

Ilona Pawełoszek-Korek

Wybrane problemy wykorzystania ontologii w systemach informacyjnych organizacji rozproszonych geograficznie

Ekonomiczne Problemy Usług nr 67, 110-117

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

ILONA PAWEŁOSZEK-KOREK

Politechnika Częstochowska

WYBRANE PROBLEMY WYKORZYSTANIA ONTOLOGII W SYSTEMACH INFORMACYJNYCH ORGANIZACJI ROZPROSZONYCH GEOGRAFICZNIE

Wprowadzenie

Zjawisko globalizacji we współczesnym świecie powoduje, że przedsiębiorstwa coraz częściej dywersyfikują swoje działania, umiejscawiając je w różnych regionach geograficznych. Problem rozproszenia geograficznego i wynikające z niego negatywne skutki przejawiające się w trudnościach komunikacyjnych to bolączka wielu nowoczesnych organizacji.

W aspekcie poprawy skuteczności komunikacji bardzo obiecujące zdają się technologie semantyczne oparte na ontologiach. Obecnie wiele badań i publikacji koncentruje się na problemie wykorzystania ontologii w celach biznesowych, podkreślając ich zalety i możliwości. Jednakże mimo wielkiego potencjału semantycznej sieci Web i upatrywanych korzyści dla biznesu jej technologie nie są jak dotąd powszechnie stosowane.

Celem niniejszego artykułu jest prezentacja problemów związanych z wykorzystaniem technologii semantycznych opartych na ontologiach.

1. Specyfika komunikacji w organizacjach rozproszonych

Od lat 70. XX wieku, kiedy zaczęło się rozpowszechniać stosowanie sieci Internet w celach komunikacji biznesowej, wiele miejsca w literaturze poświęca się zdalnej współpracy. Dla ludzi głównym problemem zdalnej komunikacji jest to, iż staje się ona trudniejsza i uboższa w środki wyrazu. Dystans może mieć negatywny

wpływ na częstotliwość, terminowość i zawartość treści komunikatów¹, co jest związane z faktem, że media, przez które przesyła się dane, nie mają możliwości przekazywania kontekstu. Negatywny wpływ odległości na procesy komunikacji widoczny jest także w przypadku przedsiębiorstw międzynarodowych, w których problemem staje się bariera językowa i kulturowa.

Ponieważ dystans pomiędzy partnerami biznesowymi wpływa w znaczący sposób na komunikację między nimi oraz jakość współpracy, niezbędne są technologie, które pomagają eliminować konflikty wynikające z błędnego zrozumienia przekazu. W obecnych czasach zwanych „erą informacji” coraz więcej zadań komunikacyjnych powierza się komputerom. Systemy informatyczne oparte na architekturze zorientowanej na usługi oraz technologiach inteligentnych agentów programowych w imieniu ludzi komunikują się i dokonują transakcji. Istnienie wielu standardów i technologii, a także pewne uwarunkowania lokalne wynikające ze specyfiki procesów biznesowych powodują, że komunikacja między maszynami może być utrudniona lub niemożliwa z powodu pojawiających się problemów natury semantycznej.

2. Technologie semantyczne a interoperacyjność systemów

Współczesne systemy informatyczne są bardzo złożonymi rozwiązaniami, charakteryzującymi się dużą zmiennością i różnorodnością danych oraz dynamiką procesów. Heterogeniczność danych i procesów jest wynikiem skomplikowanych przepływów pracy specyficznych dla każdej jednostki organizacyjnej. Aby organizacje rozproszone mogły prowadzić niezakłóconą współpracę i wspólnie realizować procesy biznesowe, należy zastosować odpowiednie standardy komunikacji. Istnieje wiele popularnych i użytecznych standardów wymiany danych, jak EDIFACT czy HL7 (stosowany w służbie zdrowia)². Standardy te zapewniają interoperacyjność komunikujących się systemów, w tym sensie, że odbiorca i nadawca rozumieją treści przesyłanych komunikatów oraz ich cel. Standardy funkcjonują w kontrolowanych środowiskach, które zostały zaprojektowane, aby ze sobą współpracować (np. moduły systemu informatycznego) lub wymieniać pewną, ograniczoną liczbę komunikatów o z góry ustalonej strukturze i celu. Jednakże w rozwiązaniach tych brakuje interoperacyjności na poziomie procesów biznesowych. Interoperacyjność procesów biznesowych (BPI) jest to cecha pozwalająca im współpracować w sposób zautomatyzowany niezależnie od tego, kto jest właścicielem danego procesu,

¹ J.A. Espinosa, F. Armour: *Geographically Distributed Enterprise Architecting: Towards a Theoretical Framework*, w: Proceedings of the 41st Hawaii International Conference on System Sciences – 2008, www.computer.org/comp/proceedings/hicss/2008/3075/00/30750400.pdf (02.2011).

² Health Level Seven International, www.hl7.org

gdzie jest on zlokalizowany i z jakich rozwiązań technologicznych korzysta. Interoperacyjność daje możliwość automatycznego wyszukiwania procesów, łączenia, uruchamiania i synchronizowania ich działania, co pozwala na realizację określonego celu biznesowego.

Tradycyjnie komunikacja między niekompatybilnymi systemami odbywa się za pomocą oprogramowania pośredniczącego (*middleware*). Tego typu rozwiązania wymagają jednak stosowania wielu procedur tłumaczenia formatów danych, a ich skomplikowanie wzrasta wraz z liczbą i heterogenicznością komunikujących się stron.

Odpowiedzią na niedostatki systemów EDI i oprogramowania *middleware* są technologie semantyczne. Koncepcja ich działania zakłada, że pewien obszar rzeczywistości lub dziedzina wiedzy są w możliwie kompleksowy sposób opisane w postaci zrozumiałej dla komputerów. Mechanizmy wnioskuje potrafią porównywać otrzymane dane z modelem wiedzy, który posiadają, i odpowiednio je interpretować. Zastosowanie takiego rozwiązania w znacznym stopniu może rozszerzyć funkcjonalność systemów elektronicznej wymiany danych, sprawić, że przesyłane informacje mogą być sformułowane w sposób elastyczniejszy, bogatszy w środki wyrazu, a zarazem bardziej zbliżony do sposobu komunikacji między ludźmi.

3. Ontologia jako podstawa wzajemnego zrozumienia

W odróżnieniu od komunikacji między ludźmi w systemach informatycznych zdanie się na intuicyjne rozumienie komunikatów nie jest możliwe. Wiedza na temat znaczenia poszczególnych terminów, pojęć i relacji musi być ściśle zdefiniowana, aby uniknąć nieporozumień³. Dlatego stosuje się różne metody zapisu wiedzy w określonych, mniej lub bardziej złożonych i ekspresyjnych, strukturach, takich jak słowniki, klasyfikacje czy ontologie. Spośród wymienionych struktur ontologie są najbardziej złożoną formą reprezentacji wiedzy, a za ich pomocą można wyrazić zarówno klasyfikacje, jak i słowniki⁴.

Słowo „ontologia” pierwotnie było używane na określenie działu filozofii zajmującego się opisywaniem natury bytu. Pierwszą definicję ontologii w ujęciu informatycznym zaproponował T. Gruber, określając ją jako „specyfikację konceptualiza-

³ I. Pawełoszek-Korek: *Rola semantyki w nowoczesnych systemach informacyjnych e-biznesu*, Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, t. 21, Bydgoszcz 2009, s. 91.

⁴ W. Abramowicz: *Filtrowanie informacji*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2008, s. 132.

cji”⁵. Inna definicja mówi, że ontologia wskazuje na podstawowe pojęcia i relacje, obejmując słownictwo danej dziedziny oraz reguły łączenia pojęć i relacji⁶.

Ontologie mogą być przydatne tam, gdzie zachodzi potrzeba komunikacji pomiędzy ludźmi lub agentami programowymi, a informacja powinna być sformalizowana i jednoznacznie zrozumiana przez komunikujące się strony. Innym ważnym powodem budowania i używania ontologii może być konieczność wielokrotnego korzystania z wiedzy dziedzinowej (np. w systemach zarządzania wiedzą).

Przytoczone definicje ontologii nie mówią na temat sposobu kodowania wiedzy. W dobie gospodarki elektronicznej oczywiste jest zastosowanie języków strukturalnych opartych na logice i sformalizowanych na tyle, aby mogły być rozumiane i przetwarzane przez komputery. Podstawą standardów zapisu wiedzy w postaci ontologii są języki RDF i OWL. RDF jest językiem służącym do definiowania metadanych dla różnych zasobów sieciowych, używany wspólnie z językiem RDFS (RDF Schema) umożliwia tworzenie prostych taksonomii. Do reprezentacji bardziej skomplikowanych relacji został opracowany język OWL i jego trzy warianty: OWL Lite, OWL DL i OWL Full. Pierwszy z nich jest najuboższy pod względem możliwości prezentacji wiedzy, ale łatwy do zastosowań w aplikacjach. Dwa kolejne są bardziej rozbudowane. OWL Full ma maksymalne możliwości prezentacji wiedzy w postaci strukturalnej, ale z powodu skomplikowania nie daje gwarancji prawidłowej przetwarzalności przez komputery.

4. Praktyczne problemy zastosowania ontologii

Zagadnienia techniczne

Technologie semantyczne oparte na wykorzystaniu ontologii zdają się bardzo obiecujące. Za ich pomocą możliwe jest stworzenie inteligentnej semantycznej sieci WWW, w której w imieniu użytkownika działają agenty programowe, wyszukują informacje, załatwiają wiele spraw, dokonują transakcji, a to wszystko z uwzględnieniem indywidualnych cech i potrzeb zleceniodawcy. Teoretycznie wszystkie zasoby w sieci Internet bądź intranetach czy ekstranetach przedsiębiorstw mogłyby zostać opisane odpowiednimi znacznikami RDF, według wybranej ontologii, i tym samym mogłyby być zdadne do przetwarzania przez maszyny. Jednakże obecnie technologie semantyczne stosuje się z powodzeniem jedynie w zamkniętych systemach informacyjnych o stosunkowo niewielkim zakresie dziedzinowym (w porównaniu z globalną siecią Internet).

⁵ T. Gruber: *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, „International Journal Human-Computer Studies” 1995, vol. 43, iss. 5–6, s. 907–928.

⁶ A. Gomez-Perez, O. Corcho, M. Fernandez-Lopez: *Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer-Verlag, Londyn 2004.

Jednym z podstawowych problemów, które wpływają na ograniczenie zastosowania technologii semantycznych opartych na ontologiach, jest złożoność tych technologii. Języki służące do definiowania ontologii oraz zapisu wiedzy – różne odmiany RDF i OWL – są dość skomplikowane, opanowanie ich przez człowieka zajmuje wiele czasu i wymaga odpowiedniego przygotowania merytorycznego, jak chociażby znajomości zasad logiki, wiedzy na temat opisywanej dziedziny, precyzji w wyrażaniu myśli. Zarówno reprezentacja graficzna (grafy), jak i zapis w postaci znaczników są trudne do opanowania przez użytkowników.

Im więcej faktów jest zakodowanych w danej ontologii, tym jest ona teoretycznie bardziej przydatna. W praktyce jednak rozbudowane dokumenty OWL są trudne do interpretacji zarówno przez człowieka, jak i maszyny. Języki OWL i RDF są zaprojektowane tak, aby ludzie mogli je zrozumieć, jednakże czytanie dokumentów zapisanych w ten sposób jest mozolne. Należy także podkreślić, że zadania wymagające dużego nakładu pracy, a jednocześnie specjalistycznej wiedzy są najdroższe w realizacji.

Popularyzacja standardów sieci semantycznej wymaga zastosowania języków znacznikowych na szeroką skalę. Im więcej dokumentów będzie opisanych znacznikami, tym większa ilość danych będzie użyteczna i potencjalnie łatwiej dostępna w celu automatycznego wnioskowania. Opisywanie dokumentów znacznikami jednak wciąż najskuteczniej jest wykonywane przez ludzi. Istnieją wprawdzie narzędzia ułatwiające proces anotacji semantycznej dokumentów HTML (np. RDFAuthor⁷), lecz nie są w stanie go wyręczyć w tym uciążliwym zadaniu. Inną, jak dotąd nierozwiązaną do końca kwestią, pozostaje dodawanie anotacji do dokumentów bardziej skomplikowanych niż HTML, np. zapisanych w formatach edytorów tekstów, grafiki i multimediów czy schematów technicznych.

Istnieją próby stworzenia algorytmu, który z tekstu w języku naturalnym generuje dokumenty RDF, niestety programy tego typu są dalekie od doskonałości, większość z nich przetwarza jedynie język angielski. Jeśli stworzono by na tyle inteligentny algorytm, aby umiał on zrozumieć tekst pisany w języku naturalnym i tworzyć do niego semantyczne anotacje z trafnością porównywalną z pracą człowieka, to istnienie tego algorytmu stawiałoby pod znakiem zapytania potrzebę dokonywania takich anotacji. Skoro istniałyby programy, które potrafią interpretować tekst w języku naturalnym, który nie posiada anotacji, to ich tworzenie stałoby się zbędne⁸.

Problem standaryzacji

Ustanowienie języków RDF i OWL standardami w budowie aplikacji semantycznych tylko w niewielkiej części rozwiązuje problem standaryzacji. Za pomocą tych języków można bowiem stworzyć wiele ontologii i baz wiedzy opisujących

⁷ www.aksw.org/Projects/RDFauthor (02.2011).

⁸ T. Gruber: *Toward Principles...*, *op. cit.*

jakąś dziedzinę życia, ale w różnym kontekście, przy odmiennych założeniach i konwencjach zapisu.

Istnieje potrzeba zastosowania rozwiązania w postaci translatora, który będzie mapował koncepcje stosowane w różnych ontologiach i oceniał ich równoznaczność. Na przykład zawartość znacznika RDF oznaczającego numer edycji książki może mieć różną postać: edycja druga, edycja II, 2 edycja itd. Tak niewielkie różnice nie byłyby z pewnością nawet zauważone przez człowieka, jednakże stanowią problem dla maszyn.

Jak widać, posiadanie dokumentów z semantycznymi anotacjami nie zawsze prowadzi do jednoznacznej ich interpretacji. Zatem skutecznym rozwiązaniem wydaje się narzucenie z góry pewnych standardów opisu oraz skonstruowanie interfejsów użytkownika zgodnych z tym standardem, które uniemożliwiłyby wprowadzenie danych w sposób błędny.

Ponieważ komunikacja między partnerami biznesowymi w różnych organizacjach zachodzi na wielu płaszczyznach i może dotyczyć bardzo szerokiego spektrum tematów, z tego punktu widzenia słuszne wydaje się opracowanie jak największej liczby standardów. Tutaj jednak problemem staje się fakt, który już w 1933 roku zauważył znany matematyk K. Godel⁹: Żaden logiczny system nie może być jednocześnie zarówno spójny, jak i kompletny. Duża liczba standardów stwarza niebezpieczeństwo, że wymagania w nich zawarte mogą być sprzeczne, na przykład opis danego obiektu w jednym kontekście może być prawdziwy, a w innym ten sam zapis może być błędny.

Problemy reprezentacji wiedzy i interpretacji

Dziedzina reprezentacji wiedzy skupia się na zastosowaniu logiki i ontologii do konstruowania komputerowo przetwarzalnych modeli pojęć z wybranej dziedziny. J.F. Sowa wyróżnia trzy klasy problemów w dziedzinie reprezentacji wiedzy¹⁰:

Logika. Różne implementacje są oparte na odmiennych rodzajach logiki i nie zawsze dają się między sobą przetłumaczyć w prosty sposób – niektóre informacje mogą być utracone lub zmodyfikowane.

Ontologia. Różne systemy mogą używać różnych nazw dla tych samych encji lub tych samych nazw dla różnych obiektów. Czasami zadanie udowodnienia, że są one tożsame, może być trudne lub niemożliwe.

Przetwarzanie. Nawet kiedy nazwy i definicje są identyczne, pewne szczególne implementacyjne mogą spowodować, że ta sama wiedza zachowuje się różnie w różnych systemach. W niektórych implementacjach kolejność wprowadzania reguł i danych może mieć wpływ na efekt wnioskowania i rezultat przetwarzania.

⁹ www.en.wikipedia.org/wiki/Kurt_G%C3%B6del (02.2011).

¹⁰ J.F. Sowa: *Building, Sharing, and Merging Ontologies*, www.jfsowa.com/ontology/ontoshar.htm#s5 (02.2011)

Czasami prosty proces wnioskowania, który działa bez problemu w jednym systemie, w innym zawiesi się w niekończącej się pętli.

Chociaż te trzy aspekty reprezentacji wiedzy stwarzają różne problemy, są one współzależne. To, jakie słowa są używane na określenie obiektów świata rzeczywistego często jest uwarunkowane czynnikami historycznymi i kulturowymi. Zdarza się, iż w różnych językach stosuje się odmienną hierarchię pojęć opisujących te same obiekty. Aby skutecznie wyeliminować konflikty semantyczne, należałoby stworzyć mechanizmy radzenia sobie z niepewnością przez maszyny. W pewnym stopniu problem ten można rozwiązać, wykorzystując teorię zbiorów rozmytych w celu uzupełnienia modeli wiedzy zapisanych w formacie RDF o dane rozmyte. Jednakże to powoduje uzyskanie dokumentów o jeszcze bardziej skomplikowanej strukturze oraz konieczność konstruowania mechanizmów wnioskujących opartych na nowych regułach.

Podsumowanie

Wizja semantycznej sieci, w której funkcjonują inteligentne programy realizujące w imieniu użytkowników zadania komunikacji biznesowej, jest bardzo interesująca i stanowi szerokie pole do przedsięwzięć badawczych i rozważań teoretycznych. W ostatnich dziesięcioleciach coraz częściej realizowane są próby stosowania rozwiązań opartych na prostych formach reprezentacji wiedzy do wspomagania działania różnych aplikacji (wyszukiwarek, portali, platform *e-commerce* 'owych). Dla rozpowszechnienia tych technologii podstawowe znaczenie ma pokonanie trudności natury technicznej, standaryzacyjnej oraz biznesowej. Ich rozwojowi sprzyja stale zwiększające się grono naukowców, programistów oraz wizjonerów zainteresowanych semantyczną siecią Web.

Przy opracowywaniu i wdrażaniu nowoczesnych technologii bardzo istotne jest poradie ze strony biznesu, zwłaszcza dużych firm zainteresowanych poprawą swojego funkcjonowania. Należy zauważyć, że problematyka semantyki oraz ontologii nie jest łatwa i przyjazna, a efekty jej zastosowania nie zawsze są oczywiste i wymierne, dlatego trudno jest ją skutecznie promować w środowiskach biznesowych.

Literatura

1. Abramowicz W.: *Filtrowanie informacji*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2008, s. 132.
2. Espinosa J.A., Armour F.: *Geographically Distributed Enterprise Architecting: Towards a Theoretical Framework*, w: Proceedings of the 41st Hawaii International

- Conference on System Sciences – 2008, źródło internetowe: 02.2011, www.computer.org/comp/proceedings/hicss/2008/3075/00/30750400.pdf
3. Gomez-Perez A., Corcho O., Fernandez-Lopez M.: *Ontological Engineering with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer-Verlag, Londyn 2004.
 4. Gruber T.: *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, „International Journal Human-Computer Studies” 1995, vol. 43, iss. 5–6.
 5. Pawełoszek-Korek I.: *Rola semantyki w nowoczesnych systemach informacyjnych e-biznesu*, Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, t. 21, Bydgoszcz 2009.
 6. Sowa J.F.: *Building, Sharing, and Merging Ontologies*, źródło internetowe: 02-2011 <http://www.jfsowa.com/ontology/ontoshar.htm#s5>
 7. www.aksw.org/Projects/RDFauthor 02-2011
 8. www.en.wikipedia.org/wiki/Kurt_G%C3%B6del (02.2011).
 9. www.hl7.org Health Level Seven International (02.2011).

SELECTED PROBLEMS OF EXPLOITING ONTOLOGIES IN INFORMATION SYSTEMS OF GEOGRAPHICALLY DISPERSED ORGANIZATIONS

Summary

Globalization often drives enterprises to run their businesses in geographically dispersed environments. The problem of geographic dispersion often causes negative side effects on communication processes. Semantic technology based on ontologies is considered to be the very promising challenge in the area of improving communication. The aim of this paper is to present selected problems occurring while using ontologies, that impede proliferation of semantic technologies among business organizations.

Translated by Ilona Pawełoszek-Korek