

Marek Nahotko

ONTOLOGIA I ONTOLOGIE – OD FILOZOFII ARYSTOTELESA DO SZTUCZNEJ INTELIGENCJI

[ONTOLOGY AND ONTOLOGIES – FROM ARISTOTLE
TO ARTIFICIAL INTELLIGENCE]

Abstrakt: W artykule przedstawiono znaczenie terminu *ontologia* w wybranych dziedzinach: filozofii, bibliotekarstwie i informacji naukowej, sztucznej inteligencji, przetwarzaniu języka naturalnego oraz w Semantycznym Webie.

ONTOLOGIE – SEMANTYCZNY WEB – ZASTOSOWANIA ONTOLOGII

Abstract: In the paper it is shortly described the meaning of *ontology* in different disciplines: philosophy, librarianship and scientific information, Artificial Intelligence, natural language processing and Semantic Web.

ONTOLOGIES – ONTOLOGIES APPLICATIONS – SEMANTIC WEB

*

* * *

Termin *ontologia* pojawił się w literaturze filozoficznej w XVII w. W filozofii ontologia (inaczej metafizyka) jest obok epistemologii podstawowym działem tej nauki.

W wielu dyscyplinach nauki poza filozofią, a także w działaniach praktycznych, coraz częściej stosowane są ontologie (w liczbie mnogiej, w odróżnieniu od filozofii, gdzie termin *ontologia* stosuje się w liczbie pojedynczej), definiowane jako jawna specyfikacja konceptualizacji danej dziedziny [Gruber 1993, s. 199], służące formalizacji wiedzy. Konceptualizacja ta jest identyfikacją obiektów hipotetycznie istniejących w świecie oraz relacji pomiędzy nimi. Obiekty rozumiane są jako rzeczy zarówno konkretne, jak i abstrakcyjne.

Dr Marek NAHOTKO, adiunkt w Instytucie Informacji Naukowej i Bibliotekoznawstwa UJ. Adres elektr.: nahotko@inib.uj.edu.pl

Najważniejszym zadaniem tak rozumianych ontologii jest umożliwienie wzajemnej komunikacji i rozpowszechniania wiedzy. Ogólnie mówiąc, celem stosowania ontologii jest integracja modeli w różnych dziedzinach w spójną całość dla ułatwienia dystrybucji informacji pomiędzy ich użytkownikami. Zadanie to realizowane jest podczas modelowania, projektowania i programowania ontologii, a także w rozproszonej architekturze systemów wieloagentowych. Dzięki wykorzystaniu ontologii informacja związana z daną dziedziną może być opisana w sposób umożliwiający posługiwanie się nią za pomocą różnych narzędzi dla zaspokojenia potrzeb różnych użytkowników.

Filozofia

Ontologia jest teorią bytu – dziedziną filozofii, która zajmuje się naturą i organizacją rzeczywistości. Próbuje odpowiedzieć na pytania, takie jak: „co istnieje”, „czym charakteryzuje się byt”, badając strukturę rzeczywistości, problematykę związaną z pojęciami bytu, istoty, istnienia, przedmiotów, przyczynowości, czasu, przestrzeni, konieczności i możliwości.

Spśród wielu rozmaitych kategorii ontologicznych i metafizycznych najstarsze i najbardziej znane są kategorie Arystotelesa [Woźniak 2000, s. 65–66]. Definiował on ontologię jako naukę o bytach jako takich. Inaczej niż nauki szczegółowe, które badają określoną klasę bytów i tworzą ich opis, ontologia jest wiedzą, która rozważa byt jako byt wraz z tym, co przysługuje mu w sposób istotny [Stróżewski 2003, s. 24]. W takim sensie ontologia stara się odpowiadać na pytanie „co to jest byt” lub – mówiąc bardziej obrazowo – „jakie cechy są wspólne dla wszystkich bytów”. Można więc powiedzieć, że filozofia interesuje się ontologią jako określonym systemem kategorii dla pewnej wizji świata. System taki nie jest zależny od stosowanego języka.

Takie pojmowanie ontologii dziś nazywane jest „ontologią ogólną”, w odróżnieniu od różnych „ontologii specjalnych” lub „regionalnych” (np. biologicznych, społecznych). Oprócz tego wyróżnia się ontologię formalną, która jest definiowana jako dziedzina, zajmująca się określeniem warunków prawdopodobieństwa istnienia obiektu w ogólności oraz wymagań, które muszą być spełnione przez naturę każdego obiektu. Według Niccoli Guarina i Pierdaniele Giaretty, ontologia formalna może być zdefiniowana jako systematyczny, formalny, aksjomatyczny rozwój logiki wszystkich form i trybów bytów [Guarino, Giaretta 1995]. Wynika z tego, że nie skupia się ona zbyt mocno na egzystencji określonych obiektów, lecz raczej na ścisłym opisie ich form bytu, tzn. na cechach strukturalnych.

Współcześnie jednym z najbardziej znanych badaczy ontologii dla reprezentacji wiedzy jest John Sowa [Sowa 2001; 2000a; 2000b]. Jego prace zostały zainspirowane poglądami filozofów Charlesa Sandersa Pierce’a i Alfreda Northa Whiteheada, pionierów logiki symbolicznej.

Według Sowy, przedmiotem ontologii jest badanie kategorii rzeczy istniejących lub mogących istnieć w określonej domenie [Sowa 2000a]. Wynik takich badań, zwany ontologią, jest katalogiem typów rzeczy, o których uważa się, że istnieją w domenie **D**, z perspektywy osoby używającej języka **J**, w celu mówienia o **D**. Logika nieinterpretacyjna, taka jak rachunek predykatów, grafy konceptualne lub KIF (*Knowledge Interchange Format*),

jest ontologicznie neutralna. Wymaga braku ograniczeń zakresu dziedzinowego lub określenia możliwych sposobów charakteryzowania dziedziny. Logika sama z siebie nie mówi niczego o niczym, jednak połączenie logiki z ontologią tworzy język, za pomocą którego można wyrazić relacje pomiędzy jednostkami (*entity*) w danej dziedzinie.

Bibliotekarstwo i informacja naukowa

Jak pisze Christopher Welty, bibliotekarstwo od dawna dysponuje trzema rodzajami ontologii, które mają fundamentalne znaczenie dla rozwoju informacji cyfrowej i Webu. Są to:

- Ontologia katalogu kartkowego, która pozwoliła na zdefiniowanie elementów meta-danych (typu: autor, tytuł, wydawca). Ontologie tego typu rozwijane są od setek lat.
- Ontologia bibliograficzna, definiująca strukturę rekordu bibliograficznego dla wszelkiego rodzaju publikacji. Autor określa wiek takiej ontologii na sto lat.
- Ontologia przedmiotowa, która powoduje podział piśmiennictwa na dyskretne działy. Ten typ ontologii liczy sobie około 150 lat i przez cały ten czas ewoluował [Welty 1998; 2000].

Ogólnie stwierdzić można, że ontologie w bibliotekarstwie służą głównie organizacji informacji bibliograficznej (tradycyjne opracowanie formalne zbiorów) oraz modelowaniu wiedzy uniwersalnej i/lub dziedzin szczegółowych (opracowanie rzeczowe zbiorów).

Zastosowanie ontologii w bibliotece cyfrowej opisuje Nenad Stojanovic [Stojanovic 2003]. Przedstawił on system stworzony na wzór funkcjonowania biblioteki: użytkownicy przedstawiają zapytania, przeszukując zasoby wiedzy. Tak zwany agent biblioteczny (oprogramowanie), korzystając z całej dostępnej informacji, ontologii danej dziedziny oraz wiedzy na temat sposobów postępowania poprzednich użytkowników, służy pomocą wyszukującym. W oparciu o analizy różnego rodzaju agent – prowadząc z użytkownikiem dialog interaktywny – wspomaga go w tworzeniu bardziej efektywnych scenariuszy wyszukiwawczych. W szczególności, w oparciu o podstawowy słownik (ontologię) i zawartość zasobów informacji, agent mierzy stopień niejednoznaczności zapytania użytkownika, a w przypadku wysokiej wieloznaczności zapytania sugeruje ponowne sformułowanie zapytania.

Sztuczna inteligencja

Jak pisze Barbara Sosińska-Kalata, w informatyce badania dotyczące technologii ontologicznych kontynuują nurt badawczy w dziedzinie teorii baz danych oraz deklaratywnych modeli reprezentacji wiedzy w systemach sztucznej inteligencji (*Artificial Intelligence, AI*). Ich główną dyrektywą metodologiczną jest rozwijanie teorii przez konstruowanie abstrakcyjnego i sformalizowanego modelu świata [Sosińska-Kalata 2004]. Sztuczna inteligencja jest zarówno dziedziną badawczą, jak i wynikiem prowadzonych w jej ramach badań: jest to część informatyki, zajmująca się badaniem sztucznych systemów realizujących działania, które zwykle uważa się za wymagające inteligencji, takie jak wnioskowanie. Istnieją dwa przeciwstawne paradygmaty sztucznej inteligencji, obejmujące z jednej strony podejście

kolektywne (*connectionism*), oparte na symulacji elementarnych aktorów* naturalnych i rozwijające powstawanie kolektywnych zachowań inteligentnych we współdziałaniu, oraz podejście symboliczne (*symbolicism*), oparte na jawnym, formalnym rozwoju systemów symbolicznych i związanych interpretacji w celu reprezentacji wiedzy formalnej i symulacji sensownych operacji na niej.

W informatyce ontologie zajmują się raczej znaczeniem niż istnieniem, którym zainteresowana jest filozofia. Jak pisał Guarino, w AI ontologia dotyczy inżynierskiego artefaktu, składającego się ze szczegółowego słownika, służącego opisowi określonej rzeczywistości, oraz zestawu jawnych założeń, dotyczących intencyjnego znaczenia terminów słownika [Guarino 1998].

Ontologie są nowym elementem AI, który niedawno wykazał swoją przydatność jako narzędzie konceptualne, służące do modelowania wiedzy. Stwarzają one koherentną podstawę w formie uzgodnionego słownika pojęciowego do budowy opisów i aktów komunikacyjnych. Ontologia jest narzędziem unifikującym różne punkty widzenia i służy jako podstawa, umożliwiająca komunikację pomiędzy ludźmi, pomiędzy ludźmi i systemami oraz pomiędzy systemami. Ma więc pełnić rolę wspólnego, powszechnie rozumianego języka. Stąd w kontekście AI można opisać ontologię programu komputerowego przez zdefiniowanie zestawu terminów reprezentacji. W takiej ontologii definicje łączą nazwy jednostek w uniwersum dyskursu (np. klasy, relacje, funkcje lub inne obiekty) z tekstem czytelnym dla człowieka, opisującym znaczenie nazw i zawierającym formalne aksjomaty, które uściślają interpretację i wykorzystanie tych terminów. Formalnie można stwierdzić, że ontologia jest wyrażeniem teorii logicznej [Gruber 1994].

W świecie fizycznym sytuacyjność, lokalność i rozproszenie są nieuniknioną częścią rzeczywistości. Rozproszona sztuczna inteligencja stara się „zrozumieć” zasady, będące podstawą zachowania jednostek sztucznej inteligencji (agentów) oraz ich oddziaływań w celu stworzenia sensownych zachowań systemu wieloagentowego.

Ontologie są uważane za sposób na usuwanie wieloznaczności. Mogą reprezentować znaczenie różnych treści, wymienianych w systemach informacyjnych. Im bardziej rozbudowywane są inteligentne systemy informacyjne, tym większą rolę we wnioskowaniu odgrywa wiedza ogólna o rzeczach i ich kategoriach. Jeżeli chce się, aby maszyny zachowywały się inteligentnie i zrozumiale, wiedza ta musi być im przekazana.

Przetwarzanie języka naturalnego

Jedną z dziedzin, które uzależnione są od bogatej reprezentacji wiedzy, jest przetwarzanie języka naturalnego (*natural language processing – NLP*). Zastosowania NLP pozwalają ludziom na komunikowanie się z systemami komputerowymi za pomocą języka naturalnego (zwykle ograniczanego) – zarówno za pomocą tekstu, jak i głosu. Jednym z głównych

* Aktor – agent, który może działać autonomicznie, realizując zamierzone cele. Agent natomiast, to narzędzie realizujące usługę na rzecz jego właściciela [Wiederhold 1996].

problemów NLP są trudności z kontrolowaniem terminologii i pojęć, stosowanych przez niewykształconego użytkownika końcowego.

Ontologie przydatne są w NLP w dwojaki sposób. Przede wszystkim wiedza o domenie pełni często zasadniczą rolę w zapewnieniu jednoznaczności. Dobrze zaprojektowana ontologia domeny stanowi podstawę dla reprezentacji wiedzy domeny. Dodatkowo ontologia domeny wspomaga identyfikację kategorii semantycznych używanych do zrozumienia dyskursywności danej dziedziny. W tym przypadku ontologie odgrywają rolę słownika pojęć. Ogólnie mówiąc, dla NLP niezbędne jest posiadanie zarówno ogólnej ontologii wysokiego poziomu, jak i ontologii specyficznej dla danej dziedziny, służącej opisowi tej dziedziny. Przykładami takich ontologii, służących zrozumieniu języka, są CYC, WordNet i Sensus. NLP jest jedną z tych dziedzin, w których następuje bardzo szybki rozwój zastosowań ontologii [Kiryakov, Simov 2000]. Nawet CYC, który początkowo tworzony był z myślą o zastosowaniu w systemach wiedzy, dla zapewnienia im wspólnej wiedzy o świecie, okazał się bardziej użyteczny w obszarze badań nad językiem naturalnym.

Ontologie mogą także być wykorzystywane w komunikacji między maszyną a człowiekiem, tzn. dla celów generacji wyrażen języka naturalnego. W tym przypadku problemem z punktu widzenia maszyny nie jest unikanie wieloznaczności, gdyż należy spodziewać się, że komputer posiada kontrolę nad informacją, którą chce przedstawić użytkownikowi. Ontologie mogą jednak przydać się do tworzenia bardziej „eleganckich” wyrażen.

Pomiędzy NLP a technologiami związanymi z ontologiami istnieje relacja zwrotna w tym sensie, że nie tylko ontologie są przydatne w NLP, ale techniki NLP mogą być stosowane do odszukiwania pojęć ontologicznych w dokumentach tekstowych, tzn. NLP służy automatyzacji budowy ontologii, która w końcowym zastosowaniu może nie mieć nic wspólnego z NLP.

Semantyczny Web

Idea Semantycznego Webu przedstawiona została przez Tima Bernersa-Lee. Według niego, obecnie to wciąż ludzie, a nie maszyny, w pełni korzystają z zasobów Webu. Przyczyną tego jest fakt, iż większość informacji stworzono z myślą o jej wykorzystaniu i interpretacji przez ludzi. Struktura informacji jest więc nieodpowiednia do przetwarzania maszynowego. W dzisiejszym Webie maszyny mogą analizować informację w celu uzyskania danych o jej strukturze i rutynowego przetwarzania, takiego jak odszukanie linków lub słów kluczowych itp., ale nie mogą rozumieć semantyki przetwarzanych treści. Semantyczny Web ma rozwiązać ten problem [Berners-Lee, Hendler, Lassila 2001].

Celem Semantycznego Webu jest rozdzielenie danych od układu tradycyjnych stron Webu i uzyskanie dwóch różnych źródeł, dzięki czemu użytkownicy mający dostęp do strony Webu będą widzieć źródło danych w odpowiednim dla siebie układzie, podczas gdy maszyna, uzyskując dostęp do strony, będzie otrzymywać tylko źródło danych. Same dane muszą pozwolić maszynie zrozumieć znaczenie danych i przeprowadzić wnioskowanie na ich temat.

Komputery będą zdolne do zrozumienia znaczenia terminów, opisujących dane (metadanych) dzięki wykorzystaniu ontologii. Te zaś opisują znaczenie metadanych, dzięki czemu możliwe będzie użycie zasad wnioskowania dla uzyskania logicznych wyrażeń na ich temat. Problem stanowią relacje synonimii, homonimii i polisemii, niesprawiające na ogół ludziom większych kłopotów, a zupełnie nieczytelne dla maszyn. Wykorzystanie ontologii do opisu terminów może pomóc rozwiązać ten problem. Maszyny będą wykorzystywały ontologie jako narzędzie do zrozumienia znaczenia terminów, a moduł wnioskowania wydedukuje odpowiednią informację z danych zawartych w ontologii.

Wykorzystane źródła i opracowania

- Berners-Lee, T.; J. Hendler; O. Lassila, dok. elektr. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*. http://sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 [odczyt: 10.01.2006].
- Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition* Vol. 5 (2), p. 199–220.
- Gruber, T.R., dok. elektr. (1994). What is an Ontology? <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> [odczyt: 10.01.2006].
- Guarino, N.; P. Giaretta (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. [In:] *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. Ed. N.J.I. Mars. Amsterdam: IOS Press, p. 25–32.
- Guarino, N. (1998). Formal Ontologies and Information Systems. [In:] *Proceedings of FOIS, Trento, Italy*. Ed. N. Guarino. Amsterdam: IOS Press, p. 3–15.
- Kiryakov, A.; K. Simov (2000). A Comparison Between Upper Cyc Ontology and EuroWordnet Top Ontology. [In:] *Ontologies and Lexical Knowledge Bases OntoLex 2000*. Eds. K. Simov, A. Kiryakov. Sofia: OntoText Lab., p. 190–199.
- Sosińska-Kalata, B. (2004). Ontologia i wyszukiwanie informacji. [W:] *Przestrzeń informacji i komunikacji społecznej*. Red. M. Kocójowa. Kraków: Wydaw. UJ, s. 84–89.
- Sowa, J. (2000a). Knowledge representation: logical, philosophical and computational foundations. Pacific Grove: Brooks/Cole.
- Sowa, J. (2000b). Ontology, metadata and semiotics. [In:] *Conceptual structures: Logical, linguistic and computational issues. Lecture notes in AI* No. 1867. Berlin: Springer Verl., p. 55–81.
- Sowa, J., dok. elektr. (2001). Ontology. <http://www.jfsowa.com/ontology> [odczyt: 01.09.2006].
- Stojanovic, N., dok. elektr. (2003). On the role of a Librarian Agent in Ontology-based Knowledge Management Systems. [In:] *Proceedings of the Workshop Ontologie-basiertes Wissensmanagement*. Luzern, 2–4 April. Eds. Y. Sure, H.-P. Schnurr. <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-68/> [odczyt: 10.01.2006].
- Stróżewski, W. (2003). Ontologia. Kraków: „Aureus”; „Znak”.
- Welty, C., dok. elektr. (1998). Towards an Ontology for Library Modalities. <http://www.cs.vassar.edu/faculty/welty/papers/multi-modal98.pdf> [odczyt: 01.09.2006].
- Welty, C., dok. elektr. (2000). Towards a Semantics for the Web. Dagstuhl Symposium on Semantics for the Web. Dagstuhl, Germany. <http://www.cs.vassar.edu/faculty/welty/papers/dagstuhl-2000.pdf> [odczyt: 01.09.2006].
- Wiederhold, G. (1996). Glossary: Intelligent Integration of Information. *Journal of Intelligent Information Systems* Vol. 5, p. 101–112.
- Woźniak, J. (2000). Kategoryzacja: studium z teorii języków informacyjno-wyszukiwawczych. Warszawa: Wydaw. SBP.