

ILONA PAWEŁOSZEK-KOREK

Politechnika Częstochowska

STANDARDY SIECI SEMANTYCZNEJ W ZARZĄDZANIU WIEDZĄ ORGANIZACJI

Wprowadzenie

Współczesna rzeczywistość charakteryzuje się nadmiarem informacji, dostępnej w różnych postaciach (elektronicznej, papierowej, przekazów medialnych). Informacje pochodzą z wielu heterogenicznych źródeł, mają zatem zróżnicowaną jakość i przydatność w procesach decyzyjnych. Informacja, która jest użyteczna w procesie podejmowania decyzji staje się wiedzą – najbardziej poszukiwanym zasobem współczesnej cywilizacji.

Wyszukiwanie informacji potrzebnych do wykonywania zadań złożonych jest czasochłonne i wprowadza pracowników we frustrację, szczególnie w przypadku braku dostępu do dobrze zorganizowanej i łatwej w obsłudze infrastruktury umożliwiającej korzystanie z wiedzy¹.

Zadaniem nowoczesnych systemów wyszukiwawczych jest inteligentne wspomaganie użytkownika polegające na dostarczaniu informacji we właściwym kontekście i formie. Temu zapotrzebowaniu wychodzi na przeciw koncepcja opracowana przez Tima Bernersa-Lee określana mianem sieci semantycznej².

¹ W.M. Gruszewski, I.K. Hejduk, *Zarządzanie wiedzą w przedsiębiorstwach*, Warszawa 2004, s. 42.

² T. Berners-Lee, J. Hendler, O. Lassila, *The semantic web*, „Scientific American” 2001, vol. 279(5).

Celem niniejszego artykułu jest analiza możliwości zastosowania sieci semantycznej w zarządzaniu wiedzą organizacji oraz prezentacja wybranych standardów umożliwiających praktyczną realizację tego zadania.

1. Specyfika zarządzania wiedzą we współczesnych organizacjach

Osiągnięcia technologii informacyjnej w postaci rozbudowanych systemów komputerowych działających w sieciach kablowych i bezprzewodowych przestają być nowością, a stają się nieodzownym standardem dla wielu sprawnie działających przedsiębiorstw³. Zatem jedynie posiadana wiedza i innowacyjne jej wykorzystanie mogą zapewnić przewagę konkurencyjną przedsiębiorstwa.

Zarządzanie wiedzą odnosi się do procesów gromadzenia, tworzenia, odkrywania i dostarczania wiedzy oraz technologii informacyjnej wykorzystywanej w tych procesach i jest niezbędnym czynnikiem stwarzającym szansę przetrwania organizacji w konkurencyjnym środowisku globalnej gospodarki rynkowej.

Współcześnie ważną kwestią zarządzania wiedzą jest dobór odpowiednich rozwiązań technologii informacyjnej, dostosowanych do potrzeb organizacji, sposobu jej działania, stylu zarządzania i kultury. Systemy zarządzania wiedzą, funkcjonujące od lat w przedsiębiorstwach i instytucjach, bazujące na relacyjnych bazach danych i archiwach tekstowych dokumentów, mają wiele niedociągnięć, do których zaliczyć można między innymi:

- Wyszukiwanie informacji za pomocą słów kluczowych często daje nierелеwantne wyniki zawierające terminy o innym znaczeniu niż to, którego oczekuje użytkownik. Nie ma możliwości wyszukiwania kontekstowego.
- Brakuje możliwości automatycznego agregowania informacji z różnych źródeł i wnioskowania na podstawie analizy tekstów. Użytkownik musi sam wyekstrahować pożądaną wiedzę, przeglądając wyniki wyszukiwania i tworząc zestawienia.
- Aktualizacja informacji słabo ustrukturyzowanych (głównie w postaci tekstu) jest trudnym i czasochłonnym zadaniem, zwłaszcza w przypadku dużych zbiorów.

³ I. Pawełoszek-Korek, *Zarządzanie wiedzą organizacji w świetle trendu Web 2.0*, w: *Współczesne problemy zarządzania przedsiębiorstwami w gospodarce rynkowej*, Bielsko-Biała 2008, s. 255.

- Wiele informacji nie ma postaci tekstowej, lecz multimedialną (sekwencje wideo, dźwięki, obrazy), tradycyjne systemy wyszukiwawcze nie potrafią jednak rozpoznać znaczenia informacji w postaci nietekstowej.

Ważną rolę w efektywnym wykorzystaniu zarówno wiedzy jawnej (skodyfikowanej), jak i ukrytej (niesformalizowanej) odgrywają intranety. W przypadku wiedzy jawnej, technologia ta dostarcza wszechobecnego interfejsu, za pomocą którego wiedza zostaje zorganizowana po relatywnie niskim koszcie, przy użyciu otwartych standardów. Przeniesienie informacji z postaci papierowej do intranetu przynosi również korzyści, między innymi w wymiarze szybkiej aktualizacji, a zatem dokładności. Problemem, który należy rozważyć, jest sposób dostarczania właściwej informacji do właściwych osób we właściwym czasie. Wiedzę jawną można postrzegać na wiele sposobów: jako informację we właściwym kontekście lub informację, która może prowadzić do efektywnego działania. W przypadku wiedzy ukrytej można korzystać z narzędzi intranetowych, w celu połączenia ludzi o podobnych zainteresowaniach lub problemach, zachęcając ich do dialogu i skorzystania z możliwości wymiany wiedzy ukrytej⁴.

W ostatnich latach w Internecie pojawił się trend określany jako Web 2.0, którego główną cechą jest wykorzystanie narzędzi informatycznych w celu budowania społeczności użytkowników sieci, współtworzących zasoby wiedzy. Istnieje wiele technologii charakterystycznych dla trendu Web 2.0, do najpopularniejszych można zaliczyć: wiki, sieci kontaktów, zakładki społecznościowe, blogi, podcasty i videocasty, które zyskują coraz większą popularność wśród użytkowników Internetu. W kręgach praktyków zarządzania wiedzą szybko dostrzeżono zalety tych narzędzi, które przy odpowiednim zastosowaniu sprzyjają nawiązywaniu kontaktów międzyludzkich i tworzeniu społeczności wymiany wiedzy. Rozwiązania oparte na narzędziach Web 2.0 bardzo odpowiadają swoim charakterem specyfice procesów pozyskiwania wiedzy, które są ze swej natury niestrukturalne i różnią się od siebie (podobnie jak sposoby myślenia i nabywania wiedzy przez poszczególne osoby). Gromadzenie i publikowanie informacji w postaci czytelnej dla człowieka przestało już być wystarczającą funkcjonalnością narzędzi Web 2.0, ich ciągły rozwój dąży w kierunku zapewnienia informacji o odpowiedniej strukturze i formacie zrozumiałym także dla maszyn, które wspomagają użytkowników w procesie pozyskiwania wiedzy.

⁴ J. Davies, D. Fensel, F. Harmelen, *Towards the Semantic Gap*, Springer 2007, s. 2.

2. Sieć semantyczna

Obecnie popularne systemy wyszukiwawcze działające w sieci Web bądź w środowiskach lokalnych przyjmują od użytkownika zapytanie w postaci zbioru słów kluczowych i w odpowiedzi zwracają listę odnośników do dokumentów, które te słowa zawierają. Zatem użytkownik dostaje dane, które następnie musi przejrzeć, przeanalizować i odpowiednio zestawić, aby uzyskać informację bądź wiedzę, którą chce zdobyć.

Uwzględniając ilość i różnorodność zasobów informacyjnych, z jakimi stykają się współcześni użytkownicy systemów informacyjnych, łatwo zauważyć, iż tradycyjne systemy wyszukiwawcze mają wiele niedostatków. W odpowiedzi na narastające potrzeby informacyjne użytkowników powstała koncepcja sieci semantycznej opracowana przez Tima Bernersa-Lee. Jest ona narzędziem, które pomaga organizować, przeszukiwać i zestawiać dane w odpowiednim kontekście, tak aby tworzyły znaczące z punktu widzenia użytkownika informacje bądź dostarczały precyzyjnej odpowiedzi na konkretne pytania.

W tradycyjnym modelu wyszukiwania za pomocą słów kluczowych ważne dokumenty mogą zostać nieodnalezione, a jednocześnie lista wyników wyszukiwania może być pełna dokumentów nierelevantnych. Zatem, aby uzyskać pożądane informacje, wyszukiwanie wymaga powtarzania z różnym zestawem słów kluczowych i odfiltrowania nieistotnych wyników. Problem wyszukania relewantnej informacji staje się jeszcze trudniejszy w przypadku informacji w postaci obrazów bądź dźwięków, których systemy wyszukiwawcze nie potrafią zinterpretować.

W wypadku systemu opartego na koncepcji sieci semantycznej istnieje z kolei możliwość udostępnienia użytkownikowi informacji wzbogaconej metadany, które mogą być interpretowane i przetwarzane przez maszyny. Zadaniem sieci semantycznej jest udzielenie bezpośredniej odpowiedzi na zapytanie użytkownika, wynikającej z analizy i interpretacji znalezionych dokumentów.

Aby działanie sieci semantycznej było możliwe, system wyszukiwawczy musi posiadać wiedzę na temat:

- profilu użytkownika opisującego jego cechy, sytuację bądź preferencje,
- treści zapytania,
- treści i struktury przeszukiwanych dokumentów.

Sieć semantyczna potrafi powiązać ze sobą wymienione informacje, tak aby udzielić odpowiedzi na zapytanie użytkownika w sposób właściwy, odpowiadający jego aktualnym potrzebom informacyjnym.

Sieć semantyczna składa się z węzłów i wątków. Węzły stanowią reprezentację obiektów czy też pojęć z pewnej dziedziny, natomiast wątki odzwierciedlają relacje pomiędzy węzłami. W przypadku sieci obrazującej wiedzę organizacji, niektóre punkty węzłowe mogą stanowić odzwierciedlenie zasobów informacyjnych (takich jak dokumenty, bazy, repozytoria, archiwa), inne zaś ich twórców. Poszczególne punkty węzłowe są powiązane relacjami, które odwzorowują wzajemne powiązania pomiędzy zasobami. Dwa węzły może łączyć więcej niż jedna relacja, należy jednak zauważyć, iż węzły mogą mieć różne znaczenia, zgodnie z modelem, na jakim się opierają.

3. Elementy sieci semantycznej

Jak dotąd nie ma jednoznacznie uznanej definicji sieci semantycznej, jednakże analizując jej budowę, pełnione funkcje i cele działania można stwierdzić, że sieć semantyczna jest nośnikiem wiedzy oraz narzędziem umożliwiającym jej pozyskiwanie.

Pojęcie „wiedza” może być różnie postrzegane i rozumiane, zależnie od dyscypliny, pod której kątem ma być zdefiniowane. Jedna z definicji odpowiadająca ujęciu informatycznemu mówi, iż wiedza jest zbiorem informacji dającym możliwość wyciągania wniosków na podstawie przesłanek⁵. W takim ujęciu, aby powstała wiedza, niezbędne są dwa czynniki: baza danych (zbiór opisów określonych faktów przy użyciu pewnego modelu danych) i baza wiedzy (zawierająca dane oraz reguły wnioskowania na podstawie danych). Koncepcja sieci semantycznej jako nośnika wiedzy doskonale odzwierciedla tę definicję.

Organizowanie sieci semantycznej odwzorowującej relacje pomiędzy zasobami informacyjnymi polega na stworzeniu powiązań pomiędzy dokumentami poprzez analizę i klasyfikację zasobów (w sposób manualny bądź zautomatyzowany). Każdy dokument powinien zostać opisany za pomocą odpowiedniego schematu metadanych mówiących na temat związku treści dokumentu z określonymi tematami lub obiektami. Sieć semantyczna składa się zatem z następujących elementów:

⁵ *Efektywne zarządzanie a sztuczna inteligencja*, red. A. Baborski, Wrocław 1994, s. 20.

- dokumentów opisujących właściwości obiektów świata rzeczywistego,
- ontologii, czyli zbioru reguł opisujących typy zasobów, jakie mogą wystąpić w sieci, i relacje, jakie mogą łączyć poszczególne obiekty,
- mechanizmów wnioskujących – używanych w celu interpretowania zależności oraz sprawdzania przynależności zasobu do określonej klasy,
- interfejsu użytkownika pozwalającego wprowadzać zapytania do sieci semantycznej i wyświetlać ich wyniki w postaci zrozumiałej dla odbiorcy.

Wymienione elementy są z reguły projektowane z wykorzystaniem pewnych standardów głównie określających sposób zapisu danych. Mają one na celu zapewnienie interoperacyjności projektowanych rozwiązań. Zastosowanie standardów jest obecnie szczególnie istotne, systemy informatyczne przedsiębiorstw projektowane są bowiem jako architektury otwarte pozwalające na wymianę danych i usług pomiędzy partnerami biznesowymi.

Interfejsy użytkownika są projektowane przy użyciu różnych narzędzi programistycznych i języków programowania w zależności od tego, w jakim środowisku ma działać aplikacja sieci semantycznej. Obecnie najczęściej wybieranym interfejsem użytkownika jest przeglądarka internetowa wyświetlająca formularze zapytań (XHTML bądź XML) i wyniki. Rozwiązanie takie pozwala zrealizować aplikację jako część portalu korporacyjnego i udostępniać przez firmowy intranet bądź sieć globalną w zależności od potrzeb i rozproszenia geograficznego użytkowników.

W kolejnym podpunkcie zostaną zaprezentowane najczęściej używane standardy wykorzystywane do budowy sieci semantycznych, tak w środowisku sieci Internet, jak i na potrzeby wewnętrznych systemów zarządzania wiedzą w przedsiębiorstwach, opartych na intranetach.

4. Wybrane standardy sieci semantycznej

Nowoczesny standard wymiany danych i projektowania aplikacji w Internecie, który ostatnio zyskał dużą popularność, to Rozszerzalny Język Znaczników – XML (*eXtensible Markup Language*). Jest to metajęzyk służący do opisywania zawartości dokumentów elektronicznych za pomocą znaczników. Jego zaletą jest zrozumiały zarówno dla maszyn, jak i ludzi sposób zapisu. Ponadto XML (w przeciwieństwie do powszechnie stosowanego do projektowania witryn WWW języka HTML) nie ma z góry ustalonego zbioru znaczników, lecz pozwala pro-

gramiście je tworzyć i nadawać im dowolne nazwy i atrybuty. Elastyczność i prostota XML sprawiła, że szybko stał się standardem, który posłużył jako postawa do stworzenia wielu innych języków i zdefiniowania reguł zapisu, jakie mają w nich obowiązywać. Języki pochodzące od XML zwane są aplikacjami, natomiast zapis reguł obowiązujących w danym języku znacznikowym zwany jest definicją typu dokumentu – DTD (*Document Type Definition*).

XML posłużył także do stworzenia standardów zapisu wiedzy w sieciach semantycznych. Podstawą ich funkcjonowania są dwa języki RDF i OWL wraz z ich odmianami tworzonymi na potrzeby określonych zastosowań.

Technologie, które stanowią podstawę języka XML, a zatem i budowy sieci semantycznych, to standard kodowania znaków Unicode oraz identyfikatory URI (*Uniform Resource Identifiers*). Unicode jest uniwersalnym standardem kodowania znaków, który pozwala, aby dokumenty XML mogły zawierać znaki narodowe. Natomiast URI są używane jako unikatowe identyfikatory pozwalające odnaleźć opisy poszczególnych zasobów w sieci semantycznej. Przykładem zasobów mogą być opisy różnych obiektów świata rzeczywistego (jak autor, tytuł, dokument, przedmiot), o których informacje mają być udostępniane użytkownikom sieci semantycznej.

RDF (*Resource Description Framework*) jest językiem ogólnego przeznaczenia do reprezentowania informacji w sieci. Jest to popularny model specyfikacji służącej do prezentowania metadanych dokumentu (znaczenia terminów i koncepcji używanych do tworzenia przypisów w zasobach Internetu) w formie, którą komputery mogą łatwo przetwarzać. Celem RDF jest dostarczenie sposobu kodowania i mechanizmu interpretacji, aby zasoby mogły być opisane czytelnie dla określonego oprogramowania⁶.

RDF jest ustalonym standardem reprezentacji wiedzy w sieci semantycznej. W połączeniu z pochodzącym od niego standardem RDFSchemata (RDFS) tworzą język reprezentacji prostych ontologii i baz wiedzy funkcjonujących w sieciach. RDFS jest używany do definiowania ontologii lub słownictwa, natomiast RDF do definiowania bazy wiedzy. Słownictwo RDFS jest definiowane przy użyciu RDF, zaletą tego podejścia jest wykorzystanie jednego języka reprezentacji zarówno do ontologii, jak i bazy wiedzy. Motywacją do stworzenia standardu RDF była w pierwszej kolejności potrzeba stworzenia opisów mówiących o zasobach w sieci, stąd nazwa standardu⁷.

⁶ B. Leuf, *The Semantic Web: Crafting Infrastructure for Agency*, West Sussex 2006, s. 44.

⁷ C.D. Walton, *Agency and the Semantic Web*, Oxford 2007, 20.

Dokumenty RDF stanowią opis zasobów dostępnych w sieci, jednakże funkcjonowanie sieci semantycznej wymaga, aby informacje były uporządkowane zgodnie z określonymi zasadami, według których maszyny mogą interpretować dane i wnioskować.

Sposobem na umożliwienie maszynom rozpoznawania znaczenia przetwarzanych informacji jest zastosowanie ontologii, które pozwalają sformalizować wiedzę w sposób zrozumiały zarówno dla człowieka, jak i maszyny. To umożliwia użytkownikowi wyszukiwanie informacji na podstawie jej znaczenia, nie zaś składni. Kluczowym problemem jest zatem zdefiniowanie standardów reprezentujących strukturę informacji, czyli ontologię.

Obecnie OWL (*Web Ontology Language*) jest standardem rekomendowanym przez W3C. OWL to rodzina języków reprezentacji wiedzy służących do modelowania ontologii dla sieci semantycznej. Historycznie język ten ewoluował ze standardu DAML+OIL, który powstał przez połączenie europejskiego standardu OIL (*Ontology Inference Layer*) z jego amerykańskim odpowiednikiem DAML (*DARPA Agent Markup Language*). OWL jest językiem oferującym szerokie możliwości modelowania ontologii w różnych odmianach: tezaurusów, taksonomii, schematów klasyfikacji. Pozwala także tworzyć specyfikacje zależności hierarchicznych pomiędzy klasami i ograniczenia, modelować atrybuty i powiązania. Te własności sprawiają, że możliwe jest wnioskowanie na podstawie informacji, które nie są jawnie zapisane w ontologii⁸.

Kolejnym nowym standardem rekomendowanym przez konsorcjum W3C od stycznia 2008 roku jest język SPARQL (W3C, 2008), który pozwala na konstruowanie zapytań kierowanych do struktur sieci semantycznej tworzonej przez dokumenty RDF. Idea tego standardu przypomina SQL znany od wielu lat jako skuteczny język zapytań stosowany w relacyjnych bazach danych. SPARQL jest ciągle w fazie rozwoju, główne wysiłki grup programistów pracujących nad nim skupiają się na zaprojektowaniu jak najbardziej wydajnych mechanizmów zapytań.

Portale działające w sieci Web lub intranetach konstruowane są głównie przy użyciu języków XHTML i XML, jednakże ten sposób zapisu, bardzo wygodny dla projektantów WWW, sprawia, że informacje nie mogą być bezpośrednio interpretowane jako zasoby sieci semantycznej. Wychodząc naprzeciw temu problemowi, opracowano kolejny standard, jakim jest format znaczników GRDDL (*Gleaning Resource Descriptions from Dialects of Languages*). Umoż-

⁸ M. Ehring, *Ontology Alignment Bridging the Semantic Gap*, Springer 2007, s. 15.

liwia on użytkownikom wyekstrahowanie dokumentu RDF z dokumentów w formacie XML i XHTML. Odbywa się to przy wykorzystaniu arkuszy stylów XSLT. GRDDL jest rekomendowany przez W3C od września 2007 roku⁹. GRDDL transformuje dane zawarte na stronach WWW do formatu, który może być wykorzystany bezpośrednio w innych aplikacjach.

W prostszych rozwiązaniach, mających na celu automatyzację pobierania niektórych danych o określonej strukturze semantycznej z witryn WWW przez aplikacje, takie jak przeglądarki, programy pocztowe, kalendarze, harmonogramy zadań, stosuje się tak zwane mikroformaty. Jest to zestandaryzowany sposób opisu elementów dokumentu XML, polegający na odpowiednim nazwaniu znaczników i dodaniu do nich specjalnej klasy lub atrybutu.

Jak pokazuje praktyka, semantyczny Internet może iść w parze z trendem Web 2.0¹⁰. Tworzenie treści przez społeczności internautów i wzbogacanie ich o relacje i opisy zasobów zrozumiałe dla maszyn może w znacznym stopniu przyczynić się do zwiększenia użyteczności informacji zawartych w portalach wiki, blogach i różnego rodzaju serwisach społecznościowych, które coraz częściej stają się platformami wymiany wiedzy.

Zakończenie

Powszechne zastosowanie standardów sieci semantycznych do budowy i porządkowania zbiorów danych pozwoliłoby zwiększyć efektywność ich wykorzystania przez wszystkie jednostki i organizacje korzystające z systemów zarządzania wiedzą tak w sferze naukowo-badawczej, administracyjnej, jak i biznesowej.

Nowoczesne podejście do zasobów informacyjnych powinno się opierać na koncepcji systemowej, wychodząc od dobrze znanej maksymy, mówiącej, iż całość to więcej niż tylko suma części. Każdy system jest bowiem celowo zorganizowanym zbiorem obiektów i relacji, jakie między nimi zachodzą¹¹.

⁹ W3C Completes Bridge Between HTML/Microformats and Semantic Web, <http://www.w3.org/2007/07/grddl-pressrelease>.

¹⁰ Przykładem może być projekt DBpedia, którego celem jest dostarczenie narzędzi do ekstrakcji ustrukturyzowanej informacji z zasobów Wikipedii. DBpedia pozwala zadawać skomplikowane zapytania do Wikipedii oraz łączyć jej zasoby z innymi zbiorami danych w sieci (DBpedia, 2008).

¹¹ *Wstęp do systemów informacyjnych zarządzania w przedsiębiorstwie*, red. A. Nowicki, Częstochowa 2005, s. 14.

Dzięki stworzeniu sieci semantycznej uzyskać można zwiększoną wartość zasobów informacyjnych, gdyż zostają one wzbogacone o relacje. Traktowanie zasobów informacyjnych w sposób semantyczny prowadzi do lepszego zrozumienia wewnętrznych powiązań i zależności wielu zjawisk i elementów otoczenia oraz do odkrywania dotąd nieznanych bądź nie do końca rozpoznanych relacji.

Literatura

- Berners-Lee T., Hendler J., Lassila O., *The semantic web*, „Scientific American” 2001, vol. 279(5).
- Davies, J., Fensel, D., Harmelen F., *Towards the Semantic Web: Ontology-Driven Knowledge Management*, Chichester 2003.
- Efektywne zarządzanie a sztuczna inteligencja*, red. A. Baborski, Wrocław 1994.
- Ehrig M., *Ontology Alignment Bridging the Semantic Gap*, Springer 2007.
- Grudzewski W.M., Hejduk I.K., *Zarządzanie wiedzą w przedsiębiorstwach*, Warszawa 2004.
- <http://dbpedia.org/About>, lipiec 2008.
- Leuf B., *The Semantic Web: Crafting Infrastructure for Agency*, West Sussex 2006.
- Pawełoszek-Korek I., *Zarządzanie wiedzą organizacji w świetle trendu Web 2.0*, w: *Współczesne problemy zarządzania przedsiębiorstwami w gospodarce rynkowej*, red. H. Howaniec, W. Waszkielewicz, Bielsko-Biała 2008.
- SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>, lipiec 2008.
- W3C Completes Bridge Between HTML/Microformats and Semantic Web, <http://www.w3.org/2007/07/grddl-pressrelease>.
- Walton C.D., *Agency and the Semantic Web*, Oxford 2007.
- Wstęp do systemów informacyjnych zarządzania w przedsiębiorstwie*, red. A. Nowicki, Częstochowa 2005.

SEMANTIC NETWORK STANDARDS IN ORGANIZATIONAL KNOWLEDGE MANAGEMENT

Summary

The paper presents the idea of semantic network and possibilities to exploit it in contemporary knowledge management systems. Selected standards of knowledge repre-

sentations in semantic networks also has been described. Special attention has been paid to knowledge management tools based on social software often referred to as Web 2.0.

Translated by Ilona Pawełoszek-Korek

