład *MORZYSKO* [silne bóle brzucha, zapalenie kiszek łączone też z zapaleniem wyrostka robaczkowego], określenia pomocnicze, na przykład *ROBAKI* [robaczyca] lub terminy szczegółowe, na przykład *TRUDNOŚCI W ODDYCHANIU* [duszenie]. W wykazie nazw roślin wprowadzono na oznaczenie relacji odpowiedniości odsyłacze „zob.”, na przykład *ANYŻ* zob. biedrzyca, a na oznaczenie relacji kojarzeniowych odsyłacz „zob. też”, na przykład *JEŻYNA* zob. też *BORÓWKA CZARNA*.

Jak już sygnalizowałem, z wcześniej opracowanego gniazdowego wykazu terminów wygenerowano słownik języka słów kluczowych. Słownik słów kluczowych stanowi jeden z elementów sztucznego języka słów kluczowych dla etnologii. Ma on pełnić funkcję użytkową, służąc do opisu i wyszukiwania zgromadzonych we wspomnianym instytucie materiałów etnograficznych. Przyjęto alfabetyczny układ słów kluczowych. Są nimi przede wszystkim rzeczowniki, lecz pojawiają się również przymiotniki. Przymiotniki uznano za słowa kluczowe wtedy, gdy nie istnieją odpowiadające im rzeczowniki, na przykład *CIEPŁY*. Słowa kluczowe wyróżniono dużymi literami. Terminy niebędące słowami kluczowymi podano małymi literami. „Dziwna”, jak na tego typu jednostki leksykalne, postać niektórych słów kluczowych uwarunkowana jest specyfiką terminologii reprezentowanej dziedziny, na przykład *WÓDKA TRZYMANA W USTACH TRUPA*. Jest to rodzaj frazy kluczowej. Wielu słowom kluczowym towarzyszą uzupełnienia i dopowiedzenia pełniące różne funkcje. Specyfika tego typu dookreśleń wynika przede wszystkim z „ludowego” źródła terminologii etnologicznej. Umieszczone w nawiasach okrągłych wyrażenia stanowią integralną część słowa kluczowego i wraz z nim powinny znaleźć się w charakterystykach wyszukiwawczych, na przykład *KOLKA* (objaw). Niektórym słowom kluczowym towarzyszą umieszczone w nawiasach kwadratowych definicje i wyjaśnienia precyzujące zakres ich znaczenia, na przykład *BŁAŻEJKA* [świeczka] i/lub łączliwości, na przykład *ZAPALENIE WÓDKI* [pijaństwo], określenia współczesne, na przykład *PRZESILENIE* [podźwignięcie], dookreślenia informujące o zawartości klasy, na przykład *GWOŹDZIK* [przyprawa]. Terminy o tym samym zakresie połączono odsyłaczem „zob.”, kierującym do jednostki leksykalnej przyjętej w budowanym języku – na przykład: chybnięcie zob. *PRZEŁAMANIE* (dotyczy to również pospolitych nazw chorób, na przykład: ból zob. *KOŁTUN*) lub, gdy wyrażenie ludowe było nazwą różnych chorób, do różnych jednostek leksykalnych – na przykład: gościec zob. *KOŁTUN*, *REUMATYZM*. Odsyłacz ten jest wykładnikiem stosowanej w języku tzw. preferencji leksykalnej.

Omawiany język jest przykładem takiej podwójnej prezentacji systemu leksykalnego języka słów kluczowych. Zilustruję to fragmentem układu gniazdowego terminów wybranych kategorii kultury (*Układ gniazdowy terminów i słownik słów kluczowych wybranych kategorii kultury. Medycyna ludowa*. Pod red. Cz. Robotyckiego, W. Babika. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego 2005, s. 30–32.):

(...)

SUCHOTY, *astma, dera, gruźlica, nędza*

.. *ETIOLOGIA*

- ... BAWIENIE SIĘ Z KOTEM
- ... CZAR
- ... DZIEDZICZENIE
- ... POŁKNIĘCIE SIERŚCI
- ... POŁKNIĘCIE WŁOSA
- ... POŁKNIĘCIE KOCIEGO WŁOSA
- ... PRZEZIĘBIENIE
- ... TĘSKNOTA ZA PIERSIĄ

.. *RODZAJE*

- ... SUCHOTY GALOPOWE, *suchoty ostre, suchoty śmiertelne*
- ... SUCHOTY GARDLANE
- ... SUCHOTY PŁUCNE, *suchoty wewnętrzne*
- ... SUCHOTY ŚWIATOWE
- ... SUCHOTY ŻOŁĄDKOWE

.. *OBJAWY*

- ... BEZSENNOŚĆ
- ... BÓL W PIERSIACH
- ... BRAK APETYTU
- ... CHUDNIĘCIE, *schnięcie*
- ... GORĄCZKA
- ... KASZEL

.. *LEKI I SPOSOBY LECZENIA*

- ... KĄPIEL
 - GROCHOWINY
 - KAWAŁKI PNIA DO RAĄBANIA MIĘSA
 - KLUSKI PSZENNE
 - ODWAR Z DZIEWANNY
 - ODWAR Z GAŁĘZI DĘBU
 - ODWAR Z PODRÓŻNIKA
 - OSKROBINY [z ziemniaków]
 - POMYJE
 - Z KOTEM
 - Z PSEM
 - ZIEMIA Z CMENTARZA
- ... PICIE
 - ODWAR Z BAGNA
 - ODWAR Z KORZENIA ŻYWOKOSTU
 - ODWAR Z LIŚCI PODBIAŁU
 - ODWAR Z PĘDÓW SOSNY
- ... SMAROWANIE, NACIERANIE
 - ŁÓJ ZE ŚWIECY
 - TŁUSZCZ, SADŁO RAKA
 - TŁUSZCZ, SADŁO ZAJĘCZE
- ... WKŁADANIE CHOREGO DO PIECA
- ... WKŁADANIE CZASZKI KONIA DO ŁÓŻKA

... ZAMAWIANIE
... ZAŻEGNYWANIE

.. CZAS LECZENIA
.. MIEJSCE LECZENIA
.. OSOBA LECZĄCA

SZKARLATYNA, ból gardła, dyfteryt, odra, płonica, żaba

.. ETIOLOGIA
... PRZEZIĘBIENIE
... ZŁAMANIE ZAKAZU

.. OBJAWY
... BÓL GARDŁA
... KROSTY
... NAROŚLA POD JĘZYKIEM
... ZACZERWIENIENIE SKÓRY

.. LEKI I SPOSOBY LECZENIA
... OKŁADY
... SIEMIEŃ LNIAŃ
... ŻABA
... PĘDZŁOWANIE GARDŁA
... MIÓD
... NAFTA
... SPIRYTUS
... PICIE
... MIÓD
... ODWAR Z KWIATU LIPY
... POCIERANIE
... CZERWONE SUKNO
... USUSZONA ŻABA
... SÓL
... SMAROWANIE, NACIERANIE ATRAMENTEM
... ZAMAWIANIE
... ZAŻEGNYWANIE

.. CZAS LECZENIA
.. MIEJSCE LECZENIA
.. OSOBA LECZĄCA
(...)

Przytoczony tu układ hierarchiczny ujawnia strukturę paradygmatyczną języka, która różni się od relacji paradygmatycznych w języku naturalnym (terminologii), zachodzących między równokształtnymi ze słowami kluczowymi terminami języka naturalnego. Oto fragment leksyki omawianego języka słów kluczowych¹¹⁴ (s. 154–155):

¹¹⁴ W. Babik, E. Duszeńko-Król, M. Golonka-Czajkowska, Z. Libera, M. Maj, Cz. Robotycki, S. Trebunia-Staszek: *Słownik słów kluczowych*. [W:] *Układ gniazdowy terminów i słownik słów kluczowych*

(...)
SUCHOTY
SUCHOTY GALOPOWE
SUCHOTY GARDLANE
suchoty ostre zob. SUCHOTY GALOPOWE
SUCHOTY PŁUCNE
suchoty śmiertelne zob. SUCHOTY GALOPOWE
SUCHOTY ŚWIATOWE
suchoty wewnętrzne zob. SUCHOTY PŁUCNE
SUCHOTY ŻOŁĄDKOWE
suchoty zob. ASTMA
SUCHY
suchy ból zob. REUMATYZM
SUKNO
SUPEŁKI
SUROWY
SUSZONY
SWĘDZENIE
SYROP
SZALEJ
SZAŁWIA
SZARE MYDŁO
szarpanie zob. SKURCZE MIĘŚNI
SZCZAW
SZCZEBLE DRABINY
SZCZUPIENIE
SZCZOTKA DO CZESANIA LNU
SZCZUPAK
SZCZYPY SMOLNE
SZELEŻNIK
SZKARLATYNA
szkarlatyna zob. ODRA
SZKORBUT
SZMATA
SZMATKA PO OBMYSIUCIU ZWŁOK
szmatka zob. SZMATA
SZNUREK
SZPIŁKA
SZTUCZNE ODDYCHANIE
SZYJA
SZYSZKI
SZYSZKOJAGODY
ŚCINANIE
ŚCINANIE BRZYTWĄ [gest symboliczny]
ŚCINANIE SIERPEM [gest symboliczny]
ŚCISKANIE
ŚCISKANIE PALCA

wych wybranych kategorii kultury. Medycyna ludowa. Pod red. Cz. Robotyckiego i W. Babika. Kraków: Wydawnictwo UJ 2005, s. 137–159.

ślad zob. ZNAMIE
ŚLAZ
ŚLEDŹ
ŚLEPNIĘCIE Z KURAMI
ŚLINA
ŚLINIENIE SIĘ
ŚLINOTOK
ŚLIWKA
ŚMIETANA
ŚNIEG
ŚPIĄCZKA
śpiączka zob. CHOLERA
ŚWIECA
ŚWIECA Z OŁTARZA
ŚWIERK
ŚWIERZB
świerzbiczka zob. ŚWIERZB
ŚWIERZBIENIE
ŚWIETLIK
ŚWIEŻY
ŚWIĘCONY
ŚWINIA
świniaki zob. WRZODY
TABAKA
TALERZ
tarczycza zob. WOLE
TARTY
TASIEMKA
TATARAK
TATARKA
TERPENTYNA
TĘSKNOTA ZA PIERSIĄ
tłuczenie zob. KONWULSJE, PADACZKA
TŁUCZONY
TŁUSTE JEDZENIE
TŁUSTY
(...)

W słowniku słów kluczowych – jak można zauważyć – między terminami wprowadza się mniej relacji znaczeniowych. Na ogół nie tworzy się struktur hierarchicznych terminów, poprzestając na wyszczególnieniu synonimów.

Wszystkie uwagi poczynione w stosunku do poprzedniej grupy omawianych słowników należy odnieść również do tej grupy. Można dostrzec wyraźne dążenie do rozszerzenia zakresu i repertuaru niespecyfikowanych relacji paradygmatycznych, co wynika przede wszystkim z szerszego niż dotychczas uwzględnienia relacji asocjacyjnych. Warto też zauważyć, że w językach słów kluczowych opartych na jednojęzycznych słownikach brakuje słowników słów kluczowych charakteryzujących się całkowicie jawnym sposobem przejścia z języka naturalnego na język słów kluczowych (i odwrotnie). Ten sposób przejścia występuje w słownikach o słow-

nictwie sztucznym, a w przypadku języków słów kluczowych w słownikach dwu- i wielojęzycznych.

Słowniki słów kluczowych o strukturze gniazdowej są bliskie internetowym ontologiom, które stają się elementem tzw. semantycznego Webu. Ontologie stanowią rodzaj rozbudowanej sieci semantycznej reprezentującej pojęciową strukturę wiedzy zawartej w zasobach Internetu. W węzłach tej sieci są umieszczane różnojęzyczne wyrażenia (nazwy i terminy). Węzły są wiązane za pomocą odpowiednich relacji w grupy kategorialne i gniazda semantyczne. W ten sposób słowa kluczowe stają się lingwistycznym narzędziem filtrowania informacji bez udziału człowieka.

Twierdzę, że w zaprezentowanych słownikach nieklasycznych języków słów kluczowych z założenia jest prezentowana struktura paradygmatyczna języka słów kluczowych. Znaczenie słów kluczowych rozumiane jest jako relacja wewnątrzsystemowa danego języka słów kluczowych, zachodząca pomiędzy jednostkami tego systemu językowego, choć w części bazująca na kompetencji jego użytkowników w zakresie języka naturalnego, dlatego przy korzystaniu z tych słowników trzeba posługiwać się nie tylko informacją o znaczeniu, wynikającą z kompetencji językowej użytkownika konkretnego języka naturalnego, ponieważ ta jest systemowo modyfikowana. Języki tego typu odwołują się przede wszystkim do wiedzy językowej i pozajęzykowej oraz kompetencji użytkownika w zakresie danego języka słów kluczowych, którego leksykę reprezentują, co stanowi ważny element tożsamości tych języków.

Słowniki nieklasycznych języków słów kluczowych charakteryzują się jawnym przejściem z języka naturalnego na język słów kluczowych, gdyż *explicite* występują w nich wyrażenia języka naturalnego w funkcji metainformacyjnej, powiązane ze słowami kluczowymi relacją odpowiedniości wyszukiwawczej, co jest charakterystyczne dla tej odmiany języków słów kluczowych.

3.1.3.1.5. Wielojęzyczne słowniki słów kluczowych

Osobną grupę problemów językowych języka słów kluczowych generują wielojęzyczne słowniki słów kluczowych. W Polsce pierwszym i jedynym ośrodkiem, który budował tego typu słowniki słów kluczowych, było Centrum Informacji Naukowej i Biblioteka Szkoły Zdrowia Publicznego Akademii Medycznej w Krakowie. W ośrodku tym na początku lat dziewięćdziesiątych XX wieku zbudowano następujące wielojęzyczne słowniki słów kluczowych: *Słownik słów kluczowych zdrowia publicznego (francusko-polski)* (1993), *Słownik słów kluczowych zdrowia publicznego (polsko-francuski)* (1993), *Słownik słów kluczowych zdrowia publicznego (polsko-niemiecki)* (1994). Były one wynikiem współpracy tego ośrodka z *Ecole Nationale de la Santé Publique* we Francji. Słownikiem bazowym i wzorcowym był *Index des mots-clés de la base de données de l'ENSP au 8 février 1990*. Słowniki te zaliczyłem do grupy słowników słów kluczowych z systemami odesłań (zob. rozdział 1.3.1.2), w których uwzględniono „podstawowe” relacje słownikowe. Nie zostały one opublikowane, gdyż stanowiły materiał roboczy w budowie wielojęzycznego tezauryusa zdrowia publicznego. Niestety taki tezaurus nie został zbudowany, a sam ośrodek

zaprzestał działalności w momencie włączenia biblioteki do systemu bibliotecznego VTLS, obecnie VIRTUA.

Za interesujące i godne zwrócenia uwagi uważam rozwiązanie zastosowane w systemie Slavistik-Portal (<http://www.slavistik-portal.de/>). Do indeksowania dokumentów używa się w nim jednocześnie kilku języków. Podstawę do tworzenia zasobu leksykalnego języka słów kluczowych stanowi język niemiecki. System udostępnia interfejs użytkownika w językach niemieckim, angielskim i rosyjskim. Charakterystyka wyszukiwawcza dokumentów jest tworzona w języku niemieckim. Oprócz tego wykorzystuje się język rosyjski do tworzenia słów kluczowych w formie komplementarnej charakterystyki treściowej, co jest podyktowane zakresem systemu, który obejmuje szeroko rozumianą slawistykę.

Zasygnalizowane przedsięwzięcia budowy wielojęzycznych słowników słów kluczowych stanowią tylko wybrane przykłady narzędzi wielojęzycznego dostępu i wyszukiwania informacji. W tego typu narzędziach nie tłumaczy się wyrażenń pobieranych z poszczególnych języków na wyrażenia pozostałych języków, lecz łączy się je relacjami równoważności. Stosowana tu technika tzw. mapowania (ang. *mapping*), to jest odwzorowywania jednego systemu w drugim, zapewnia kompatybilność i współdziałanie różnych systemów. Takie systemy zapewniają równoległe przeszukiwanie zasobów za pomocą języków słów kluczowych opartych na leksyce kilku różnych języków etnicznych.

Opisane rozwiązania stanowią przykład narzędzi spójnego i wielojęzycznego przeszukiwania zasobów informacyjnych. Nurt ten obejmuje zarówno prace nad multilingwistycznymi językami katalogów OPAC i systemami informacji bibliograficznej wykorzystującymi słownictwo paranaturalne, jak też prace nad nowoczesnymi narzędziami semantycznej organizacji zasobów Internetu.

W podsumowaniu rozważań o jawnych listach/słownikach słów kluczowych zwracam uwagę na ich wartość jako słownictwa ujednoliconego i kontrolowanego, które można wykorzystać w Internecie, zwłaszcza w odpowiednich polach metadanych, o czym już była mowa w pierwszej części książki.

Wprawdzie większość omówionych języków słów kluczowych stanowi przykład odejścia od modelu klasycznego języka słów kluczowych, ale mogą okazać się bardzo przydatne ze względu na już gotowe zasoby ujednoliconego słownictwa, które to zasoby w coraz większym stopniu wykorzystuje się w najnowszych systemach wyszukiwania informacji (Hartley 2009).

Kończąc rozważania nad grupą jawnych słowników słów kluczowych, sygnalizuję, że obecnie planuje się prace nad kolejnymi słownikami słów kluczowych z dziedziny etnologii (Uniwersytet Jagielloński).

3.1.3.1.6. Słowa kluczowe w indeksach rzeczowych, systemach pełnotekstowych, katalogach bibliotecznych (OPAC) i bibliotekach cyfrowych

Języki słów kluczowych znajdują zastosowanie również w tworzeniu nowych typów indeksów nazywanych niekonwencjonalnymi, permutacyjnymi lub też indeksami słów kluczowych bądź tytułowymi. Historycznie pierwszym systemem opartym na koncepcji uwzględniającej kontekst słów kluczowych był system KWIC (Key-Word-in-Context), nazywany też indeksem permutacyjnym, opracowany pod koniec lat pięćdziesiątych XX wieku przez H.P. Luhna (Luhn 1960). W połowie lat siedemdziesiątych XX wieku koncepcja ta została rozwinięta w brytyjskim systemie PRECIS (PREserved Context Indexing System), opracowanym dla British National Bibliography przez Dereka Austina, oraz hinduskim systemie POPSI (Postulate-based Permuted Subject Indexing).

Języki słów kluczowych stosowane w indeksach KWIC i KWOC uwzględniają zasady permutacji, czyli eksponowania kolejnych elementów fraz leksykalnych. Słowa kluczowe są prezentowane w kontekście (na przykład w indeksie KWIC), bez kontekstu (na przykład w indeksie KWOC) i/lub w tytule (na przykład w indeksie KWIT). Przykład może stanowić baza danych GOSPODARKA, tworzona przez Bibliotekę Główną Uniwersytetu Ekonomicznego w Krakowie, która utrzymuje bazę wzorcowych słów kluczowych w języku polskim i języku angielskim, zwaną e-słownikiem słów kluczowych, zintegrowaną ze zbiorem głównym bazy GOSPODARKA. Do tej pory jest to słownik wyłącznie języka słów kluczowych w funkcji języka indeksowania. Fragment tego słownika w formie indeksu KWIC przedstawia następująca lista:

Informacja marketingowa	26
Informacja masowa	2
Informacja poufna	4
Informacja przestrzenna	2
Informacja publiczna	18
Informacja rynkowa	29
Informacja statystyczna	25
Informacja techniczna	2
Informacja w podejmowaniu decyzji	24
Informacja w przedsiębiorstwie	40
Informacyjna funkcja sprawozdań finansowych	3
Jakość informacji	5
Jawność informacji	5
Komputerowy system informacyjny	4
Logistyczne systemy informacyjne	4
Manipulacje informacją	4
Marketingowy system informacyjny	8
Ochrona informacji niejawnych	9
Polityka informacyjna	19
Potrzeby informacyjne	11

Prawo do informacji	13
Proces informacyjny	7
Promocja jako przekaz informacyjny	4
Przepływ informacji między organizacjami	3
Przepływ informacji w organizacji	14
Przepływ informacji w łańcuchu dostaw	5
Rynek informacji finansowej	2
Społeczeństwo informacyjne	66
System informacji gospodarczej	8
Baza GOSPODARKA:	8
System Informacji Przestrzennej = <i>Geographic Information System (GIS)</i>	
System informacji statystycznej	6
System informacyjny	72
System informacyjny w przedsiębiorstwie	54
System informacyjny w zarządzaniu	12
Baza GOSPODARKA:	281
Technologia informacyjna = <i>Information Technology (IT)</i>	

Zwracam uwagę na sposób prezentacji słów kluczowych, wykorzystujący zasadę permutacji, czyli eksponowania kolejnych elementów fraz leksykalnych. Towarzystwo im liczby oznaczają ich frekwencję. Jest to przykład systemu opartego na stosowaniu jednostek leksykalnych języka naturalnego (bez jakiegokolwiek normalizacji) w charakterze słów kluczowych do indeksowania dokumentów w postaci indeksu permutacyjnego. Przy takim zastosowaniu leksyki języka naturalnego do współzrędnego indeksowania dokumentów w charakterystyce wyszukiwawczej używane są różne formy gramatyczne jednych i tych samych słów i zestawień słownych, a także synonimy i homonimy.

Mika Käiki z University of Tampere (Finlandia), zainspirowany indeksem KWIC zbudowanym przez Luhna, zaproponował tzw. fKWIC (frequency-based Keyword in Context Index) (Käiki 2006). Klasyczny indeks KWIC stanowi formę konkordancji, tj. indeksu słów, w którym każde ze słów jest wyświetlane wraz z jego kontekstem. Słowa kluczowe użyte w wyszukiwaniu stanowią w nim bazę dla indeksu, podczas gdy pozostałe słowa z tytułu stanowią dla niego kontekst. Istotą nowej propozycji przedstawionej przez Käkiego jest generowanie indeksu fKWIC za pomocą algorytmu o charakterze statystycznym.

Zajmując się implementacją języka słów kluczowych, trzeba uwzględnić również wprowadzone w latach sześćdziesiątych XX wieku katalogi OPAC i ich dynamiczny rozwój w latach osiemdziesiątych i dziewięćdziesiątych ubiegłego stulecia. Niewątpliwie jest to, oprócz systemu WWW, nowy obszar zastosowań języków słów kluczowych¹¹⁵. Katalogi biblioteczne tzw. pierwszej generacji funkcjonowały na zasadzie prostych prekoordynowanych indeksów, generujących frazy kluczowe wybrane ze struktury opisu bibliograficznego, głównie z tytułów dokumentów. Katalogi

¹¹⁵ Już na początku lat osiemdziesiątych XX wieku w USA zauważono w katalogach OPAC tendencję spadkową wyszukiwania przedmiotowego na rzecz słów kluczowych.

następnych generacji znacznie rozszerzyły swoje możliwości. I tak, druga generacja umożliwiła wyszukiwanie według słów kluczowych, zaś kolejna generacja także wyszukiwanie według słów kluczowych w obrębie hasła przedmiotowego, a następnie rozszerzenie wyszukiwania według słów kluczowych na pozostałe wybrane elementy opisu, na przykład tytuły dokumentów, nazwy serii, adnotacje zawartościowe (spisy treści), abstrakty (głównie streszczenia autorskie), a nawet całe teksty dokumentów, czy tzw. odpowiedniki słowne symboli klasyfikacyjnych. Liczba pól przeszukiwanych metodą słów kluczowych jest różna w różnych bibliotekach.

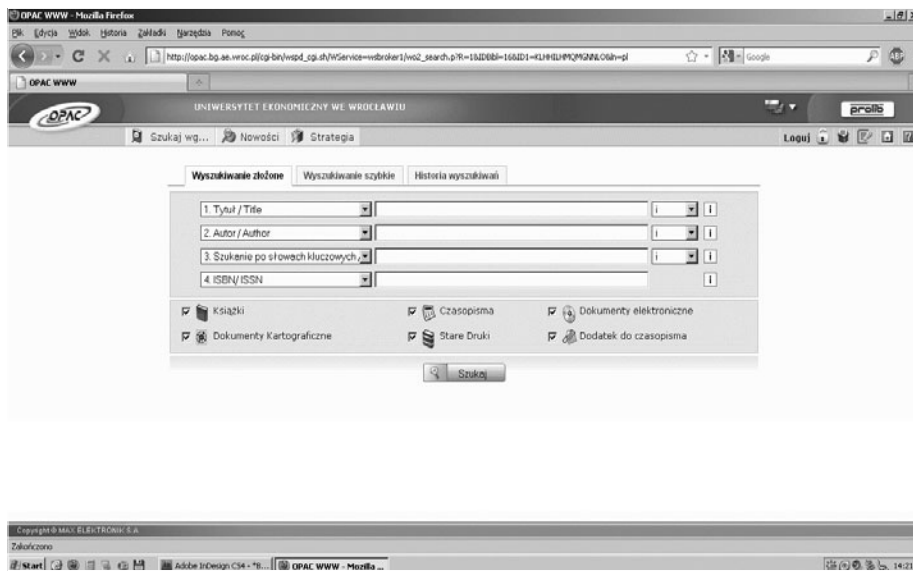
W tradycyjnym systemie katalogów kartkowych większość bibliotek prowadziła katalog systematyczny lub działowy. Po ich skomputeryzowaniu najczęściej stosowane są hasła przedmiotowe i słowa kluczowe. Duża część bibliotek tworzy kilka uzupełniających się charakterystyk rzeczowych, wykorzystując jednocześnie różne języki informacyjno-wyszukiwawcze. Mamy wtedy do czynienia z polireprezentacją informacji¹¹⁶. Związana z tym jest, widoczna w ostatnich latach, tendencja do uszczegółowienia haseł przedmiotowych, na przykład tworzonych w Bibliotece Narodowej.

Komputeryzacja bibliotek wymusza ujednoczenie metod opracowania zbiorów, w tym konieczność ujednoczenia haseł formalnych i określenia języka informacyjno-wyszukiwawczego stosowanego w opisie dokumentów. W tym bardziej przydatne okazują się nowsze generacje katalogów OPAC, czyli tzw. OKAPI (Online Keyword Access to Public Catalog), które wykorzystują ankietową metodę indeksowania, oferującą stałą długość tekstu opisu dokumentu. Tu pytanie użytkownika jest poddawane analizie i pewnej obróbce przez system. Ten sposób indeksowania za pomocą języka słów kluczowych może być użyteczny również w innych typach systemów informacyjno-wyszukiwawczych.

W Polsce o językach stosowanych w komputerowych katalogach bibliotecznych pisała Jadwiga Sadowska (Sadowska 2001). Jej zdaniem, wyszukiwanie za pomocą języka słów kluczowych jest obecnie najbardziej znanym sposobem wyszukiwania w bibliotecznych katalogach komputerowych. Trzeba jednak zauważyć, że słowem kluczowym jest tu powszechnie nazywane dowolne słowo wybrane z charakterystyki formalnej lub rzeczowej dokumentu. Podjąłem się weryfikacji tej tezy.

Okazuje się, że oprócz możliwości wybierania słów z tytułu dokumentu lub z poszczególnych pól opisu bibliograficznego, niewiele bibliotek w swoich OPAC-ach oferuje możliwość wyszukiwania według słów kluczowych. Przykładem takiej biblioteki może być Biblioteka Główna Uniwersytetu Ekonomicznego we Wrocławiu (rys. 7).

¹¹⁶ Istotę polireprezentacji informacji stanowi intencjonalna redundacja opisu, to jest zasada wielokrotnego, równoległego opisywania różnymi językami tych samych obiektów i ich atrybutów. Por. też Larsen, Ingwersen 2005.



Rys. 7. Język słów kluczowych w katalogu OPAC (źródło: <http://opac.bg.ae.wroc.pl>)

Niewielki zakres stosowania języków słów kluczowych w OPAC-ach potwierdzają badania Zofii Pokusińskiej i Teresy Dębickiej, opublikowane w 2004 roku¹¹⁷. Stosowanie języka słów kluczowych jako doraźnego rozwiązania zalecano w bibliotekach kościelnych zrzeszonych w Federacji FIDES. W tym celu nawet opracowano *Słownik słów kluczowych z teologii* (Klimek, oprac. 2001). Jednak większość bibliotek, po początkowej euforii związanej z możliwością wyszukiwania w OPAC-u za pomocą słów kluczowych, zrezygnowała z tego pomysłu, gdyż wymagało to dodatkowych prac nad indeksowaniem książek za pomocą słów kluczowych. Prace nad językami słów kluczowych ujawniły, że są to zagadnienia wymagające pracy całych zespołów ludzkich i współpracy wielu instytucji. Porzucono więc tę możliwość, uzupełniając wyszukiwanie za pomocą haseł przedmiotowych nieujednoliconymi słowami generowanymi automatycznie z poszczególnych pól opisu bibliograficznego/katalogowego, głównie z pola „tytuł dokumentu”, a także z pola „opis przedmiotowy”. Takie indeksy nieujednoliconych słów – nazywane indeksami słów kluczowych – są tworzone automatycznie, m.in. w OPAC-u Biblioteki Jagiellońskiej w Krakowie i innych bibliotekach posługujących się systemem bibliotecznym VIRTUA.

W katalogach OPAC, w których stosuje się obecnie język słów kluczowych, funkcjonują dwie odmiany funkcjonalne tego języka. W języku indeksowania jest on znormalizowany, a jego słownik jest jawny, zwykle w postaci tzw. kartoteki wzorcowej, zaś w języku wyszukiwawczym opiera się na słowniku niejawnym (indeksowanie swobodne) lub jawnym, z wykorzystaniem wspomnianej kartoteki

¹¹⁷ Wprawdzie badania, o których tu mowa, dotyczyły stosowania UKD w katalogach komputerowych bibliotek akademickich, ale przy okazji autorki zebrały i ujawniły informacje o innych językach stosowanych w tych katalogach.

wzorcowej. Użytkownicy tych katalogów mają więc do dyspozycji dwa rodzaje słowników: słowniki *explicite* (jawne) w postaci indeksów, *de facto* list jednostek leksykalnych do wyboru użytkownika, oraz słowniki *implicite* (niejawne) w indeksowaniu swobodnym.

Języki słów kluczowych są stosowane zarówno w indeksowaniu, jak i wyszukiwaniu informacji także w systemach pełnotekstowych i bibliotekach cyfrowych, współpracujących z systemami automatycznego indeksowania, bazujących przede wszystkim na wyszukiwaniu pełnotekstowym. Podstawą automatycznego indeksowania są odpowiednie pola metadanych dokumentu elektronicznego, które są przeznaczone do zapisu informacji o treści dokumentu w postaci zestawu słów kluczowych lub wyrażeń języków o słownictwie kontrolowanym albo cały tekst dokumentu.

W piśmiennictwie naukowym problematyce tej poświęcono już sporo prac. Ustalono, że rozmieszczenie informacji w artykułach jest zróżnicowane. Wprawdzie abstrakty są najbardziej nasycone „słowami kluczowymi”, ale pełny tekst – chociaż ma zdecydowanie mniejsze nagromadzenie słów kluczowych – zawiera zdecydowanie więcej informacji relewantnych dla użytkownika, które zazwyczaj muszą być odczytywane z kontekstu. Obecnie dominuje pełnotekstowe wyszukiwanie informacji za pomocą słów kluczowych. Jego podstawy są rozwijane w ramach programu Text REtrieval Conference (TREC). Prowadzone w tym zakresie badania są od 1992 roku stymulowane przez serię dorocznych konferencji roboczych znanych pod nazwą TREC Workshops¹¹⁸.

Słowa kluczowe są też wykorzystywane do opisu i wyszukiwania informacji graficznej, na przykład za ich pomocą opisuje się i wyszukuje różne kategorie zdjęć opisanych słowami kluczowymi w formie etykiet zawierających takie nazwy, jak na przykład: rodzina, sport, szkoła. Wiele systemów gromadzących tego rodzaju informacje (na przykład iPhoto2) posiada cechę wyszukiwania na podstawie słów kluczowych lub dowolnego słowa, czy wyrażenia znajdującego się w komentarzu (por. <http://www.apple.com.pl/products/iphoto/organize.html>).

Koncepcja słów kluczowych stała się na tyle nośna, że podjęto próbę jej zastosowania w analizie i odkrywaniu zawartości dźwiękowych dokumentów audialnych, to jest przeniesienia jej na kod wokalny. Propozycję taką przedstawili Lie Lu i Alan Hanjalic (Lu, Hanjalic 2006). Autorzy ci zaproponowali zastosowanie do sprawnego i dokładnego opisywania i katalogowania takich dokumentów swoistych „dźwięków kluczowych”, charakterystycznych dla danego dokumentu. W początkowych badaniach nad wykrywaniem dźwięków kluczowych wykorzystywano nadzorowaną analizę dokumentów i metody klasyfikacyjne, takie jak ukryte modele Markowa (ang. Hidden Markov Models – HMM), które pozwoliły na znalezienie dziesięciu kluczowych elementów dźwiękowych. Podobnie działa Maszyna Wektorów Nośnych (ang. Suport Vector Machines), która wykrywa w grach sportowych takie dźwięki kluczowe, jak uderzenie piłki i gwizdy czy też syreny i strzały z pistoletu dla celów

¹¹⁸ Ten cykl konferencji jest organizowany przez instytucje amerykańskie: National Institute of Standards and Technology (NIST) oraz Defence Advanced Research Projects Agency (DARPA). Zob. też: <http://trec.nist.gov>.

indeksowania filmów. Inicjatywa ta pokazuje, że pesymizm A.I. Czernego (Czerny 1978) w kwestii wykorzystania w języku informacyjno-wyszukiwawczym innego kodu niż kod graficzny był nieuzasadniony.

3.1.3.2. Klasyczny język słów kluczowych a słowa kluczowe w systemie WWW

Wykorzystanie języka słów kluczowych w Internecie polega na generowaniu za pomocą tego języka struktur ahierarchicznych, w których słowa kluczowe stanowią źródło leksyki bezpośrednio dla indeksowania obiektów informacyjnych, a pośrednio dla tworzenia nazw punktów dostępu. Język słów kluczowych zastosowany do odwzorowania treści obiektów informacyjnych w systemie wyznacza formę jego wyrażeń. Wynika to ze względów praktycznych, to jest możliwości wykorzystania jednostek leksykalnych tego języka w formie równokształtnej z wyrażeniami języka naturalnego do bezpośredniego oznaczania zakresu nazwy punktu dostępu/kłucza wyszukiwawczego. Język ten jest przydatny tylko w generowaniu narzędzi dostępu z wykorzystaniem płaskiej (ahierarchicznej) organizacji punktów dostępu. Są to zbiory jednostek leksykalnych tworzących indeks alfabetyczny, którego elementy stanowią wejścia do zbioru wyszukiwawczego. Każde z nich jest węzłem dla odesłania hipertekstowego – jego aktywacja powoduje wyświetlenie listy charakterystyk wyszukiwawczych obiektów informacyjnych, w których użyto ich do zindeksowania tych dokumentów. Ahierarchiczna budowa tego typu narzędzi charakteryzuje organizację wyrażeń, w której nie wykorzystuje się żadnej relacji hierarchicznej, a pojęcia reprezentowane przez jednostki leksykalne są traktowane jako współrzędne.

W Internecie język słów kluczowych funkcjonuje przeważnie jako klasyczny język słów kluczowych, tzn. w wersji swobodnych słów kluczowych. Pełni on funkcję zasobu leksykalnego. Tego typu indukcyjne podejście w językach informacyjno-wyszukiwawczych powoduje, że jest ono dopasowane do stopnia rozproszenia elementów treści w polu tematycznym systemu. Wadą tego typu podejścia jest mała spójność całego układu, gdyż operuje on wyłącznie słowami kluczowymi wyodrębnionymi na podstawie zasady *literary warrant*. Tego typu struktury nie spełniają wymogów stawianych podziałowi logicznemu (rozłączności i adekwatności), a jedyną relacją organizującą jest relacja współwystępowania, która dostarcza bardzo szerokich ram interpretacyjnych.

Wyszukiwanie w Internecie odbywa się przeważnie za pomocą swobodnych wyrażeń derywowanych z tekstu języka naturalnego (ang. *free-text searching*). W cyberprzestrzeni model języka słów kluczowych jako języka sztucznego, o którym wspomniałem w drugiej części rozprawy, nie bardzo przystaje do tej (nie)rzeczywistości. Tu słowa kluczowe jako narzędzie indeksująco-wyszukujące to jednostki leksykalne języka naturalnego w funkcji metainformacyjnej, przystosowane do potrzeb zautomatyzowanego wyszukiwania informacji. Cały aparat indeksowania i wyszukiwania nie należy do języka, lecz do systemu i został tu rozbudowany o wiele dodatkowych opcji, zależnych od aktualnych możliwości technologii komputerowej, a nie od sa-

mego języka. Szczególnie odnosi się to do zastosowania języka słów kluczowych jako narzędzia wyszukiwawczego, gdyż w Internecie samo indeksowanie polega raczej na „zbieraniu” słownictwa, niż na dokonywaniu zmian w jego pisowni czy łączeniu poszczególnych słów kluczowych w zdania.

Dostęp do informacji w systemie WWW za pomocą słów kluczowych polega na grupowaniu charakterystyk wyszukiwawczych w odpowiednie kategorie treści, zagnieżdżone w strukturze hipertekstu. Układ kategorii jest ahierarchiczny, uporządkowany alfabetycznie. Uruchomienie powiązania hipertekstowego dla określonego węzła w takiej strukturze powoduje wyświetlenie jedynie listy CHWD przyporządkowanych do danej kategorii, bez powiązanych kategorii treściowych, bo takie w tej strukturze nie istnieją. Istnieją tylko w strukturach hierarchicznych i hybrydowych.

Słowa kluczowe przyjmują w systemach internetowych najczęściej postać jedno- lub dwuwyrzowych jednostek leksykalnych. W przypadku wielowyrzowych nazw własnych lub nazw przedmiotów wykorzystuje się formę najbardziej rozpowszechnioną. Wykorzystanie tego języka w odwzorowywaniu treści elementów zbioru informacyjnego polega na zbudowaniu komunikatu za pomocą słów kluczowych, które pozostają względem siebie równoważne i które bezpośrednio identyfikują elementy treści dokumentów internetowych. Charakterystyczną regułą tworzenia tekstów w języku słów kluczowych jest szyk swobodny.

W Internecie wykorzystuje się, choć rzadko, również kontrolowane słowa kluczowe. Przykładem takiego języka jest język słów kluczowych International Bibliography of the Social Science (IBSS) systemu INTUTE (nauki społeczne).

Zdecydowana większość narzędzi wyszukiwawczych (wyszukiwarek) w Internecie w trybie prostym wyszukiwania informacji, oprócz możliwości wprowadzania pytania informacyjnego w postaci ciągu swobodnych słów kluczowych, nie udostępnia użytkownikowi żadnych dodatkowych opcji jego uszczegóławiania. Efektywność wyszukiwania zależy wówczas zarówno od poprawnego sformułowania pytania, jak i wiedzy użytkownika o systemie. Tylko w niewielu systemach występuje sprzężenie słownika słów kluczowych (lub słownika innego języka) z mechanizmem wyszukiwawczym, co pozwala użytkownikowi pobierać wyrażenia z takiego zasobu do budowy pytania informacyjnego. Jest to zależne od stosowanego w danym systemie interfejsu użytkownika dla modułów wyszukiwania informacji w trybie prostym i trybie zaawansowanym. Takie rozwiązanie wprowadzono na przykład w EBSCO HOST 2 i Google'ach, gdzie zapewniono użytkownikowi możliwość formułowania pytania w postaci ciągu niekontrolowanych słów kluczowych lub pobierania jednostek leksykalnych w takiej formie, w jakiej występują w słowniku stosowanego języka słów kluczowych. Zwracam uwagę na możliwość wystąpienia tu tzw. fałszywej koordynacji, która polega z jednej strony na rozdzielaniu elementów składowych terminu wielowyrzowego, zaś z drugiej strony na odszukaniu wystąpień terminów w całej strukturze CHWD, nie zaś w polach przeznaczonych do prezentacji tego rodzaju informacji.

Analiza sposobów wykorzystywania języka słów kluczowych w organizacji zasobów informacji, w tym w Internecie, pozwoliła na identyfikację sposobów wykorzystania tego typu języka do porządkowania i organizacji dostępu do dokumentów

w systemie WWW poprzez jego adaptację do środowiska hipertekstowego, nastawionego na szersze uwzględnianie relacji kojarzeniowych w sieciowej strukturze. Języki te stanowią niezbyt zróżnicowaną klasę języków informacyjno-wyszukiwawczych. Opierają się na nieco innym postrzeganiu rzeczywistości niż pozostałe typy języków, w tym języki klasyfikacyjne, które – jak już zauważono¹¹⁹ – wywodzą się jeszcze z XIX wieku i są reprezentowane przez największe i jednocześnie najczęściej stosowane klasyfikacje, jak: Klasyfikacja Dziesiąta Deweya, Uniwersalna Klasyfikacja Dziesiąta i Klasyfikacja Biblioteki Kongresu USA, a ich postrzeganie rzeczywistości nie przystaje do potrzeb współczesnego człowieka. Duża popularność języka słów kluczowych w organizacji zasobów Internetu wynika, jak już sygnalizowałem, z większej jego elastyczności, a więc łatwości w asymilacji terminologii nowych obszarów wiedzy, braku sztywnych zasad podziału logicznego oraz mniejszego stopnia skomplikowania niż w przypadku innych języków informacyjno-wyszukiwawczych.

W dziedzinowych językach słów kluczowych można dostrzec wyraźną tendencję opierania tych języków wyłącznie na systemach terminologicznych poszczególnych dziedzin, co przejawia się w traktowaniu systemu terminologicznego jako podstawy systemu leksykalnego danego języka słów kluczowych.

Języki swobodnych słów kluczowych, zwane też autorskimi lub odautorskimi słowami kluczowymi, charakteryzują się niekonsekwencją w wyborze słownictwa. Tłumaczy się to silnym dążeniem do wybierania tylko tych słów kluczowych, które są w tytułach dokumentów. Odbywa się to jednak kosztem spójności systemu leksykalnego. Sądzę, że spowodowane jest to nie ułomnością samego narzędzia, lecz nieefektywnym jego wykorzystaniem w wielu systemach i serwisach internetowych.

Ważnym problemem języków słów kluczowych w Internecie jest zagadnienie prezentacji tego języka. Obecnie technologie komputerowe pozwalają na prezentację leksyki w każdej formie, stosownie do wymagań użytkownika oraz umożliwiają stałą aktualizację danych, co przy użyciu metod konwencjonalnych było trudne do realizacji z uwagi na wzajemne uzależnienie poszczególnych danych od siebie. Słowniki słów kluczowych zawierają tylko te jednostki leksykalne, które są niezbędne do opisu i wyszukiwania informacji w systemie, dlatego występuje dążenie do ich redukcji.

3.1.4. Wnioski z przeprowadzonych analiz

Z dotychczasowych rozważań w tej części rozprawy można wysnuć wniosek, że w obecnie stosowanych językach słów kluczowych w praktyce najczęściej rezygnuje się z metodologicznego postulatu czystego języka słów kluczowych w wersji klasycznej na rzecz różnych odmian języka słów kluczowych, odbiegających od modelu klasycznego, ale za to bardziej odpowiadających potrzebom jego użytkowników. Języki takie nazwałem nieklasycznymi językami słów kluczowych.

¹¹⁹ Takie badania nad wykorzystaniem klasyfikacji w organizacji zasobów Internetu prowadzono w ramach projektu DESIRE (Development of a European Service for Information on Research and Education) pt. *Rola schematów klasyfikacyjnych w opisie i dostępie do zasobów Internetu*.

Dokonana analiza rzuciła trochę więcej światła na istniejące realizacje języka słów kluczowych. Wynika z niej, że w systemach informacyjno-wyszukiwawczych stosowane są dwa warianty języków słów kluczowych:

- język słów kluczowych z zerową normalizacją słów kluczowych języka naturalnego, a więc używanie do indeksowania dokumentów wyrażen i związków wyrazowych języka naturalnego;
- język słów kluczowych o dużym stopniu formalizacji, tj. z ujednoliceniem i normalizacją słów kluczowych języka naturalnego.

Należy zauważyć, że niektóre języki słów kluczowych zostały wzbogacone o pewne cechy innych typów języków informacyjno-wyszukiwawczych. Dotychczasowe rozważania prowadzą więc do stwierdzenia, że stosowanych w Internecie słów kluczowych nie można wrzucać do jednego worka, gdyż mogą to być zarówno wyrażenia języka naturalnego w funkcji metainformacyjnej, jak i jednostki leksykalne sztucznego języka słów kluczowych.

Różnice między poszczególnymi językami słów kluczowych wynikają przede wszystkim z:

- przyjętego sposobu wyboru wyrażen ze słownictwa języka naturalnego, co sprowadza się do wprowadzenia do jednego języka słów kluczowych wyrażen pomijanych w innym języku słów kluczowych;
- stosowania właściwych tylko dla danego języka zaleceń metodycznych określających m.in. zasady gramatyczne i pragmatyczne dotyczące budowy ze słów kluczowych zdań i tekstów (tworzenie charakterystyk i instrukcji wyszukiwawczych).

Wobec bogactwa języka naturalnego decyzja o reprezentowaniu przez daną nazwę treści dokumentów wymaga rozpatrzenia jej zakresu oraz zakresów nazw z jej bliskiego otoczenia semantycznego.

Analiza struktury wybranych słowników słów kluczowych ujawniła, że w języku słów kluczowych relacja ekwiwalencji (odpowiedniości) wyszukiwawczej jest najczęściej ustanawiana w następujących przypadkach:

- różnych form ortograficznych;
- różnych form gramatycznych;
- wyrażen archaicznych i wyrażen używanych współcześnie;
- antonimów i wyrażen komplementarnych;
- form inwersyjnych;
- wyrażen wieloelementowych, których drugi lub kolejny element może być używany przez użytkowników w wyszukiwaniu;
- synonimii i quasi-synonimii, w tym wyrażen o różnej etymologii i należących do różnych podsystemów (specjalistyczny, żargonowy, gwarowy, potoczny) języka;
- zapożyczeń;
- gdy wyrażenie reprezentuje pewną całość lub rodzaj (klasę nadrzędną), a wyrażenie języka naturalnego – część tej całości lub gatunek (klasę podrzędną);
- gdy duża część komponentów semantycznych wyrażenia języka naturalnego jest identyczna z konotacją przyjętą dla wyrażenia języka informacyjno-wyszukiwawczego.

Zasygnalizowane „możliwości” relacji ekwiwalencji wyszukiwawczej świadczą o tym, że w praktyce jest ona bardzo zróżnicowana. Próbując zdefiniować ten związek, można stwierdzić jedynie, że jest on spełniony wtedy i tylko wtedy, gdy iloczyny denotacji i iloczyny konotacji elementów relacji nie są zbiorami pustymi. Różna jest też użyteczność tych przypadków odpowiedniości. Większość z nich redukuje bogactwo języka naturalnego do potrzeb klasycznego języka słów kluczowych. Dotyczy to w szczególności aktualnie używanej leksyki.

Za pomocą podanych przykładów starałem się pokazać naturę słowników słów kluczowych i ich funkcje w zautomatyzowanych systemach wyszukiwania informacji, a przede wszystkim wagę i znaczenie specyfikowanych relacji odpowiedniości (szerzej: przejścia z języka naturalnego na język słów kluczowych) dla procesów indeksowania i wyszukiwania informacji. Analiza ujawniła stopniowe komplikowanie się struktur słowników tych języków. Wnioski z obserwacji zachowań użytkowników samodzielnie szukających informacji są tak samo ważne.

Spojrzenie na implementowane języki słów kluczowych pod kątem ich struktury i funkcji pozwala stwierdzić, że przeważnie mamy do czynienia z klasycznymi językami słów kluczowych, ale występują również struktury „wzbogacone”, tj. z rozbudowaną paradygmatyką (rozszerzenie repertuaru relacji paradygmatycznych, elementy fasetyzacji, struktura gniazdowa), co czyni te języki nieklasycznymi. Dzięki temu stają się one lingwistycznym narzędziem o znacznie większych możliwościach niż klasyczne języki słów kluczowych, a zarazem są bardziej dostosowane do potrzeb ich rzeczywistych użytkowników oraz specyfiki danej dziedziny.

Rola języków słów kluczowych w systemach hipertekstowych polega na wykorzystywaniu ich leksyki do identyfikacji i organizacji węzłów hipertekstowych pełniących funkcje wyszukiwawcze. Węzłem są najczęściej jednoelementowe wyrażenia języka naturalnego, reprezentujące wymiar treści dokumentów. Organizacja punktów dostępu za pomocą słów kluczowych przyjmuje formę prostych struktur w postaci alfabetycznych wykazów słów kluczowych. Do budowy tych struktur wykorzystuje się słowniki słów kluczowych jako źródło leksyki dla nazw kategorii tematycznych.

Podział pola tematycznego na pierwszym poziomie (w większości języków słów kluczowych jest to jedyny poziom) wyznacza tzw. elokwencję całej struktury, sposób i „filozofię” traktowania całej rzeczywistości dokumentacyjnej – w językach słów kluczowych jest to wieloaspektowe i nieostre traktowanie jej elementów także jako nakładających się na siebie. Dzięki wykorzystywaniu relacji kojarzeniowych w organizacji leksyki tych języków, powstałe w ich wyniku struktury mogą być traktowane jako sieć o niespecyfikowanych relacjach kojarzeniowych między słowami kluczowymi. Jest to bardzo użyteczne w procesach wyszukiwania informacji, gdyż pozwala rozszerzyć możliwości przeglądania całej struktury, a więc wyznacza dodatkowe możliwości wyszukiwawcze i umożliwia „przypadkowe” odkrywanie informacji (ang. *serendipity*). W zasadzie pomiędzy słowami kluczowymi występują niespecyfikowane relacje tematyczne. Mają one słabą siłę semantyczną i są tylko zakresowe. Sieć niespecyfikowanych relacji kojarzeniowych umożliwia stworzenie odpowied-

niej matrycy ścieżek dla użytkownika systemu informacyjno-wyszukiwawczego. Jest to podejście indukcyjne, na zasadzie *literary warrant*.

Poddane analizie słowniki słów kluczowych to struktury przeważnie ahierarchiczne, a zatem i równoważnościowe struktury monorelacyjne. Jako takie dają możliwość tworzenia indeksów rzeczowych, formalnych i mieszanych. Tego typu indeksy mają formę alfabetycznych wykazów jednego rodzaju wyrażen. Zwykle jest to alfabetyczny wykaz słów kluczowych. Takie indeksy zawierają również jednostki leksykalne w postaci nazw własnych: osobowe, korporatywne oraz geograficzne. Dla tej grupy słów kluczowych zwykle stosuje się odrębną metodę kontroli słownictwa, którą stanowią kartoteki wzorcowe normalizujące formy językowe, na przykład nazw języków etnicznych. Specyfika takich narzędzi kontrolnych zakłada wykorzystanie równoważnościowych struktur organizujących zbiór informacyjny. Tym samym porządek wyrażen pełniących funkcje wyszukiwawcze ma charakter formalny i wykorzystuje układ alfabetyczny. Wykorzystanie języka słów kluczowych w dostępie do zbioru polega na traktowaniu jego słownika przede wszystkim jako źródła słownictwa w procesie automatycznego wyboru nazw dla punktów dostępu. Tego typu narzędzia opierają się na generowaniu konstrukcji pola semantycznego, która pełni funkcję swoistego rodzaju mapy konceptualnej, dającej użytkownikowi możliwość wglądu w dystrybucję tematów w obiektach informacyjnych kolekcji. Konstrukcje ahierarchiczne (płaskie) są ubogie, gdyż charakterystyka dokumentu jest przyporządkowywana w nich tylko do alfabetycznego indeksu słów kluczowych, w którym porządek słów kluczowych ma formę alfabetyczną, a nie logiczną. Brak podziału logicznego (rozłącznego i adekwatnego) powoduje sytuację, w której na tym samym poziomie wyodrębnia się elementy treści o różnym stopniu szczegółowości, co powoduje pewnego rodzaju niespójność systemu leksykalnego języka słów kluczowych.

Prezentacja słów kluczowych w porządku alfabetycznym nie dostarcza wprawdzie złożonych zależności semantycznych, jednak siłą tego narzędzia jest możliwość dynamicznego modyfikowania zakresu pytania informacyjnego poprzez wykorzystanie do łączenia jednostek leksykalnych algebry Boole'a. Należy tu jeszcze wspomnieć o możliwości wariantu hybrydowego, w którym nie występują określone zależności strukturalne. Jego podstawową zaletą jest duża elastyczność, która jest rezultatem wielowymiarowej strukturalizacji pola semantycznego języka słów kluczowych. Oparcie punktów dostępu na fasetowym modelu organizacji wiedzy stanowi w tym języku jednak pewnego rodzaju novum.

Wprowadzanie do list słów kluczowych oraz indeksów coraz to nowych słów na podstawie terminów używanych „chwilowo” przez autorów, a nawet w pewnych okresach rozpowszechnionych, ułatwia wprawdzie doraźne poszukiwania, równocześnie jednak staje się źródłem synonimów, jak również homonimów pochodzących stąd, że różni autorzy używają tych samych wyrażen w różnych znaczeniach. Ważnym problemem jest ponadto uogólnianie bądź wyszczególnianie; słowa wyszczególniające ułatwiają często przydział, utrudniają jednak poszukiwania prac syntetyzujących.

Omówione w tym rozdziale wyniki przeprowadzonej analizy języków słów kluczowych potwierdzają szeroki (i znaczenie szerszy niż w innych językach) oraz coraz

bardziej wzrastający stopień uwzględniania w tych językach relacji kojarzeniowych. Przejawem tej tendencji są występujące w językach słów kluczowych struktury oparte wyłącznie na relacjach kojarzeniowych. „Stopień, rzetelność i umiejętność korzystania z tego typu metadanych – jak twierdzi B. Sosińska-Kalata – najczęściej pozostawia jednak wiele do życzenia, toteż w praktyce większość systemów indeksujących ignoruje je, poddając analizie statystycznej tekst dokumentu” (Sosińska-Kalata 2004, s. 89).

Analiza pozwoliła zauważyć pewien wpływ innych paranaturalnych języków na języki słów kluczowych, na przykład w niektórych słownikach słów kluczowych stosuje się obecnie model organizacyjny tezaury, w innych rozwinięto zasady kontroli słownictwa stosowane w języku haseł przedmiotowych, a w jeszcze innych – w semantycznej organizacji leksyki wykorzystano koncepcje języków klasyfikacyjnych, szczególnie klasyfikacji fasetowej.

Budowa i zastosowanie struktur językowych konkretnych języków słów kluczowych jako narzędzia organizacji i dostępu do zbiorów jest złożonym problemem, wymagającym odrębnych badań. Moim zamiarem było tylko wskazanie najważniejszych sposobów wykorzystania języków słów kluczowych w różnych typach systemów wyszukiwania informacji, zaś głównym celem rozważań była identyfikacja tych elementów, które określają tożsamość języków słów kluczowych.

Niewątpliwie jedną z charakterystycznych cech rozwoju języków słów kluczowych jest doskonalenie semantycznej organizacji ich słownictwa. W tym celu wykorzystuje się przede wszystkim metody organizacji fasetowej słownictwa oraz rozszerza repertuar relacji kojarzeniowych. Słowniki słów kluczowych coraz częściej stają się elementem interfejsu użytkownika oraz narzędziem nawigowania po jego zasobach. Prezentacja słownictwa w postaci alfabetycznego wykazu słów kluczowych ma istotne znaczenie dla zastosowań języka słów kluczowych. Skonstruowane i wdrażane w ostatnich latach indeksy słów kluczowych przybierają różne formy, ujawniając dwa podstawowe typy relacji między słowami kluczowymi, jakimi są ekwiwalencja wyszukiwawcza oraz relacje kojarzeniowe. Są też kategorialne listy słów kluczowych, z reguły uporządkowane alfabetycznie. Jednak coraz częściej indeksy słów kluczowych – podobnie jak tezaury – przystosowywane są do nowych funkcji, które bardziej wiążą je z wyszukiwaniem niż indeksowaniem dokumentów. W przypadku realizacji funkcji wyszukiwawczych eksponowana jest przede wszystkim ich funkcja słownika pojęciowego, ułatwiającego identyfikację form wyrażen najlepiej reprezentujących pojęcia i tematy stanowiące przedmiot wyszukiwania informacji. W tego rodzaju zastosowaniach słowników/zbiorów słów kluczowych zanika koronna dotąd funkcja normatywna i przekładowa oraz traci sens zróżnicowanie statusu słowa kluczowego i wyrażenia języka naturalnego. Wszystkie słowa kluczowe mogą być użyte w wyszukiwaniu. Dzięki relacji ekwiwalencji lub synonimii wyszukiwawczej użytkownik jest informowany o tym, że znaczenie niektórych z nich może być reprezentowane także przez inne wyrażenie/formę językową. Słowniki słów kluczowych w takiej postaci, wykorzystywane do wyszukiwania informacji, stają się bliskie lub niemal identyczne z alfabetycznym wykazem form nominatywnych słów/wyrażeń języka naturalnego, a w szczególności alfabetycznym wykazem

terminów, odbiegając, a właściwie wracając do ukształtowanego w wyszukiwaniu informacji, bazującego na języku unitermów, modelu słownika słów kluczowych.

Analiza stosowanych w Internecie słów kluczowych potwierdza tezę, że nie są one aż tak mocno obciążone tradycją jak inne języki paranaturalne (języki haseł przedmiotowych i języki deskryptorowe). Należałoby zastanowić się, czy dokument-obiekt fizyczny należy dalej uważać za optymalną jednostkę opisu i wyszukiwania. Obecnie w rzeczywistości bibliotek cyfrowych stają się dostępne fragmenty, porcje treści, z których część na pewno jest warta zaopatrzenia we własną identyfikację metadanową z wykorzystaniem języka słów kluczowych.

Szczegółowa leksyka nieklasycznych języków słów kluczowych, a w konsekwencji stałe zwiększanie rozmiarów ich słowników, często znacznie przekraczających zasób aktywnego słownika przeciętnego użytkownika języka i jego możliwości akwizycji informacji, oraz częsta ingerencja w denotację i konotację wyrażen języka naturalnego jest zabiegiem nietrafnym i mało skutecznym, gdyż nie prowadzi do poprawy relewancji, lecz przeciwnie – do zwiększenia szumu informacyjnego i powstawania ciszy informacyjnej. W tej sytuacji ogromne koszty budowy takich słowników (a także indeksowania w takich językach) są niewspółmierne do efektów, gdyż część słownictwa nie jest i nie będzie wykorzystywana.

Ten nurt działań praktycznych i teoretycznych jest zbieżny z pracami nad innymi tradycyjnymi językami informacyjno-wyszukiwawczymi, gdyż języki te próbowano rozwijać tak samo jak pozostałe. Jest on przejawem wiary w możliwość stworzenia kontrolowanego i sterowalnego komputerowego systemu wyszukiwania informacji. Nie dostrzega się jednak niebezpieczeństwa, jakie stanowi fakt, że tego typu systemy na skutek niekontrolowanego wzrostu stają się niesterowalne i zgodnie z teorią chaosu przekształcają się w systemy samoregulujące, których działanie nie zawsze jest podporządkowane przede wszystkim potrzebom użytkowników i ich możliwościom percepcyjnym.

Pomocnym narzędziem staje się tu Internet, gdzie wykorzystuje się swobodne słowa kluczowe jako wyrażenia języka naturalnego w funkcji informacyjnej, których nie trzeba specjalnie tworzyć oraz współrzędne indeksowanie i wyszukiwanie informacji. Jest to lingwistyczne narzędzie, którym od dawna już się posługujemy. Dzięki Internetowi języki te wracają do swej pierwotnej, klasycznej formuły.

3.1.5. Propozycja zestawu kryteriów opisu języków słów kluczowych

Pomimo istnienia wielu różniących się języków słów kluczowych, do tej pory nie opracowano ich typologii. Typologia jest sposobem porządkowania elementów zbioru, według którego każdy z nich kwalifikowany jest do pewnego typu elementów na podstawie relacji podobieństwa. Stworzenie typologii wymagałoby wcześniejszego zebrania w miarę pełnych informacji o wszystkich (a przynajmniej większości) istniejących językach słów kluczowych, co jest niemożliwe, a następnie przypisania ich do wydzielonych typów. Ze względu na szczupły materiał badawczy w tym

przedmiocie nie jest to możliwe w ramach niniejszej rozprawy. Wyodrębnione w tym rozdziale cechy konstytutywne i indywidualne badanych języków słów kluczowych okazały się jednak na tyle cennym materiałem, że podjąłem próbę stworzenia na ich podstawie zestawu kryteriów opisu tej klasy języków. Wziąłem pod uwagę następujące kryteria:

1. Zakres pola semantycznego:

- język słów kluczowych dziedzinowy;
- język słów kluczowych wielodzinowy.

2. Charakter słownictwa (rodzaj notacji):

- język słów kluczowych o słownictwie paranaturalnym;
- język słów kluczowych o słownictwie mieszanym¹²⁰.

3. Stopień kontroli słownictwa/jednostek leksykalnych:

- język słów kluczowych o słownictwie kontrolowanym – wariant kontrolowany (lista wzorcowa słów kluczowych);
- język słów kluczowych o słownictwie częściowo kontrolowanym – wariant częściowo kontrolowany (lista wzorcowych słów kluczowych, lista wzorcowa słów niekluczowych, stop-lista¹²¹);
- język słów kluczowych o słownictwie niekontrolowanym – wariant niekontrolowany (słownictwo swobodne).

4. Związek z systemem informacyjno-wyszukiawczym:

- język słów kluczowych oparty na słowniku praktycznym;
- język słów kluczowych oparty na słowniku teoretycznym.

5. Charakter słownika:

- język słów kluczowych ze słownikiem jawnym;
- język słów kluczowych ze słownikiem częściowo jawnym;
- język słów kluczowych ze słownikiem niejawnym (zadaniem instrukcją).

6. Model struktury pola semantycznego:

- język słów kluczowych klasyczny (płaska struktura pola semantycznego);
- język słów kluczowych nieklasyczny (m.in. z elementami fasetyzacji, fasetowy, oparty na kategoryzacji leksykalnej lub układzie gniazdowym).

¹²⁰ Zawierają one elementy sztucznych kodów, na przykład wzory chemiczne.

¹²¹ Podane formy kontroli słownictwa różnią się między sobą. Lista wzorcowych słów kluczowych to nieskończony zbiór „takich i tym podobnych” wyrażeń, które mogą być słowami kluczowymi. Lista wzorcowa słów niekluczowych to nieskończony zbiór „takich i tym podobnych” wyrażeń, które nie mogą być słowami kluczowymi. Stop-lista zawiera skończony zbiór wyrażeń, które słowami kluczowymi być nie mogą.

7. Stopień autonomiczności słownictwa:

- język słów kluczowych o słowniku własnym;
- język słów kluczowych o słowniku hybrydowym (włączającym inne układy słownictwa).

8. Sposób przejścia z języka naturalnego na język słów kluczowych:

- język słów kluczowych o słowniku, w którym to przejście jest niejawne;
- język słów kluczowych o słowniku, w którym to przejście jest częściowo jawne;
- język słów kluczowych o słowniku, w którym to przejście jest jawne.

9. Typ stosowanej gramatyki:

- język słów kluczowych z tzw. gramatyką zerową;
- język słów kluczowych z gramatyką częściowo pozycyjną.

10. Funkcja w systemie¹²²:

- język słów kluczowych w funkcji języka indeksowania;
- język słów kluczowych w funkcji języka wyszukiwawczego;
- język słów kluczowych w funkcji języka informacyjno-wyszukiwawczego.

11. Liczba języków naturalnych stanowiących podstawę słownictwa:

- język słów kluczowych oparty na słowniku jednojęzycznym;
- język słów kluczowych oparty na słowniku wielojęzycznym.

12. Status języka słów kluczowych w systemie informacyjno-wyszukiwawczym:

- język słów kluczowych stanowiący etap w budowie języka deskryptorowego;
- język słów kluczowych będący językiem docelowym.

13. Osoba tworząca charakterystyki:

- język autorskich słów kluczowych;
- język nieautorskich słów kluczowych.

14. Generacje systemów wyszukiwania informacji (metody wyboru słów kluczowych):

- język słów kluczowych I generacji („intelektualne” (przez człowieka) wybieranie słów kluczowych);
- język słów kluczowych II generacji (automatyczne wybieranie słów kluczowych);
- język słów kluczowych III generacji („inteligentne” wybieranie słów kluczowych).

15. Metoda tworzenia/budowy języka słów kluczowych:

- język słów kluczowych budowany metodą indukcyjną;
- język słów kluczowych budowany metodą dedukcyjną;
- język słów kluczowych budowany metodą mieszaną.

¹²² Por. rozdział 2.3.1.

16. Natura języka/słów kluczowych:

- sztuczny język słów kluczowych;
- język naturalny w funkcji metainformacyjnej.

Przedstawiony zestaw kryteriów opisu języków słów kluczowych, odwzorowujący istotne elementy tożsamości języka słów kluczowych, uwzględnia zarówno elementy struktury tych języków, jak i ich funkcje. Do elementów struktury języka należą: metody budowy języka i wyboru słów kluczowych, liczba języków, zakres i charakter pola semantycznego oraz jego model, charakter i stopień autonomiczności słownictwa, stopień kontroli i koordynacji jednostek leksykalnych, sposób przejścia na język naturalny i odwrotnie, typ stosowanej gramatyki. Do funkcji języków słów kluczowych zaliczyłem: funkcje w systemie i związek z systemem informacyjno-wyszukiwawczym. Wariant hybrydowy został wyodrębniony z uwagi na brak możliwości przyporządkowania niektórych z analizowanych języków słów kluczowych do pozostałych wariantów. Nie ma podstaw do opracowania jego modelowej wersji z uwagi na heterogeniczność strukturalną języków słów kluczowych kwalifikujących się do tego typu.

W przypadku każdego zestawu kryteriów najwięcej kontrowersji zwykle budzi sam zestaw. Przy wyborze kryteriów podziału zawsze więc decydują względy pragmatyczne. Dobry zestaw to taki, który jest poprawny ze względu na kryteria wewnętrzne i optymalny ze względu na jego funkcjonalność. Zaproponowany zestaw kryteriów, będący rezultatem rozważań zamieszczonych w tym rozdziale, może być przydatny w strukturalno-funkcjonalnym opisie języków słów kluczowych, gdyż pozwala ujawnić wewnętrzne zróżnicowanie struktur tych języków oraz pełnionych przez nie funkcji, a także pokazać związki tych języków z językiem naturalnym.

3.2. Wybrane problemy pragmatyki słów kluczowych

3.2.1. Pragmatyka języka słów kluczowych: mentalna i formalna

W opisie pragmatyki¹²³ języków słów kluczowych zamierzam zwrócić uwagę wyłącznie na te problemy, które dotyczą bezpośrednio języków słów kluczowych i korespondują z problemami poruszonymi w rozprawie.

Problematyka badań dotyczących posługiwania się językami informacyjno-wyszukiwawczymi w praktyce dotychczas rzadko była podejmowana zarówno w światowym (Hutchins 1978), jak i w polskim piśmiennictwie poświęconym tym

¹²³ Pragmatyka języka od przełomu lat siedemdziesiątych i osiemdziesiątych XX wieku jest rozumiana jako dziedzina językoznawstwa zajmująca się problematyką zasad poprawnego użycia języka w konkretnych sytuacjach komunikacyjnych. Por. EJO 1995.

językom (Ćwiekowa 1988; Hys 2009). Brak też opracowań dogłębnie analizujących pragmatykę języków słów kluczowych. Zresztą część reguł pragmatycznych języka słów kluczowych nie jest jawna i nie została dotychczas sprecyzowana. Tymczasem sposób posługiwania się językiem słów kluczowych jest uwarunkowany nie tylko strukturą tego języka, ale i sytuacją problemową użytkownika, określoną przez właściwości jego wiedzy i języka, której projekcja ujawnia się w procesie wyszukiwania informacji.

Pragmatyka języka słów kluczowych opiera się na indeksowaniu współrzędnym, przez które rozumie się – powtarzam, bo to ważne – charakteryzowanie treści dokumentów lub opis potrzeb informacyjnych za pomocą współrzędnych, niezależnych, dających się swobodnie zestawiać elementarnych jednostek leksykalnych, przeważnie mających postać wyrazów i połączeń wyrazowych zaczerpniętych z języka naturalnego, nazywanych w przypadku języków słów kluczowych słowami kluczowymi. W ten sposób treść dokumentu podlega w procesie indeksowania dekompozycji na zbiór oddzielnych słów kluczowych (proces analityczny). Słowa kluczowe tworzące charakterystykę wyszukiwawczą dokumentu lub instrukcję wyszukiwawczą zapytania traktowane są jako wzajemnie niezależne i równorzędne, czyli współrzędne. Każda taka jednostka leksykalna może pełnić funkcję klucza wyszukiwawczego, a więc staje się samodzielnym elementem wyszukiwawczym bez względu na jej pozycję w tekście. Jednostki te mogą znajdować się w dowolnym miejscu tekstu, a więc być wymieniane w dowolnej kolejności.

Język słów kluczowych umożliwia wykładnię pojęciową wyszczególniającą elementy treści dokumentu, która jest wyrażana w postaci nazw w sformalizowanym języku etnicznym lub w sztucznym języku słów kluczowych. Szczegółowe instrukcje tworzenia słów kluczowych i ich wykorzystania w indeksowaniu dokumentów zwykle dotyczą warstwy leksykalno-semantycznej (doboru słownictwa). Kolejność słów kluczowych (warstwa syntaktyczna) jest dowolna. Płaszczyzny te, istotne przy budowie zdań i tekstów w językach słów kluczowych, oprócz punktu widzenia semantyki, leksykologii i syntaktyki, przeważnie są modyfikowane przez aspekty pragmatyczne, uwzględniające potrzeby użytkowników tego języka. Obok instrukcji wyboru słów kluczowych tworzy się słowniki (wykazy) słów kluczowych, które mają zapewnić większy lub mniejszy stopień kontroli słownictwa tych języków.

W systemach dokumentacyjnych indeksowanie jest rodzajem rzeczowego opracowania dokumentów. Jego rezultatem w języku słów kluczowych jest zbiór słów kluczowych, które wyrażają elementy treści dokumentu bez podziału na główne, poboczne lub szczegółowe. Każde ze słów kluczowych, wchodzące w skład opisu systemowego, jest autonomiczne, tj. wyraża pewien element treści niezależnie od innych. Treść dokumentu odwzorowuje zestaw słów kluczowych, które są oddzielone od siebie, w zależności od systemu, za pomocą specjalnych znaków, zwanych separatorami. Separatorami są zwykle następujące znaki: ., ,, :, ;, -, \$, %, &, *, podwójna spacja. Jednak nigdy zbiór słów kluczowych nie wyraża myśli autora tak dokładnie, jak tekst w języku naturalnym, gdyż język słów kluczowych nie posiada rozbudowanej składni, która jest narzędziem wyrażania relacji sytuacyjnych. Składnia języka

słów kluczowych jest sztuczna i opiera się tylko na relacji współwystępowania. Oto przykład opisu dokumentu w języku słów kluczowych.

(B. Sosińska-Kalata: *Języki informacyjno-wyszukiawcze. Współczesne tendencje w badaniach i zastosowaniu*. [W:] *Spółczesność informacyjna i jej technologie*. Pod red. B. Sosińskiej-Kalaty, K. Materskiej, W. Glińskiego. Warszawa: Wydawnictwo SBP 2004 s. 79–91)

Słowa kluczowe:

Języki informacyjno-wyszukiawcze, systemy klasyfikacyjne, tezaury, opracowanie rzeczowe literatury pięknej, jednostka analizy i opisu treści, wielojęzyczność, organizacja zasobów Internetu

Stąd charakterystykę wyszukiawczą w języku słów kluczowych można przedstawić w postaci modelowej jako:

$$0 = x_1, x_2, x_3, \dots, x_n$$

gdzie:

0 – charakterystyka wyszukiawcza,

x_n – słowa kluczowe.

W zaprezentowanym modelu ewidentnie widać brak zróżnicowania elementów treści. Model tego języka różni się od modelu języka haseł przedmiotowych, gdzie wyróżnia się przedmioty dokumentu, ich ujęcia, a także formy prezentacji. Te kategorie w ogóle nie mają zastosowania w języku słów kluczowych. Ograniczone możliwości wyrażania relacji sytuacyjnych (poza relacją współwystępowania) w tym języku rekompensują narzędzia wyszukiawcze należące do repertuaru narzędzi systemu informacyjno-wyszukiawczego, do których zaliczana jest m.in. algebra Boole'a i operatory kontekstowe. Teksty języka słów kluczowych opierają się więc nie na relacjach systemowych, lecz wyłącznie na okrojonych do minimum relacjach sytuacyjnych, tj. relacjach współwystępowania.

Opis rzeczowy dokumentu składa się ze słów kluczowych w języku naturalnym lub w języku paranaturalnym, odwzorowującym treść dokumentu/informacji poprzez określenie jej elementów. Słowa kluczowe, o których tu mowa, to dominujące w tekście wyrażenia (rzeczowniki), najczęściej terminy, podane w przyjętej w danym języku formie, które charakteryzują to, o czym jest mowa w tekście. W toku analizy wyrażenia dominujące w tekście są traktowane jako słowa kluczowe, gdyż zakłada się, że wyrażają one pojęcia kluczowe tekstu dokumentu. Słowa kluczowe są wprawdzie już na etapie ich wyboru pozbawiane kontekstu, ale kontekst polegający na ich współwystępowaniu zachowują w CHWD i IWZ. Następnie stosuje się swoistego rodzaju „inżynierię leksykalną”, polegającą na nadawaniu tym wyrażeniom odpowiedniej formy właściwej dla języka słów kluczowych, określonej w instrukcji indeksowania.

Charakterystyka wyszukiwawcza dokumentu w języku słów kluczowych ma charakter statyczny, co oznacza, że tekst ten, będący rezultatem procesu indeksowania, nie ulega zmianom. Sposób opracowywania charakterystyki wyszukiwawczej jest zależny od odmiany języka słów kluczowych. Inaczej wygląda w języku swobodnych słów kluczowych, a inaczej w języku kontrolowanych słów kluczowych, opartym na słowniku jawnym czy częściowo jawnym. W pierwszym przypadku kluczową rolę odgrywają instrukcje indeksowania, w drugim zaś – już gotowe albo powstające, tzn. in statu nascendi, listy/wykazy/indeksy/słowniki słów kluczowych.

Utworzony ze słów kluczowych tekst w postaci charakterystyki wyszukiwawczej dokumentu w nieco inny sposób opisuje treść dokumentu niż inne języki informacyjno-wyszukiwawcze. Słowa kluczowe opisujące treść dokumentu stanowią jego kondensat, a ich liczba jest zależna m.in. od głębokości indeksowania, przy czym maksymalnie głębokie indeksowanie zapewnia kompletność wyszukiwania, ale równocześnie zmniejsza jego trafność i powoduje wysoki stopień szumu informacyjnego, co sprawia, że wydane dokumenty zawierające informacje na dany temat są marginalnie z nim związane, czyli nierelevantne. Nie ma tu jednak prostej zależności, bo wymienione parametry zależą od paradygmatyki języka. A więc nie jest to ogólna prawidłowość, chociaż w językach słów kluczowych tak jest. W innych językach informacyjno-wyszukiwawczych głębokie indeksowanie może zwiększać trafność wyszukiwania.

Twórcy języka słów kluczowych przeważnie nie formułują reguł pragmatyki tego języka. Czynią to zazwyczaj użytkownicy korzystający z systemu, formułując odpowiednie instrukcje indeksowania, czyli opracowywania rzeczowego. Taka instrukcja definiuje sposób stosowania języka informacyjno-wyszukiwawczego w opisie i wyszukiwaniu dokumentów i jest elementem pragmatyki danego języka. Określa ona m.in. sposób stosowania składni języka słów kluczowych, na przykład sposób oddzielania poszczególnych słów kluczowych separatorem, zasady sporządzania instrukcji wyszukiwawczych oraz postulaty działań pragmatycznych, do których zalicza się adekwatność i jednolitość indeksowania (Hys 2009).

Z tego powodu materiałem badawczym na potrzeby tego rozdziału uczyniłem istniejące w tym zakresie w języku polskim instrukcje indeksowania oraz inne dokumenty zawierające informacje, jak poprawnie posługiwać się konkretnymi językami słów kluczowych. Należą do nich – oprócz nielicznych instrukcji indeksowania – wstępy do słowników słów kluczowych, rozdziały w podręcznikach dotyczących języków informacyjno-wyszukiwawczych oraz uwagi w artykułach omawiających te języki. Wnioski z tej analizy to określenie zasad/wytycznych dotyczących pragmatyki ogólnej języków słów kluczowych oraz pragmatyk szczegółowych poszczególnych języków słów kluczowych. Istniejące opisy pragmatyki języka słów kluczowych mają wyłącznie wartość historyczną, gdyż obecnie reguły te muszą mieć charakter formalny.

Ponieważ każda nazwa lub termin mogą być użyte w funkcji słowa kluczowego, istnieje duże prawdopodobieństwo, że dokumenty o takiej samej lub podobnej treści mogą mieć formalnie przypisane różne opisy rzeczowe, gdy zawierają różne określenia tych samych elementów treści lub ich własności. Język słów kluczowych nie pozwala całkowicie wyeliminować z opisów dokumentów wyrażen synonimicznych

i bliskoznacznych ani wyrażen wieloznacznych, czyli homonimów. W rezultacie informacje dotyczące tych samych treści dokumentów są w systemie rozproszone.

W zasadzie kolejność słów kluczowych w charakterystyce/instrukcji wyszukiwawczej jest dowolna, nieuporządkowana. Istnieją jednak instrukcje, w których ta kolejność jest z góry określona. O kolejności słów kluczowych w charakterystyce i/lub instrukcji wyszukiwawczej może decydować ich trafność, na przykład na początku umieszcza się słowa kluczowe najbardziej charakterystyczne dla treści dokumentu/informacji, następnie, po separatorze, słowa kluczowe na drugim stopniu trafności. Innym sposobem ustalania trafności słów kluczowych w charakterystyce i instrukcji wyszukiwawczej jest przypisywanie im wag, zwanych współczynnikami wagowymi. W przypadku metody pozycyjnej instrukcja z góry podaje kolejność poszczególnych słów kluczowych (oddzielanych od siebie ustalonymi z góry znakami – separatorami), określając aspekt lub rolę słów kluczowych, na przykład, na początku charakterystyki powinno stać słowo kluczowe oznaczające przedmiot badań, po nim słowo kluczowe oznaczające metodę badań, wyniki badań, zastosowania wyników badań w praktyce. Kolejność słów kluczowych może też mieć charakter formalny, tzn. być alfabetyczna. Wymienione tu odstępstwa dotyczą języków słów kluczowych, które nie mieszczą się w definicji klasycznego języka słów kluczowych.

Słowa kluczowe w odróżnieniu od deskryptorów nie zawierają żadnych znaków specjalnych, interpunkcyjnych i alfanumerycznych. W celu wyróżnienia graficznego – w zależności od przyjętej konwencji – są pisane wersalikami, kursywą, zaczynają się od dużej litery. Liczba słów kluczowych, tam gdzie nie jest limitowana, jest zależna jedynie od zakresu i wagi informacji zawartych w dokumencie. Słowa kluczowe podlegają opracowaniu leksykograficznemu według obowiązujących zasad ogólnych i/lub szczegółowych zawartych w odpowiedniej instrukcji indeksowania.

Prezentacja samego języka słów kluczowych ma ułatwić użytkownikowi poruszanie się po meandrach słownictwa. W polskim piśmiennictwie problem ten jest w zasadzie ignorowany. Języki słów kluczowych dysponują jedynie mniej lub bardziej rozbudowanym systemem odsyłaczy. W niektórych odmianach tego języka system odsyłaczy ograniczono do minimum. Godny uwagi jest stosowany w wersjach komputerowych indeks permutacyjny. Ze względu na rozmaite przyzwyczajenia użytkowników do określonego porządku elementów, indeks permutacyjny jest jedynym sensownym rozwiązaniem wobec alternatywy mnożenia liczby odsyłaczy. Słownictwo języka słów kluczowych często cechuje duża liczba terminów wieloelementowych. Nie zawsze udaje się zastosować ogólną zasadę układu elementów, na przykład szyk naturalny czy jeden stosowany konsekwentnie typ szyku przestawnego. Osobnym zagadnieniem pozostaje problem nazw wieloelementowych, których składniki nie są oddzielnymi wyrazami. Nazwy te nie mogą być uwzględnione w indeksie permutacyjnym. Prawie nigdy nie tworzy się dla nich odsyłaczy. Problem ten jest szczególnie ważny dla terminów, dla których przyjęto „angielską” kolejność elementów.

Słownictwo języka słów kluczowych powinna cechować – podobnie jak poprawną wypowiedź w języku naturalnym – konsekwencja i jednorodność przyjętych form. Dotyczy to nie tylko zgodności z normami formułowania terminologii specjalistycz-

nej. W języku słów kluczowych trzeba przestrzegać również ogólnych norm poprawnościowych języka polskiego, nawet jeśli w języku specjalistycznym pojawiają się, podobnie jak w języku mówionym, formy niepoprawne. Jest to ważne dla dalszych rozważań oraz opisu wyników badań własnych.

Dalej zajmę się pragmatyką indeksowania i pragmatyką wyszukiwania, które często są włączane do pragmatyki języka, lecz w rzeczywistości są zależne przede wszystkim od potrzeb użytkowników systemu oraz możliwości systemu wyszukiwawczego, a tylko w części od możliwości danego języka. Zasadność ich omawiania w tym miejscu może budzić pewne wątpliwości. Omówię je jednak, traktując jako narzędzia niwelujące deficyty pragmatyki języka słów kluczowych.

Skupię się teraz na problemach dotyczących używania słów kluczowych najpierw w indeksowaniu¹²⁴, a potem w wyszukiwaniu informacji. Zarówno podczas indeksowania, jak i wyszukiwania za pomocą słów kluczowych powstają przeważnie te same problemy, które dotyczą przede wszystkim trudności ze wskazaniem odpowiednich słów znaczących, co jest w dużym stopniu subiektywne¹²⁵.

3.2.2. Pragmatyka indeksowania za pomocą słów kluczowych

Zaletą indeksowania współrzędnego jest możliwość tworzenia wieloaspektowej charakterystyki dokumentu za pomocą swobodnie kombinowanych pojedynczych elementów. Według Teuna A. van Dijka do opisu treści dokumentu powinno się używać 5–20 słów kluczowych (średnio 8–12), a według Briana C. Vickery'ego –

¹²⁴ Metodyka indeksowania obejmuje głównie zagadnienia głębokości (szczegółowości) i szerokości (kompletności) indeksowania, które określane są jako parametry indeksowania. Głębokość indeksowania to stopień ogólności/szczegółowości wyrażania treści dokumentu/informacji przez zestaw słów kluczowych. Chodzi tu nie tyle o liczbę słów kluczowych, która – wprawdzie deklarowana jako maksymalne x – powinna być wystarczająca, lecz o wnikliwość opracowania dokumentu/informacji przejawiającą się w odpowiedniej specyfikacji elementów treści dokumentu, wszystkich albo tylko najważniejszych, uwzględnianiu charakteru potrzeb potencjalnych użytkowników systemu. Przez szczegółowość indeksowania rozumiem wprowadzanie do charakterystyk wyszukiwawczych słów kluczowych na tym samym poziomie ogólności co użyte w pracy. Gwarantuje ją pobieranie słów kluczowych bezpośrednio z tekstu dokumentu, a nie ich uogólnianie. Przez szerokość indeksowania rozumiem odzwierciedlenie w sposób wyczerpujący i pełny istotnych elementów treści dokumentu. Kompletność indeksowania jest związana z liczbą słów kluczowych tworzących charakterystykę wyszukiwawczą danego dokumentu. Jednak duża liczba słów kluczowych może zależeć od szczegółowości, a nie od kompletności. Kompletność indeksowania określa zwykle odpowiednia instrukcja indeksowania dla określonego systemu, w której bierze się pod uwagę potrzeby potencjalnych użytkowników systemu informacyjno-wyszukiwawczego.

¹²⁵ W 2005 roku Jarosław Pacek przeprowadził interesujące badania wybranych stron internetowych związanych z informacją naukową. Przedmiotem badań była zawartość i jakość trzech tagów: TITLE – tytuł, DESCRIPTION – opis, streszczenie i KEYWORDS – słowa kluczowe, których obecność i poprawność słusznie przyjął za niezmiernie ważne dla procesu wyszukiwania informacji. Za najważniejsze przyczyny trudności w posługiwaniu się słowami kluczowymi uznał stosowanie słów nieznaczących, tj. nierelevantnych dla użytkownika, słów o małym ładunku treści oraz używanie w roli słów kluczowych mało precyzyjnych wyrażań wielowyrzowych, niewielką głębokość i małą szerokość indeksowania, a także ich niewłaściwy wybór (Pacek 2010).

6–9 słów kluczowych¹²⁶. Liczbę możliwych kombinacji słów kluczowych wyraża następujący wzór:

$$2^k - 1 = c$$

gdzie:

k – liczba słów kluczowych,

c – liczba możliwych kombinacji.

Liczba kombinacji umożliwia rozbudowywanie zakresu odwzorowywanej treści dokumentu oraz stwarza możliwość skojarzeń wieloznacznych na skutek indeksowania współrzędnego, co nie pozwala na uzyskanie wysokiego współczynnika trafności, chyba że, tak jak w indeksowaniu pozycyjnym, wysunie się na pierwsze miejsce najważniejsze słowo kluczowe – tzw. hierarchizowanie kolejnością – przez co następuje dookreślenie danego elementu treści. Idea tego języka polega na tym, że słowa kluczowe mają być narzędziem dostępu do dokumentu na poziomie mikro, to jest powinny mieć stosunkowo wąski zakres.

Przejście z języka naturalnego na język informacyjno-wyszukiawczy następuje w procesie charakteryzowania dokumentu za pomocą wyrażeń języka informacyjno-wyszukiawczego, a dokładniej w momencie przekładu charakterystyki słownej dokumentu, utworzonej ze słów kluczowych występujących *explicite* lub *implicite* w tekście charakteryzowanego dokumentu, na jednostki leksykalne stosowanego języka¹²⁷. Słowa kluczowe, będące elementami charakterystyki słownej dokumentu, nie są wyrażeniami żadnego języka informacyjno-wyszukiawczego. Są to wyrażenia języka naturalnego, i – podobnie jak jednostki języka informacyjno-wyszukiawczego – służą do wyrażania metainformacji. Słów będących elementami charakterystyki słownej nie należy mylić ze słowami kluczowymi tworzącymi gotową listę słów kluczowych. Te ostatnie są bowiem elementarnymi jednostkami leksykalnymi sztucznego języka słów kluczowych.

W związku z tym w procesie indeksowania słownikowego (normatywnego) mamy do czynienia z nieidentycznością lub nie w pełni identycznością transformacją tekstu charakterystyki słownej dokumentu i nieidentycznością transformacją jego funkcji, polegającą na tym, że oprócz funkcji metainformacyjnej tekst ten zaczyna pełnić funkcję wyszukiawczą.

W zasadzie warstwę podkładową w procesie indeksowania stanowią słowa kluczowe języka naturalnego. Jak pisze E. Ścibor, problem ten „można rozpatrywać z dwóch punktów widzenia: 1. odpowiedniości między językiem naturalnym a językiem informacyjno-wyszukiawczym, wyrażonych *explicite* (jawnie) lub *implicite*

¹²⁶ Liczba słów kluczowych pokazuje zróżnicowanie i relatywne proporcje elementów treści danego dokumentu. Podane liczby dotyczą bardzo starych zaleceń, uwarunkowanych środkami technicznymi. Obecnie nie mają już sensu.

¹²⁷ Problemy związane z charakteryzowaniem dokumentów za pomocą języka słów kluczowych występują także przy charakteryzowaniu pytań informacyjnych. W celu uproszczenia wywodów przyjmuję w tym tekście, że wszystko to, co dotyczy indeksowania dokumentów, odnosi się także do pytań informacyjnych, których obecnie nie indeksuje się, gdyż użytkownik od razu podaje słowa kluczowe.

(niejawnie) w słowniku języka informacyjno-wyszukiwawczego, 2. charakteryzowania dokumentów za pomocą słów kluczowych, stanowiącego jeden z etapów procesu indeksowania. W pierwszym wypadku mamy do czynienia ze »strukturalnym« aspektem zagadnienia, w drugim przypadku – z jego aspektem technologicznym. Oba aspekty zresztą są ściśle ze sobą związane, gdyż słownik języka słów kluczowych jako tablica przejścia z języka naturalnego na język słów kluczowych (a więc w przypadku przejścia *explicite* – także wyrażenia języka naturalnego, połączone relacją odpowiedniości wyszukiwawczej z określonymi wyrażeniami języka słów kluczowych) ma umożliwić i maksymalnie ułatwić przekład charakterystyki słownej dokumentu, złożonej z »naturalnych« słów kluczowych, na jego charakterystykę wyszukiwawczą składającą się ze sztucznych jednostek leksykalnych zwanych słowami kluczowymi. Rozpatrując zagadnienie podkładowej warstwy słów kluczowych z pierwszego, tj. strukturalnego punktu widzenia, można przyjąć, że »klasą równobliiskoznacznych słów kluczowych« jest grupa wyrażen języka naturalnego o tym samym lub pokrewnym znaczeniu, powiązanych relacją odpowiedniości wyszukiwawczej z jedną i tą samą jednostką leksykalną języka słów kluczowych, która – według terminologii V.A. Moskoviča – jest »przedstawicielem« danej grupy wyrażen w danym języku słów kluczowych. Sposób wyrażenia relacji odpowiedniości wyszukiwawczej między jednostką leksykalną języka słów kluczowych a przyporządkowanymi jej wyrażeniami języka naturalnego (»słowami kluczowymi«) jest zależny od typu słownika języka słów kluczowych» (Ścibor 1999, s. 45–46).

W przypadku języka swobodnych słów kluczowych oba etapy nakładają się. Tworzenie charakterystyki słownej dokumentu jest etapem końcowym procesu indeksowania. Staje się ona zarazem (równocześnie) charakterystyką wyszukiwawczą dokumentu/pytania użytkownika. Wybrane swobodnie z tytułu i/lub tekstu dokumentu wyrażenia języka naturalnego uzyskują od razu, po odpowiednim ujednoczeniu na zasadzie przyjętej konwencji, status słów kluczowych, używanych w wyszukiwaniu informacji.

W tekstach języków słów kluczowych poziom szczegółowości pojęć współtworzących wiedzę zawartą w dokumencie podlega redukcji do grupy pojęć uznanych przez indeksującego za najistotniejsze. Wyższy stopień unifikacji formy wyrażen reprezentujących to samo pojęcie występuje w tekstach języków o słownictwie kontrolowanym, które przeważnie obejmują wyłącznie wyrażenia języka naturalnego w formie prostej lub rozwiniętej frazy nominalnej, określane jako słowa kluczowe. Słowom takim nadawana jest standardowa postać gramatyczna, zwykle zbieżna z formą hasła, przyjętą w tradycji leksykograficznej określonego języka naturalnego. W zbiorach terminów indeksowych, derywowanych z tekstów języka naturalnego, forma wyrażen reprezentujących to samo pojęcie podlega znacznej zmienności zderminowanej przez reguły gramatyczne języka naturalnego.

Specyfikację ról sytuacyjnych pojęć reprezentowanych w systemowym opisie dokumentu umożliwia gramatyka pozycyjna wykorzystywana w językach słów kluczowych opartych na systemach tzw. indeksowania fasetowego (ankietowego, rastrowego, pozycyjnego). Języki słów kluczowych o gramatyce częściowo pozycyjnej określają za pomocą relatorów typ związków sytuacyjnych między desygnatami wyodrębnio-

nych elementów treści dokumentów. Przykładem tego typu modeli relacyjnych („relatorowych”) są schematy tekstów generowane za pomocą systemów relatorów Perreaulta i indeksowania relacyjnego Farradane’a. Obie propozycje powstały w połowie lat sześćdziesiątych XX wieku, ale dopiero obecnie odnoszą swoisty sukces.

System relatorów Jeana M. Perreaulta został opracowany jako rozwinięcie podstawowego zestawu wykładników relacji sytuacyjnych stosowanego w UKD. Przewidziano w nim jedynie identyfikację dwóch typów związków między pojęciami reprezentującymi elementy treści dokumentu: współwystąpienia niezależnego, czyli alternatywy i współwystąpienia zależnego, czyli koniunkcji (Perreault 1994). Zaproponowany przez Perreaulta system obejmuje 120 typów relacji uporządkowanych w czterech klasach: związki hierarchiczne (*subsumptive*), determinujące (*determinative*) i porządkujące (*ordinal*) oraz ogólne związki logiczne (*logical*). System ten może być wykorzystany jako narzędzie identyfikacji związków sytuacyjnych w dowolnym języku informacyjno-wyszukiwawczym, również w języku słów kluczowych. Jako takie uniwersalne narzędzie został zaprojektowany system indeksowania relacyjnego Jasona E.L. Farradane’a (Farradane 1967). Model ten powstał na podstawie teorii procesów poznawczych Joya Paula Guilforda (por. rozdział 1.3.1).

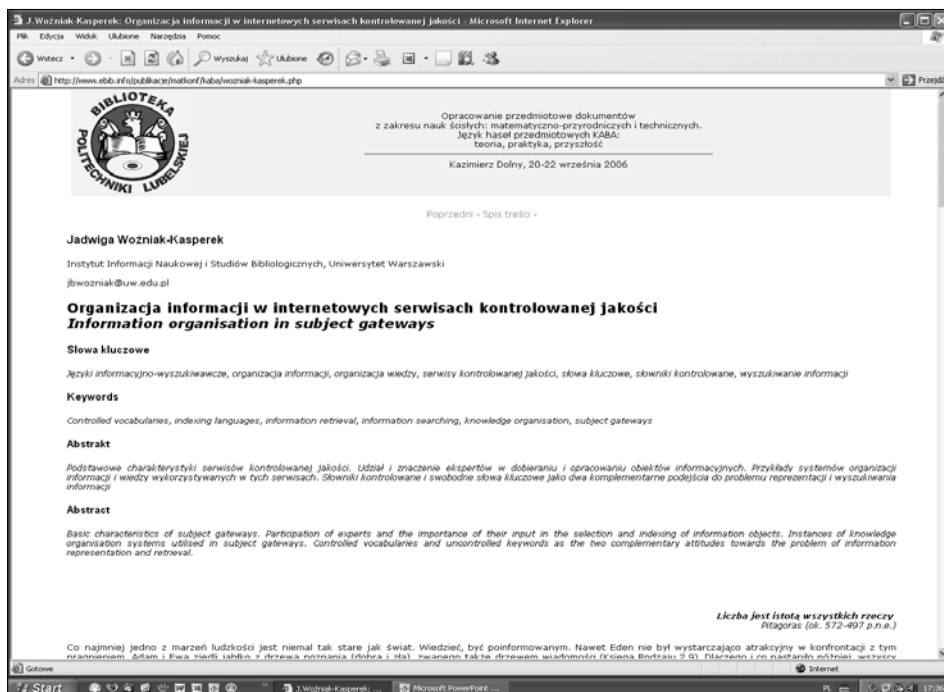
W charakterystykach i instrukcjach wyszukiwawczych klasycznych języków słów kluczowych nie występuje synonimia – na etapie indeksowania dokonuje się bowiem tzw. redukcji elementów treści (skupienie redukcyjne słów kluczowych), polegającej na tym, że spośród wyrażeń uznanych za ekwiwalenty wybiera się jedno, a pozostałe wyrażenia odrzuca się. Oczywiście całość systemowo załatwia słownik wzorcowych słów kluczowych, w którym pozostałe wyrażenia/słowa kluczowe łączy się z tym wyrażeniem odsyłaczami całkowitymi, na przykład: bibliologia zob. księgoznawstwo.

Na sposób formułowania i prezentacji słów kluczowych wpływają uznane lub przyjęte w danym środowisku/miejscu wzorce formułowania, redakcji i liczby słów kluczowych (odpowiednie instrukcje). Literatura przedmiotu dotycząca języków słów kluczowych określa sposoby przetwarzania materiału językowego zgodnie z ustalonymi normami językowymi i społecznymi. Poszczególne „autorskie” języki swobodnych słów kluczowych opierają się na powszechnie przyjętych zasadach pragmatyki ogólnej języków słów kluczowych oraz różniących się instrukcjach zawierających pragmatykę szczegółową danego języka. Różnice te dotyczą: liczby słów kluczowych, ich formy zewnętrznej, porządku w opisie systemowym oraz sposobów ich oddzielania od siebie. Na przykład systemowe opisy prac licencjackich i magisterskich w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego (IINiB UJ) są prezentowane w postaci tekstów stanowiących sekwencję współwystępujących swobodnych słów kluczowych. Są one generowane na podstawie „Instrukcji Instytutu INiB UJ: prace magisterskie” (<http://www.inib.uj.edu.pl>), która na s. 2 zawiera następujące wytyczne z zakresu pragmatyki szczegółowej: „Słowa kluczowe (max. 5 słów kluczowych w języku polskim oraz angielskim WERSALIKAMI, ułożonych w porządku alfabetycznym, oddzielanych od siebie pauzami)”. Identyczne wymagania są kierowane do autorów materiałów konferencyjnych publikowanych przez IINiB UJ ([175](http://www.inib.uj.edu.pl/konger/ks/inst_</p></div><div data-bbox=)

dla_autorow.htm). Liczbę pięciu słów wymieniają również podobne instrukcje (ang. *guide for authors*), na przykład instrukcja dotycząca przygotowania tekstów do publikacji w czasopiśmie „Neurocomputing” (wydawnictwo Elsevier) (<http://authors.elsevier.com/GuideForAuthors.html?PubID=505628&dc=FGA>)¹²⁸. Zwracam też uwagę na instrukcję znacznie różniącą się od dotychczas omówionych. Instrukcja ta dotyczy stosowania języka słów kluczowych w opisie prac magisterskich w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Wrocławskiego (<http://www.ibi.uni.wroc.pl/>). Zawiera ona następujące wytyczne: „Słowa kluczowe (ułożone w porządku rzeczowym według wagi poruszanych problemów) – do 2 linijek”. Jednak nie jest to już indeksowanie współrzędne. W celu zobrazowania występującej w tym względzie różnorodności można jeszcze przytoczyć przykłady wytycznych w przedmiocie słów kluczowych zamieszczane w instrukcjach wydawniczych, na przykład instrukcja wydawnicza Uniwersytetu Łódzkiego dopuszcza użycie do dziesięciu słów kluczowych. Warto wspomnieć również o indeksowaniu fasetowym, którego przykładem jest instrukcja zamieszczona w „Materials Research Bulletin” (vol. 37, no 10, 2002), opierająca się na kontrolowanej liście słów kluczowych i dopuszczająca wybór maksimum pięciu słów kluczowych z dostępnej listy i przedstawienie ich w postaci: A.alloys, A.metals, C.X-ray diffraction: tutaj litera poprzedzająca słowo kluczowe jest nazwą kategorii danego słowa (por. rozdział 3.1.3.1.3).

Aby zgłębić omawianą problematykę, poddałem analizie w zakresie stosowania autorskich słów kluczowych opisy dokumentów zamieszczone w materiałach konferencyjnych z Kazimierza Dolnego, opublikowanych w „Elektronicznym Biuletynie Informacyjnym Bibliotekarzy” (<http://www.ebib.info/publikacje/matkonf/kaba/>). Publikacja składa się z tekstów dziewiętnastu referatów bibliotekarzy i pracowników naukowych z tej dziedziny. Poszczególne dokumenty zawierają następujące elementy: autor/zy, tytuł, abstrakt, słowa kluczowe w języku polskim i w języku angielskim, tekst dokumentu. Słowa kluczowe miały być formułowane na podstawie zamieszczonej na stronie internetowej instrukcji dotyczącej języka swobodnych słów kluczowych (tzw. słów autorskich). Materiał badawczy ilustruje rys. 8.

¹²⁸ Interesujące informacje o wymaganiach dotyczących słów kluczowych w bazach danych INSPEC (Information Science for Physics, Engineering, and Computing), CAB (Current Agriculture Bibliography), ISTA (Information Science and Technology Abstracts) oraz LISA (Library and Information Science Abstracts) zawiera praca Gil-Leiva i Alonso-Arroyo (2007).



Rys. 8. Elementy opisu artykułu w publikacji elektronicznej
(źródło: <http://www.ebib.info/publikacje/matkonf/kaba/wozniak-kasperek.php>)

W celu przeprowadzenia analizy słów kluczowych sporządziłem następującą listę użytych słów kluczowych, stanowiącą słownik zadany instrukcją indeksowania:

Analiza dokumentu	Język haseł przedmiotowych KABA
Automaty klasyfikacji	Język haseł przedmiotowych RAMEAU
Badania użytkowników katalogu komputerowego OPAC	Język KABA
Bazy wiedzy semantyczne	Języki informacyjno-wyszukiwawcze
Biblioteka Jagiellońska	Jhp KABA
Biblioteka Narodowa	KABA
CCS	KABA i jhp Biblioteki Narodowej
Chemia	Kartoteka haseł wzorcowych (khw)
Doświadczenia bibliotek stosujących KABA	Kartoteka haseł wzorcowych
Geoinformacja	Kartoteka haseł wzorcowych NUKAT
Geoinformatyka	Katalog NUKAT
Geomatyka	Katalogi OPAC
GIS	Katalogowanie przedmiotowe
Hasło przedmiotowe	Katalogowanie rzeczowe
Informatyka	Kontrola rekordów
Integracja kartotek	LCSH
Integracja kartotek haseł wzorcowych NUKAT	Listy określników
Język haseł przedmiotowych	Literatura piękna
	Logi transakcyjne

Math Classification	Serwisy kontrolowanej jakości
Nauki komputerowe	Sesja użytkownika
Numery kontrolne	Składnia jhp RAMEAU
Obiekt techniczny	Słowa kluczowe
Określniki chronologiczne	Słownictwo języka haseł przedmiotowych (jhp)
Określniki formalne	Słownictwo matematyczne
Określniki geograficzne	Słowniki kontrolowane
Określniki rzeczowe	Systemy informacji geograficznej
Ontologie	Tematy techniczne
Opracowanie rzeczowe	Terminologia jhp RAMEAU
Organizacja informacji	Terminologia naukowa
Organizacja wiedzy	Typologia haseł wzorcowych
Prace zbiorowe	Wątki wyszukiwawcze
Prekoordynacja	Współkatalogowanie
Procedury	Wyszukiwanie informacji
Przeformułowania zapytań	Zmiany w jhp KABA
Semantyka jhp RAMEAU	

Pomimo klarownych wytycznych dotyczących opisu rzeczowego artykułów w 37% prac (siedem prac), zresztą recenzowanych, udało mi się zidentyfikować następujące błędy o charakterze formalnym:

1. W dwóch przypadkach ewidentnie pomyłono język słów kluczowych z językiem haseł przedmiotowych i charakterystykę rzeczową sporządzono nie za pomocą słów kluczowych, lecz haseł przedmiotowych, co ujawnia kłopoty z identyfikacją poszczególnych języków paranaturalnych;
2. W dwóch przypadkach charakterystykę dokumentów opracowano w języku naturalnym, o czym świadczą następujące sformułowania w polu „Słowa kluczowe”: „Dotychczasowe tworzenie haseł dla nazw geograficznych w NUKAT (analiza przedmiotu)”; „Aktualne uwarunkowania: format MARC21, integracja bibliotek”; „Proponowane rozwiązania i wnioski”; „KABA i jhp Biblioteki Narodowej”; „Doświadczenia bibliotek stosujących KABA”; „Zmiany w jhp KABA”. Potwierdza to kłopoty użytkowników z identyfikacją języka słów kluczowych i odróżnianiem go od języka naturalnego;
3. W wielu przypadkach mamy do czynienia ze zróżnicowanym stopniem postkoordynacji słów kluczowych;
4. Stosowano różne sposoby (formy) nazywania tego samego przedmiotu dokumentu, na przykład: „Język haseł przedmiotowych KABA”, „Język KABA”, „KABA”.

Wymienione rodzaje błędów świadczą o niskim stopniu znajomości języka słów kluczowych. Moim zadaniem nie była ocena jakości indeksowania z uwzględnieniem takich parametrów jak szerokość i głębokość indeksowania. Mogłaby ona ujawnić dalsze niekonsekwencje procesu indeksowania. Podobną analizę zbioru słów kluczowych, tyle że znacznie wcześniej, przeprowadziła Jadwiga Choroszuszyzna (Choroszuszyzna 1968).

Ważny zespół zagadnień w ramach problematyki indeksowania za pomocą słów kluczowych dotyczy indeksowania automatycznego, opartego na metodach formalnych, gdzie analiza tekstu, wybór i przekład na język słów kluczowych odbywa się

za pomocą odpowiednich programów komputerowych (metod automatycznych). W procesie automatycznego wyboru słów kluczowych stosuje się dwie metody:

- wybór oparty na liście wzorcowych słów kluczowych,
- wybór oparty na liście wzorcowych słów niekluczowych.

Listy wzorcowe słów kluczowych są stosowane tylko tam, gdzie terminologia dziedziny, dla której tworzone są indeksy, została uporządkowana i ujednoczona. Metoda ta jest stosowana stosunkowo rzadko i przeważnie w odniesieniu do wąskich dziedzin wiedzy ze względu na to, że porównywanie tytułów i słów wzorcowych przy zbyt rozbudowanych listach wzorcowych znacznie wydłuża proces indeksowania. Bardziej rozpowszechniona jest metoda listy wzorcowej słów niekluczowych oraz tzw. stop-lista.

W celu obiektywizacji procesu indeksowania tworzy się odpowiednie programy, które analizują wskazane teksty i kierując się częstotliwością występowania słów i/lub ich formą albo miejscem w zdaniu, wybierają je automatycznie. Nie zawsze jest to widoczne dla użytkownika systemu. Obecnie w celu nadania dokumentom słów kluczowych używa się coraz częściej programów komputerowych typu „tools”, które umożliwiają automatyczną obróbkę informacyjną. Jednym z nich jest Keyword Discovery Tool. Istnieje też indeksowanie półautomatyczne, w którym dokumentalista zaznacza w tekście słowa kluczowe. Następnie są one odczytywane automatycznie przez program i wprowadzane do słownika. Jest to rodzaj indeksowania derywacyjnego wspomaganego komputerem.

Języki słów kluczowych służą też do opracowania rzeczowego dokumentów sieciowych, zwanych też sieciowymi obiektami informacyjnymi, będąc jednocześnie jednym z elementów standardów metadanych¹²⁹, zapewniających ich identyfikację według kryteriów treściowych. Podstawą automatycznego indeksowania, zwykle opartego na wykorzystaniu metod statystycznych, są pola metadanych, w które wyposażony jest dokument. Istnieje wiele typów metadanych. Część z nich jest przeznaczona do kodowania informacji o treści dokumentu w postaci ciągu słów kluczowych albo wyrażen języków o słownictwie kontrolowanym. Jednak większość systemów indeksujących nie bierze ich pod uwagę przy dokonywaniu analizy statystycznej tekstu dokumentu.

W dobie taniej i coraz szybciej wzrastającej pojemności pamięci komputerów indeksowanie stało się rutynowym zbieraniem słownictwa przez stworzone specjalnie do tego celu oprogramowanie, działające jako część składowa wyszukiwarek internetowych. Wyszukiwarki operujące na tekście dokumentu same dokonują identyfikacji słów kluczowych. Większość miejsc w sieci jest dostępna poprzez strony WWW. Nawet jeśli są to programy dostępne poprzez usługę FTP czy pliki dźwiękowe udostępniane bezpośrednio z komputera, w którego pamięci są przechowywane, to prowadzące do nich odnośniki znajdują się na stronach Web. Powoduje to, że proces

¹²⁹ Za metadane uważa się zestaw logicznie powiązanych z dokumentem elektronicznym usystematyzowanych informacji opisujących ten dokument i umożliwiających jego wyszukiwanie, kontrolę, zrozumienie i długoterwale przechowywanie (projekt Rozporządzenia MSWiA w sprawie niezbędnych elementów struktury dokumentów elektronicznych, <http://bip.mswia.gov.pl/download.php?s=4&id=564>).

indeksowania sprowadza się do przeglądania przez odpowiedni program wyszukiwarki stron WWW, co daje możliwość dotarcia do niemal wszystkich źródeł w sieci.

Jak już wspomniałem w rozdziale 1.2, strona WWW składa się z dwóch podstawowych części: HEAD i BODY, które tworzone są w języku HTML lub jemu pokrewnych. Dodatkowo umieszczane są na stronach tzw. aplety (ang. *apply*) lub skrypty pisane w językach postskryptowych. Mają one za zadanie zwiększenie możliwości usług świadczonych poprzez Internet, na przykład prowadzenie działalności e-commerce. Oferują tworzenie wielkich baz danych (głównie artykułów na sprzedaż) dostępnych dla wszystkich użytkowników Internetu. Język HTML oraz skrypty odpowiedzialne są za formę, w jakiej strona ukazuje się w oknie przeglądarki po wpisaniu jej adresu, natomiast cała treść zawarta na stronie jest redagowana w edytorze tekstu i umieszczona między znacznikami języka HTML. Część HEAD jest odpowiedzialna za wstępną, ważniejszą dla indeksowania i wyszukiwania część strony. Jest ona typowym opisem bibliograficznym. Znajduje się w niej adres, tytuł, opis strony kodowej (informacje o użytych językach, a co za tym idzie także czcionce), informacje o autorze, opis dokumentu (abstrakt) oraz słowa kluczowe. Ta część strony jest bardzo ważna dla przeszukującego sieć programu, co później znajduje wyraz w nadawanych poszczególnym częściom strony wartościom w rankingu. Odpowiedni program zbiera zawarte w każdym z pól wyrazy, indeksując je jako słowa kluczowe. Uwzględnia także zawartość części BODY i wszystkich znajdujących się w niej słów, ale w ogólnym rankingu przypisuje im zdecydowanie mniejszą wartość. Słowa kluczowe są zbierane do bazy danych, będącej otwartym i wciąż rozrastającym się słownikiem. Operacje na słowach kluczowych zachodzące w bazie dokładnie odpowiadają czynnościom przeprowadzanym na elementarnych jednostkach leksykalnych w klasycznym modelu języka słów kluczowych. Każde słowo ma swoją wagę w zależności od częstości, z jaką pojawia się w bazie i na tej podstawie jest mu nadawana wartość początkowa. Do niej dodawana jest wartość określana przez miejsce występowania słowa kluczowego na stronie i w tym przypadku jest ona zależna od pomysłu konstruktora wyszukiwarki. Z reguły najwyższe wartości przypisuje się słowom kluczowym występującym w części HEAD w polach: TITLE, DESCRIPTION, KEYWORDS. Wszystkie pola w nagłówku mają status metadanych. Dwa ostatnie noszą nazwę znaczników META i nie ukazują się na witrynie. Zawartość META-kodu DESCRIPTION jest widoczna pod tytułem strony, wśród wyników wyszukiwania. Słowa kluczowe z części BODY otrzymują z reguły niższe wartości, zależne od miejsca, w którym się pojawiają i ich częstości. Jeżeli wyraz ma stać się słowem kluczowym, to wskazane jest umieszczenie go w pierwszym akapicie tekstu w części BODY (Babik 2007).

Ważną cechą indeksowania automatycznego jest liczenie częstości występowania poszczególnych słów kluczowych w dokumencie lub w całym zbiorze dokumentów, a także ustalanie, w ilu dokumentach zbioru występuje dane słowo kluczowe. Zwykle przyjmuje się założenie, że słowa występujące częściej (ale niezbyt często!) są trafniejsze i na tej podstawie przypisuje się im odpowiednie współczynniki wagowe. Wagą jest liczba wystąpień słowa w dokumencie, jednak najczęściej współczynnik wagowy wyraża stosunek częstości występowania danego słowa w dokumencie

lub zbiorze do ogólnej liczby słów w dokumencie lub do wystąpień tego słowa we wszystkich dokumentach zbioru. Współczynniki te umożliwiają w procesie wyszukiwania informacji ustalenie, z jakim prawdopodobieństwem dany dokument/informacja odpowiada pytaniu informacyjnemu.

Rozważania dotyczące automatycznego indeksowania dokumentów za pomocą słów kluczowych zakończę zasygnalizowaniem istnienia rozszerzonej metody indeksowania opartej na programie KeyGraph, pozwalającym badać występowanie terminów wraz z ich kontekstem, opisanej przez trzech badaczy japońskich: Naohiro Matsumurę, Yukio Ohsawę i Mitsuru Ishizukę (Matsumura, Ohsawa, Ishizuka 2003). Zasada działania tego programu opiera się na wykorzystaniu procesów poznawczych zachodzących w mózgu człowieka i związanym z tym efektem torowania (ang. *priming effect*). Polega on na tym, że terminy używane przez autora jakiegoś tekstu są ugruntowane w jego świadomości i wywołują określone skojarzenia. Terminy te tworzą w mózgu powiązania nazywane siecią aktywacji i są w efekcie włączane do tekstu w tym kontekście. W sieci uaktywnianych jest kilka terminów podstawowych, a następnie od tych terminów rozwijane są powiązania do kolejnych pokrewnych terminów. Opracowano nawet specjalny algorytm PAI, który symuluje efekt torowania, powstający w czasie czytania w umyśle odbiorcy tekstu. Algorytm ten przetestowano na materiale, który stanowiły artykuły naukowe i referaty konferencyjne. Przeprowadzony eksperyment pokazał, że PAI może „wylapywać” reprezentatywne słowa kluczowe niezależnie od częstości ich występowania w tekście.

Z przeprowadzonych rozważań oraz przyjętej w rozprawie szerokiej definicji języka słów kluczowych wynika, że opis dokumentu, generowany na podstawie indeksowania automatycznego obejmującego analizę frekwencji i ujednoczenie form wyrażań oraz ewentualnie ich ważenie na podstawie własności tekstowych, jest tekstem języka słów kluczowych. Język, w którym tworzy się charakterystykę słowną dokumentu, jest bardzo podobny do tego, którym operują użytkownicy w wyszukiwaniu i tego, który staje się językiem reprezentacji treści tekstu po ich automatycznym indeksowaniu.

Ważna w indeksowaniu, zarówno automatycznym jak i przeprowadzanym przez człowieka (intelektualnym), jest wspomniana już stop-lista, czyli lista (baza danych) słów uznanych za nieznaczące. Taka lista jest dostępna w Internecie. Wyrażenia znajdujące się na niej nie są używane w indeksowaniu z powodu zbyt częstego pojawiania się w tekście bądź z powodu ich kategorii leksykalnej. Należą do nich przede wszystkim takie części mowy jak: przedimki, zaimki, spójniki oraz niosące z sobą niską wartość informacyjną wyrazy typu: „wpływ”, „zagadnienie”, „proces”. Mogą one natomiast pojawiać się we frazach, czyli wśród kilku słów kluczowych ujętych w cudzysłów i stanowiących nierozzerwalną całość. Znając metody indeksowania wykorzystywane przez wyszukiwarki i stosując się do nich, można znacznie zwiększyć wartość, jaka zostanie przyznana stronie w rankingu, a to wszystko za sprawą umiejętnie użytych i odpowiednio umiejscowionych słów kluczowych.

Zajmę się teraz katalogami internetowymi, czyli spisami zawartości dokumentów znajdującymi się w sieci, tworzonymi nie przez komputer, lecz przez indeksatorów. O ile „komputer” zbiera wszelkie słowa kluczowe pojawiające się w dokumencie i na

podstawie zadanego algorytmu nadaje im określoną wartość, o tyle proces indeksowania przez człowieka opiera się na jego doświadczeniu w tej kwestii. Oczywiście, sam proces gromadzenia słownictwa jest wspomagany przez odpowiednie programy, ale wybór słów znaczących należy do człowieka. Musi on dokładnie przejrzeć dokument i ustalić dziedzinę lub dziedziny, o jakich traktuje. Następnie zgodnie z tym przyporządkowaniem wybiera z dokumentu lub ze słownika odpowiednie słowa kluczowe, tak aby stanowiły one w miarę dokładne odzwierciedlenie poszczególnych elementów treści dokumentu. Tu właśnie tkwi zasadnicza różnica między indeksowaniem automatycznym a indeksowaniem intelektualnym. Programy komputerowe dokonują indeksowania według słów kluczowych wartościowanych na podstawie zadanego algorytmu, natomiast indeksatorzy według relewantnych elementów treści, opierając się na własnym doświadczeniu.

Indeksowanie intelektualne jest indeksowaniem intuicyjnym, w którym analiza tekstu, wybór słów kluczowych i ewentualny przekład na język słów kluczowych opierają się na wiedzy i intuicji człowieka, co ogranicza możliwości wyszukiwawcze i kieruje użytkownika na tory myślowe, którymi kroczył przed nim indeksujący, antycypując sposób zachowania się wyszukiwujących.

Przykładem narzędzia językowego wspomagania indeksowania stron internetowych może być polski słownik językowy będący efektem pracy firmy TiP Sp. z o.o. z Katowic¹³⁰, który powstaje od końca lat osiemdziesiątych. Jego zawartość (co do form podstawowych) jest porównywalna ze słownikiem Witolda Doroszewskiego. Odpowiedni program komputerowy rozpoznaje za jego pomocą m.in. wyrazy złożone (biało-czerwony), liczebniki rzymskie (MCMLXXVII), skróty (por., prof., dr, kmdr), skrótowce (PO, PZU), przymiotniki o charakterze liczebnikowym (tysiąćpięćsetpięćdziesięciokilometrowy). Każda forma wyrażenia jest porównywana ze wzorcem odpowiedniej odmiany, formy pochodne wynikają z rodzaju części mowy. Słownik ten jest w pełni fleksyjny, co oznacza, że dla każdego wyrazu zawiera:

- zbiór form hasłowych wyrazu;
- zbiór reguł tworzenia wyrazów i poszczególnych ich form;
- zbiór reguł przekształcających formę hasłową w zbiór form fleksyjnych konkretnego wyrazu;
- procedurę rozpoznawania wyrazów w tekście.

Jako kryterium dodawania nowych jednostek do słownika przyjęto kryterium statystyczne oparte na częstości ich występowania w tekstach. Można więc nie znaleźć w nim niektórych neologizmów, które pojawiły się w języku polskim. Oprócz tego słownik zawiera nazwy własne, wśród których jest większość znanych imion, nazw geograficznych polskich i światowych, nazwisk historycznych (por. www.szukaj.onet.pl). Trudności nie stanowi też wyszukiwanie słów w tym słowniku, ponieważ dzieje się to w sposób czysto mechaniczny na zasadzie rozpoznawania znaków przez odpowiednio skonstruowany program komputerowy. Dzięki temu za pomocą słów kluczowych, używając dowolnej ich liczby, można opisać wszelkie dokumenty ukazujące się w Internecie. Co do liczby słów kluczowych, przyjmuje się, że dookreśle-

¹³⁰ Zob.: <http://www.tip.net.pl/>.

nie poprzez użycie dużej liczby słów kluczowych daje lepsze rezultaty¹³¹. Z drugiej strony, użycie unikalnych słów powoduje, że nie spełniają one swojej funkcji. Wiele katalogów internetowych ogranicza liczbę słów kluczowych do 25, tak więc opis zawartości strony powinien zmieścić się w tym limicie, chociaż większa liczba i lepsze doprecyzowanie daje lepsze rezultaty.

Dotychczasowe rozważania uzupełniam informacją, że w Internecie także funkcjonują wyszukiwarki, które indeksują nie tylko tytuły, ale i pełne teksty dokumentów, tj. teksty artykułów naukowych i publikacji dostępnych w Internecie. Darmową wyszukiwarką działającą w ten sposób jest na przykład Google Scholar (www.scholar.google.com).

Reprezentacja tekstów języka naturalnego za pomocą zbioru wyizolowanych wyrażeń derywowanych z tekstu lub arbitralnie wyznaczanych słów kluczowych często prowadzi do powstawania fałszywych koordynacji, którym w systemach tradycyjnych zwykle zapobiega się poprzez wykorzystanie w indeksowaniu prekoordynowanych związków wyrazowych ustalonych przez indeksującego lub przez konstruktora systemu leksykalnego języków o słownictwie kontrolowanym. Większość metod automatycznej identyfikacji terminów opiera się na technikach statystycznych. Zwykle odwołują się one do dwóch parametrów formalnych: frekwencji współwystąpień pary wyrazów oraz stopnia ich sąsiedztwa (odległości) w tekście. Metody te jednak nie są niezawodne i nie zabezpieczają przed wszystkimi fałszywymi koordynacjami oraz nie zapobiegają generowaniu związków wyrazowych o niskiej reprezentatywności dla treści dokumentu. Przykładem może być eksperymentalny system automatycznego indeksowania frazowego opracowany przez Gerarda Saltona dla Online Computer Library Center (OCLC). Do identyfikacji i generowania reprezentatywnych fraz kluczowych wykorzystuje on analizę syntaktyczną (Salton 1988). System ten generuje z tekstu w języku naturalnym znaczniki frazowe dla istotnych elementów tekstu (śródtytuły, zdania części wprowadzającej, zdania części konkluzyjnej) i ekstrahuje z nich frazy nominalne. Frazy te następnie przekształca na zespół fraz kluczowych tworzących reprezentację tekstu, wykorzystywanych w wyszukiwaniu w bazach danych. Stosowanie wyłącznie analizy syntaktycznej nie zapobiega powstawaniu fraz nieużytecznych dla reprezentowania treści dokumentów. Stąd w systemie Saltona analizę struktur frazowych tekstu w języku naturalnym uzupełniono analizą semantyczną.

Od niedawna w technologii Web 2.0 do charakterystyki tradycyjnych i internetowych obiektów cyfrowych stosuje się tzw. folksonomie¹³². Pozwalają one w nowy sposób interpretować relacje między użytkownikiem a zasobami internetowymi oraz usługami oferowanymi przez ten system. Użytkownik ma tu większą możliwość

¹³¹ W Internecie zdarza się, że autorzy opisują dokumenty za pomocą zbyt wielu słów kluczowych o wątpliwej relewancji i niskiej (ze względu na liczbę słów) wadze wyszukiwawczej. J. Woźniak-Kasperek podaje przykład dokumentu zaindeksowanego za pomocą aż 330 słów kluczowych (por. Woźniak 2000).

¹³² "Folksonomies (a term made up from folk and taxonomy) allows Internet users to index digitizer documents in order to access them again more easily through a system of key words, known as tags" (Le Deuff 2006). Folksonomia (termin utworzony z dwóch wyrazów: *folk* i *taxonomy*) pozwala użytkownikom Internetu indeksować dokumenty elektroniczne w celu ułatwienia dostępu za pomocą słów kluczowych zwanych tagami).

wpływania na liczbę i rodzaj udostępnianych mu informacji, stając się nie tylko aktywnym konsumentem, lecz również aktywnym twórcą. Wykorzystując intelektualną aktywność użytkownika, folksonomie angażują go do charakterystyki tych obiektów, tj. tworzenia tekstów w postaci ciągu niekontrolowanych (swobodnych) słów kluczowych reprezentujących treść dokumentu. Folksonomie są wspólnym narzędziem katalogowania dokumentów graficznych, dźwiękowych, audiowizualnych, hipertekstowych, a także tradycyjnych. Wykaz użytych słów kluczowych tworzy indeks o strukturze hipertekstowej w postaci tzw. chmury tagów, powiązanych z terminem wyszukiwawczym (skojarzeniami, terminami w innych językach). Każde z użytych słów kluczowych otrzymuje status węzła hipertekstowego, którego aktywacja powoduje wyodrębnienie podzbioru charakterystyk obiektów cyfrowych. W ten sposób powstaje jakby równoległa (tym razem z punktu widzenia użytkownika) charakterystyka danego obiektu internetowego, będąca przejawem społecznego klasyfikowania obiektów cyfrowych zasobów WWW.

Opis jednostki katalogowej przez użytkownika stanowi więc wartość dodaną do całego opisu. Mamy tu do czynienia z pewnego rodzaju polireprezentacją informacji. Z jednej strony jest to zobiektywizowany i sformalizowany opis bibliograficzny wraz z charakterystyką wyszukiwawczą, tworzoną i wyrażaną w jednym lub wielu językach informacyjno-wyszukiwawczych, z drugiej zaś strony jest to ciąg słów kluczowych, w istocie wybranych wyrażen języka naturalnego w funkcji metainformacyjnej jako produkt percepcji dokumentu przez użytkownika.

Analiza leksykalna zbioru słów kluczowych tworzących chmurę/zestaw tagów, przeprowadzona przez Louise F. Spiteri (Spiteri 2007) z Toronto (Kanada), ujawniła duże nacechowanie emocjonalne oraz subiektywizm w reprezentacji cech i form kolekcji, a przede wszystkim niekonsekwencje w stosowaniu określonej formy słów kluczowych. Subiektywizm i formalna niespójność nie stanowią jednak przeszkody do uznania tego narzędzia za wartościowe źródło informacji.

Wśród języków słów kluczowych folksonomie wyróżniają się tagami. Olivier Le Deuff określił je jako „słowa kluczowe, kategorie nazw lub metadane (...) tag jest po prostu swobodnym słowem kluczowym. Ponieważ tagi nie są tworzone przez specjalistów do spraw informacji, nie stosują się do wskazówek formalnych. To oznacza, że mogą być skategoryzowane przy pomocy jakiegokolwiek słowa określającego związek między źródłem online i pomysłem użytkownika” (Le Deuff 2006). Aby stworzyć tag opisujący dany dokument, autor może wybrać którekolwiek z nieskończonej liczby słów i to niezależnie od stopnia jego trafności. Tak więc słowo kluczowe tego typu zależy od stopnia kultury użytkownika i stopnia opanowania języka.

Olivier Le Deuff twierdzi ponadto, że „folksonomia jest jednocześnie tym, co najlepsze i tym, co najgorsze w zarządzaniu informacją”. Folksonomie mają więc zarówno zalety, ale też i wady. Do pierwszych Le Deuff zalicza:

- możliwość współtworzenia systemów przez amatorów, co w przyszłości może stać się pierwszym poziomem porządkowania Internetu;
- możliwość tworzenia rankingów najpopularniejszych tagów, co pozwala określać najchętniej odwiedzane zasoby WWW i dzięki temu poznać szybkość rozprzestrzeniania się informacji w Internecie;

- w aspekcie językowym folksonomia pozwala zauważyć, jakie pola leksykalne są dominujące (najczęściej stosowane);
- indeksowanie przez użytkowników nie generuje kosztów finansowych.

Le Deuff, powołując się na autorytety z dziedziny nauki o informacji oraz ekonomii, do wad tego typu systemów zalicza:

- „astroturfing”, czyli formę promocji polegającą na tworzeniu nieprawdziwych informacji zachwalających produkty oraz manipulowanie wynikami rankingów;
- możliwość mylnego utożsamiania tagów z „prawdziwymi” metadanymi;
- absolutną dowolność w doborze słownictwa;
- słowa kluczowe stosowane w folksonomii nie pojawiają się w jakimkolwiek godnym zaufania teaurusie;
- nieprawidłowe tworzenie systemów przez absolutnych amatorów;
- utożsamianie wyników rankingów z pertynencją.

Krytyczne uwagi pod adresem folksonomii warto uzupełnić opinią wspomnianej już L.F. Spiteri. Wśród zalet folksonomii wymienia ona to, że pojęcia są zapisywane w języku zrozumiałym dla ogółu, że powstają nowe skojarzenia, że istnieje możliwość wybrania najlepszych znaczników, że procedura jest elastyczna, zaś za wady uważa niejednoznaczność, niejednorodność, wykorzystanie homonimii i synonimii, błędy ortograficzne i niestosowanie się do reguł serwisów (Spiteri 2007).

Sposobem na unikanie niekorzystnych skutków folksonomii jest odpowiednia edukacja użytkowników Internetu. Efektem takiej edukacji – zdaniem Le Deuffa – powinno być stosowanie następujących reguł indeksowania za pomocą tagów:

- używanie liczby mnogiej dla określenia kategorii (liczba mnoga jest bardziej właściwa, gdyż kategorie mogą zawierać różne zmienne);
- niestosowanie wielkich liter, chyba że inaczej zrozumienie słowa byłoby niemożliwe;
- obserwacja i używanie stosowanych już zasad indeksowania stron i sieci.

Mechanizmy folksonomii są wykorzystywane do organizacji i udostępniania informacji w tematycznych portalach społecznościowych w Internecie i stanowią ważne źródło informacji o jego zasobach. Do tej grupy należy m.in. serwis społecznościowy KO-SWICKI (<http://ko-swicki.eurekster.com/>), który działa w ramach brytyjskiego oddziału International Society for Knowledge Organization (ISKO) (<http://www.iskouk.org/>).

Niewątpliwie folksonomie i tagowanie¹³³ należą do nowych metod opisu dokumentów w Web 2.0. Metody te polegają na określaniu elementów treści dokumentu przez ich użytkowników za pomocą samodzielnie swobodnie tworzonych słów kluczowych. Zjawisko to, nazywane w języku angielskim „social bookmarking”, jest także ważne, ale i niebezpieczne w procesach wyszukiwania informacji, gdyż spowodować może spadek jakości informacji – trudno bowiem oczekiwać od wszystkich użytkowników znajomości standardów opisu. Można temu w pewnym stopniu zapobiec poprzez stosowanie prostych i jasnych formularzy oraz słownictwo znor-

¹³³ Tagowanie – nadawanie przez użytkowników elementom kolekcji etykiet w postaci swobodnych słów kluczowych.

malizowane (kontrola słownictwa). Pozostaje jednak problem wiarygodności wprowadzanych danych (selekcja słów kluczowych).

Trzeba pamiętać, że problem identyfikacji słów kluczowych na podstawie automatycznej analizy tekstów nie wiąże się tylko z Internetem. De facto metodami stosowanymi dziś w indeksowaniu zasobów internetowych w części lingwistycznej posługiwano się już od początku lat siedemdziesiątych XX wieku w systemach pełnotekstowych. To samo dotyczy narzędzi wyszukiwawczych wykorzystujących algebrę Boole'a i operatory kontekstowe.

3.2.3. Pragmatyka wyszukiwania za pomocą słów kluczowych

Większość uwag odnoszących się do procesu indeksowania daje się przez analogię odnieść do wyszukiwania informacji, z tym że uwagi te trzeba odnosić do sposobów formułowania instrukcji wyszukiwawczej. I tak, wyszukiwujący informację również określa przedmiot/treść, tyle że swoich zainteresowań, potrzeby informacyjnej, luki informacyjnej. Różnica polega na tym, że użytkownik tworzy opis swojej potrzeby, który to opis ma charakter dynamiczny, tzn. zmienia się w trakcie procesu wyszukiwania informacji i może być ciągle modyfikowany w zależności od sytuacji oraz dostępnych w systemie narzędzi i środków wyszukiwawczych, na przykład algebry Boole'a czy metod obcinania końcówek, które to narzędzia nie należą do języka słów kluczowych, aczkolwiek go wspomagają.

Słowa kluczowe są stosowane w alfabetycznych wyszukiwarkach współpracujących z systemami automatycznego indeksowania posługującymi się szeroką gamą technik wyszukiwania pełnotekstowego. Zasady ich działania opierają się na przekonaniu – zresztą charakterystycznym dla pierwszego okresu rozwoju Internetu – że najbardziej efektywną organizację i wyszukiwanie informacji zapewniają programy indeksowania automatycznego i techniki wyszukiwania wzorowane na doświadczeniach systemów pełnotekstowych. Dostęp do Internetu zależy od zasad konstrukcji pytania informacyjnego oraz kryteriów wyszukiwawczych przyjętych w danym systemie. Jest najczęściej realizowany w trybie uproszczonym, w którym użytkownik werbalizuje swoją potrzebę informacyjną i przedstawia ją w postaci ciągu swobodnych słów kluczowych. W tym trybie użytkownicy zaawansowani dodatkowo wykorzystują narzędzia działania na zbiorach, m.in. algebrę Boole'a (na przykład w wyszukiwarce Google).

Zmiany, jakie są dokonywane w słowach kluczowych, aby mogły one sprawnie funkcjonować i być efektywne w wyszukiwaniu informacji, prześledzę na przykładzie budowy ich poszczególnych poziomów. W warstwie elementarnych znaków graficznych używa się dodatkowych znaków (z reguły interpunkcyjnych) do zastąpienia symboli, których użytkownik nie jest pewien, wtedy gdy chodzi o poprawną pisownię, np. p?zenica = pszenica. Częściej symbol ten używany jest w maskowaniu, kiedy to użytkownik ogranicza się tylko do rdzenia wyrazu, a jego końcówki pozostawia do wyboru narzędziu wyszukiwawczemu. Pozwala to uzyskać w ramach jednego zapytania szereg słów kluczowych o wspólnym temacie, mających zróżnicowaną

wartość informacyjną. Wiele wyszukiwarek taką funkcję przypisuje symbolowi „?”, stawianemu po temacie wyrażenia. Przykładem zastosowania „?” może być: bibliot? = biblioteka, biblioteczny, bibliotekarz, bibliotekarski, bibliotekoznawczy. Niektóre z wyszukiwarek oferują możliwość ustalenia liczby znaków, która może pojawić się po symbolu „?”, co jest podawane za pomocą cyfr 1–9, np. bibliot?5 = bibliotekarz, biblioteczny. Inne wyszukiwarki stosują symbol „*”, który określa liczbę nieznanymi znaków (do pięciu) występujących po temacie, np. bibliot* = biblioteka, biblioteczny, bibliotekarz, ale słowa: „bibliotekarski” i „bibliotekoznawczy” oraz wszystkie, które mają na końcu więcej niż pięć znaków, nie ukażą się już na liście wyników. Różnica polega na funkcji nadawanej każdemu z tych symboli i zależy od konstruktorów wyszukiwarki. Z reguły takie informacje znajdują się w systemach pomocy.

Bardzo często użytkownicy stosują symbol „*” w miejsce końcówki liczby mnogiej poszukiwanych przez siebie wyrazów, uważając, że jest to sposób na zwiększenie szybkości poszukiwań. Nie znając do końca funkcji tego symbolu, otrzymują więcej wyników nierelevantnych do swego zapytania. Symbol „?” jest używany nie tylko w znaczeniu „wszystkie możliwe końcówki danego tematu”, lecz także jako „znak zastępujący jedną i tylko jedną literę lub cyfrę”. W tej funkcji występuje z reguły w środku wyrazu. Oto przykład: prz?szłość = przyszłość, ale również przeszłość. Podobnie jest w przypadku symbolu „*”, który może zastępować do pięciu znaków, np. sy*yjski = syryjski, syberyjski.

W wyszukiwaniu ważne jest także, czy słowa zapisywane są z użyciem tzw. małych i wielkich liter. Istotne jest to na przykład w wyszukiwaniu akronimów. Niektóre z wyszukiwarek stosują użyteczny wskaźnik „u/c”, np. aids u/c = AIDS, lub onz u/c = ONZ. Jeżeli wyszukiwarce zostanie podane słowo zaczynające się wielką literą, to zostanie przez nią uwzględnione, czego skutkiem będzie zawężona liczba wyników i większa relewancja, natomiast używanie małych liter przy wyszukiwaniu dokumentów za pomocą nazw własnych, bez uwzględnienia wielkich liter, daje skutek odwrotny do zamierzonego.

Słowa kluczowe bywają też opatrywane tzw. wagami, zależnie od przypisywanej im wartości. Im słowo kluczowe rzadziej występuje w języku, i jednocześnie niesie dużą wartość informacyjną, tym jego waga jest większa. W praktyce, mając na względzie użytkowników, próbuje się to wypośredkować. Używanie słów wyłącznie ze słownika wyrazów obcych może spowodować, że strona nie pokazuje się wśród wyników wyszukiwania, ponieważ użyte zbyt unikalne słowa nie zostaną ponownie zastosowane w ramach kwerendy. Z drugiej strony, nie mogą to być słowa używane zbyt często, a w żadnym przypadku umieszczone na stop-liście, co powoduje pominięcie ich w procesie indeksowania. Wyszukiwarki internetowe stosują zerową kontrolę w znakowaniu kojarzeniowym, czyli wybierają słowa bezpośrednio z tekstu bez zwracania uwagi na poprzednio użyte słowa kluczowe. Generuje to dużą liczbę synonimów, homonimów oraz polisemię słów kluczowych. Problem ten jest wciąż ważny dla stosujących metodę indeksowania swobodnego. Starają się oni te niedoskonałości językowe wyeliminować. W indeksowaniu automatycznym istnienie synonimii i homonimii umożliwia uzyskanie większej liczby słów kluczowych, wśród których wyższą wartością można wyróżnić te rzadziej używane.

Na kolejnym, ostatnim poziomie języka słów kluczowych następuje łączenie słów kluczowych w zdania. Prawdziwa siła i realna przyczyna używania słów kluczowych w wyszukiwaniu informacji w Internecie leży w możliwościach ich kombinowania. Stosuje się tu dwie metody: metodę operatorów boole'owskich¹³⁴ oraz metodę operatorów pozycyjnych.

W Internecie dba się o łatwość dostępu do poszczególnych informacji, co umożliwiają narzędzia systemowe. Tworzone są różnego rodzaju indeksy, spisy treści, środki do wyszukiwania kontekstowego, a także w szerokim zakresie stosowane są techniki hipertekstowe. W wyszukiwaniu informacji wykorzystuje się złożone formuły logiczne oraz technikę warunków „rozmytych” (na przykład operator bliskości NEAR). Wyszukiwanie informacji historycznych ułatwia narzędzie TIMELINE, a kontrolę zależności geograficznych – zintegrowany ATLAS.

W polskim piśmiennictwie krytyczne uwagi dotyczące wyszukiwania dokumentów w Internecie za pomocą słów kluczowych znaleźć można m.in. w artykule Thomasa J. Froehlich'a pt. *Kryteria oceny systemów wyszukiwawczych w Internecie* (Froehlich 2000). Podniósł on m.in. takie problemy, jak wykorzystywanie skrótów terminów wyszukiwawczych i związany z tym szum informacyjny, stosowanie fraz kluczowych, czyli wyrażeń złożonych z więcej niż jednego słowa, słowa wymagane lub zabronione, słowa nieistotne, które powinny być zignorowane przy wyszukiwaniu.

W wyszukiwarkach internetowych są stosowane różne szczegółowe zasady formułowania instrukcji wyszukiwawczych. Wśród nich jest możliwość formułowania pytań prostych, rozumianych jako ciąg słów kluczowych, analizowanych później indywidualnie, albo tzw. pytań zaawansowanych, czyli złożonych, wykorzystujących operatory logiczne. Pytania złożone są jednak często budowane błędnie, gdyż użytkownicy mechanicznie przenoszą interpretacje spójników języka naturalnego na operatory logiczne, co w rzeczywistości nie jest uprawnione, stąd pytanie o x i y faktycznie jest pytaniem o x **lub** y , na przykład: pytanie o oferty wakacyjnego wypoczynku we Włoszech i w Grecji faktycznie są pytaniem o wypoczynek we Włoszech **lub** w Grecji. Tylko w rzadkich przypadkach może chodzić użytkownikowi o x i y jednocześnie, to znaczy w tym wypadku o wypoczynek w obu krajach w trakcie tego samego urlopu. Równie częstym błędem jest formułowanie wyrażeń logicznych w oknie, którego zawartość jest interpretowana jedynie jako ciąg wyrazów, a wszelkie symbole i nawiasy są ignorowane albo traktowane jak każdy inny wyraz języka naturalnego. Ze względu na stosowanie różnych strategii wyszukiwawczych wyrażenia w głównym oknie wyszukiwarki bywają zamieniane na instrukcje wyszukiwawcze w postaci alternatywy (na przykład AltaVista, Excite), albo, coraz częściej, na koniunkcję (na przykład Google, Netoskop). Rezultaty wyszukiwań są więc nieporównywalne.

W przypadku języka polskiego dodatkowy problem stwarza fleksja i znaki diakrytyczne. Rozpoznawanie form fleksyjnych wyrazów języka polskiego rozwiązano

¹³⁴ Algebra Boole'a pozwala na specyfikację relacji sytuacyjnych. Robi się to w tekstach opisujących potrzeby informacyjne (język wyszukiwawczy). W takich opisach nie podaje się relacji systemowych, lecz wyłącznie relacje sytuacyjne.

w wyszukiwarce portalu ONET. Rozpoznaje ona nie tylko różne formy gramatyczne szukanych słów, lecz także potrafi znaleźć ich bliskoznaczne odpowiedniki. Jednak w niektórych wyszukiwarkach najprostszym, ale i naiwnym, rozwiązaniem problemu znaków diakrytycznych w języku polskim jest ich pomijanie. W konsekwencji powoduje to nierozróżnialność m.in. takich wyrażen jak „sądownictwo” i „sadownictwo”.

Ranking dokumentów w odpowiedzi jest ustalany przez wyszukiwarki według odpowiednich kryteriów. Brane są pod uwagę takie elementy, jak: długość adresu strony; liczba odsyłaczy do stron z innych stron świadcząca o popularności strony; zawartość odpowiedniego wyrażenia w polach poprzedzonych metaznacznikami: TITLE, DESCRIPTION, KEYWORDS; słowa znajdujące się w tekście; odległość wyrazów. Występujące czasami tzw. linki sponsorowane najczęściej wprowadzają użytkownika w błąd.

W cyberprzestrzeni czołowe miejsce w rankingu zajmują portale internetowe, które tworzą tysiące stron. Ilustrację tego może stanowić portal The ESPN Sports (<http://espn.com>), w skład którego wchodzi ponad 100 000 stron, mających kilka milionów wejść dziennie (<http://www.starwave.com/starwave/about.espn.html>). Konkurencja z takimi gigantami staje się niemożliwa, stąd sposobem na umieszczenie strony wyżej w rankingu bywa użycie rzadziej stosowanego słowa kluczowego. Wprawdzie pociąga to za sobą mniejszą liczbę odwiedzin, ponieważ większość internautów używa najpopularniejszych słów, ale mimo wszystko pozwala na zauważenie tej witryny w sieci i dotarcie do niej użytkownika, który – jeżeli strona będzie mu odpowiadała – to do niej i tak dotrze.

Kolejność stron internetowych w odpowiedzi w przypadku wyszukiwarek na ogół nie świadczy ani o ich relewantności, ani o jakości. Internet przenosi ciężar troski o jakość odpowiedzi (relewantność wyszukanych stron) z mechanizmu indeksowania na mechanizm wyszukiwania, czyli na użytkownika, który nie zawsze jest świadomy faktycznych możliwości informacyjno-wyszukiwawczych sieci. Ważną rolę odgrywają też kompetencje problemowe, językowe i informatyczne użytkowników systemu informacyjno-wyszukiwawczego, a także ich nawyki heurystyczno-wyszukiwawcze.

Klasyczny język słów kluczowych nie pozwala na odwzorowywanie informacji o zdarzeniach opisanych w dokumentach. W języku wyszukiwawczym odwzorowywanie związków zachodzących między pojęciami oraz między reprezentującymi je wyrażeniami redukowane jest do formuł boole’owskich, w których rodzaj połączeń wyrażonych przez operatory logiczne lub kontekstowe wyznacza sam użytkownik na podstawie własnej wiedzy.

Badania prowadzone nad sposobem wyszukiwania informacji za pomocą języka słów kluczowych z użyciem operatorów Boole’a wykazały małą efektywność tego wyszukiwania. Przyczyny są następujące:

- słowa kluczowe, występujące w tekstach dokumentów, mają różne znaczenia, zależne od kontekstu i dziedziny wiedzy;
- te same słowa w różnych wypowiedziach oznaczają różne pojęcia;
- różne słowa mogą oznaczać te same pojęcia.

Obecnie istnieją już narzędzia i programy, które pozwalają użytkownikowi, nawet gdy nie posługuje się fachowym językiem wyszukiwawczym, znaleźć w systemie informacyjno-wyszukiwawczym to, co chce, na przykład platforma EBSCO HOST 2 oferuje nowy interfejs użytkownika, dostępny od lipca 2008 roku, który dostarcza takich opcji, jak „poszukuję wszystkich słów”, możliwość inteligentnego tekstu (Smart Text), zaś Google – „podobne do”. Komputer sam przypisuje słowom kluczowym odpowiednie wagi lub pozwala użytkownikowi nadawać wagi poszczególnym słowom kluczowym. Użytkownik ma też możliwość określania stopnia relewancji, tj. stopnia powiązania danego słowa kluczowego z treścią dokumentu/pytania informacyjnego.

Prowadzone na świecie badania pokazują, że wiele problemów wyszukiwawczych wynika z charakteru samego języka, na przykład język angielski, który obecnie dominuje w wyszukiwaniu informacji, generuje wiele wyrażen wieloznacznych, natomiast w przypadku języka francuskiego jest mniej problemów z wyszukiwaniem informacji w Internecie. Jeszcze inaczej jest w języku niemieckim (por. Schmetzke 1998). W języku polskim występują problemy charakterystyczne dla języków fleksyjnych.

Wyszukiwanie informacji za pomocą języka słów kluczowych sprowadza się do poprawnej werbalizacji potrzeby informacyjnej i wykorzystywania przez użytkownika mechanizmów wyszukiwawczych systemu. Jakość wyników wyszukiwania zależy przede wszystkim od jakości stosowania języka informacyjno-wyszukiwawczego (w trybie indeksowania uogólniającego lub w trybie indeksowania wyszczególniającego). Tego dowiodły eksperymenty cranfieldzkie¹³⁵. Przy ocenie istotne jest określenie, co rozumiemy przez jakość wyszukiwania. Najczęściej jakość wyszukiwania utożsamia się z relewancją techniczną¹³⁶ albo z pertynencją¹³⁷.

Należy mieć świadomość, że języki słów kluczowych pozwalają uzyskać odpowiedzi zarówno na szerokie, przekrojowe, jak i ukierunkowane dyscyplinarnie pytania informacyjne. Jest to także odpowiednie narzędzie wyszukiwawcze dla pytań informacyjnych dotyczących zagadnień bardzo szczegółowych. Zależy to od rozbudowy konkretnego języka słów kluczowych.

W języku wyszukiwawczym najczęściej wykorzystuje się do indeksowania pytań informacyjnych wyrażenia i związki wyrazowe języka naturalnego bez normalizacji i eliminowania ich niejednoznaczności. Przykładem takiego systemu jest indeks permutacyjny. W celu zapewnienia dostatecznej jakości wyszukiwania informacji, przy formułowaniu instrukcji wyszukiwawczych niejednoznaczność użytych wyrażen i związków wyrazowych, a także różnorodność ich form gramatycznych jest rekompensowana za pomocą wspomnianych operatorów algebry Boole’a, operatorów

¹³⁵ Dotyczyły one badań porównawczych różnych typów języków informacyjno-wyszukiwawczych. Najbardziej znane są dwa eksperymenty przeprowadzone w Cranfield (Wielka Brytania) w latach sześćdziesiątych XX wieku (por. Richmond 1963; Lancaster, Mills 1964; Cleverdon 1967).

¹³⁶ Relewancja techniczna to „relacja, która w zbiorze wyszukiwawczym, na przykład w zbiorze charakterystyk wyszukiwawczych dokumentów, danego systemu informacyjno-wyszukiwawczego wyznacza taki podzbiór informacji, że dla każdej z nich spełniony jest warunek nie mniejszej zgodności (podobieństwa) z instrukcją wyszukiwawczą niż ustalona dla danego systemu” (Bojar 2002, oprac., s. 231).

¹³⁷ Pertynencja, relewancja pragmatyczna to relacja wyznaczająca stopień użyteczności informacji relewantnej technicznie i/lub semantycznie ze względu na zainteresowania użytkownika (por. Bojar 2002, oprac.).

arytmetycznych – na przykład nakładających warunek wartości danego pola (dokumenty wydane po roku 2000) – oraz operatorów pozycyjnych. Te ostatnie służą do określania odległości między wybranymi wyrażeniami, tj. ich położenia względem siebie w poszukiwanych tekstach, na przykład w tytułach dokumentów, abstraktach, hasłach przedmiotowych. Funkcję taką pełni również maskowanie.

Obecnie w wyszukiwaniu informacji można zauważyć tendencję nieujednolniczenia formy gramatycznej słów kluczowych. Jest to spowodowane wzrastającymi możliwościami technicznymi w Internecie, co sprowadza się do technik lewo- i/lub prawostronnego obcinania/maskowania. Jednostki leksykalne używane w wyszukiwaniu są często maskowane lewo- i prawostronnie o określoną liczbę znaków, co pozwala na wyszukiwanie wyrażen różniących się tylko końcówkami (różne przypadki, formy przymiotnikowe) lub przedrostkami. Są też systemy, które dopuszczają możliwość maskowania wewnątrz wyrazu. Ten sposób postępowania jest szczególnie użyteczny w systemach, w których stosuje się wyszukiwanie za pomocą swobodnych słów kluczowych. Podnosi on bowiem znacznie wskaźnik kompletności. Omawiany sposób postępowania jest wspomagany możliwością przeglądania (wertowania) indeksu i wyboru odpowiedniej formy wyrażenia. Mechanizm maskowania wykorzystuje również stosowane w wielu systemach tzw. szybkie przeszukiwanie (ang. *Quick Search*), które polega na wpisaniu przez użytkownika systemu tylko ośmiu liter. Wspomniane formy są stosowane także w katalogach OPAC.

W wyszukiwaniu dokumentów niezastąpione okazują się wyszukiwarki internetowe. Ich pierwszym zadaniem jest wprowadzanie do bazy danych słów kluczowych ukazujących się w dokumentach publikowanych w sieci, co oznacza, że słownik słów kluczowych nieustannie rozrasta się, w miarę jak odpowiednie programy (wyszukiwarki) przeczesują sieć w poszukiwaniu nowego słownictwa. Drugim zadaniem wyszukiwarek jest rozpoznawanie wagi wyszukanych słów i stosownie do tego nadawanie im odpowiednich liczb stanowiących miarę niesionej informacji. Trzecim i najistotniejszym dla użytkownika zadaniem wyszukiwarek jest wydawanie poszukiwanych dokumentów na podstawie kwerendy złożonej ze słów kluczowych. Tu właśnie kryje się powód, dla którego język słów kluczowych tak dobrze funkcjonuje w strukturze Internetu. Jest to reguła konkatenacji, pozwalająca na prezentację słów kluczowych w dowolnej kolejności (permutacja), co jest łatwym zadaniem dla każdego. Tak więc język słów kluczowych znalazł szerokie zastosowanie w Internecie dzięki otwartemu słownikowi oraz regule konkatenacji.

Używanie słów kluczowych w Internecie ma również pewne wady. Szukanie według słów kluczowych wymaga bowiem, aby wpisane słowa rzeczywiście wystąpiły w tytule, opisie lub nagłówku strony, a więc słowa użyte przez pytającego powinny być dokładnie takie same, jakich użył nadawca. „Falszywe tropy”, czyli pojawiające się niekiedy tysiące nierelevantnych stron, to właśnie przejaw braku relewantności podanego słowa kluczowego do treści dokumentu/pytania informacyjnego lub uznania relewantności słów kluczowych za kwestię drugoplanową, a w niektórych wypadkach wręcz marginalną.

Automatyzacja wyszukiwania dokumentów wymaga takich sposobów ich wyrażania i rozpoznawania, aby było to możliwe bez udziału aktu ich rozumienia, a więc

wymaga pewnego formalizmu. Jest to szczególnie ważne, gdy proces wyszukiwania, czyli odczytywania informacji (w tym wyrażania pytań informacyjnych) ma być dokonywany przez komputer. Ponieważ treść zapytań jest zazwyczaj zbyt bogata, aby ją wyrazić, często trzeba by użyć wielu wyrażań. Badania przeprowadzone w USA na jednym miliardzie zapytań skierowanych do wyszukiwarki Altavista pokazują jednak, że jedna czwarta użytkowników poszukując informacji, wprowadza tylko jeden termin, a przeciętnie nie więcej niż dwa terminy. Nietrudno zauważyć, że niemal 80% użytkowników nie używa w ogóle żadnych operatorów. Reasumując, większość poszukuje informacji poprzez pojedyncze słowa i frazy (Jansen, Spink 2005, 2006).

Interesujące badania kwestionariuszowe dotyczące wyszukiwania za pomocą słów kluczowych w katalogu online biblioteki Uniwersytetu Oklahomy przeprowadził w 1995 roku Charles R. Hildreth (Hildreth 1997). Celem tych badań było ustalenie, czy typowi użytkownicy katalogu rozumieją podstawowe zasady wyszukiwania za pomocą słów kluczowych. Dodatkowo chodziło o określenie częstości wyszukiwania za pomocą słów kluczowych w porównaniu z tradycyjnymi formami wyszukiwania w katalogu alfabetycznym i rzeczowym. Dane zebrano na podstawie składającego się z sześciu pytań kwestionariusza i zapisów przebiegu wyszukiwania komputerowego w systemie bibliotecznym NOTIS. Pierwsze pytanie dotyczyło doświadczenia użytkowników w wyszukiwaniu za pomocą słów kluczowych, drugie – rodzaju pierwszego wyszukiwania, trzecie – wyszukiwania za pomocą jednowyrazowych słów kluczowych, czwarte – wielowyzrazowych słów kluczowych, piąte i szóste – dotyczyło rodzaju preferowanej pomocy. Uzyskano 228 kwestionariuszy. Otrzymane dane pozwoliły na stwierdzenie, że użytkownicy najczęściej wyszukują informację za pomocą słów kluczowych (92,5% badanych). Często są to wyszukiwania nieudane, a większość użytkowników nie rozumie działania systemu. W odpowiedzi na drugie pytanie uzyskano informację, że dla 54,2% respondentów pierwsze wyszukiwanie było właśnie przez słowa kluczowe. Na drugim miejscu znalazło się wyszukiwanie poprzez temat (33% badanych). Najmniej osób zadeklarowało wyszukiwanie przez tytuł dokumentu (7,5%), a przez autora tylko 5,3%. Tylko 36,1% respondentów wiedziało, że wyszukiwanie za pomocą jednowyrazowych słów kluczowych obejmuje wszystkie pola rekordu, chociaż odpowiedzi na inne pytania wskazywały, iż ponad połowa (53,2%) zdawała sobie sprawę, iż wyszukiwanie za pomocą słów kluczowych nie ogranicza się tylko do jednego pola. W odpowiedzi na czwarte pytanie uzyskano informację, że 33,8% badanych używa wielowyzrazowych słów kluczowych. W razie trudności z wyszukiwaniem prawie połowa badanych (47,8%) oczekiwała pomocy bibliotekarza, tylko 21% preferowało instrukcje drukowane bądź wyświetlane na monitorze. W nawiązaniu do tych badań warto wspomnieć o badaniach ankietowych przeprowadzonych w 2006 roku w Bibliotece Jagiellońskiej w Krakowie. Wyszukiwanie według słów kluczowych znalazło się dopiero na czwartym miejscu, po wyszukiwaniu według autora, wyszukiwaniu według tytułu i wyszukiwaniu według haseł przedmiotowych.

Oparcie wyszukiwania informacji na zadawaniu pytań w języku naturalnym oferuje wyszukiwarka Ask Jeeves (<http://www.ask.com>). Przyjęta metoda jest stosowa-

na tylko przez tę wyszukiwarke i – jak podają rankingi wyszukiwarek – nie jest zbyt popularna wśród internautów. Być może jest to spowodowane niską relewancją oraz czasochłonnością poszukiwań. Po każdym pytaniu bowiem wyszukiwarka podaje listę wyników, które nie zawsze są odzwierciedleniem kwerendy użytkownika, oraz wyświetla interakcyjne menu, w którym podane jest użyte pytanie informacyjne wraz z możliwymi kontekstami. Dopiero po wybraniu odpowiedniego kontekstu wyniki wyszukiwania są zadowalające. Daje się jednak zauważyć tendencję do umożliwienia użytkownikowi jak najprostszej formy zdobycia informacji, która następnie najłatwiej przenoszona jest przez język naturalny. Przyszłość to porozumiewanie się z systemem informacyjno-wyszukiwawczym przy użyciu głośno wypowiedzianych zdań w języku naturalnym.

Rola języka słów kluczowych w internetowych mechanizmach wyszukiwawczych może też polegać na wykorzystaniu jego słownika w budowie pytania informacyjnego. Mechanizmy wyszukiwawcze z reguły dostarczają użytkownikowi możliwości wyboru kryteriów wyszukiwawczych, stosownie do cech reprezentowanych przez pola w strukturze formatu opisu. Wyszukiwanie zawsze odbywa się w języku indeksowania, chociaż nie zawsze jest to język wejścia użytkownika do systemu. Jeżeli w indeksowaniu posługujemy się słownikiem jawnym, a więc używamy języka słów kluczowych opartego na słowniku, to słownik ten może również pełnić funkcję wyszukiwawczą w systemie. Wykorzystanie języka słów kluczowych w wyszukiwaniu informacji polega na stworzeniu użytkownikowi przez system możliwości pobierania słów kluczowych ze słownika do budowy pytania informacyjnego. Użytkownik może „manualnie” pobierać poszczególne słowa kluczowe i łączyć je za pomocą przyjętych w systemie technik wyszukiwawczych. Operacja ta może odbywać się również automatycznie. Integracja mechanizmów wyszukiwawczych ze słownikiem słów kluczowych umożliwi automatyczne tłumaczenie pytania informacyjnego na ciąg słów kluczowych równokształtnych z ich odpowiednikami w słowniku słów kluczowych. W tym procesie jest niezbędna redukcja synonimii wyszukiwawczej. Do tego służą tzw. pierścienie synonimów, wiążące relacją ekwiwalencji określoną grupę terminów, ale wyłącznie na potrzeby wyszukiwania informacji¹³⁸. Nie są one związane z językiem, lecz jest to narzędzie systemowe. Taka struktura jest przydatna w tzw. ekstensji zapytania (ang. *query extension*), uzupełnia bowiem wariantami językowymi wprowadzane przez użytkownika słowa kluczowe. Zwalnia też użytkownika z powinności posiadania niezbędnej wiedzy o formie słów kluczowych. Pierścienie synonimów leksykalnych (jawne i niejawne) są tworzone nie tylko przez wyrażenia oznaczające pojęcia, ale i akronimy nazw własnych oraz ich pełne formy (na przykład PZU – Państwowy Zakład Ubezpieczeń), różną pisownię nazw własnych, terminy naukowe i popularnie stosowane czy też błędne i poprawne zapisy tej samej nazwy. Jest to jedno z rozwiązań problemu tzw. synonimii wyszukiwawczej. Rozbudowane odsyłacze w słowniku słów kluczowych stanowią doskonały materiał do tworzenia tego typu narzędzi.

¹³⁸ Pomysł ten pojawił się już w 1999 roku w moim artykule (Babik 1999a).

Przedstawiona charakterystyka wykorzystania słów kluczowych w organizacji i dostępie do zasobów WWW jest z założenia niekompletna, gdyż immanentną cechą tego środowiska oraz funkcjonujących w nim narzędzi jest ciągła zmienność. Nowe koncepcje, które pojawiają się w związku z Web 2.0, permanentnie ewoluują, pozostawiając w tyle wdrożenia nowych mechanizmów wyszukiwawczych uwzględniających także słowa kluczowe. Przytoczone przykłady wykorzystania języka słów kluczowych w Internecie miały na celu wyłącznie zasygnalizowanie głównych mechanizmów dostępu do Internetu za pomocą słów kluczowych.

W rezultacie przeprowadzonych w tym rozdziale rozważań możemy dostrzec pewną względność klasycznego języka słów kluczowych. Czym innym jest język słów kluczowych dla indeksatora, a czym innym dla użytkownika. Jeżeli czegoś nie ma w języku indeksowania, to nie może być uwzględniane w wyszukiwaniu, chyba że dany system informacyjno-wyszukiwawczy jest dodatkowo wyposażony w jakiś język pośrednik¹³⁹. Nie zawsze istnieje tego świadomość, o czym może świadczyć fakt włączania do tego języka na przykład pierścieni synonimów. Bez względu na przyjętą formę słowa kluczowego użytkownik zawsze odnajdzie poszukiwany dokument, jeśli tylko system dysponuje odpowiednio bogatym aparatem powiązań paradygmatycznych.

3.3. Wyszukiwanie w języku naturalnym. Czy to już koniec języków słów kluczowych?

W kontekście dotychczasowych rozważań nad językami słów kluczowych rodzą się następujące pytania: Czy omawiana klasa języków słów kluczowych rozwija się? Czy nieklasyczne języki słów kluczowych można uznać za przejaw rozwoju tej klasy języków? Jaki los czeka języki słów kluczowych w przyszłości? Na te pytania postaram się odpowiedzieć w toku dalszych rozważań.

Niewątpliwie klasa paranaturalnych języków informacyjno-wyszukiwawczych jest bardzo zróżnicowana. Dotyczy to również języków słów kluczowych. Związki tej klasy języków informacyjno-wyszukiwawczych z językiem naturalnym zasługują na rozpatrzenie osobno w odniesieniu do każdego z typów. W nawiązaniu do rozdziału 2.4.1 i 3.2.2 w tym miejscu skupię się na języku słów kluczowych, choć niejednokrotnie będę odnosić się do innych języków paranaturalnych.

¹³⁹ Problem języka pośrednika już próbowano rozwiązać na różne sposoby. Jedną z oryginalnych prób były kody semantyczne oparte na wykorzystaniu semantyki składnikowej.

3.3.1. Technologia semantyczna – wyszukiwanie za pomocą języka naturalnego

Nie ma wątpliwości, że technologia komputerowa i Internet przeobrażają podstawy informacji, wiedzy i edukacji. Za jeden z bardzo ważnych instrumentów tego przeobrażenia była i jest uważana wyszukiwarka o nazwie Google. Jej działanie opiera się na algorytmach statystycznych, które jakoś informacji próbują ustalić na podstawie ich „popularności”¹⁴⁰. Tymczasem informacja wartościowa nie zawsze jest najpopularniejsza i odwrotnie. Poza tym systemy te działają retrospektywnie, chociaż w mniejszym stopniu niż tradycyjne. Potrzebują czasu na sporządzanie odsyłaczy i ich zebranie. Próbuje się temu zaradzić m.in. poprzez stosowanie odpowiednich technik optymalizacji (pozycjonowania) wyszukiwania. Dzięki temu informacja jest stale podatna na komercjalizację i manipulację. Wyjście poza statystykę wymaga wręcz rewolucyjnego podejścia i przełomu technologicznego, polegającego na określaniu jakości informacji poprzez wykorzystanie w działaniu wyszukiwarek technologii semantycznej właściwej dla Web 3.0. Proponuje ona bardziej elastyczną organizację informacji, jak również nowe formy analizy i eksploracji danych, bazujące na inżynierii ontologicznej i uczeniu maszynowym. Ma umożliwić komputerom „zrozumienie” całych fraz i fragmentów tekstu. Staje się to coraz bardziej realne dzięki pracom nad tzw. rozszerzoną rzeczywistością wirtualną. Osiągnięcie przez komputery zdolności do ujmowania języków w algorytmy komputerowe wymaga stworzenia odpowiednich ontologii, które nie sprowadzają się tylko do słownika znaczeń i synonimów. Ontologia bowiem to mapa powiązanych pojęć, odwzorowująca różnego rodzaju relacje językowe i pozajęzykowe. Jednym z lingwistycznych narzędzi budowania takich map są relacje kojarzeniowe. Dzięki ich szerokiemu uwzględnieniu można przekształcić słowniki języków paranaturalnych, w tym języka słów kluczowych, w odpowiednie mapy pojęciowe. Budowanie takich ontologii jest bardzo skomplikowane, ale możliwe. Podjęły się tego zadania m.in. firmy Hakia, Cognition Search czy Lexxe. Jednym z elementów budowanych przez te firmy ontologii są odpowiednie leksykony. Mają być narzędziami umożliwiającymi osiągnięcie semantycznej precyzji w obsłudze języków naturalnych. Technologia semantyczna próbuje określić jakość informacji. Czyni to poprzez wykrywanie/specyfikację bogactwa powiązań pojęciowych w danym tekście.

Wyszukiwarki oparte na technologii semantycznej, które są zdolne do generowania odpowiedzi w języku naturalnym, to m.in. Ask.com, Powerset (wyszukiwarka semantyczna dla Wikipedii), Semantifind (wyszukiwarka semantyczna dla Google), Cognition, Eeggi, AnswerBus.com. Ta ostatnia, chociaż odpowiada na pytania zadawane w sześciu językach, na razie nie spełnia oczekiwań użytkowników, gdyż generuje zestaw odnośników do stron bardzo luźno skojarzonych z zadanymi pytaniami. Do tej grupy zaliczam również, nastawioną obecnie przede wszystkim na

¹⁴⁰ Pozycja strony w systemie Google nie zależy, jak w innych wyszukiwarkach, od frekwencji szukanych słów kluczowych, ale od oceny danej strony dokonanej przez w pełni automatyczny algorytm PageRank. Ocena ta jest dokonywana na podstawie liczby linków prowadzących do niej od innych stron.

pytania w języku angielskim, wyszukiwarę Hakia.com, która jest tworem nowojorskiej firmy Hakia¹⁴¹ (<http://www.searchengines.za.pl/hakia.htm>). Jej twórcami są Riza Berkan, specjalista od sztucznej inteligencji, logiki zbiorów rozmytych i systemów informacyjnych oraz Pentti Kouri, ekonomista nowojorski (<http://en.wikipedia.org/Wiki/Hakia>). Założona w 2004 roku firma Hakia podkreśla, że jej produkt jest pierwszą wyszukiwarą udzielającą odpowiedzi, z wykorzystaniem analizy językowej, na pytania zadawane przez użytkowników w sposób naturalny. Dzięki temu nie jest konieczne skracanie wpisywanego hasła do najistotniejszych wyrazów. Zgodnie z założeniami jej twórców, wyszukiwarka ta będzie koncentrować się na udzielaniu „gotowych odpowiedzi”, a nie na prostym odszukiwaniu ciągu słów kluczowych. Ma to być więc inteligentna wyszukiwarka internetowa, wyposażona w translator całych zdań, dzięki czemu wyszukiwanie będzie oparte na „znaczeniu”. Wyszukiwarka ta będzie zawierać mechanizmy samouczenia się, podnoszące trafność odpowiedzi w miarę zadawania pytań. Ma to też być narzędzie interaktywne, przełamujące bariery językowe, co jest możliwe w przypadku rezygnacji z jednej logiki języka, którą posługują się dość mechanicznie obecnie funkcjonujące wyszukiwarki internetowe. Hakia na razie nie obsługuje polskich znaków. Od października 2008 roku firma rozpoczęła współpracę z bibliotekarzami i profesjonalistami w dziedzinie zarządzania informacją. Mają oni pomóc w optymalnym szeregowaniu linków powiązanych z danymi słowami kluczowymi. Do projektu Hakii pod koniec 2008 roku włączyła się polska firma Prokom Software, która bierze udział m.in. w opracowywaniu kolejnych leksykonów.

Czy słowa kluczowe mogą pomóc w rozwiązywaniu problemu jakości informacji? Tak, jeśli umieści się je w odpowiednich strukturach, na przykład mapach pojęciowych, wzbogaconych o rozszerzony repertuar relacji kojarzeniowych, na co pozwala podejście sieciowe do tych języków. Wówczas słowa kluczowe, dzięki bogactwu relacji, mogą stać się nie tylko narzędziami selekcjonowania i wartościowania w procesach indeksowania i wyszukiwania informacji, lecz przede wszystkim narzędziem selekcjonowania i wartościowania wiedzy w procesach wyszukiwania wiedzy. Przyszłość sieciowego wyszukiwania tkwi w semantyce.

Nawet pobieżna obserwacja zachowań informacyjnych użytkowników Internetu dowodzi, że coraz częściej formułuje się opinię o rychłym upadku wszelkich form kontroli słownictwa na rzecz (w domyśle!) swobodnych słów kluczowych. Słowa kluczowe są powszechnie wykorzystywane w wyszukiwaniu informacji w systemie WWW, co nie znaczy, że nie ma tam słownika. Wyszukiwarka, zadająca użytkownikowi pytanie typu „Czy chodziło Ci o...?”, z pewnością dysponuje jakąś „wiedzą” słownikową. Praktyka biblioteczna i informacyjna pokazuje, że „wiedza słów kluczowych” nie może i nie powinna być jedyną wiedzą systemu.

W tym miejscu sygnalizuję istnienie narzędzia wyszukiwawczego stworzonego przez inżynierów telekomunikacji brytyjskiej. Składa się ono z dwóch głównych elementów: ontologii hybrydowej oraz modułu wyszukiwania opartego na słowach

¹⁴¹ Program ten jest nastawiony na przekształcanie tekstu nie w listy frekwencyjne, lecz w sieć semantyczną.

kluczowych z indeksowaniem wzbogaconym o automatyczną klasyfikację tekstu. Ontologia została zbudowana z myślą o inżynierach, jednak ambicją Géryego Ducatela, Zhana Cui i Bena Azvine'a jest wdrożenie tej technologii do serwisu samoobsługowego online (Ducatel, Cui, Azvine 2006). Autorzy traktują ontologię jako sposób opisu interesującej ich domeny, wraz z informacją o związkach między pojęciami oraz ich właściwościami. Podstawowym zadaniem jest analiza pytania w języku naturalnym poprzez rozbicie go na poszczególne frazy oraz stworzenie tzw. map pojęć. W wyniku podziału powstają frazy lub mniejsze jednostki, które mogą zostać dopasowane do poszczególnych pojęć. Dla każdego pojęcia w ontologii istnieje lista fraz kluczowych, spokrewnionych z danym pojęciem. Każda fraza posiada wartość rzędu 0–1, która wyraża poziom relewancji danej frazy w stosunku do pojęcia. Opisano sposoby wykrywania przez system słów kluczowych, odkrywania nowych słów kluczowych oraz przypisywania im wag za pomocą odpowiednich algorytmów. Stworzono też narzędzie służące do przetwarzania pytań w prawdziwe słowa kluczowe. Zdaniem wymienionych autorów brakującym ogniwem między użytkownikiem a informacją, której potrzebuje, jest sieć semantyczna.

Łatwo zauważyć, że zaimplementowany w większości wyszukiwarek prosty model wyszukiwania dokumentów oparty na wyszukiwaniu za pomocą słów kluczowych, w związku z dynamicznym rozwojem Internetu i wzrostem jego zasobów, przestał spełniać efektywnie swoje zadania. Mechanizm stosowany w tym modelu opiera się głównie na porównywaniu częstości występowania słów kluczowych w dokumentach, zatem nie uwzględnia w większości przypadków relacji między terminami ani znaczenia samych słów kluczowych, poddając analizie tylko formalny kształt słowa kluczowego. Proponowane rozwiązanie, mogące usprawnić wyszukiwanie za pomocą słów kluczowych, opiera się na uwzględnieniu relacji semantycznych między słowami kluczowymi. Ich zdaniem, należy stworzyć lub wykorzystać już istniejące bazy zawierające informacje o związkach semantycznych między wyrażeniami, aby na tej podstawie można było określić dla każdego konkretnego słowa kluczowego sieć relacji (na przykład synonimii, homonimii) wspomagających proces wyszukiwania relewantnych dokumentów.

Wspomniany mechanizm wyszukiwania miałby polegać na tym, że zadane przez użytkownika pytanie wyszukiwawcze (w postaci słowa kluczowego) zostanie przetworzone przez odpowiedni algorytm, który określi sieć relacji dla tego słowa kluczowego i będzie analizował dokumenty pod względem częstości występowania nie tylko słowa kluczowego użytego przez użytkownika, lecz także powiązanych z nim słów kluczowych, nadając każdemu inną wartość: rzeczownikom – największą, czasownikom i przymiotnikom – mniejszą. Taki mechanizm umożliwiłby ocenę rzeczywistej relewancji każdego dokumentu, eliminując dokumenty, w których dane słowo wprawdzie pojawia się wielokrotnie, ale na przykład w nietypowym kontekście. Pozwoliłby również odnaleźć dokumenty relewantne, w których jednak słowo kluczowe użyte przez użytkownika nie pojawia się fizycznie. Przez tradycyjne wyszukiwarki, jeśli te nie mają ciągów synonimicznych, takie dokumenty są pomijane. Wykorzystując mechanizm relacji, algorytm wyszukiwania pozwoliłby „zorientować się”, czy dokument zawiera interesujące informacje. Zaproponowany mechanizm

przewycięża słabości tradycyjnego wyszukiwania opartego na słowach kluczowych i z powodzeniem mógłby zostać wykorzystany w systemach informacyjno-wyszukiwawczych.

Wewnątrz nowej, wyłaniającej się struktury sieci Internetu semantycznego używanie tradycyjnych internetowych narzędzi wyszukiwawczych, opartych przede wszystkim na słowach kluczowych dostarczanych przez użytkowników, staje się niewystarczające. Zdaniem hiszpańskich specjalistów José A. Royo, Eduarda Mena, Jorga Bernad i Arantzy Illarramendi nowe metody powinny bazować na słowach kluczowych, ale relewantnych dla użytkownika i opierać się na liście tzw. „czystych słów kluczowych”, dostarczanych przez samych użytkowników (Royo, Mena, Bernad, Illarramendi 2005). W tym celu próbuje się wykorzystać – wspomniane już – ontologie, które dostarczają innych znaczeń (szerszego kontekstu) użytkownikom słów kluczowych oraz umożliwiają im odkrywanie relacji między nimi. Ułatwia to tworzenie listy słów kluczowych precyzyjnie określających potrzeby i oczekiwania użytkownika. Użytkownik potrzebuje bowiem dokumentów, a nie słów kluczowych. Chodzi o to, aby system wyszukiwał tylko to, czego potrzebuje i czego oczekuje użytkownik. Ograniczyłoby to liczbę przypadkowych wyników wyszukiwania.

Pilnym zadaniem staje się też przekład języka słów kluczowych użytkownika na słowa kluczowe wykorzystywane w powszechnie stosowanych systemach wyszukiwawczych. Ma to pozwolić na stworzenie tzw. listy słów semantycznych, mających to samo znaczenie i sens co słowa kluczowe stosowane przez użytkowników zaczerpnięte z języka naturalnego. Chodzi też o to, aby system wychwytywał synonimy, aby użytkownik nie musiał „uczyć się” słownika słów kluczowych, lecz mógł używać własnego języka i uzyskiwać zadowalające efekty wyszukiwania. Taki system powinien z pełnych zdań i zwrotów, wpisywanych przez użytkownika, wychwytywać i przejmować nie tylko słowa kluczowe, lecz także relacje, które występują pomiędzy nimi i nadają znaczenie zastosowanym zwrotom.

Wraz z rosnącą liczbą informacji dostępnych w światowych sieciach informacyjnych problemem wymagającym pilnego rozwiązania staje się bariera językowa. Użytkownicy oczekują informacji bez względu na to, gdzie i w jakim języku jest ona dostępna. Sprawilo to, że od kilku lat dużo uwagi poświęca się rozwijaniu narzędzi i technologii dostępu do informacji wyrażanej w różnych językach etnicznych (Gey, Kando, Peters 2005). Jednym z możliwych rozwiązań tego problemu jest budowa wielojęzycznych słowników języków informacyjno-wyszukiwawczych, w tym wielojęzycznych słowników słów kluczowych, co umożliwi użycie jednej kwerendy w celu wyszukania informacji w zbiorach różnojęzycznych (ang. *cross-language retrieval*).

Dostęp do zasobów w jedenastu językach zapewnia już wielojęzyczny tezaurs EUROVOC, który jest wykorzystywany przez Parlament Europejski. Pomysł jego użycia w generowaniu słów kluczowych przedstawił Ralf Steinberger (2001). Propozycje rozwiązania problemu wielojęzyczności zaproponował też Timothy C. Craven (2004), a poza językiem słów kluczowych – Hsin-Chang Yang, Chung-Hong Lee i Ding-Wen Chen (2009).

3.3.2. W jakim kierunku zmierza język słów kluczowych?

Język słów kluczowych wykazuje ścisłe więzi z konkretnym językiem etnicznym. Uwzględnia wyrażenia z języka nauki i działalności zawodowej, a także z innych sfer życia i działalności człowieka. Zasadniczo unika wyrażen metaforycznych i opisowych. Raczej wystrzega się sformułowań pochodzących z odmian terytorialnych i środowiskowych języka naturalnego. Dąży do formalizacji wyrażen używanych do indeksowania i wyszukiwania dokumentów. Użytkownik najczęściej nie zna reguł formalizacji wyrażen pochodzących z języka naturalnego, a nierzadko też – otoczenia semantycznego tych wyrażen¹⁴².

Poszczególne języki słów kluczowych z wielu powodów, które sygnalizowałem, różnią się między sobą, niekiedy nawet znacznie, ale – jak sądzę – udało mi się wyodrębnić i określić (bardziej lub mniej precyzyjnie) cechy konstytutywne tego języka tworzące wspólny „rdzeń”, zawierający cechy stanowiące o istocie tych języków, oraz próbnie zakreślić granice klasycznego języka słów kluczowych.

Znaczący wpływ na ocenę jakości języka słów kluczowych przez użytkownika mają dwa czynniki:

- możliwy do zaakceptowania poziom sztuczności tego języka;
- specjalistyczna wiedza użytkownika.

Rozwój tych języków, a zwłaszcza ich internetowe implementacje wyraźnie pokazują, że w praktyce języki te systematycznie zmierzają w kierunku redukcji ich sztuczności, rozumianej jako rozbieżność semantyczno-leksykalna pomiędzy językiem słów kluczowych a językiem naturalnym. Granica między tak rozumianą ich „sztucznością” a „naturalnością” jest dość płynna. Zwykły użytkownik języka słów kluczowych posługuje się nim w sposób intuicyjny i zapewne nie odczuwa sztuczności tego języka. To ogromne podobieństwo do języka naturalnego czyni go paraturalnym.

W przypadku klasycznych języków słów kluczowych, podobnie jak w przypadku innych języków o słownictwie paranaturalnym, w sposób szczególny trzeba dążyć do tego, aby znaczenia poszczególnych słów kluczowych zbyt daleko nie odbiegały od znaczeń odpowiadających im wyrażen języka naturalnego. Użytkownik, odnajdując w słowniku „znajome” mu słowo kluczowe, skłania się do przypisania mu takiego samego lub bliskiego znaczeniu, jakie ma ono w języku naturalnym (lub w idiolektie użytkownika).

Nadmierna sztuczność semantyczno-leksykalna języka słów kluczowych, wyrażająca się w drastycznej rozbieżności systemu semantycznego języka użytkownika i języka słów kluczowych, może prowadzić do obniżenia efektywności wyszukiwania oraz swoistego rodzaju ucieczki użytkowników od takich języków. Dodatkową trudnością jest fakt, że słowa kluczowe, podobnie jak wyrazy w języku naturalnym, mają znaczenia słownikowe i realne, czyli te, w jakich używa ich każdy, kto posługuje

¹⁴² Stąd – niedopuszczalna w pytaniach użytkownika i w słownikach w tym języku – praktyka wprowadzania w roli jednostek leksykalnych rozbudowanych konstrukcji językowych spotykanych w języku polskim, na przykład „Integracja kartotek haseł wzorcowych NUKAT”.

je się danym językiem naturalnym. W przypadku języka słów kluczowych i innych języków o notacji paranaturalnej, znaczenia jednostek leksykalnych można traktować dwuaspektowo:

- z punktu widzenia użytkownika, który, posługując się słowem kluczowym, rozumie je tak samo jak analogiczne wyrażenie języka naturalnego;
- z punktu widzenia indeksatora, który, wprowadzając określone słowa kluczowe do charakterystyki wyszukiwawczej dokumentu, używa tych wyrażeń w funkcji metainformacyjnej w formie dostosowanej do wymagań konkretnego systemu wyszukiwania informacji. Naturalność notacji języka słów kluczowych wprawdzie sprzyja jego otwartości i gościnności¹⁴³, ale nie rozwiązuje (a niekiedy nawet przesłania) problemu zmian treści wyrażeń językowych (przy zachowaniu niezmiennych ich formy).

W literaturze zachodniej słowa kluczowe są uznawane za jednostki leksykalne języka naturalnego i nie pisze się tam o procedurze zmiany statusu tego typu jednostek leksykalnych (Jansen, Spink, Pfaff 2000). W polskim piśmiennictwie dotyczącym języków informacyjno-wyszukiwawczych takie rozróżnienie jest praktykowane. Moim zdaniem, słowa kluczowe są, zarówno w sensie genetycznym, jak i strukturalnym, jednostkami leksykalnymi języka naturalnego, tyle że użytymi w funkcji metainformacyjnej. A użycie wyrażeń języka naturalnego w specjalnej funkcji, wykraczającej poza standardowe funkcje tych języków, powoduje, że otrzymują one (podobnie jak w przypadku terminologii) tzw. wartość naddaną, w postaci dodatkowej lub tylko akcentowanej referencji, co różni je od pozostałych wyrażeń języka naturalnego, dlatego w polskim piśmiennictwie nazywa się je paranaturalnymi jednostkami leksykalnymi, czego nie stosuje się w piśmiennictwie zachodnim. Jest to więc różnica nie tylko terminologiczna, dotyczy bowiem funkcji językowych.

Sposób posługiwania się słowami kluczowymi jest uwarunkowany nie tylko strukturą tego języka, ale i sytuacją problemową użytkownika, określoną przez właściwości jego wiedzy i języka, której projekcja ujawnia się w procesie wyszukiwania informacji. Posługujący się językiem swobodnych słów kluczowych opierają się na swojej intuicji, tymczasem potrzebna jest też odpowiednia wiedza językowa i pozajęzykowa (specjalistyczna), której często nie mają.

W zastosowaniach języków słów kluczowych ujawniają się zarówno subiektywne podejścia do interpretacji ustalanych przepisów, jak i trudności w przestrzeganiu ich założeń. Wprowadzane do struktur paradygmatycznych odsyłacze, głównie pod wpływem języka haseł przedmiotowych, znajdują coraz szersze zastosowanie. Wykorzystywanie w tych językach relacji kojarzeniowych świadczy o podejściu sieciowym w językach informacyjno-wyszukiwawczych.

Język słów kluczowych powinien stanowić naczynie połączone z repertuarem techniki wyszukiwawczych w systemie (wagi, maskowanie, operatory Boole'a, operatory kontekstowe). Tymczasem, jak pokazują badania, operatory Boole'a są używane bardzo rzadko (Jansen, Pooch 2001). W systemie informacyjno-wyszukiwaw-

¹⁴³ Gościnność języka, zwana też produktywnością języka, to cecha języka umożliwiająca tworzenie nowych wyrażeń bez naruszania systemu tego języka (por. Bojar oprac. 2002).

czym wyszukiwanie za pomocą języka słów kluczowych najczęściej nie ogranicza się do konkretnego pola, lecz pozwala na korzystanie z możliwości przeszukiwania wszystkich pól. Obserwuje się też zjawisko wykorzystywania słów kluczowych i algebry Boole'a do selekcjonowania informacji w obrębie pól haseł przedmiotowych katalogów online. Istnieje również tendencja wzbogacania opisu dokumentu o dodatkowe elementy, jak adnotacje lub analizy dokumentacyjne, terminy pobierane ze spisów treści oraz indeksów przedmiotowych. Współczesne systemy wyszukiwania informacji zmiernają w kierunku systemów pełnotekstowych, w których istnieje możliwość wyszukiwania pełnotekstowego. Zmieniło się też miejsce słownika słów kluczowych w systemie. Jest integrowany z całym systemem, co nadaje nową jakość temu narzędziu, zmieniając jego funkcję i częściowo zawartość. Słowniki słów kluczowych wzbogacają indeks zbioru wyszukiwawczego, a zbiór wyszukiwawczy wzbogaca słownik, na przykład o liczniki wystąpień słów kluczowych. Ułatwia to konstruowanie odpowiednich strategii wyszukiwawczych.

Problem tożsamości i granic języka słów kluczowych stał się żywotny głównie pod wpływem Internetu, a właściwie systemu WWW z jego słowami kluczowymi. Informatyzacja procesów informacyjnych pozwala zupełnie inaczej niż dotychczas spojrzeć na takie zjawiska językowe występujące w języku słów kluczowych, jak synonimia i homonimia. Do tej pory relację synonimii omawiano przeważnie w kontekście jej eliminacji w językach informacyjno-wyszukiwawczych. Usilnie starano się, aby w ogóle nie zaistniała w systemie. Obecnie widoczna jest tendencja do przenoszenia całego ciężaru z eliminacji występowania tego zjawiska na rzecz eliminacji jego skutków. Odpowiedzialność za to spoczywa nie na języku, lecz na systemie, podobnie jak kwestia gramatyki i koordynacji jednostek, który jest odpowiedzialny za koordynację jednostek leksykalnych (por. zjawiska pre- i postkoordynacji w językach informacyjno-wyszukiwawczych). Na temat tego zjawiska pisałem już w artykule pt. *Synonimia i homonimia – czy naprawdę niepożądane w nowoczesnych systemach informacyjno-wyszukiwawczych?* (Babik 1999a). Zarówno w indeksowaniu, jak i w wyszukiwaniu dopuszczalna staje się synonimia, pod warunkiem, że sam system, a nie język, dokonuje koordynacji jednostek synonimicznych. Dzięki temu systemy te stają się bardziej przyjazne dla użytkownika.

Możliwości dostępu do informacji cyfrowej, jakie stwarza Internet, zwłaszcza dla jego nieprofesjonalnych użytkowników, skłaniają do rezygnacji z wyrafinowanych narzędzi wyszukiwawczych, wymagających specjalnego przygotowania, na rzecz prostych narzędzi, które można stosować intuicyjnie i niekoniecznie profesjonalnie. Takim lingwistycznym narzędziem wyszukiwania informacji są właśnie słowa kluczowe. Większość użytkowników Internetu tak naprawdę nie wie, co to jest słowo kluczowe i słusznie utożsamia go z wyrażeniami, tj. wyrazami i terminami języka naturalnego. Odpowiedź na pytanie, czy używane w Internecie słowa kluczowe są wyrażeniami języka naturalnego czy języka sztucznego, nie jest dla użytkownika ważna, ważny jest wynik wyszukiwania. Granice języka słów kluczowych określają aktualne możliwości wykorzystania języka naturalnego w systemach wyszukiwania informacji.

Zdaniem Jennifer Rowley, czynnikami, które zadecydują o stopniu wykorzystania (użycia) języka naturalnego w wyszukiwaniu informacji, są:

- kontekst wyszukiwawczy (ang. *searching environment*) dotyczący dziedziny, trybu wyszukiwania;
- wiedza dziedzinowa użytkownika i jego doświadczenie w zakresie wyszukiwania informacji;
- dostępne w systemie metody i strategie wyszukiwawcze;
- natura wyszukiwania, w tym kompletność i szczegółowość wyszukiwania, typ wyszukiwania (Rowley 1994).

W tym kontekście odpowiedź na pytanie, czy to już koniec języków słów kluczowych, narzuca się sama. Nieklasycznych języków słów kluczowych tak.

Zaprezentowane w tej części – a dotyczące języków i słowników słów kluczowych – rozwiązania są modyfikowane i dostosowywane do lokalnych lub nawet indywidualnych potrzeb i wymagań indeksatorów oraz użytkowników.

W trzeciej części opracowania przedstawiłem przykładowe implementacje języka słów kluczowych w systemach tradycyjnych i w Internecie oraz problemy określania tożsamości nieklasycznych języków słów kluczowych. Celem tej części rozprawy była też weryfikacja tezy o końcu języków słów kluczowych. Niektóre zaprezentowane rozwiązania mogą budzić wątpliwości, czy mamy do czynienia jeszcze z językiem słów kluczowych, czy wchodzimy już na grunt języka naturalnego albo języka deskryptorowego.

W stosowanej obecnie w Polsce i za granicą terminologii języków informacyjno-wyszukiwawczych „słowo kluczowe” najczęściej jest definiowane wąsko, tj. zgodnie z tradycyjnym ujęciem dokumentacyjnym, wywodzącym się jeszcze z początku lat pięćdziesiątych XX wieku i terminologii, która powstawała na użytek języków unitermowych. Jak wcześniej zaznaczyłem, zakres używania słów kluczowych uległ znacznemu poszerzeniu, przede wszystkim ze względu na potrzeby sieciowego środowiska informacyjnego. Wiele powstałych na przełomie XX i XXI wieku, a także tworzonych obecnie słowników/indeksów słów kluczowych nie spełnia już zwyczajowych wymagań tego typu słowników i jest budowanych na wzór słowników języka naturalnego, a przede wszystkim słowników terminologicznych, ba, nawet słowniki terminologiczne próbuje się wykorzystywać w funkcji słownika wyszukiwawczego (Babik 1995). Warto też zauważyć, że często rezygnuje się z używania nazwy „języki słów kluczowych” na rzecz samych „słów kluczowych”.

W tej części starałem się zwrócić uwagę na podstawowe różnice związane z uwarunkowaniami metodologicznymi języka słów kluczowych, zależnymi od zastosowania go w różnych rodzajach systemów oraz do opisu różnych źródeł, a także zbiorów o odmiennych strukturach dziedzinowych i formalnych. Różnice występują ponadto z powodu różnych typów języka etnicznego, w którym jest redagowany, i głębokości indeksowania. Internet jako nowe sieciowe środowisko informacyjne oferuje nowe możliwości technologiczne coraz szerszego wykorzystania tego języka.

Niewątpliwie użytkownicy tego języka mają w tym znaczny udział. Możliwości te często błędnie przypisuje się samemu językowi, twierdząc, że zmieniają one klasyczny język słów kluczowych w twór nieklasyczny, a nawet hybrydowy.

Rozwój języka słów kluczowych, dokonujący się zależnie od potrzeb jego konkretnych zastosowań w poszczególnych krajach, a także ustaleń metodycznych w instrukcjach indeksowania w tym języku, ukształtowały teorię języka słów kluczowych w wersji języków narodowych. Uogólnienie praktyki stosowania tego języka, badania porównawcze różnych jego odmian oraz cały dorobek teoretyczny w tym zakresie pozwoliły mi stworzyć zarys ogólnej teorii języka słów kluczowych. Opiera się on na przyjętej w rozprawie rozszerzonej definicji języka słów kluczowych.

Teoria języka słów kluczowych rozwijała się przede wszystkim jako teoretyczna płaszczyzna dla tego języka, różnych jego odmian i aplikacji w systemie informacyjno-wyszukiwawczym. Język słów kluczowych powstawał spontanicznie w wersji swobodnych słów kluczowych. Rozwijał się więc niejednolicie, gdyż punktem wyjścia – podobnie jak większości języków informacyjno-wyszukiwawczych – była i jest praktyka stosowania tego języka przede wszystkim do charakteryzowania treści prac naukowych (ale nie tylko), zależna od konkretnego języka etnicznego. Teoria języka słów kluczowych rozwija się równolegle do rozwoju zastosowań tego języka. Obejmuje problemy semantyczno-leksykalne i semantyczno-syntaktyczne. Niewątpliwie jednym z aspektów tej teorii jest określenie elementów tożsamości języka słów kluczowych. W przypadku tych języków trudno mówić o precyzji – wszystko jest rozmyte i w związku z tym nie można wyznaczyć jakiegokolwiek granic. Pomimo poszerzenia zakresu terminu „język słów kluczowych” analiza istniejących implementacji tego języka nie pozwoliła na udzielenie pełnej odpowiedzi na postawione w rozprawie pytanie o tożsamość i granice tego języka. Udało się jedynie określić tzw. *core* (rdzeń), zawierający zbiór cech konstytutywnych definiujących ten język. Skoro nie da się precyzyjnie ustalić granic tego języka, to jest to trochę „dziwny” język. Jedną z przyczyn jest niejednorodność natury tej klasy języków oraz „indywidualne” określanie przez projektantów granic języka w każdym systemie. Powoduje to, że pomimo istniejącego *core* tożsamość tych języków często jest rozumiana intuicyjnie. Nie da się więc obecnie już obronić tezy, że istnieje coś takiego jak język swobodnych słów kluczowych. Sam fakt używania takiego sformułowania bynajmniej nie oznacza, że istnieje taki język. Faktycznie w Internecie on nie istnieje. Owszem, istnieją słowa kluczowe jako jednostki/wyrażenia języka naturalnego w funkcji metainformacyjnej, przystosowane do potrzeb wyszukiwania w zautomatyzowanych systemach informacji, operujące na relacjach w języku naturalnym dzięki wykorzystaniu przez użytkowników kompetencji językowej w zakresie języka naturalnego. Wyszukiwanie w Internecie jest wyszukiwaniem w języku naturalnym. Słowa kluczowe to pewna forma wykorzystania języka naturalnego w procesach wyszukiwania informacji. Pierwotną formą języka słów kluczowych jest język swobodnych słów kluczowych, który w swej istocie opiera się na słowach kluczowych będących jednostkami leksykalnymi języka naturalnego. Później próbowano z tego zrobić język kontrolowanych (znormalizowanych) słów kluczowych. Wprawdzie język swobod-

nych słów kluczowych jest w pewnym stopniu „znormalizowany”, ale w dalszym ciągu jest to język naturalny w funkcji metainformacyjnej.

Nazwy „język (swobodnych, kontrolowanych) słów kluczowych” utrzymują się przede wszystkim ze względów marketingowych. Po prostu przyzwyczajono się do stosowania tych nazw na określenie słów kluczowych używanych w wyszukiwaniu. Przyszłość wyszukiwania za pomocą słów kluczowych leży więc w rękach każdego internauty.

Sztuczność tego języka pojawiła się dopiero na pewnym etapie jego rozwoju, to znaczy wtedy, gdy zaczęto tworzyć (projektować) języki/słowniki słów kluczowych nie metodą indukcyjną, jak w przypadku (języka) swobodnych słów kluczowych, lecz inną. Język swobodnych słów kluczowych nigdy nie był i nie jest językiem sztucznym, podobnie jak swobodne słowa kluczowe w Internecie.

W (języku) swobodnych słów kluczowych (nazwanym tak umownie, chociaż nie jest to samoistny język) od samego początku używa się odpowiednio ujednoczonych wyrażeń języka naturalnego w funkcji metainformacyjnej i wyszukiwawczej. Dopiero potem starano się z niego stworzyć język, który byłby w większym lub mniejszym stopniu znormalizowany, stąd mowa o kontrolowanych słowach kluczowych. Z językiem słów kluczowych mamy do czynienia wtedy, gdy istnieje (jest ustalona) pewna konwencja etykietowania tekstu: jak przy użyciu słów kluczowych albo fraz nominalnych wyrażać elementy treści dokumentu. Język słów kluczowych od początku był najmniej znormalizowany, bo nie podawało się żadnych definicji (znaczenie przejmowano z języka naturalnego) ani nie nakładało się żadnej struktury. Przejmowało się tylko terminy, które często były używane, ale nadal był to język naturalny w funkcji metainformacyjnej. Obecnie słowa kluczowe są formą wykorzystywania języka naturalnego w procesach zautomatyzowanego wyszukiwania informacji. Ich status językowy w Internecie dobitnie charakteryzuje określenie „chmura tagów”. Reguły pragmatyki tego języka do tej pory nie zostały skodyfikowane.

Od rozważań teoretycznych do praktyki

Celem tych rozważań była próba określenia tożsamości języka słów kluczowych i charakterystyka jego natury poprzez określenie granic pomiędzy językami słów kluczowych (czyli po prostu słowami kluczowymi) a językiem naturalnym oraz innymi paranaturalnymi językami informacyjno-wyszukiwawczymi. Do tego celu potrzebne było ustalenie, co rozumie się przez słowa kluczowe i język słów kluczowych, a także określenie granic między językami deskryptorowymi, językami haseł przedmiotowych i językami słów kluczowych.

Biorąc pod uwagę istniejącą już praktykę w zakresie stosowanej nomenklatury, poszerzyłem zakres terminu „język słów kluczowych”, tak aby obejmował również nieklasyczne języki słów kluczowych. Było to niezbędne, gdyż istniejące definicje słownikowe nie obejmują tej grupy języków słów kluczowych, przez co pozostawałyby poza zakresem moich rozważań. Tak szerokie rozumienie słów kluczowych pozwala objąć także wykorzystywanie słów kluczowych w folksonomiach, w standardach metadanych oraz w ontologiach¹⁴⁴. Wszystkie te struktury są zbliżone do języka swobodnych słów kluczowych powiązanych wewnątrznie nie tylko relacjami asocjacji. Egzemplifikacji związanych z tym problemem dokonałem na materiale słów kluczowych stosowanych we współczesnych dokumentacyjnych systemach wyszukiwania informacji.

Ważnym aspektem moich rozważań było zwrócenie uwagi na to, co wnosi do języka słów kluczowych sam użytkownik i co jest niezbędne, aby języki te „odpowiadały” użytkownikowi, w tym pokazanie, że organizacja ich słownictwa ma swoje podłoże w strukturze pamięci semantycznej, która to struktura powinna być odwzorowywana w leksyce języków słów kluczowych. Sieci relacji w mózgu powinna odpowiadać sieć relacji w języku słów kluczowych, odwzorowywana m.in. w słowniku jawnym i niejawnym (mentalnym) tego języka. Niezbędne okazało się również odniesienie do tych właściwości języka słów kluczowych, które są wypełniane/uzupełniane przez użytkownika w postaci jego wiedzy o języku i rzeczywistości pozajęzykowej, a więc kompetencji językowych i merytorycznych.

Słowa kluczowe są jednostkami leksykalnymi najprostszej formy paranaturalnych języków informacyjno-wyszukiwawczych, w planie wyrażania równokształtne z wy-

¹⁴⁴ Obecnie modne ontologie zawierają wpisane w paradygmatykę języka szeroko rozumiane relacje hierarchiczne, mereologiczne i asocjacyjne, w tym intra- i intertekstualne.

rażeniami języka naturalnego, stanowiącego bazę do generowania słów kluczowych. W planie treści zakłada się, że użytkownik tego języka, na podstawie swojej kompetencji językowej w zakresie odpowiedniego języka naturalnego, sam domyśli się znaczenia danego słowa i sam odkryje, że znaczenie słów kluczowych odpowiada równoznacznym z nimi wyrażeniom języka naturalnego w funkcji metainformacyjnej.

Bodźcem powstania i siłą napędową rozwoju języka słów kluczowych jest działalność informacyjna, a więc praktyka. To na tej bazie rodzą się ciągle nowe rozwiązania, coraz bardziej przystosowujące tę klasę języków do konkretnych potrzeb różnych kategorii użytkowników systemów wyszukiwania informacji. W związku z automatyzacją wyszukiwania informacji przyjmowano dotychczas (za językoznawstwem) frekwencyjność jako podstawę do automatyzacji budowy wszelkich modeli słów kluczowych. Okazało się jednak, że nie da się zastosować jednego, uniwersalnego a zarazem skutecznego, narzędzia generowania słów kluczowych, lecz potrzeba ich wiele. Do tego niezbędne są metody jakościowe.

Dotychczasowe podejście w językach informacyjno-wyszukiwawczych, zorientowane na system, umożliwiło określenie roli języka wyłącznie w ramach struktury systemu, ograniczając rozważania nad językami słów kluczowych do elementów strukturalnych, co w niemałym stopniu redukowało badania nad otoczeniem systemu, w tym nad pragmatyką języków informacyjno-wyszukiwawczych. Przyjęcie podejścia sieciowego, jako pewnego rodzaju „filozofii” oraz określonego punktu widzenia badanego języka, znacznie poszerzyło możliwości eksploracji tych języków. Pozwoliło mianowicie dostrzec powiązania języka informacyjno-wyszukiwawczego z procesami mentalnymi oraz dynamikę wyszukiwania informacji, a w wewnętrznych rozwiązaniach metodycznych tych języków uznać za czynnik decydujący specyfikę potrzeb użytkowników informacji będących pod presją sieci i innych układów kooperacyjnych.

We współczesnym piśmiennictwie można spotkać krytyczne, ale też i powierzchowne, opinie na temat efektywności języków słów kluczowych; stosowne badania prowadzono już w latach sześćdziesiątych XX wieku (tzw. eksperymenty cranfieldzkie). Mimo pewnych mankamentów języki słów kluczowych są najbardziej zbliżone do języka naturalnego. Pozostawiają dużą swobodę w zakresie wyboru i ustanawiania słów kluczowych, zapewniają możliwość wielowariantowego wyrażania treści dokumentów i formułowania pytań informacyjnych. Języki te wykorzystują przede wszystkim terminologię, stanowiącą podzbiór słownika języka naturalnego. Z innymi językami tej grupy łączą je te same problemy, czyli kłopoty „ujednociania” (standaryzacji) słownictwa, trudności gramatyczne i semantyczne (wieloznaczność słownictwa i tekstów), co utrudnia interpretację relacji paradygmatycznych i syntagmatycznych. Języki słów kluczowych oferują nieuporządkowane wprawdzie podejście „folksonomiczne”, ale za to zrozumiałe.

Scalając przedstawione tu analizy, uważam za celowe przypomnienie najważniejszych konstatacji dotyczących tożsamości tych języków. Na tej podstawie podejmuję też próbę sformułowania konkluzji uogólniających, wskazujących na możliwe kierunki ich doskonalenia.

Języki słów kluczowych mają niejednorodną naturę. Niejednorodność tej klasy języków wyraża opozycyjna skala: klasyczne języki swobodnych słów kluczowych, w których słowa kluczowe to wyrażenia języka naturalnego w funkcji metainformacyjnej – nieklasyczne, sztuczne języki kontrolowanych słów kluczowych, w których konstrukcja języka wyprzedza jego użycie. Ma to poważne konsekwencje zarówno w polu semantycznym tych języków, jak i w pełnionych funkcjach, a to w konsekwencji kształtuje tożsamość tego języka. Współczesne języki słów kluczowych nie dążą już do ideału „czystości” formalnej, tj. do wersji klasycznej. Pojawia się więc nieklarowność granic między językami słów kluczowych a innymi językami paraturalnymi oraz językiem słów kluczowych a językiem naturalnym. W przypadku języka słów kluczowych granice te są ciągle poszerzane.

Niekontrolowany lub nie zawsze kontrolowany rozwój tych języków doprowadził do powstania różnych i znacznie różniących się odmian języka słów kluczowych, usiłujących sprostać coraz bardziej wyrafinowanym potrzebom użytkowników systemów wyszukiwania informacji. Obecnie współistnieje kilka odmian języka słów kluczowych, co świadczy o niejednorodności natury tego języka. Oprócz tego istnieje tendencja przybliżania ich struktur do struktur innych języków, zwłaszcza do języków deskryptorowych i języków haseł przedmiotowych, zwana hybrydyzacją. Zjawisko to obszernie omówiłem w artykule pt. *O zjawisku hybrydyzacji we współczesnych językach informacyjno-wyszukiwawczych* (Babik 2010).

Pole semantyczne języków słów kluczowych jest niejednorodne i zależy od odmiany języka słów kluczowych. Pole semantyczne języka swobodnych słów kluczowych – podobnie jak języka naturalnego – w swej strukturze składa się z przenikających się wzajemnie i zespalających się pól elementarnych, których granice są zmienne, zaś pole semantyczne języka kontrolowanych słów kluczowych ma po części charakter mozaikowy. Języki te pozwalają więc uzyskać pełny – jakkolwiek „poszatkowany”, a więc niespójny – obraz rzeczywistości dokumentacyjnej, przede wszystkim dlatego, że struktury tych języków nie uwzględniają relacji hierarchicznych, zapewniających spójność systemu leksykalnego języka oraz dokładność wyszukiwania. Uwzględniają natomiast relacje kojarzeniowe (w tym pokrewieństwa), zapewniające kompletność wyszukiwania informacji. Regułą porządkującą w słownikach słów kluczowych wyznacza alfabetyczny układ tych słów, zaś w opisie systemowym dokumentu – porządek losowy lub inny porządek ustalony instrukcją indeksowania.

Stopień zbliżenia struktur semantycznych języka słów kluczowych do odpowiednich struktur w języku naturalnym wywiera istotny wpływ na dokonywaną przez użytkowników ocenę przyjazności stosowanych języków. Trudność stanowi jednak odmiennność kategoryzacji rzeczywistości przez poszczególne języki etniczne, stanowiące bazę dla języków słów kluczowych. Języki słów kluczowych za swoją prostotę leksykalną i gramatyczną płacą semantyką. Sądzę, że ten problem wymaga podjęcia dalszych badań, zwłaszcza nad odmiennością kategoryzacji rzeczywistości w poszczególnych językach etnicznych. Istnieje duża rozbieżność pomiędzy tym, co oferują obecnie stosowane języki słów kluczowych, a „ofertą” ludzkiego mózgu i możliwościami technicznymi współczesnych systemów komputerowych. Tu bardzo przydatne są wyniki badań psycholingwistycznych, szczególnie w ustalaniu

relacji między strukturą języka, przede wszystkim jego systemem semantycznym, a strukturą wiedzy jego użytkowników oraz problemu przetwarzania informacji językowych. Istotną cechą języków słów kluczowych, przez niektórych uważaną za ich wadę, jest ich związek z konkretnymi językami naturalnymi, co ogranicza ich używanie w systemach informacyjno-wyszukiwawczych i powoduje niekompatybilność systemów. Uporanie się z tą wadą jest możliwe dzięki wykorzystaniu lingwistyki komputerowej.

Języki słów kluczowych reprezentują ahierarchiczny wariant organizacji informacji, którego charakterystyczną cechą jest odwzorowywanie treści dokumentu bez jej kategoryjnego różnicowania, jak w przypadku wariantu hierarchicznego czy hybrydowego (aspektowo-przedmiotowego). Alfabetyczny wykaz punktów dostępu, który stanowią słowa kluczowe, nie reprezentuje żadnego porządku rzeczywistości dokumentacyjnej i z założenia nie odwzorowuje żadnych zależności między nimi. Dopiero zastosowanie fasetowej organizacji słów kluczowych, która umożliwia wielowymiarową interpretację treści dokumentu i jej aktywizację podczas wyszukiwania informacji, poprawia efektywność tego narzędzia i czyni go elastycznym.

Organizację wiedzy w tych językach określa się jako modelową w zakresie struktur sieciowych, ponieważ odpowiada podobnej organizacji treści w umyśle człowieka. Wymaga to więc lepszego wykorzystania w językach słów kluczowych różnych relacji kojarzeniowych. Sieciowa struktura pola semantycznego języka słów kluczowych jest czymś naturalnym, bo wynika z organizacji informacji w mentalnym słowniku człowieka. Dzięki możliwościom, jakie stwarzają systemy zautomatyzowane (zwłaszcza systemy inteligentne), obojętna w sensie formalnym jest kolejność elementów składowych w wielowyrazowych słowach kluczowych (już dzisiaj rozwiązują to tzw. pierścienie synonimów leksykalnych). Stopniowo też postępuje rozszerzanie synonimii wewnątrzjęzykowej na synonimię międzyjęzykową. Stosowane powszechnie w wyszukiwarkach internetowych i w systemach zautomatyzowanych rozwiązania już teraz pozwalają na dowolną kolejność zestawianych wyrażen i szerokie wykorzystywanie różnego rodzaju operatorów (m.in. logicznych i kontekstowych). Z pewnością zwiększy się też częstotliwość wykorzystywania przymiotników jako słów kluczowych, gdyż pozwala na to szerszy dostęp do jednostkowych składników wielowyrazowych słów kluczowych.

W celu zapewnienia odpowiedniej dokładności i kompletności wyszukiwania informacji przy formułowaniu instrukcji wyszukiwawczych niejednorodność użytych słów i wyrażen, a także różnorodność ich form gramatycznych jest kompensowana za pomocą narzędzi systemowych, głównie poprzez wykorzystanie algebry Boole'a i maskowanie. Daje to podstawę do twierdzenia, iż posługując się językiem słów kluczowych stosujemy język naturalny, choć w bardzo ograniczonym zakresie, zależnym od możliwości technicznych systemu.

Niewątpliwą zaletą tej klasy języków jest fakt, że umożliwiają łatwy dostęp do treści dokumentów poprzez indeksowanie i wyszukiwanie swobodne oparte na indeksowaniu współrzędnym. Słowa kluczowe stanowią mianowicie rodzaj wrót dostępu rzeczowego do treści (tematyki) zasobów informacyjnych. Języki te zapewniają nie tylko możliwość wyszukiwania na podstawie omawianych w dokumentach

elementów treści, lecz także – w zależności od ich słownika – według kategorii reprezentujących tematy lub nawet dziedziny, do których należy treść dokumentów.

Posługując się językiem słów kluczowych, niezależnie od jego odmiany, wykorzystuje się, w mniejszym lub większym stopniu, nie tylko słownictwo języka naturalnego w jego funkcji metainformacyjnej (ograniczone do odpowiednich kategorii słownictwa) stosownie do aktualnych możliwości stworzonych przez konkretny system, ale i kompetencję językową w zakresie języka naturalnego, której elementami są wiedza o języku i rzeczywistości pozajęzykowej, a więc i paradygmatykę, a także część gramatyki (ograniczonej do określonych form gramatycznych). Stanowi to podstawę do twierdzenia, że posługujemy się językiem naturalnym w jego funkcji metainformacyjnej, aczkolwiek w ograniczonym (tzn. jeszcze niepełnym) zakresie.

Paradygmatyka języka swobodnych słów kluczowych jest „naturalna”, co nie znaczy że jest identyczna¹⁴⁵ i nie odbiega od paradygmatyki języka naturalnego, z zastrzeżeniem, iż oba rodzaje języków odwzorowują zupełnie różne rzeczywistości. W przypadku języków słów kluczowych opartych na szerokiej definicji tych języków często znaczenie słów kluczowych jest współwyznaczane przez sztuczną paradygmatykę tych języków, nieraz znacznie odbiegającą od paradygmatyki danego języka naturalnego.

W toku rozważań doszedłem też do wniosku, że język słów kluczowych to nie jeden język, lecz dwa języki o różnej naturze, które łączą słowa kluczowe, pełniące te same funkcje, to jest funkcję metainformacyjną i wyszukiwawczą, a więc pod względem funkcjonalnym takie same. Różnica w ich naturze polega na tym, że język swobodnych słów kluczowych jest językiem naturalnym, zaś język kontrolowanych słów kluczowych jest językiem sztucznym, gdyż jest projektowany. Uprawniony jest też wniosek, że jest to jeden język naturalny w funkcji języka informacyjno-wyszukiwawczego w dokumentacyjnym systemie wyszukiwania informacji, którego wyrażenia w funkcji metainformacyjnej są wykorzystane w systemie wyszukiwania informacji, przy czym stopień ich wykorzystania i normalizacji jest zróżnicowany w zależności od konkretnego języka słów kluczowych – od zerowej normalizacji, przez częściową, po całkowitą.

W językach słów kluczowych odwzorowywanie wiedzy o zdarzeniach opisanych w dokumentach jest redukowane do wskazania wyłącznie grupy pojęć istotnych dla tej wiedzy, rezygnuje się natomiast z odwzorowywania związków i ról sytuacyjnych. Języki te pozwalają na upraszczanie i przyspieszanie procedur wyszukiwania, ale w znikomym stopniu umożliwiają odwzorowanie struktur wiedzy publicznej, która powinna ułatwić użytkownikowi proces wyszukiwawczy. Można sądzić, że podejście techniczne zdominowało podejście poznawcze. Mimo częściowej zbieżności organizacji wiedzy w strukturach kognitywnych i modelach realizowanych za pomocą języków słów kluczowych w systemach dokumentacyjnych podstawowe problemy efektywności w wyszukiwaniu informacji pozostają nadal nierozstrzygnięte. Przyczyną może być niewystarczająca znajomość zasad pragmatyki przez posługujących się tymi językami.

¹⁴⁵ Nie jest identyczna, gdyż ma też paradygmatyczne relacje dziedzinowe oparte na wiedzy, zwane też relacjami tematycznymi, których język naturalny nie posiada. Relacje w języku słów kluczowych dotyczą zawierania się zbiorów.

Charakterystyczna dla tej klasy języków redukcja wszystkich struktur do prezentacji zbioru terminów odwzorowujących pojęcia kluczowe jest zbyt uproszczona, aby uznać ją za satysfakcjonującą. Jednak sprostanie wymogom odwzorowywania polihierarchicznej i polirelacyjnej struktury pojęciowej wiedzy prowadziłyby do nadmiernej komplikacji, co przekracza możliwości tych języków i znajduje się poza zakresem ich możliwości strukturalnych oraz funkcjonalnych. Doświadczenie wskazuje bowiem, że nadmierna komplikacja struktur języka często stoi w sprzeczności z jego przyjaznością dla użytkowników. Uproszczenie struktur językowych w językach słów kluczowych nie idzie więc w parze z ich mocą semantyczną.

Z tych rozważań i analiz wynika kilka sugestii w zakresie doskonalenia istniejących zasobów słów kluczowych:

- Do podniesienia skuteczności języków słów kluczowych w procesach opisu i wyszukiwania informacji w systemach dokumentacyjnych przyczynić się może wykorzystanie wiedzy o strukturach poznawczych człowieka. Ważne jest to, co wnosi do języka słów kluczowych sam użytkownik – jego wiedza o języku i o rzeczywistości pozajęzykowej, konstytuujące kompetencję językową. Korzystanie z ustaleń psycholingwistyki oraz lingwistyki kognitywnej, dotyczących czynnego i biernego zasobu systemu leksykalnego używanego przez użytkownika języka, a także struktury pamięci semantycznej uważam za niezbędne.
- W nawiązaniu do zaprezentowanych w rozprawie wyników badań nad pamięcią, pozytywne byłoby praktyczne wykorzystanie sieciowych modeli pamięci w organizacji systemu leksykalnego języka słów kluczowych. Języki te powinny nadal pozostać (dla użytkownika) prostymi i przejrzystymi, a cały aparat „techniczny” i „formalny”, konstytuujący funkcję kontrolną tego języka, powinien – podobnie jak w językach programowania – zostać „ukryty” w systemie informacyjno-wyszukiwawczym. Oznacza to, że zbliża się nieuchronny koniec nieklasycznych języków słów kluczowych.
- Należy dążyć do rozwiązań, dzięki którym systemy leksykalno-semantyczne języków słów kluczowych będą miały jak najbardziej naturalny charakter, będą zbliżone w możliwie jak największym stopniu do sposobów myślenia i wyszukiwania informacji przez człowieka w języku naturalnym. Owa naturalność musi tu być rozumiana umownie, gdyż nie ma jednego „naturalnego” obrazu świata wszystkich użytkowników systemu, a jego wykorzystanie jest zdeterminowane przez aktualne możliwości danego systemu wyszukiwania informacji.

Istotnym wyzwaniem tej klasy języków jest utrzymanie optymalnego poziomu „trudności”, a więc prostoty tego języka. Stopień skomplikowania języka ma istotny wpływ na efektywność jego funkcjonowania. Użytkownik tego języka, znajdując znane mu wyrażenie w języku słów kluczowych, skłania się do przypisania mu znaczenia bliskiego znaczeniu, jakie ma to wyrażenie w języku naturalnym. W językach słów kluczowych, które opierają się na płaskiej strukturze pola semantycznego, lub mających przewagę cech struktury płaskiej, ciężar identyfikacji relewantnych jednostek leksykalnych spoczywa na użytkowniku. Języki słów kluczowych są więc najbardziej elastycznymi językami informacyjno-wyszukiwawczymi. Elastyczność języka słów

kluczowych oznacza zaś większą niż w innych językach zdolność dostosowywania się do zmian, jakie zachodzą w systemach informacyjno-wyszukiwawczych.

Organizacja wiedzy semantycznej najczęściej przyjmuje postać sieci asocjacyjnych lub semantycznych, w których wyodrębnia się zestaw relacji semantycznych odpowiadających podstawowym relacjom paradygmatycznym wyróżnianym w strukturze systemów leksykalnych języka naturalnego. Złożoność i wieloaspektowość organizacji i zapisu wiedzy w systemie poznawczym człowieka próbuje się odwzorowywać w językach słów kluczowych, stosując zasady wieloaspektowego podziału dyscyplinarnego i kategoryjnego (fasetowość).

Języki słów kluczowych pojawiły się na wczesnym etapie zastosowania komputerów w wyszukiwaniu informacji, ale wiele problemów dotyczących automatycznego opracowywania rzeczowego pełnotekstowych dokumentów za pomocą słów kluczowych wykorzystywanych przy przeszukiwaniu zasobów elektronicznych jest nadal nierozwiązanych, chociaż przedstawione tu wyniki badań świadczą o wzrastającym stopniu uwzględniania w strukturach języka słów kluczowych relacji kojarzeniowych. Większe uwzględnienie relacji kojarzeniowych pozwala zmniejszyć szum informacyjny i poprawić współczynnik relewancji.

Jak do tej pory, wykorzystywane środki techniczne miały niewielki wpływ na tradycyjne języki informacyjno-wyszukiwawcze. W bazach danych, które powstały w oparciu o wydawnictwa abstraktowe, nadal stosuje się języki deskryptorowe i klasyfikacje fasetowe. Systemy online w większości przejęły więc języki, których używano dotychczas. Środki techniczne ułatwiają ich stosowanie; umożliwiają m.in. jednoczesne stosowanie kilku różnych języków, mają też istotny wpływ na konstrukcję modułu komunikacji użytkownika z systemem. Tradycyjne języki informacyjno-wyszukiwawcze ewoluują w kierunku coraz większej komplikacji ich słowników, przejawiającej się w stale powiększających się ich zasobach leksykalnych i coraz większym komplikowaniu ich systemów paradygmatycznych. Prawdziwą rewolucję w językach słów kluczowych wywołał Internet, który spowodował skok w przeciwnym kierunku, a mianowicie odwrót od skomplikowanych i rozbudowanych struktur nieklasycznych języków słów kluczowych do języka naturalnego, używanego w wyszukiwarkach internetowych w funkcji metainformacyjnej w postaci swobodnych słów kluczowych.

Istota i funkcje języka słów kluczowych nadal pozostają takie same, więc trudno mówić o rozwoju tego języka. Zwiększają się natomiast możliwości jego użycia, które kompensują deficyty, gdyż zwiększają możliwości systemów informacyjno-wyszukiwawczych, co sprawia wrażenie, w istocie będące złudzeniem, że rozwija się sam język.

Wzory zachowań informacyjno-wyszukiwawczych użytkowników języka słów kluczowych, wypływające z korzystania z Internetu, nie powinny być ignorowane przez teoretyków ani przez praktyków. Uważam, że znaczny potencjał tych języków tkwi jeszcze w parametrach ergonomiczno-użytkowych systemów, takich jak interfejs użytkownika i miejsce słowników słów kluczowych w systemie, oraz rozbudowie aparatu metainformacyjnego systemu informacyjno-wyszukiwawczego, m.in. poprzez sieć odsyłaczy. Obserwacja zachowań użytkowników systemów wyszukiwania

informacji pokazuje, że wyszukiwanie informacji według ogólnie zdefiniowanych obszarów zainteresowania, którego nie zapewniają wyszukiwarki oparte na indeksowaniu statystycznym, jest dla użytkowników co najmniej równie ważne, jak wyszukiwanie dokumentów za pomocą szczegółowych nazw elementów treści dokumentu.

W optymalizacji działania tych systemów mogą być przydatne badania struktury definicji słowników terminologicznych, w tym wykorzystanie metody ciągów definicyjnych, która pozwala ujawnić paradygmatyczną strukturę kompetencji językowej użytkowników języka, a także metody gniazd leksykalnych, odwzorowujących relacje kojarzeniowe. Metoda ciągów definicyjnych może być wykorzystana do generowania wyjściowej struktury systemu leksykalnego (Babik 1996a), jak i polepszenia wyników wyszukiwania informacji w systemach wykorzystujących język słów kluczowych (Głazek 1987, 1989). Przydatne tu są również słowniki frekwencyjne, ułatwiające ustalenie optymalnego zasobu leksykalnego słowników tych języków.

Przyszłość języka słów kluczowych, a właściwie słów kluczowych – moim zdaniem – jest integralnie związana z ciągłym rozszerzaniem w wyszukiwaniu informacji zakresu wykorzystywania języka naturalnego, z czym wiąże się m.in. uwzględnianie w większym stopniu niż dotychczas różnych rodzajów relacji kojarzeniowych, co nie narusza (istotnej dla tej klasy języków) płaskiej struktury pola semantycznego. Relacje kojarzeniowe nie są bowiem relacjami hierarchicznymi, co wynika ze sposobów magazynowania i przechowywania informacji w pamięci semantycznej, są wykorzystywane w procesach wyszukiwania dzięki aktywowaniu przez użytkowników kompetencji językowej w zakresie języka naturalnego. Specyfika relacji kojarzeniowych polega bowiem na tym, że są to relacje sytuacyjne. Poszerzenie repertuaru relacji asocjacyjnych zwiększy wielowariantowość nawigacji po systemie leksykalnym. Języki informacyjno-wyszukiwawcze o najbardziej złożonej organizacji systemu leksykalnego/słownikowego dążą do identyfikacji i dokładnej specyfikacji możliwie wszystkich rodzajów relacji pojęciowych, zapewniając każdemu reprezentowanemu pojęciu bogaty kontekst relacyjny, definiujący sposób jego interpretacji. W ten sposób powstaje swoistego rodzaju sieć semantyczna. Oznacza to, że nazwa „język słów kluczowych” zniknie, a pozostaną tylko słowa kluczowe jako wyrażenia języka naturalnego w funkcji metainformacyjnej, używane w wyszukiwaniu informacji. Wyszukiwanie informacji odbywać się będzie za pomocą języka naturalnego i/lub języków sztucznych, obejmujących pozostałe języki stosowane w tym zakresie.

Rozwój sieci semantycznych jest związany z badaniami nad komputerowym przetwarzaniem języka naturalnego. Obecnie już wiele wyszukiwarek internetowych wykorzystuje sieć semantyczną – opartą na najnowszych teoriach psycholingwistycznych, dotyczących metod porządkowania wiedzy o języku w pamięci człowieka – o nazwie WordNet¹⁴⁶, opracowaną w Cognitive Science Laboratory Princeton University (USA).

Słowa kluczowe, jako nazwy węzłów w sieci, wykorzystuje się również w tzw. mapach myśli (ang. *think-maps*) i mapach semantycznych (ang. *semantic maps*),

¹⁴⁶ Leksykalna baza danych dla języka angielskiego jest dostępna pod adresem: <http://www.cogsci.princeton.edu/~wn/>. Na temat tej sieci istnieje już obszerne piśmiennictwo.

uważanych za najnowsze rodzaje systemów organizacji wiedzy¹⁴⁷. Technologia *think-maps* opiera się na idei uczenia się poprzez tworzenie tzw. map umysłu (ang. *mind-maps*), czyli wielowymiarowych układów pojęć, które mają odwzorowywać naturalne procesy myślowe w umyśle człowieka. *Think-maps* są narzędziem przekształcającym statyczny zbiór różnych informacji, opisywanych za pomocą słów kluczowych, w strukturę dynamiczną, przede wszystkim dzięki uwzględnieniu różnorodnych relacji kojarzeniowych między poszczególnymi informacjami. W ten sposób powstają semantyczne mapy słów kluczowych.

Oparcie języka słów kluczowych na paradygmacie sieciowym stwarza możliwość wykorzystania tego języka w nawigacji jako jednej ze strategii wyszukiwawczych, umożliwiając niesekwencyjne przeglądanie informacji i zbioru informacyjnego, w celu wyszukiwania informacji. Zapewnia to udział słów kluczowych w mechanizmach wyszukiwawczych systemów hipertekstowych poprzez wykorzystanie relacyjnej (choć płaskiej) organizacji pola semantycznego tego języka do wyznaczania struktury nawigacji według zawartości zasobów informacyjnych. Do rozwiązania pozostaje problem wielojęzycznego dostępu do informacji, w tym kwestie „dobrego” tłumaczenia słów kluczowych w Internecie.

Komunikowanie się za pomocą języków słów kluczowych opiera się na operowaniu skrótami myślowymi, których etykietami są właśnie słowa kluczowe bazujące na krótkich nazwach i skojarzeniach mentalnych. Umiejętność tworzenia różnorodnych, formalnie i treściowo, skojarzeń w postaci słów kluczowych staje się szczególnie ważna w społeczeństwie informacji i wiedzy.

Największy wpływ na współczesną popularność języka słów kluczowych mają dwa czynniki: (1) łatwość operowania słowami kluczowymi, słusznie utożsamianymi przez użytkowników z wyrażeniami języka naturalnego; (2) ekonomiczność opracowania zasobów informacyjnych za pomocą słów kluczowych, związana z możliwością względnie łatwej automatyzacji oraz łatwością przyswojenia sobie zasad indeksowania w tym języku, umożliwiającą scedowanie opracowania opisów na użytkowników: autorów, wolontariuszy, społeczność internetową. Komputer i możliwości łatwego porozumiewania się przez Internet i telefonię komórkową, głównie za pomocą SMS, więc w postaci krótkich tekstów, sprawiają, że „filozofia” języków słów kluczowych staje się coraz bliższa współczesnemu człowiekowi poszukującemu łatwych w użyciu narzędzi wyszukiwawczych. Język słów kluczowych proponuje bardzo uproszczony dialog w systemach online, a więc stosunkowo łatwy (również do nauczania się), ale za to obciążony pewną dozą „niedomówień”, „niedookreśleń” i szumem informacyjnym. Język słów kluczowych, podobnie jak bardzo mu bliski, bo wzorowany na słowach kluczowych (słowach kluczach) język SMS, charakteryzuje się taką samą oszczędnością słownictwa w zdaniach i w tekstach. Jest to przejawem tendencji do upraszczania struktur i reguł posługiwania się tym językiem, ale i zwiększenia stopnia wykorzystania kompetencji językowych użytkowników języka naturalnego.

Z uwagi na fakt stosunkowo prostej struktury tych języków nastąpiła marginalizacja zagadnień dotyczących pragmatyki, ze szkodą dla samego języka. Zagadnienia

¹⁴⁷ Por. Garshol 2004. Przekazywanie wyników wyszukiwania w postaci mapy myśli lub wyszukiwanie na linii czasu jest już dostępne w amerykańskiej wersji Google.

te w niniejszym opracowaniu zostały tylko zasygnalizowane. Powinny stać się przedmiotem osobnych i pogłębionych analiz.

Tekst tej rozprawy stanowi rodzaj wprowadzenia do teorii języka słów kluczowych, porządkując dotychczasowy stan wiedzy oraz uzasadniając potrzebę i nakreślając kierunki dalszych badań. Jest to zasadne i niezbędne dla kontynuowania badań nad tym językiem, które powinny dotyczyć m.in. następujących kwestii:

- jakości reprezentacji dokumentów, w tym jakości odwzorowania treści dokumentu w języku słów kluczowych;
- jakości indeksowania i wyszukiwania w języku słów kluczowych;
- kształcenia umiejętności posługiwania się językiem słów kluczowych;
- zakresu wykorzystywania oraz ewolucji tych języków w poszczególnych systemach wyszukiwania informacji;
- społecznych aspektów rozwoju i otoczenia języka słów kluczowych.

Dominujące obecnie podejście do słów kluczowych i języków słów kluczowych charakteryzuje wyraźny pragmatyzm. Języki słów kluczowych wychodzą naprzeciw potrzebom opisu dokumentów, wypływającym z interdyscyplinarności współczesnych badań naukowych.

Możliwości i zakres stosowania języka naturalnego w wyszukiwaniu informacji w systemach zautomatyzowanych stale się zwiększają. Zakres stosowania języka naturalnego w procesach wyszukiwania informacji zależy od aktualnego poziomu rozwoju technologii informatycznych. Należy pamiętać, że funkcjonowanie języka naturalnego jest możliwe tylko przy aktywnym udziale w procesie wyszukiwania informacji człowieka, który w tym procesie wykorzystuje całą swoją wiedzę o świecie i o języku.

W przyszłości będzie dominowało wyszukiwanie informacji za pomocą słownictwa języka naturalnego, zapewne w znacznie szerszym zakresie niż obecnie. W 2003 roku naukowcy z Massachusetts Institute of Technology ogłosili, że w niezbyt odległej przyszłości będziemy porozumiewać się z komputerem za pomocą mowy. Do tego, aby swobodnie porozumiewać się z komputerem, niezbędne jest jeszcze rozwiązanie szeregu problemów dotyczących komputerowej interpretacji tekstu oraz zbudowanie algorytmów rozumienia mowy, to jest przekształcających mowę w tekst. Oczekiwania, że słowa kluczowe, bądź słowa z tekstu, będą wystarczającym automatycznym narzędziem wyszukiwawczym w bazach danych i w zasobach internetowych, spełniają się w coraz większym stopniu. Dostęp do nich będzie efektywniejszy, gdy w organizacji struktur wyszukiwawczych w większym zakresie stosowane będą relacje intertekstualne, ponieważ umożliwiają one prowadzenie wyszukiwań zarówno szerszych, jak i węższych, dodając konteksty do używanych terminów wyszukiwawczych. Dotyczy to również obecnie stosowanej metodyki indeksowania.

Znikną więc granice pomiędzy „sztucznymi” słowami kluczowymi a słowami kluczowymi w języku naturalnym. Językiem wyszukiwawczym będzie zapewne język naturalny, ale w znacznie większym zakresie niż obecnie. Nie oznacza to jednak, że znikną problemy, o których tu pisałem, dotyczące (języka) słów kluczowych. Ich rozwiązanie pozostanie w gestii systemu, zaś rozwiązania zostaną „ukryte” w samych systemach informacyjno-wyszukiwawczych. Problemy więc nie znikną, a ich rozwiązanie stanie się niezbędne do efektywnego funkcjonowania samych systemów.

Słowa kluczowe stopniowo będą wypierać inne języki stosowane w wyszukiwaniu informacji. Ich generowanie będzie miało zarówno charakter autorski, jak i społecznościowy oraz automatyczny.

Przedstawione kłopoty z tożsamością i identyfikacją języków słów kluczowych oczywiście nie wyczerpują całego zagadnienia. Pomimo podjętych prób dokonania pewnych rozstrzygnięć, zdaję sobie sprawę, że nie udzieliłem odpowiedzi na wszystkie nasuwające się w związku z tym pytania i wątpliwości, a i zaproponowane rozstrzygnięcia mogą być dyskusyjne.

Współczesne języki słów kluczowych, służące wyłącznie opisowi i wyszukiwaniu tekstowemu, próbuje się integrować z narzędziami wyszukiwania informacji nie tylko dokumentacyjnych systemów informacyjno-wyszukiwawczych. Ciągłe trwają prace nad uniwersalnym wyszukiwaniem, które mogłoby pozwolić na łączenie potencjału ludzkiego i sztucznej inteligencji, tworzyć bogate semantycznie środowiska i oferować łatwiejszy dostęp do informacji elektronicznej, nie tylko dla ekspertów i fascynatów, lecz także dla osób, dla których nowoczesne technologie komunikacyjne są jeszcze względnie obce.

W obszarze automatycznego indeksowania nie zostały do tej pory rozwiązane wszystkie problemy automatycznej analizy tekstu i ekstrakcji słów kluczowych. Badania te są nadal na etapie wstępnym. Przyśpieszyłoby je opracowanie metod generowania tzw. wiedzy niejawnej (ang. *implicite knowledge*). Głównym wyzwaniem jest wygenerowanie metod pozwalających na odczytywanie znaczenia na poziomie pełnotekstowym. Pewien postęp, przynajmniej w odniesieniu do rozumienia znaczenia terminów i małych fragmentów dokumentów, stanowi wiązanie technik stosowanych w badaniach nad wyszukiwaniem informacji ze sztuczną inteligencją (sieciami neuronowymi) i lingwistyką komputerową. Cel, jaki stoi przed tymi badaniami, to wypracowanie semantyki pełnego tekstu. Jest to jeszcze dalekie do osiągnięcia. Włożony wysiłek oraz koszty formalizacji analizy składniowej i doczepienia do tego wartości informacyjnych tak naprawdę byłyby niewspółmierne do uzyskanych efektów, dlatego trzeba zdać się na rozwiązania informatyków. O zakresie wykorzystania języka naturalnego w wyszukiwaniu informacji decyduje i go ogranicza poziom rozwoju technologii komputerowej (por. Strzałkowski 1995; Carballo, Strzałkowski 2000; Tait 2005). Automatyczne indeksowanie może natomiast znacząco wspomóc indeksowanie intelektualne, zwłaszcza że gigantycznie zwiększa się liczba zróżnicowanych strukturalnie, nieustrukturalizowanych, heterogenicznych i nieustannie zmiennych zasobów internetowych.

Słowa kluczowe stanowią przykład wykorzystania języka naturalnego jako języka informacyjno-wyszukiwawczego w wyszukiwaniu informacji. Języki słów kluczowych dzięki nowym możliwościom systemów wyszukiwania informacji w Internecie zatoczyły koło, wracając do swojej pierwotnej postaci, to jest do swobodnych słów kluczowych, będących w istocie wyrażeniami języka naturalnego. Warto zwrócić uwagę, że dynamicznie rozwijają się zastosowania słów kluczowych. Zaprezentowane w rozprawie podejście to niewątpliwie próba nowego zdefiniowania słowa kluczowego i języka słów kluczowych w kontekście potrzeb internetowego środowiska elektronicznego systemów wyszukiwania informacji.

Rozwój języków informacyjno-wyszukiwawczych zawsze wyznaczały trzy najważniejsze czynniki: niekontrolowany przyrost informacji koniecznej do przetworzenia, stały wzrost i specjalizacja potrzeb informacyjnych użytkowników informacji oraz możliwości systemów w zakresie przetwarzania informacji. Jediną możliwością zapewnienia sprawności tych systemów było doskonalenie języków informacyjno-wyszukiwawczych. Wymusiło to przede wszystkim rozbudowę ich leksyki oraz ich systemów paradygmatycznych do tego stopnia, że opracowywanie rzeczowe stało się wąskim gardłem systemów informacyjno-wyszukiwawczych. Dopiero pojawienie się komputera, a wraz z nim możliwość automatyzacji operacji wykonywanych w systemach, ale dopiero po uprzedniej formalizacji stosowanych w nich systemów językowych, otworzyło nowe możliwości rozwiązania dręczących od dawna informacje naukową problemów, stając się jednocześnie nowym wyzwaniem. Urzeczywistnienie idei języków informacyjno-wyszukiwawczych wolnych od wszelkich ograniczeń stało się realne. Takimi niemal idealnymi językami okazały się języki generowane z języków naturalnych, przyjazne dla użytkownika i prawie gotowe, zwane swobodnymi słowami kluczowymi. Zapewniają one współrzędne indeksowanie oraz swobodne i łatwe indeksowanie i wyszukiwanie informacji. Poza tym są powszechnie zrozumiałe i stosowane w wyszukiwarkach internetowych, m.in. Google, będących elementem wszystkich, już z „natury” hipertekstowych, współczesnych systemów działających w Internecie. Ze swej natury jest to język naturalny w funkcji metainformacyjnej i wyszukiwawczej w postaci swobodnych słów kluczowych. Tak więc wyszukiwanie informacji w nowoczesnych systemach informacyjnych w erze cyfrowej będzie przede wszystkim procesem przetwarzania informacji zapisanych w języku naturalnym, co dzieje się już w Internecie. Inne języki informacyjno-wyszukiwawcze będą komplementarne do tego języka. Szybszy postęp w tym zakresie będzie możliwy dzięki programom komputerowym wykorzystującym efekty formalizacji języka naturalnego w obrębie semantyki i relacji paradygmatycznych, umożliwiając wyszukiwanie na pożądanym poziomie szczegółowości-ogólności, relewancji, dokładności i kompletności informacji. Ma to poważne konsekwencje dla teorii języków informacyjno-wyszukiwawczych. Cecha sztuczności tych języków przestaje być cechą dystynktywną, a więc i definicyjną.

„W informacji – jak trafnie zauważyła B. Bojar – na naszych oczach dokonuje się rewolucja: skokowe przejście od coraz bardziej wyspecjalizowanych systemów informacyjnych do uniwersalnego systemu informacyjnego, jakim jest Internet, a od coraz bardziej »wymiślnych« i rozbudowanych języków informacyjno-wyszukiwawczych powrót do używania języka naturalnego w funkcji metainformacyjnej, czego przejawem jest używanie w wyszukiwarkach słów kluczowych” (Bojar 2009, s. 22). Tekst rozprawy dowodzi, że podzielam tę opinię w całej rozciągłości. Radikalnie zmienia się układ/proporcja używania tych języków. W epoce cyfrowej język naturalny staje się „normą” w reprezentacji i wyszukiwaniu informacji. Inne języki będą używane tylko w niewielkim stopniu. Zaistnienie i obecność słów kluczowych w Internecie – moim zdaniem – stało się początkiem rewolucji w językach informacyjno-wyszukiwawczych.

Literatura

(Opisy dokumentów elektronicznych dotyczą wersji aktualnych 1 VII 2009 r.)

- Aitchinson J. (2003): *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford UK – Cambridge USA: Blackwell.
- Babik W. (1989): *Metodyka budowy słowników słów kluczowych. Wybrane zagadnienia (Na przykładzie prac Pracowni Informacji Naukowej OIN PAN w Krakowie)*. [W:] II Międzynarodowe Sympozjum nt. „Zastosowanie mikrokomputerów w inte”, Katowice 4–6 października 1989. Materiały konferencyjne. Cz. I. Katowice, s. 113–125.
- Babik W. (1994): *Słowniki słów kluczowych – budowa i możliwości ich wykorzystania w zautomatyzowanych systemach informacji*. [W:] *Sieć informacyjna w zdrowiu publicznym. Konferencja zorganizowana przez Szkołę Zdrowia Publicznego Uniwersytetu Jagiellońskiego*. Kraków, 20–21 kwietnia 1994. *Materiały konferencyjne*. Kraków: Centrum Informacji Szkoły Zdrowia Publicznego Collegium Medicum UJ.
- Babik W. (1995): *Wielofunkcyjne słowniki terminologiczne*. [W:] *Komputeryzacja bibliotek a potrzeby użytkowników*. Kraków–Łopuszna, 16–18 maja 1995. [Materiały konferencyjne]. Kraków: Biblioteka Główna Politechniki Krakowskiej, s. 41–48.
- Babik W. (1996a): *Generowanie języków informacyjno-wyszukiwawczych ze słowników terminologicznych*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego (Rozprawy habilitacyjne UJ Nr 313).
- Babik W. (1996b): *Inżynieria języka naturalnego na potrzeby języka dla systemów informacyjno-wyszukiwawczych*. [W:] Vetulani Z., Abramowicz W., Vetulani G.: *Język i technologia*. Warszawa: Akademicka Oficyna Wydawnicza PLJ, s. 66–69.
- Babik W. (1999a): *Synonimia i homonimia – czy naprawdę niepożądane w nowoczesnych systemach informacyjno-wyszukiwawczych*. „Biuletyn Instytutu Metali Nieżelaznych” (numer specjalny), s. 83–87.
- Babik W. (1999b): *Termin i jego status w systemie leksykalnym języka informacyjno-wyszukiwawczego*. „Zagadnienia Informacji Naukowej”, nr 1 (73), s. 3–14.
- Babik W. (2007): *Keywords in information retrieval systems on the Internet*. [W:] Waliński J., Kredens K., Goźdz-Roszkowski S. (ed.): *Corpora and ICT in Language Studies*. Frankfurt am Main – Berlin – Bern – Bruxelles – New York – Oxford – Wien (Lodz Studies in Language, vol. 13. Ed. by B. Lewandowska-Tomaszczyk), s. 203–211.
- Babik W. (2010): *O zjawisku hybrydyzacji we współczesnych językach informacyjno-wyszukiwawczych*. [W:] *Informacja i język*. Pod red. W. Babika i D. Pietruh-Reizes (W druku).
- Bańkowska A.J., Sokołowska W. (1969): *Niektóre problemy wyboru słów kluczowych*. „Aktualne Problemy Informacji i Dokumentacji”, vol. 14, nr 6, s. 16–23.

- Barthes R. (1999): *S/Z*. Warszawa: Wydaw. KR.
- Bean C.A., Green R. (2001): *Relevance Relationships*. [W:] Bean C.A., Green R. (ed.): *Relationships in the Organization of Knowledge*. Dordrecht – Boston – London: Kluwer Academic Publishers, s. 115–132.
- Beling G., Port P., Strohl-Goebel H., ed. (2006): *Terminologie der Information und Dokumentation*. Wiesbaden – Frankfurt am Main: Dinges & Frick GmbH.
- Belkin N.J. (1984): *Cognitive models and information transfer*. “Social Science Information Studies”, vol. 4, no 2–3, s. 111–129.
- Belkin N.J. (1990): *The cognitive viewpoint in information science*. “Journal of Information Science: Principles and Practice” vol. 16, no 1, s. 11–16.
- Bertram J. (2005): *Einführung in die inhaltliche Erschliessung. Grundlagen–Methoden–Instrumente*. Würzburg: ERGON Verlag.
- Bielicka L.A., Ścibor E. (1982): *Języki informacyjne. Rodzaje i zastosowanie w działalności informacyjnej*. Warszawa: CINTE.
- Bojar B., red. (1993): *Słownik encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych*. Red. nauk. B. Bojar. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Bojar B., oprac. (2002): *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych*. Warszawa: Wydawnictwo SBP.
- Bojar B. (2004): *Pole informacyjne, dokumentacyjne, tematyczne, semantyczne czy leksykalne*. „Zagadnienia Informacji Naukowej”, nr 1 (83), s. 3–11.
- Bojar B. (2005): *Językoznawstwo dla studentów informacji naukowej*. Warszawa: Wydawnictwo SBP.
- Bojar B. (2006): *Jeden język – różne światy*. „Acta Filologia”, nr 32, Uniwersytet Warszawski, Wydział Neofilologii, Warszawa, s. 5–15.
- Bojar B. (2007): *Informacja naukowa – czy to już koniec?* „Praktyka i Teoria Informacji Naukowej i Technicznej”, nr 2, s. 3–13.
- Bojar B. (2009): *Języki informacyjno-wyszukiwawcze wczoraj, dziś... czy jutro?* „Zagadnienia Informacji Naukowej”, nr 1 (93), s. 3–24.
- Borger Z., Kuflik T., Shoval P., Shapira B. (2001): *Automatic keyword identification by artificial neural networks compared to manual identification by users of filtering systems*. “Information Processing and Management”, vol. 37, no 2, s. 187–198.
- Börner K., Sanyal S., Vespignani A. (2007): *Network science*. “Annual Review of Information Science and Technology”, vol. 41 (Ed. by B. Cronin). Medford, New Jersey: Information Today, Inc., s. 537–607.
- BS 8723 (2005): *Structured vocabularies for information retrieval*.
- Buckland M. (1999): *Vocabulary as a central concept in library and information science*. [W:] COLIS3: *Digital Libraries: Interdisciplinary Concepts, Challenges and Opportunities. Proceedings of the Third International Conference on the Conceptions of the Library and Information Science*. Zagreb, s. 3–12.
- Bush V. (1945): *As we may think*. “Atlantic Monthly”, nr 176 (1), s. 101–108 + [dok. elektr.]. Tryb dostępu: <http://www.theatlantic.com/doc/194507/bush>.
- Carballo J.P., Strzałkowski T. (2000): *Natural language information retrieval: Progress report*. “Information Processing and Management”, vol. 36, no 1, s. 155–178.
- Chlewiński Z. (1991): *Kształtowanie się umiejętności poznawczych: identyfikacja pojęć*. Warszawa: Wydawnictwo Naukowe PWN.
- Chmielewska-Gorczyca E. (1995): *Funkcje tezaury w systemie informacyjno-wyszukiwawczym*. „Zagadnienia Informacji Naukowej”, nr 1–2 (65–66), s. 3–17.

- Chmielewska-Gorczyca E. (1977): *Indeksy permutacyjne*. Warszawa: CINTE.
- Chmielewska-Gorczyca E. (1991): *Język wyszukiwawczy a potrzeby informacyjne użytkowników*. „Zagadnienia Informatyki Naukowej”, nr 1(58), s. 3–39.
- Choroszuszyzna J. (1968): *Analiza zbioru słów kluczowych*. Warszawa: CINTE.
- Choroszuszyzna J. (1970): *Metody i technika wyboru słów kluczowych stosowane przy opracowywaniu indeksu KWOC do wydawnictwa „Przegląd Piśmiennictwa Zagadnień Informatyki”*. [W:] *Indeksy słów kluczowych i spisy bibliograficzne sporządzane maszynowo na przykładzie Indeksu KWOC do wydawnictwa „Przegląd Piśmiennictwa Zagadnień Informatyki”*. *Materiały pokonferencyjne*. Warszawa, s. 43–64.
- Chowdhury G.G. (1999): *The Internet and information retrieval research: a brief review*. “Journal of Documentation”, vol. 55, no 2, s. 209–225.
- Chu H. (2003): *Information Representation and Retrieval in the Digital Age*. Medford, New Jersey: Information Today, Inc.
- Cieszowska T., Ścibor E., Żeromska F. (1969): *Niektóre problemy selekcji słów kluczowych do indeksu KWOC dla zagadnień informatyki*. „Aktualne Problemy Informatyki i Dokumentacji”, vol. 14, nr 1, s. 15–18.
- Cleverdon C.W. (1967): *The Cranfield tests on index language devices*. “ASLIB Proceedings” vol. 19, no 6, s. 173–194.
- Collins A.M., Quillian M.R. (1969): *Retrieval time from semantic memory*. “Journal of Verbal Learning and Verbal Behaviour”, no 8, s. 240–248.
- Collins A.M., Loftus E.F. (1975): *A spreading activation theory of semantic processing*. “Psychological Review”, vol. 82, no 6, s. 407–428.
- Costello J.C. (1961): *Uniterm indexing principles, problems and solutions*. “American Documentation”, vol. 12, no 1, s. 20–26.
- Craven T. (2004): *Variations in use of meta tag keywords by web pages in different languages*. “Journal of Information Science”, vol. 30, no 3, s. 268–279.
- Craven T. (2005): *Web authoring tools and meta tagging of page descriptions and keywords*. “Online Information Review”, vol. 29, no 2, s. 129–138.
- Czerny A.I. (1978): *Wstęp do teorii wyszukiwania informacji*. Warszawa: OIN PAN.
- Ćwiekowa J. (1988): *Opracowanie tematyczne piśmiennictwa. Język hasel przedmiotowych. Katalog rzeczowy*. Warszawa: PWN.
- Ducatel G., Cui Z., Azvine B. (2006): *Hybrid Ontology and Keyword Matching Indexing System*. “WWW06 World Wide Web Conference”. Intrawebs workshop. Edinburgh.
- (EJO) *Encyklopedia językoznawstwa ogólnego*. (1995). Pod red. K. Polańskiego. Wrocław: Ossolineum.
- Farradane J. (1967): *Concept organization and information retrieval*. “Information Storage and Retrieval”, vol. 3, no 4, s. 297–314.
- Frants V.I., Shapiro J., Voiskunskii V.G. (1997): *Automated Information Retrieval: Theory and Methods*. San Diego: Academic Press.
- Froehlich T.J. (2000): *Kryteria oceny systemów wyszukiwawczych w Internecie*. „Zagadnienia Informatyki Naukowej”, nr 2 (76), s. 3–20.
- Gabryś-Barker D. (2005): *Aspects of Multilingual Storage, Processing and Retrieval*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Gardin J.C. (1966): *Éléments d'un modèle pour la description des lexiques documentaires*. „Bulletin des Bibliothèques de France”, vol. 11, no 5, s. 171–182.
- Garshol L.M. (2004): *Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all*. “Journal of Information Science”, vol. 30, no 4, s. 378–391.

- Gey F.C., Kando N., Peters C. (2005): *Cross-Language Information Retrieval: the way ahead*. "Information Processing and Management", vol. 41, iss. 3, s. 415–431.
- Gil-Leiva I., Alonso-Arroyo A. (2007): *Keywords given by authors of scientific articles in database descriptors*. "Journal of the American Society for Information Science and Technology", vol. 58, no 8, s. 1175–1187.
- Głazek D. (1987): *Metoda polepszenia efektywności wyszukiwania informacji w systemie stosującym język swobodnych słów kluczowych*. Wrocław: Biblioteka Główna i Ośrodek Informacji Naukowo-Technicznej Politechniki Wrocławskiej (Raport serii SPR Nr 98).
- Głazek D. (1989): *Możliwość zwiększenia efektywności wyszukiwania informacji w języku swobodnych słów kluczowych*. „Zagadnienia Informacji Naukowej”, nr 1 (54), s. 125–138.
- Głowacka E. (1994): *Badania efektywności języków informacyjno-wyszukiwawczych (komunikat z badań)*. [W:] *Komputeryzacja bibliotek. Materiały konferencji 24–26 maja 1993 r., Toruń*. Pod red. B. Ryszewskiego. Toruń: Uniwersytet Mikołaja Kopernika, s. 209–214.
- Gödert W. (1996): *Information as cognitive construction: a communication – theoretic model and consequences for information systems*. "Knowledge Organization", vol. 23, no 4, s. 206–212.
- Gross T., Taylor A.G. (2005): *What have we got to lose? The effect of controlled vocabulary on keyword searching results*. "College & Research Libraries", vol. 66, no 3, s. 212–230.
- Guilford J.P. (1978): *Natura inteligencji człowieka*. Przeł. B. Czarniawska, W. Kozłowski, J.Radzicki. Warszawa: PWN.
- Guiraud P. (1954): *Les caractères statistiques du vocabulaire. Essai de méthodologie*. Paris: Presses Universitaires de France.
- Guiraud P. (1966): *Zagadnienia i metody statystyki językoznawczej*. Przeł. M. Kniaginowa. Warszawa: PWN.
- Hanicke P. (1966): *Schlagwörterverzeichnisse – Deskriptorsysteme – Thesauri. Wie macht man das? Eine Arbeitsanleitung für die Praxis*. Berlin: Zentralinstitut für Information und Dokumentation.
- Hartley R.J. (2009): *Folksonomies to ontologies: the changing nature of controlled vocabularies*. [W:] Griffiths J.R., Craven J.(ed.): *Access, Delivery, Performance The Future of Libraries without Walls. A Festschrift to celebrate the work of Professor Peter Brophy*. London: Facet Publishing, s. 145–157.
- Hartley J., Kostoff R.N. (2003): *How useful are 'key words' in scientific journals?* "Journal of Information Science", vol. 29, no 5, s. 433–438.
- Hildreth Ch.R. (1997): *The use and understanding of keyword searching in a university online catalog*. "Information Technology and Libraries", vol. 16, no 2, s. 52–62.
- Hutchins W.J. (1978): *Languages of Indexing and Classification. A Linguistic Study of Structures and Functions*. Stevenage: Peter Peregrinus Ltd.
- Hys J. (2009): *Pojęcie pragmatyki w teorii języków informacyjno-wyszukiwawczych*. „Przegląd Biblioteczny”, R. 77, z. 1, s. 41–53.
- Ingwersen P. (1996): *Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory*. "Journal of Documentation", vol. 52, no 1, s. 3–50.
- Ingwersen P., Järvelin K. (2005): *The Turn. Integration of Information Seeking and Retrieval in Context*. Dordrecht: Springer.
- Instrukcja indeksowania prac badawczych za pomocą swobodnych słów kluczowych w Centralnym Specjalistycznym Systemie Informacji o Pracach Naukowo-Badawczych, Rozwojowych i Ekspertyzach Naukowych SYNABA*. Warszawa, Ośrodek Przetwarzania Informacji przy Komitecie Badań Naukowych, b.r.

- ISO 5963: 1985. *Documentation. Methods for examining documents, determining their subjects, and selecting indexing terms.*
- ISO 8777: 1993 *Commands for interactive text searching (CCL – Common Command Language).*
- Jadacka H. (1978): *O formie, znaczeniu i desygnacie terminu.* „Zagadnienia Naukoznawstwa”, nr 4 (56), s. 552–557.
- Jansen B.J., Pooch U. (2001): *A review of Web searching studies and a framework for future research.* “Journal of the American Society for Information Science and Technology”, vol. 52, no 3, s. 235–246.
- Jansen B.J., Spink A. (2005): *Web Search. Public Searching of the Web. Information Science and Knowledge Management.* New York – Boston – Dordrech – London – Moscow: Springer.
- Jansen B.J., Spink A. (2006): *How are we searching the World Wide Web? A comparison of nine search engine transaction logs.* “Information Processing and Management”, vol. 42, no 1, s. 248–263.
- Jansen B.J., Spink A., Pfaff A. (2000): *Linguistic Aspects of Web Queries. Proceedings of the ASIS Annual Meeting*, vol. 37, s. 169–176.
- Języki informacyjne opracowane, adaptowane i wykorzystywane w Polsce (Stan na 31.12. 1994 r.). (1994): Oprac. J. Tomasik-Beck. Warszawa: IINTE.
- Jones S., Paynter G.W. (2002): *Automatic extraction of document keyphrases for use in digital libraries: Evaluation and applications.* “Journal of the American Society for Information Science and Technology”, vol. 53, no 8, s. 653–677.
- Käki M. (2006): *fKWIC: Frequency-Based Keyword-in-Context Index for Filtering Web Search Results.* “Journal of the American Society for Information Science and Technology”, vol. 57, no 12, s. 1606–1615.
- Kilgour F.G. (1997): *Origins of coordinate searching.* “Journal of the American Society for Information Science and Technology”, vol. 48, no 4, s. 340–348.
- Kintsch W. (1974): *The Representation of Meaning in Memory.* Hillsdale, NJ: Lewrance Erlbaum.
- Koblitz J. (1968): *Schlagwortgebung in der Information und Dokumentation.* 2 Aufl. Leipzig: VEB Bibliographisches Institut.
- Kurcz I. (1987): *Język a reprezentacja świata w umyśle.* Warszawa: PWN.
- Kurcz I. (2000): *Psychologia języka i komunikacji.* Warszawa: Wydawnictwo Naukowe „Scholar”.
- Kurek-Kokocińska S. (1994): *Warszawska szkoła teorii języków informacyjno-wyszukiwawczych.* „Przegląd Biblioteczny”, z. 3/4, s. 269–277.
- Kurek-Kokocińska S. (1998a): *Z zagadnień terminologicznych nauki o informacji: potrzeba języka słów kluczowych.* „Zagadnienia Informacji Naukowej”, nr 2 (72), s. 76–86.
- Kurek-Kokocińska S. (1998b): *Z zagadnień terminologicznych nauki o informacji. Założenia języka słów kluczowych z dziedziny nauki o informacji.* „Zagadnienia Informacji Naukowej”, nr 1 (73), s. 15–22.
- Kurek-Kokocińska S. (2004): *Związki języka naturalnego i języków informacyjno-wyszukiwawczych – na przykładzie języka haseł przedmiotowych.* „Zagadnienia Informacji Naukowej”, nr 1 (83), s. 46–64.
- Lancaster F.W., Mills J. (1964): *Testing Indexes and Index Language Devices: The ASLIB Cranfield Project.* “Journal of the American Society for Information Science and Technology”, vol. 15, no 1, s. 4–13.

- Larsen B., Ingwersen P. (2005): *Cognitive Overlaps along the Polyrepresentation Continuum*. [W:] Spink A., Cole C. (ed.): *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, s. 43–60.
- Le Deuff O. (2006): *Folksonomies: les usagers indexent le web*. „Bulletin des Bibliothèques de France”, vol. 51, no 4, s. 66–70. [Dok. elektr.]. Tryb dostępu: <http://bbf.enssib.fr/sdx/BBF/pdf/bbf-2006-4/bbf-2006-04-0066-002.pdf>.
- Lindsay P.H., Norman D.A. (1991): *Procesy przetwarzania informacji przez człowieka. Wprowadzenie do psychologii*. Warszawa: Wydawnictwo Naukowe PWN.
- Lu L., Hanjalic A. (2006): *Towards Optimal Audio “Keywords” Detection for Audio Content Analysis and Discovery. Proceedings of the 14th Annual ACM International Conference on Multimedia*. Santa Barbara, CA – New York: ACM, s. 825–834.
- Luhn H.P. (1960): *Keyword-in-Context Index for technical literature*. “American Documentation”, vol. 11, no 4, s. 288–295.
- Lyons J. (1998): *Chomsky*. Wyd. 3 rozszerz. Warszawa: Prószyński i S-ka.
- Łobacz P., Mikołajczak-Matyja N. (2002): *Skojarzenia słowne w psycholeksykologii i onomastyce psycholingwistycznej*. Poznań: Wydawnictwo SORUS.
- Marciszewski W. (1976): *Miejsce słów kluczowych w strukturze tekstu*. „Zagadnienia Informacji Nukowej”, nr 1 (28), s. 47–71.
- Marciszewski W. (1978): *The structure of text and the distribution of key words*. „Studia Semiotyczne”, t. 8, s. 107–113.
- Matsumura N., Ohsawa Y., Ishizuka M. (2003): *PAI: Automatic indexing for extracting asserted keywords from a document*. “New Generation Computing”, vol. 21, no 1, s. 37–47.
- Mooers C.N. (1963): *The Indexing Language of an Information Retrieval System. Information Retrieval Today: Papers Presented at an Institute Conducted by the Library School and the Center for Continuation Study, University of Minnesota, Sept. 19–22, 1962*. Ed. by W. Simonton. Minneapolis: The Center, s. 21–36.
- Muraszkiewicz M. (2005): *Esej: Nowy paradygmat, czyli od systemu do sieci*. [W:] *Od informacji naukowej do technologii społeczeństwa informacyjnego*. Pod red. B. Sosińskiej-Kalaty, M. Przystek-Samokowej przy współpracy A. Skrzypczaka. Warszawa: Wydawnictwo SBP, s. 83–86.
- Nęcka E., Orzechowski J., Szymura B. (2007): *Psychologia poznawcza*. Warszawa: Wydawnictwo Naukowe PWN.
- NISO (2004) *Thesaurus development team page: Developing the next generation of standards for controlled vocabularies and thesauri*. [Dok. elektr.]. Tryb dostępu: <http://www.niso.org/committees/MT-info.html>.
- ODLIS: *Online Dictionary for Library and Information Science*. [Dok. elektr.]. Tryb dostępu: http://lu.com/odlis/odlis_k.cfm.
- Oxford Wordpower. (2002). *Słownik angielsko-polski z indeksem polsko-angielskim*. Oxford – New York: Oxford University Press.
- Pacek J. (2010): *Jakość tytułów, opisów i słów kluczowych w polskich stronach internetowych dotyczących informacji naukowej* (W druku).
- Perreault J.M. (1994): *Categories and relators: a new schema*. “Knowledge Organization”, vol. 21, no 4, s. 189–198.
- Polguère A. (2008): *Lexical systems: graph models of natural language lexicons*. [Dok. elektr.] Tryb dostępu: <http://www.springerlink.com/content/n3247132gx15w876/fulltext.html>.
- Quillian M.R. (1968): *Semantic memory*. [W:] Minsky M.L. (ed.): *Semantic Information Processing*. Cambridge: MIT Press, s. 227–270.

- Rada R. (1991): *Focus on links: a holistic view of hypertext*. "Information and Classification", vol. 18, no 1, s. 13–18.
- Ribbert U. (1992): *Terminologiekontrolle in der Schlagwortnormdatei*. „Bibliothek. Forschung und Praxis”, vol. 16, no 1, s. 9–25.
- Richmond P.A. (1963): *Review of the Cranfield Project*. "American Documentation", no 4, s. 307–311.
- Robowski J. (1974): *Języki deskryptorowe. Analiza porównawcza języków deskryptorowych i innych języków informacyjnych*. Warszawa: IINTE.
- Rowley J. (1994): *The controlled versus natural indexing languages debate revisited: A perspective on information retrieval practice and research*. "Journal of Information Science", vol. 20, no 2, s. 108–119.
- Royo J.A., Mena E., Bernad J., Illarramendi A. (2005): *Searching the Web: From Keywords to Semantic Queries. Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05)*. Vol. 2. Washington: IEEE Computer Society, s. 244–249. [Dok. elektr.]. Tryb dostępu: <http://sid.cps.unizar.es/PUBLICATIONS/POSTSCRIPTS/icita05.pdf>.
- Rudnik-Karwatowa Z. (2002): *Słowa kluczowe – elementarne jednostki leksykalne języka informacyjno-wyszukiwawczego*. [W:] *Słowo z perspektywy językoznawcy i tłumacza*. Red. nauk. A. Pstyga, K. Szcześniak. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, s. 13–18.
- Rudnik-Karwatowa Z., Mikos Z. (2007): *Nowoczesny system informacji slawistycznej: zadania, dotychczasowe wyniki i perspektywy*. „Zagadnienia Informacji Naukowej”, nr 2 (90), s. 19–40.
- Sacharnyj L.V. (1989): *Struktury kojarzeniowe w języku naturalnym i języku informacyjno-wyszukiwawczym (na przykładzie języka haseł przedmiotowych)*. „Zagadnienia Informacji Naukowej”, nr 1 (54), s. 75–102.
- Sadowska J. (2001): *Języki informacyjno-wyszukiwawcze w komputerowych katalogach bibliotecznych*. [W:] *Informacja – Wiedza – Gospodarka*. Pod red. W. Pindlowej, D. Pietruch-Reizes. Warszawa: Polskie Towarzystwo Informacji Naukowej, s. 277–285.
- Sadowska J. (2010): *Kompetencja językowa twórców i użytkowników języków informacyjno-wyszukiwawczych*. [W:] *Informacja i język*. Pod red. W. Babika i D. Pietruch-Reizes. (W druku).
- Salton G. (1975): *Dynamic Information and Library Processing*. Englewood Cliffs, New Jersey: Prentice Hall.
- Salton G. (1988): *Automatic phrase construction for the representation of text content*. "Annual Review of OCLC Research" July 1987–June 1988, s. 30–31.
- Saracević T. (1991): *Individual differences in organizing, searching and retrieving information*. [W:] Griffiths J.M. (ed.): *SIS'91: Proceedings of the American Society for Information Science (ASIS) 54th Annual Meeting*. Washington, DC – Medford, NJ: Learned Information, Inc. vol. 28, s. 82–86.
- Saussure F. de (1961, wyd. pol. 1991): *Kurs językoznawstwa ogólnego*. Przeł. K. Kasprzyk. Warszawa: PWN.
- Saussure F. de (2004): *Szkice z językoznawstwa ogólnego*. Warszawa: Wydawnictwo Akademickie DIALOG.
- Schmetzke A. (1998): *The Utility of German WWW Search Services for North American Users*. "Reference Services Reviews", vol. 26, no 1, s. 43–50.
- Sigurd B. (1975): *Struktura języka. Zagadnienia i metody językoznawstwa współczesnego*. Przeł. Z. Wawrzyniak. Warszawa: PWN.

- Simpson J., Weiner E., ed. (1989): *The Oxford English Dictionary*. Vol. 8. Oxford: Clarendon Press.
- Sitarska A. (2006): *Systemowa problematyka języków informacyjno-wyszukiwawczych: w polu widzenia nauczyciela akademickiego*. „Zagadnienia Informatyki Naukowej”, nr 1 (87), s. 60–66.
- Słownik terminologiczny informacji naukowej*. (1979): Pod red. M. Dembowskiej. Wrocław: Zakład Narodowy im. Ossolińskich.
- Sosińska B. (1989): *Relacje między planem treści i planem wyrażania w językach informacyjno-wyszukiwawczych*. Warszawa: IINTE.
- Sosińska-Kalata B. (1999): *Modele organizacji wiedzy w systemach wyszukiwania informacji o dokumentach*. Warszawa: Wydawnictwo SBP.
- Sosińska-Kalata B. (2004): *Języki informacyjno-wyszukiwawcze. Współczesne tendencje w badaniach i zastosowaniu*. [W:] *Spoleczeństwo informacyjne i jego technologie*. Pod red. B. Sosińskiej-Kalaty, K. Materskiej, W. Glińskiego. Warszawa: Wydawnictwo SBP, s. 79–91.
- Sowa J.F. (1991): *Current issues in semantic networks*. [W:] Sowa J.F. (ed.): *Principles of semantic network. Explorations in the representations of knowledge*. San Mateo: Morgan Kaufmann, s. 13–43.
- Sowa J.F. (2000): *Knowledge representation*. Pacific Grove – Albany – London – Toronto: Brooks/Cole.
- Spink A., Cole Ch. (2005): *A Multitasking Framework for Cognitive Information Retrieval*. [W:] Spink A., Cole Ch., ed. *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, s. 99–112.
- Spiteri L.F. (2007): *Structure and form of folksonomy tags: The road to the public library catalogue*. „Information Technology in Libraries”, vol. 26, no 3, s. 13–25 + [online] „Webology” 2007, vol. 4, no 2. [Dok. elektr.]. Tryb dostępu: <http://www.webology.ir/2007/v4n2/a41.html>,
- Stachurski E. (1998): *Słowa-klucze polskiej epiki romantycznej*. Kraków: Wydawnictwo Naukowe WSP.
- Strzałkowski T. (1995): *Natural Language Information Retrieval*. „Information Processing & Management”, vol. 31, no 3, s. 397–417.
- Szczepanek W. (1981): *Wybór metody wiązania języków informacyjno-wyszukiwawczych na przykładzie języka słów kluczowych i tezaury*. „Aktualne Problemy Informatyki i Dokumentacji”, nr 1, s. 7–10.
- Ścibor E., Tomasiak-Beck J. (1970): *Analiza porównawcza języków informacyjnych (wybrane zagadnienia)*. Warszawa: CIINTE.
- Ścibor E. (1982): *Typologia strukturalna języków informacyjnych*. Warszawa: IINTE.
- Ścibor E. (1999): *Wybrane zagadnienia teorii języków informacyjnych*. Olsztyn: Wydawnictwo Naukowe WSP.
- Taghva K., Borsack J., Nartker T., Condit A. (2004): *The role of manually-assigned keywords in query expansion*. „Information Processing and Management”, vol. 40, no 3, s. 441–458.
- Tait J.I. (2005): *Charting a New Course*. Natural Language Processing and Information Retrieval. Dordrecht: New York. Springer.
- Taube M. (1955): *Storage and retrieval of information by means of the association of ideas*. „American Documentation”, vol. 6, no 1, s. 1–18.
- Taube M., Gull C.D., Wachtel I. (1952): *Unit Terms in Co-ordinate Indexing*. „American Documentation”, vol. 3, no 4, s. 213–218.

- Taube M. et. al. (1953): *Studies in Coordinate Indexing*. Vol. 1. Washington, D.C.: Documentation, Inc.
- Ungurian O. (1976): *Elementy teorii języków informacyjnych*. Warszawa: OIN PAN.
- Vickery B.C., Vickery A. (1987): *Information Science in Theory and Practice*. London: Butterworths.
- Voorbij H.J. (1998): *Title keywords and subject descriptors: A comparison of subject search entries of books in the humanities and social sciences*. "Journal of Documentation", vol. 54, no 4. s. 466–476.
- Westbrook L. (2006): *Mental models: a theoretical overview and preliminary study*. "Journal of Information Science", vol. 32 s. 563–579. [Dok. elektr.]. Tryb dostępu: <http://jis.sagepub.com/cgi/content/abstract/32/6/563>.
- Will L. (2005): *Glossary of terms relating to thesauri and other forms of structured vocabulary for information retrieval* [online]. [Dok. elektr.]. Tryb dostępu: <http://www.willpowerinfo.co.uk/glossary.html>.
- Woźniak J. (1997): *Kognitywizm w informacji*. „Zagadnienia Informacji Naukowej”, nr 2 (70), s. 3–16.
- Woźniak J. (1999): *Słownik języka informacyjno-wyszukiwawczego jako słownik przekładowy*. „Semiosis Lexicographica”, III. Warszawa: Uniwersytet Warszawski, IINiSB, s. 43–50.
- Woźniak J. (2000): *Kategoryzacja. Studium z teorii języków informacyjno-wyszukiwawczych*. Warszawa: Wydawnictwo SBP.
- Woźniak-Kaspepek J. (2004): *Internet – spojrzenie z perspektywy organizacji informacji*. „Zagadnienia Informacji Naukowej”, nr 1 (83), s. 29–45.
- Wyka K. (1973): *Słowa-klucze*. [W:] *Stylistyka polska. Wybór tekstów*. Wybór, opracowanie i wstęp E. Miodońska-Brookes, A. Kulawik, M. Tatar. Warszawa: PWN, s. 153–167.
- Zacharow W.P. (2002): *Informacionnyje systemy (dokumentalnyj poisk)*. Sankt Petersburg, Sankt-Petersburskij Gosudarstvennyj Universitet.

Indeks osobowy

A

Abramowicz, Witold 217
Aitchinson, Jean 48, 217
Alonso-Arroyo, Adolfo 13, 176, 220
Anderson, John Robert 46, 50
Artowicz, Elżbieta 62
Asurbanipal 9
Austin, Derek 152
Azvine, Ben 197, 219

B

Babik, Wiesław 13, 38, 58, 95, 106, 109, 133, 145, 147, 148, 180, 193, 201, 202, 207, 212, 217, 223
Bachelard, Gaston 25
Bańkowska, Anna J. 13, 217
Barthes, Roland 93, 218
Bean, Carol A. 45, 218
Beling, Gerd 87, 218
Belkin, Nicholas J. 16, 218
Berkan, Riza C. 196
Bernad, Jorge 198, 223
Bertram, Jutta 19, 87, 116, 218
Bielicka, Lucyna Anna 13, 116, 125, 126, 218
Bojar, Bożenna 10, 19, 20, 22, 30, 44, 49, 62, 63, 64, 65, 66, 69, 70, 72, 85, 87, 104, 107, 111, 119, 127, 190, 200, 216, 218
Boole, George 188, 190, 200
Borger, Zvi 13, 218
Börner, Katy 15, 218
Borsack, Julie 13, 224
Bower, Gordon H. 46
Brownson, Helen L. 76
Buckland, Michael 10, 96, 218
Budzyk, Kazimierz 25
Bush, Vannevar 15, 83, 218
Butler, Danuta 46

C

Carballo, José Pérez 215, 218
Chen, Ding-Wen 198
Chlewiński, Zdzisław 47, 218

Chmielewska-Gorczyca, Ewa 61, 66, 89, 92, 96, 218, 219
Chomsky, Noam 54, 70
Choroszczyna, Jadwiga 13, 27, 28, 29, 178, 219
Chowdhury, G. Gobinda 36, 219
Chu, Heting 84, 219
Cieszewska, Teresa 13, 219
Cieślowski, Sławomir 26
Clarke, Stella Dextre 74
Cleverdon, Cyril William 12, 190, 219
Cole, Charles 16, 222, 224
Collins, Alan M. 17, 46, 47, 48, 50, 219
Condit, Allen 13, 224
Costello, John C. 83, 219
Cox, Nigel S.M. 31
Craven, Jenny 220
Craven, Timothy C. 13, 198, 219
Cronin, Blaise 218
Cui, Zhan 197, 219
Czarniawska, Barbara 220
Czerny, A. I. 13, 27, 89, 157, 219

Ć

Ćwiekowa, Jadwiga 30, 80, 81, 84, 168, 219

D

Davies, Ron 74
Dembowska, Maria 224
Dębicka, Teresa 155
Dijk, Teun A. van 172
Doroszewski, Witold 182
Ducatel, Géry 197, 219
Duszeńko-Król, Elżbieta 147

F

Farradane, Jason E.L. 175, 219
Fischbach, Gerda 132
Fischer, Marguerite 27
Fitowa, Alina 115, 133
Frants, Valery I. 30, 219
Froelich, Thomas J. 188, 219

G

Gabryś-Barker, Danuta 17, 219
Gardin, Jean-Claude 19, 96, 219
Garshol, Lars Marius 213, 219
Gey, Frederic C. 198, 220
Gil-Leiva, Isidoro 13, 176, 220
Gilchrist, Alan 74
Gliński, Wiesław 169, 224
Głazek, Dorota 14, 212, 220
Głowacka, Ewa 14, 220
Golonka-Czajkowska, Monika 147
Goźdź-Roszkowski, Stanisław 217
Gödert, Winfried 45, 220
Green, Rebecca 45, 218
Griffiths, José-Marie 223
Griffiths, Jullian R. 220
Grose, Michael 31
Gross, Tina 13, 220
Guilford, Joy Paul 51, 175, 220
Guiraud, Jean Pierre 25, 220
Gull, Cloyd Dake 82, 224

H

Hanicke, Peter 12, 220
Hanjalic, Alan 156, 222
Hartley, James 13, 220
Hartley, Richard J. 151, 220
Hildreth, Charles R. 192, 220
Hulme, Wyndham Edward 137
Hume, David 47
Hutchins, William John 19, 60, 167, 220
Hys, Jolanta 168, 170, 220

I

Illarramendi, Arantza 198, 223
Ingwersen, Peter 16, 154, 220, 222
Ishizuka, Mitsuru 181, 222

J

Jadacka, Hanna 57, 221
Jansen, Bernard J. 192, 200, 221
Järvelin, Kalervo 16, 220
Johnson-Laird, Philip N. 48
Jones, Steve 13, 221
Jung, Carl Gustav 25

K

Käki, Mika 153, 221
Kando, Noriko 198, 220
Karpińska, Hanna 132, 136
Kasprzyk, Krystyna 223
Kazimierz Wielki 92
Kilgour, Frederick G. 81, 221
Kintsch, Walter 48, 221
Kniaginina, Maria 220
Klimek, Piotr 155
Koblitz, Josef 12, 27, 116, 117, 221

Kostoff, Ronald N. 13, 220
Kouri, Pentti 196
Kozłowski, Waldemar 220
Kredens, Krzysztof 217
Kuflik, Tsvi 13, 218
Kulawik, Adam 225
Kuraszkiewicz, Władysław 26
Kurecz, Ida 17, 54, 55, 221
Kurek-Kokocińska, Stanisława 10, 13, 221

L

Lancaster, Frederick Wilfrid 13, 14, 150, 221
Langefors, Börje 11
Larsen, Birger 154, 222
Le Deuff, Olivier 183, 184, 185, 222
Lee, Chung-Hong 198
Lewandowska-Tomaszczyk, Barbara 217
Libera, Zbigniew 147
Lindsay, Peter H. 17, 222
Loftus, Elizabeth F. 17, 50, 219
Lu, Lie 156, 222
Lubaszewski, Wiesław 73
Luhn, Hans Peter 12, 80, 108, 152, 153, 222
Lynch, Jim 31
Lyons, John 54, 222

Ł

Łobacz, Piotr 17, 35, 222
Łysakowski, Adam 80

M

Maj, Małgorzata 147
Marciszewski, Witold 13, 222
Materska, Katarzyna 169, 224
Matsumura, Naohiro 181, 222
Matwiejczuk, Wanda 135
Mayenowa, Maria Renata 25
Mena, Eduardo 198, 223
Mickiewicz, Adam 26
Mikołajczak-Matyja, Nawoja 17, 35, 222
Mikos, Zenon 13, 223
Miller, George A. 48
Mills, Jack 13, 190, 221
Minsky, Marvin Lee 222
Miodońska-Brookes, Ewa 225
Miodunka, Władysław 46
Mooers, Calvin Natop 30, 82, 83, 84, 90, 222
Morris, Charles W. 16
Moskovič, V.A. 174
Muraszkiewicz, Mieczysław 15, 222

N

Nartker, Thomas 13, 224
Nęcka, Edward 48, 50, 222
Norman, Donald A. 17, 222
Norris, Dorothy May 80

O

Ohsawa, Yukio 181, 222
Okoń, Helena 135
Olszewska, Magdalena 117
Orzechowski, Jarosław 48, 50, 222

P

Pacek, Jarosław 172, 222
Paynter, Gordon W. 13, 221
Perreault, Jean M. 175, 222
Peters, Carol 198, 220
Pfaff, Anthony 200, 221
Pierzchalska, Grażyna 116, 132
Pietraszkiewicz, Janina 132
Pietruch-Reizes, Diana 217, 223
Pietrzyk, Tadeusz 125
Pindlowa, Wanda 223
Pisarek, Walery 26
Pokusińska, Zofia 155
Polański, Kazimierz 219
Poletyło, Mikołaj 27
Polguère, Alain 103, 222
Pooch, Udo 200, 221
Port, Peter 87, 218
Przastek-Samokowa, Maria 222
Pstyga, Alicja 223

Q

Quillian, M. Ross 46, 47, 48, 219, 222

R

Rachwałowa, Maria 26
Rada, Roy 43, 223
Radzicki, Józef 220
Rej, Mikołaj 26
Ribbert, Ulrich 19, 223
Richmond, Phyllis Allen 13, 190, 223
Robotycki, Czesław 143, 144, 145, 147, 148
Robowski, Józef 13, 27, 76, 83, 116, 223
Roszkowski, Marcin 36
Rowley, Jennifer 202, 223
Royo, José A. 198, 223
Ryszewski, Bohdan 220
Rudnicka-Fira, Elżbieta 25, 26
Rudnik-Karwatowa, Zofia 13, 57, 132, 136, 223

S

Sacharnyj, Leonid V. 67, 69, 223
Sadowska, Jadwiga 71, 154, 223
Salton, Gerard 13, 16, 183, 223
Sambor, Jadwiga 26
Sanyal, Soma 15, 218
Saracevič, Tefko 45, 223
Saussure, Ferdinand de 54, 69, 93, 223
Schmecke, Axel 190, 223
Schrettinger, Martin 81
Shapira, Bracha 13, 218

Shapiro, Jacob 30, 219
Shoval, Peretz 13, 218
Sigurd, Bengt 26, 223
Simpson, John 224
Simonton, Wesley 222
Sitarska, Anna 16, 224
Siuda, Joanna 117
Skrzypczak, Andrzej 222
Smith, George 31
Sokołowska, Wanda 13
Sosińska Barbara zob. Sosińska-Kalata Barbara
Sosińska-Kalata, Barbara 16, 19, 20, 43, 44, 46, 51, 62, 101, 163, 169, 222, 224
Sowa, John F. 46, 48, 224
Spink, Amanda 16, 192, 200, 221, 222, 224
Spiteri, Louise F. 184, 185, 224
Sprys, Zofia J. 132, 134
Stachurski, Edward 25, 26, 224
Stankiewicz, Zbigniew 132
Steinberger, Ralf 198
Strohl-Goebel, Hindburg 87, 218
Strzałkowski, Tomek 215, 218, 224
Sungren, Bo 11
Surowaniec, Józef 132, 135
Szczepanek, Wiesław 13, 224
Szczęśniak, Krystyna 223
Szumowska, Jolanta 125
Szymura, Błażej 48, 50, 222

Ś

Ścibor, Eugeniusz 13, 19, 65, 85, 95, 97, 99, 112, 116, 173, 174, 218, 219, 224

T

Taghva, Kazem 13, 224
Tait, John I. 215, 224
Tatara, Marian 225
Taube, Mortimer 12, 30, 31, 81, 82, 83, 224, 225
Taylor, Arlene G. 13, 220
Tokarski, Ryszard 46
Tomasik-Beck, Joanna 13, 85, 221, 224
Tomasz z Akwinu 80
Tomlin, Eric Walter Frederick 31
Trebunia-Staszal, Stanisława 147

U

Ungurian, Olgierd 13, 19, 32, 65, 91, 92, 225

V

Vespignani, Alessandro 15, 218
Vetulani, Grażyna 217
Vetulani, Zygmunt 73, 217
Vickery, Alina 45, 225
Vickery, Brian Campbell 45, 76, 172, 225
Voiskunskii, Vladimir G. 30, 219
Voorbij, Henk J. 12, 13, 225

W

Wachtel, Irma S. 82, 224
Waliński, Jacek 217
Wawrzyniak, Zdzisław 223
Weiner, Edmund 224
Wereszczyńska-Cisło, Barbara 66
Westbrook, Lynn 46, 225
Will, Leonard 74, 225
Wojnowska-Dawiskiba, Halina 35
Woźniak Jadwiga zob. Woźniak-Kasperek Jadwiga
Woźniak-Kasperek, Jadwiga 16, 19, 39, 52, 99, 118,
183, 225

Wyka, Kazimierz 25, 225
Wyspiański, Stanisław 25

Y

Yang, Hsin-Chang 198

Z

Zacharow, W.P. 87, 225
Zarębina, Maria 26
Zembaty-Michalakowa, Maria 26
Żeromska, Felicja 13, 219
Żmijewska, Urszula 132

Indeks rzeczowy

A

Algebra Boole'a 188–189
Algorytm PageRank 195
ALLOY 43–44
Astroturfing 185

B

BAZTECH 130

C

Charakterystyka słowna (dokumentu) 173
Charakterystyka wyszukiwawcza (dokumentu) 170
CYTBIN 129

D

Denotacja 101
Deskryptor 28, 84
Deskryptory Mooersa 82, 84
Dokumentacyjna interpretacja relacji 64–65
Drzewo Porfiriusza 48
Dźwięki kluczowe 156–157

E

Eksperymenty cranfieldzkie 12–13

F

Falszywa koordynacja 113, 183
fKWIC (frequency-based Key Word Context Index) 153
Folksonomie 183–185
Formalizacja języka 107
Funkcje języka słów kluczowych 89–90, 106
 metainformacyjna 89
 określania zakresów tematycznych 89
 organizacyjna (porządkująca) 89
 wyszukiwawcza 89

G

Gościnność (produktywność) języka 200
Gramatyka 105–106
Gramatyka generatywno-transformacyjna
 N. Chomsky'ego 54
Gramatyka zerowa 105, 119

H

Hakia.com (wyszukiwarka) 196
Hipertekst 42–43
HTML 38, 180

I

Indeks nieujednoczonych słów 155
Indeksowanie (parametry)
 głębokość (szczegółowość) 170, 172
 szerokość (kompletność) 172
Indeksowanie (rodzaje)
 automatyczne 39, 178–179, 181, 182, 183, 187, 215
 fasetowe (ankietowe, rastrowe, pozycyjne) 174
 intelektualne 182
 słownikowe (kontrolowane, normatywne) 173
 swobodne 87, 112, 126, 187
 uogólniające 168
 współrzędne (kojarzeniowe, koordynowane) 81, 168, 172
 wyszczególniające 168
Indeksowanie relacyjne J.E.L. Farradane'a 175
Infologiczna koncepcja informacji 11
Instrukcja indeksowania 125–127, 170, 175–176
Instrukcja wyszukiwawcza (dokumentu) 188

J

Jednostki leksykalne
 autosyntaktyczne 117
 syntaktyczne 117
Język deskryptorowy 114–117
Język haseł przedmiotowych 117–118
Język indeksowania 89–90, 155
Język informacyjno-wyszukiwawczy
 definicja przeznaczeniowa 106
 historia 9
 metody budowy 82
 teoria 15, 16
Język kontrolowanych słów kluczowych (nieklasyczny) 111, 158, 161
Język marketingowy 35
Język naturalny

formalizacja 107
funkcje 62
pole semantyczne 62–66
Język pośrednik 90, 194
Język słów kluczowych
a język deskryptorowy 114–117, 118, 207
a język haseł przedmiotowych 80, 84, 117–118, 207
a język naturalny 64, 66, 87, 106–109, 119, 207–209
a język unitermowy 84
a użytkownicy 206, 210
badania 12–14, 16, 17, 206
charakterystyka modelowa 85
definicje 85–86, 119
elementy 88–89, 93–106, 167
system gramatyczny 105–106
system leksykalny 93–95
system pragmatyczny 167–194
system semantyczny 100–106
system słownikowy 95–100
funkcje 63–64, 86, 87, 89–90, 167
geneza 80–85
gramatyka 105–106, 119
granice języka 108, 116, 199, 201, 207
historia 84–85
kryteria opisu 165–167
miejsce w systemie informacyjno-wyszukiaw-
czym 86
natura 86–87, 207, 209
odmiany/rodzaje 111–112, 207
pole semantyczne 63–67, 101, 102, 207–208
poziomy 90–93
prehistoria 80–84
prezentacja 72, 104, 171
projektowanie 87
przyjazność 210
struktura 88, 90–106, 167
sztuczność 86–88, 119, 199, 204, 214
teksty 74, 170
tożsamość 10
tworzenie 87
warianty 160
w Internecie 14, 157–159, 186–194
w systemach hipertekstowych 161
w wyszukiwaniu informacji 10, 81, 190, 192–193
wpływ języków paranaturalnych 163
zalety i wady 112–113, 213
Język swobodnych/niekontrolowanych słów kluczo-
wych (klasyczny) 84, 87, 94, 97, 111, 124, 157,
159, 189, 203–204
Język unitermowy 81, 83–84
Język wyszukiwawczy 89–90, 155, 190
Język wyszukiwawczy (kwerend) 74

K

Kartoteka/lista wzorcowa słów kluczowych 59, 98,
179
Kartoteka/lista wzorcowych słów kluczowych 98,
165
Katalog internetowy 181–182
Kategoryzacja 103
Kluczowość słowa 94
Kod wokalny 156–157, 193, 214
Kompetencja
gramatyczna 70
językowa 54, 70–73, 131, 138, 150, 209
leksykalna 41
semantyczna 41, 65, 70, 138
Konkatencja 88
Konkordancje 80, 153
Konotacja (nazwy, pojęcia) 101
KWIC (Key-Word-in-Context) 152
KWIT (Key-Words-in-Title) 152
KWOC (Key-Word-out-Context) 152

L

Leksykografia
dokumentacyjna 96
naturalna 96
Leksykon umysłowy (słownik mentalny) 53–56
Lista
wzorcowa słów niekluczowych 165
wzorcowych słów niekluczowych 165, 179
Literary warrant 108, 137, 157, 162

M

Mapy myśli (ang. think-maps) 213
Mapy semantyczne (ang. semantic maps) 212
Mapy umysłu (ang. mind-maps) 213
Maskowanie końcówek 186, 191
Medical Subject Headings (MeSH) 115
MEMEX 15
Metadane 179
Metoda ciągów definicyjnych 212
Model A.M. Collinsa i E.F. Loftusa 50
Model A.M. Collinsa i M.R. Quilliana 47
Model G.A. Millera i P.N. Johnsona-Lairda 48
Model W. Kinscha 48
Modele pamięci semantycznej (mentalne) 46–52
hierarchiczne 47, 48, 49
sieciowe (niehierarchiczne) 49
Model systemu poznawczego I. Kurcz 54–55
Motywacja 34–35

N

Nauka o poznawaniu (kognitywistyka) 11, 16
Neuronauka 52

O

- Odsyłacze 43, 134–135
- OKAPI (Online Keyword Access to Public Katalog) 154
- Ontologie 150, 195, 205
- Ontologia hybrydowa 196–197
- OPAC (katalog online) 153–154
- Operatory
 - algebry Boole'a 86, 189
 - arytmetyczne 191
 - kontekstowe 86
 - pozycyjne 191
- Ośrodek Informacji Naukowej PAN w Warszawie 114–115
- Pracownia w Krakowie 115, 132–133

P

- Pamięć semantyczna 46–52
- Paradygmaty w badaniach umysłu
 - neuropsychologiczny 53
 - teoriainformacyjny 53
- Paradygmaty w teorii języków informacyjno-wyszukiawczych
 - sieciowy 15–16, 76, 103, 206, 213
 - systemowy 15–16, 76, 206
- Pertynencja (relewanca pragmatyczna) 190
- Pięścienie synonimów 193
- Plany językowe
 - notacyjny 55, 61
 - treści 55, 61
 - werbalny 55
- Pole dokumentacyjne 44
- Pole informacyjne 44
- Pole kojarzeniowe (asocjacyjne) 66, 69, 73
- Pole semantyczne 44, 62–66, 101–102
 - struktura płaska (ahierarchiczna) 102
 - struktura prawie płaska 102
- Pole tematyczne 161
- Polireprezentacja informacji 154
- Pragmatyka indeksowania 172–185
- Pragmatyka języka 16, 127, 167
 - definicje 16, 89, 167
- Pragmatyka wyszukiwania 186–194
- Prekoordynacja 49, 84
- Prezentacja języka 104–105, 171
- Pozycjonowanie 34, 39, 195

R

- Relacje językowe 93
 - ekwiwalencji (odpowiedniości) wyszukiawczej 99–100, 139, 160–161
 - hierarchiczne 102–103, 135
 - interpretacja dokumentacyjna 64–65
 - kojarzeniowe (asocjacyjne) 19, 66–70, 102, 105, 135, 161–162, 212
 - synonimii wyszukiawczej 103–104, 201, 208

- syntagmatyczne 105
- sytuacyjne 169, 188, 212
- tematyczne 68, 102, 105, 161
- współwystępowania (kookurencji) 105
- Relatory J.M. Perreaulta 175
- Relewanca techniczna 190

S

- Sieć semantyczna 46, 51, 212
 - A.M. Collinsa i M.R. Quilliana 47
- Skojarzenia
 - ciągłe 69
 - kierowane 69
 - paradygmatyczne 67–68
 - swobodne 69
 - syntagmatyczne 67–68
 - tematyczne 67–68
- Słowa klucze 25–26
- Słowa kluczowe
 - a deskryptory 28–29
 - a hasła przedmiotowe 28–30, 80
 - a terminy 33, 34, 35, 58–60
 - a unitermy 82
 - a wyrażenia języka naturalnego 32, 58, 65, 75
 - ahierarchiczność 113
 - automatyczna ekstrakcja 156, 179, 180–181
 - autorskie 111, 124, 159
 - badania 12, 14, 19, 31
 - definicje 24–34, 74–75, 77
 - determinanty 45–74
 - językowe 56–74
 - psychologiczne 45–56
 - elementy istotne 32
 - funkcje 24–26, 32, 34–36, 57, 72, 89–90
 - informacyjna 35
 - metainformacyjna 32, 34
 - oznaczania 37, 72
 - promująca (promocyjna/perswazyjna/reklamowa) 35, 37
 - wyszukiawcza 32, 34
 - konteksty 23
 - jednostki leksykalne 33
 - kluczowość 94
 - konwencjonalność 61
 - kryteria wyboru 135
 - liczba 172, 183
 - narzędzia wyszukiwania 37–38
 - natura 75, 77
 - odmiany/rodzaje 24–26
 - opracowywanie leksykograficzne/inżynieria leksykalna/redakcja 91–92, 169, 174, 175
 - poziomy 186–188
 - relewantność 34
 - słowo znaczące 30, 33
 - szczegółowość 162, 164

- wartość informacyjna 94
 - w bibliotekach cyfrowych 156
 - w działalności informacyjnej 24, 27–34, 42
 - w indeksach rzeczowych 152–153, 162
 - w indeksach KWIC i KWOC 30, 152
 - w informatyce 24–25, 26
 - w Internecie (systemie WWW) 34, 36–44, 186–194
 - w językoznawstwie 24–25
 - w języku marketingowym 35
 - w katalogach bibliotecznych OPAC 153–156, 192
 - w literaturoznawstwie 26
 - w promocji i reklamie 37
 - w systemach hipertekstowych 42–44
 - w systemach informacji graficznej 156
 - w systemach pełnotekstowych 156
 - w wyszukiwarkach internetowych 36–42, 186–198
 - wyszukiwanie 37, 38, 39, 192–193
 - źródła 33, 130, 156, 157
 - Słownictwo 93–100, 106, 125, 171–172
 - Słowniki (typy)
 - hybrydowy 97
 - jawny 72, 125
 - mentalny/umysłowy 53–56, 97
 - niejawny 72
 - otwarty 95, 125
 - typologia E. Ścibora 99–100
 - w sieci 37
 - zadane instrukcją 72, 125–126
 - Słowniki (indeksy, listy, wykazy) słów kluczowych
 - a tezaurus 115
 - cechy 95–96
 - definicje 32, 119
 - formy 97
 - funkcje 96, 98, 163
 - klasyczne 124–131
 - metody budowy 95, 97, 133
 - struktura 95, 97
 - fasetowa 139–143
 - gniazdowa 143–150
 - typy 97, 99–100
 - wielojęzyczne 150–151, 198
 - z systemem odesłań 131–139
 - Stop-lista (słownik negatywny) 72, 119, 165, 181
 - Strona internetowa 38–40, 180
 - indeksowanie 39
 - struktura 38–41
 - SYNABA 124–129
 - System
 - dokumentacyjny 9
 - leksykalny 58, 93–95, 101–102
 - organizacja sieciowa 76
 - gramatyczny 105–106
 - leksykalno-semantyczny 42
 - pragmatyczny 167–194
 - semantyczny 100–105
 - słownikowy 62, 95–100
 - terminologiczny 58
 - Systemy hipertekstowe i hipermedialne 42–44
 - Supozycja informacyjna 58
 - Supozycja materialna 81, 111
 - Supozycja normalna (zwykła) 111
 - Szum informacyjny 108, 113
- T**
- Tagowanie 184–185
 - Teoria procesów poznawczych J.P. Guilforda 51–52
 - Termin 33, 57–59, 61
 - Terminologia 57, 59, 60, 61
 - Terminologiczny bank danych (bank terminów) 59
- U**
- Unified Medical Language System (UMLS) 115
 - UNITERM (system) 30–31, 81–82
 - Uniterm (jednostka leksykalna) 30, 82, 83
 - Uniterm a deskryptor 82, 83
- W**
- Web 2.0 183, 185
 - Web 3.0 195
 - Wiedza językowa (o języku) 54–56, 71
 - Wiedza o świecie (pozajęzykowa) 54–55, 71
 - Wielofunkcyjne słowniki otwarte (terminologiczne) 109–110
 - Wskaźniki wagi/współczynniki wagowe 180–181, 187
 - Wyrażenia złożone 106
 - Wyszukiwanie w języku naturalnym 195–198, 202, 214, 215

Aneks: Wykaz słowników słów kluczowych poddanych analizie

Nie obejmuje słowników dostępnych w Internecie

Alphabetical List of the Key-Words for the Clearinghouse File at IINTE. (1982): Warsaw. (Mat. powiel.).

Deskryptornyj slovar po informatike. Sostaviteli: N.A. Pašenko, E.B. Ksenofontova, B.F. Skorobogataja. Naučný redaktor: A.I. Czernyj. Moskva: VINITI.

Fischbach (1970): *Schlagwortverzeichnis Pädagogik.* Bearbeitet von G. Fischbach. Berlin: Deutsches Pädagogisches Zentralinstitut Berlin – Zentralstelle für pädagogische Information und Dokumentation.

Indeks alfabetyczny słów kluczowych (zdrowie publiczne). (1993): Kraków: Centrum Informacji Naukowej i Biblioteka. Szkoła Zdrowia Publicznego Akademii Medycznej i Uniwersytetu Jagiellońskiego. (Maszyn. powiel.).

Index des mots-clés de la base données de l'ENSP au 8 février 1990. (Maszyn. powiel.).

Klimek P., oprac. (2001): *Słownik słów kluczowych z teologii. Wydanie studyjne.* Warszawa: Federacja Bibliotek Kościelnych FIDES.

Matwiejczuk W., Okoń H. (2002): *PEDAGOG. Słownik słów kluczowych bibliotecznego systemu informacyjno-wyszukiwawczego: pedagogika opiekuńczo-wychowawcza, wychowanie przedszkolne, edukacja wczesnoszkolna.* Opole: Uniwersytet Opolski. Biblioteka.

Pierzchalska G. (1986): *Słownik naukowy słów kluczowych transportu morskiego.* Gdańsk – Słupsk – Szczecin: Wydawnictwo Instytutu Morskiego.

Pietraszkiewicz J., Stankiewicz Z., Żmijewska U. (1997): *Słownik języka informacyjno-wyszukiwawczego Biblioteki Społeczno-Ekonomicznej.* Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.

Rudnik-Karwatowa Z., Karpieńska H. (1999): *Słownik słów kluczowych językoznawstwa sławistycznego.* Warszawa: Towarzystwo Naukowe Warszawskie. PAN – Instytut Sławistyki. (Słownik dostępny również w wersji online. Warszawa 2006. http://www.ispan.waw.pl/cnis/slownik_slow_kluczowych.htm).

Schlagwortliste für die Sozialwissenschaften. Systematischer Teil. (1986): Bearbeiter: U. Becker, D. Lohmeyer, H. Schott und anderen. Bonn: Informationszentrum Sozialwissenschaften.

Schlagwortliste für die Sozialwissenschaften. English-Deutsch. (1987): Bearbeitet von H. Scott. Bonn: Informationszentrum Sozialwissenschaften.

- Słownik haseł z zakresu zdrojownictwa i dziedzin pokrewnych.* (1977): Warszawa: Zjednoczenie „Uzdrowiska Polskie”.
- Słownik słów kluczowych z ekonomii politycznej.* (1988): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: B. Kędziołek, J. Kłapkowska, M. Matuszewski, M. Nowak, R. Schott, M. Wachta, T. Włudyka. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z botaniki.* (1990): Praca zbiorowa. Wybór i opracowanie słów kluczowych: Z. Mirek, M. Nowak, J. Oleszak, B. Pietras. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, M. Nahotko). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z filozofii.* (1988): Praca zbiorowa. Wykaz podstawowych terminów przygotował F. Zawadzki. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z historii Polski (z elementami historii powszechnej).* (1992): Praca zbiorowa. Wybór i opracowanie słów: A. Fitowa, przy udziale M. Feliksa i W. Konicznego. (Kierownictwo i opracowanie metodyczne: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z językoznawstwa.* (1988): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: W. Lubaszewski, F. Sowa. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z literaturoznawstwa.* (1990): Praca zbiorowa. Wykaz podstawowych terminów przygotowała A. Lubaszewska. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z marksizmu-leninizmu.* (1989): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: F. Zawadzki, A. Komendera. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z nauki o polityce.* (1989): Praca zbiorowa. Wykaz podstawowych terminów przygotował M. Grzybowski. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z neuropsychofarmakologii.* (1990): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: B. Morawska-Nowak, J. Szczepanik. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z pedagogiki.* (1988): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: Z. Gawlina, B. Kozłowska-Reczyńska, B. Łuczyńska, E. Marcjan. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z psychologii.* (1988): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: M. Kielar, J. Siuta. (Kierownictwo i opracowanie metodyczne słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z religioznawstwa.* (1993): Praca zbiorowa. Wykaz i opracowanie słów: H. Hoffmann, W. Kowalak, T. Margul, H. Swenko. (Opracowanie metodyczne słownika: A. Fitowa, B. Kamińska-Czubała, M. Nahotko, W. Szczęch). Kraków – Warszawa (prace OIN PAN).
- Słownik słów kluczowych z socjologii.* (1989): Praca zbiorowa. Wykaz podstawowych terminów przygotowali: A. Komendera, J. Szmata. (Kierownictwo i opracowanie metodyczne

- słownika: A. Fitowa, W. Babik, M. Nahotko, W. Szczęch). Kraków: OIN PAN (Prace OIN PAN).
- Słownik słów kluczowych z metodologii nauk społecznych.* (1990): Oprac. Z.J. Sprys. Poznań: OIN PAN 1990 (Prace OIN PAN).
- Słownik słów kluczowych zdrowia publicznego (polsko-niemiecki).* (1994): Vol. 2. Kraków: Centrum Informacji Naukowej i Biblioteka. Szkoła Zdrowia Publicznego Akademii Medycznej i Uniwersytetu Jagiellońskiego. (Maszyn. powiel.).
- Słownik słów kluczowych zdrowia publicznego (polsko-francuski).* (1993): Kraków: Centrum Informacji Naukowej i Biblioteka. Szkoła Zdrowia Publicznego Akademii Medycznej i Uniwersytetu Jagiellońskiego. (Maszyn. powiel.).
- Słownik słów kluczowych zdrowia publicznego (francusko-polski).* (1993): Vol. 1. Kraków: Centrum Informacji Naukowej i Biblioteka. Szkoła Zdrowia Publicznego Akademii Medycznej i Uniwersytetu Jagiellońskiego. (Maszyn. powiel.).
- Spisok normalizovannoj leksiki po filozofii i socjologii.* (1989): Cz. 1–3. Moskwa: Meždunarodnaja Informacionnaja Sistema po Občestviennym Naukam.
- Spisok ključovych slov po farmakologii, himičeskije sredstva.* (1984): Moskva: VINITI.
- Spisok ključovych slov po farmakologii. (Farmakologija obšâ. Farmakologija nervoj sistemy).* (1987): Moskva: VINITI.
- Surowaniec J. (1993): *Słownik słów kluczowych w logopedii.* Kraków: WSP im. Komisji Edukacji Naukowej w Krakowie.
- Układ gniazdowy i słownik słów kluczowych. Etos, obrzędy, demonologia, magia.* (2002): Pod red. Cz. Robotyckiego, W. Babika. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Układ gniazdowy terminów i słownik słów kluczowych wybranych kategorii kultury. Medycyna ludowa.* (2005): Pod red. Cz. Robotyckiego, W. Babika. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Układ słów kluczowych do bazy danych o źródłach etnograficznych. (Kultura ludowa Karpat Polskich).* (1995): Pod red. Cz. Robotyckiego. Zeszyty Naukowe UJ, seria Varia, t. CCCXLIII, Kraków.

Spis rysunków

Rys. 1. Pole znaczeniowe terminu „słowo kluczowe”	29
Rys. 2. Przykład opisu dokumentu w języku HTML	41
Rys. 3. Model pamięci semantycznej według A.M. Collinsa i M.R. Quilliana (1969)	48
Rys. 4. Sieć pamięci semantycznej według modelu A.M. Collinsa i E.F. Loftus (1975)	50
Rys. 5. Porównanie słowa kluczowego z deskryptorem	117
Rys. 6. <i>Słownik słów kluczowych językoznawstwa slawistycznego</i> (wersja online)	137
Rys. 7. Język słów kluczowych w katalogu OPAC	155
Rys. 8. Elementy opisu artykułu w publikacji elektronicznej	177

Keywords

Summary

In my study, I discuss the most important properties of keywords as lexical units of the keywords language that decide of the essence and value of the language and differentiate that language from other indexing and retrieval languages. My considerations are based on the process, functional and structural approach to that matter. I apply a detailed analysis to the functional, structural, semantic and pragmatic aspects of those languages. I emphasize the changeability and development of the theoretical foundations of keyword languages. I also make an attempt at determining the identity of that class of languages, creating for them an autonomous set of descriptive criteria as a basis of the future language typology. I pay a special attention in that respect to the relativism of perceiving the keyword language nature, constituting an essential element of its identity.

By comparing the keyword languages to other indexing and retrieval languages and a natural language, I formulate a conclusion of mutual language supplementing (complementariness) in indexing and retrieval systems. My study is concluded with my remarks on the current validity of the issues and the future of keywords in contemporary indexing and retrieval systems.

This keyword language monograph is designed for a wide group of users, especially including the specialists of language and indexing and retrieval system scopes, as well as the IT and library science students, documentalists and librarians.

Keywords: Keywords. Information retrieval. Indexing. Internet. Information Organization. Indexing and retrieval languages. Documentary indexing and retrieval system.

REDAKTOR PROWADZĄCY

Lucyna Sadko

ADIUSTACJA JĘZYKOWO-STYLISTYCZNA

Mirosław Ruszkiewicz

KOREKTA

Ewa Baran

SKŁAD I ŁAMANIE

Katarzyna Mróz

Wydawnictwo Uniwersytetu Jagiellońskiego
Redakcja: ul. Michałowskiego 9/2, 31-126 Kraków
tel. 12-631-18-81, 12-631-18-82, fax 12-631-18-83