

SIEĆ SEMANTYCZNA – NOWE SPOJRZENIE NA ZARZĄDZANIE WIEDZĄ W ASPEKcie PRZEDSIĘBIORSTWA

W artykule ukazany został obecny stan oraz rozwój sieci semantycznej w ujęciu aplikacyjnym ukazując wybrane przykłady praktyczne. Zawarty został przegląd najpowszechniej używanych technologii wykorzystywanych w sieci semantycznej: trójki RDF i RDFS, język opisu ontologii OWL, język zapisu reguł SWRL, język definiowania zapytań do struktur semantycznych SPARQL, a ta także systemy wnioskujące. Dokonane zostało syntetyczne przedstawienie wybranych aplikacji semantycznych zarówno komercyjnych jak i typu Open Source. Przedstawione zostały wybrane praktyczne przykłady zastosowań sieci semantycznych.

WSTĘP

Łatwość komunikacji, tworzenia, współpracy i dzielenia się treścią, jaką użytkownikom dostarczają współczesne technologie, a także ich dostępność i niskie koszty używania niwelują bariery dostarczania, współdzielenia i rozpowszechniania treści. Powszechność wykorzystywania społecznych aplikacji, w celu dzielenia się informacją, udzielania pomocy, znajdowania ludzi o podobnym podejściu do życia, wzajemnego uczenia i przekazywania sobie wiedzy wymusza i powoduje zmiany postaw, organizacji pracy czy procesów biznesowych. To w dużym stopniu pociąga za sobą zmiany w funkcjach i możliwościach współczesnych systemów informacyjnych. Zmianie uległy także postawy i tendencje społeczne związane z pozyskiwaniem i przyswajaniem wiedzy. Odbiorca wiedzy jest jednocześnie współautorem decydującym o tym kiedy, gdzie i w jaki sposób ją zdobywa, przekształca i przyswaja. Co więcej widoczna jest tendencja związana z oczekiwaniem i dążeniem do najbardziej atrakcyjnej formy przekazywania i przyswajania informacji.

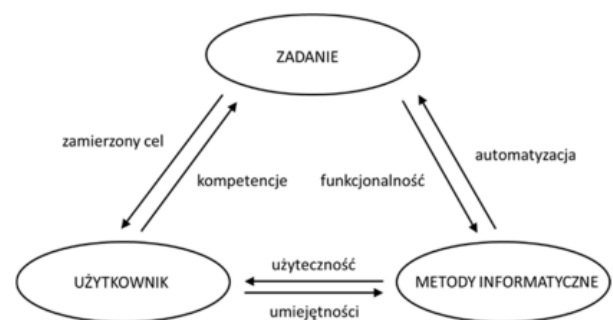
W każdej organizacji pracownik potrzebuje wiedzy na stanowisku pracy w danym momencie – tu i teraz i dokładnie takiej jaką potrzebuje. Zamiast skumulowanej porcji wiedzy potrzebne są małe jej dawki adekwatne do stanowiska pracy, na którym znajduje się pracownik budując sobie potrzebną wiedzę – dokonujący jej agregacji. Pracownik decyduje o procesie zdobywania wiedzy zgodnie ze swoimi potrzebami i możliwościami, jakie oferuje firma. W dużej mierze następuje indywidualizacja procesu dostarczania wiedzy wynikająca także z nadmiaru informacji i konieczności wsparcia pracownika w poszukiwaniu i dostępie do właściwej informacji.

Tempo i ciągle zmiany technologii, wymagań i potrzeb w miejscu pracy, powoduje zwiększające się zapotrzebowanie na informację na żądanie, online. Występująca coraz częściej automatyzacja przy wsparciu inteligentnych systemów prowadzi do zastąpienia w coraz większym stopniu zarówno pracy rąk jak i umysłów.

Zarządzanie wiedzą bez względu na zastosowaną dziedzinę powinno odbywać się w sposób uporządkowany i skuteczny. Skuteczność jest odpowiednim kryterium racjonalnego doboru metod informatycznych do wspomaganego zarządzania wiedzą, gdyż określenie skuteczności danej metody informatycznej będzie wiązało się określeniem stopnia, w jakim wsparcie nią umożliwi realizację założonych celów zarządzania wiedzą. Skuteczność ta jest funkcją

trzech zmiennych: użytkownika, zadania i metod informatycznych [1].

W związku z dynamicznie rozwijającymi się potrzebami wiedzy na żądanie dostarczonej do miejsca pracy jej odbiorcy, systemy informacyjne nieustannie ulegają funkcjonalnym przekształceniom i rozszerzeniom. Nieustannie rozwijają się technologie semantyczne, semantyczne wyszukiwanie treści oraz ciągły i konsekwentny rozwój sieci semantycznej (*Semantic Web*).



Rys. 1. Zależność pomiędzy użytkownikiem, zadaniem a środkami informatycznymi na podstawie [1].

1. SIEĆ SEMANTYCZNA

1.1. Idea sieci semantycznej

Sieć sSemantyczna to wspólny projekt kierowany przez World Wide Web Consortium odpowiedzialnej za obowiązujące standardy tworzenia treści internetowych. Idea sieci semantycznej WWW zakłada, że komputery będą „umiały” prawidłowo przetwarzać informację w sposób adekwatny do jej znaczenia. Do kodu stron internetowych dołączone są treści znaczeniowe. Istotny jest tutaj fakt, iż taka sieć korzystać ma z istniejącej infrastruktury technicznej obecnego Internetu. Zmianie ulega jedynie sposób przetwarzania przesyłanych danych polegający na przekształceniu obecnego Internetu zdominowanego przez nieustrukturyzowane dane w „sieć danych”.

Pojęcie to nie odnosi się wyłącznie do stron internetowych, sama idea ma także zastosowanie dla zbiorów danych, niekoniecznie upubliczniczonych. Prawidłowo funkcjonująca organizacja potrafi w sposób odpowiedni zarządzać gromadzonymi danymi, przekształcać je w użyteczną wiedzę, aby móc z niej skorzystać przy podej-

mowaniu decyzji biznesowych. Obecne relacyjne bazy danych nie zawsze swoimi możliwościami odpowiadają na wymagania użytkowników, zwłaszcza czerpiących dane z Internetu mające niejednokrotnie różną strukturę. Dlatego też podjęte zostały próby użycia odpowiednich struktur semantycznych i wzbogacenie tym samym systemów korzystających do tej pory z relacyjnych baz danych.

W 2001 roku Tim Berners-Lee opublikował artykuł, w którym przedstawił wizję utworzenia standardów opisywania treści, gdzie dane będą związane ze swoją semantyką [2]. Dzięki temu dostępne dane będą mogły być przetwarzane zarówno przez człowieka jak i przez specjalne programy, nazwane agentami mające na celu takie przetworzenie danych, aby wyluskać także ich znaczenie. Koncepcja sieci semantycznej jest niejednokrotnie kojarzona z „danymi czytelnymi dla maszyn” [3], „inteligentnymi agentami”, „rozproszoną bazą danych”, czy też „automatyczną infrastrukturą” [4].

Udostępnienie komputerom możliwości operowania na poziomie semantyki umożliwi stworzenie sieci, która dostarczy nowy i niespotykany poziom wyrafinowanych usług. Powstanie ogromna sieć wiedzy, wzbogacona o możliwość inteligentnego przetwarzania tych informacji przez komputery. Różne zautomatyzowane usługi będą w stanie pomóc użytkownikowi wykonać skomplikowane zadania poprzez dostęp do informacji rozumianej przez komputery. Ten proces w fazie końcowej stworzy ogromny system wiedzy, z różnego rodzaju wyspecjalizowanymi usługami wnioskującymi - systemami, które będą nam pomagały niemalże w każdym aspekcie naszego życia i staną się tak niezastąpione, jak dostęp do elektryczności dzisiaj.

1.2. Ontologie

Idea sieci semantycznej została przedstawiona w sposób nieformalizowany i wizjonerski, stąd konieczne było wykorzystanie formalnych środków, dzięki którym komputery będą potrafiły przetwarzać dane zgodnie z ich znaczeniem, wykorzystując sztuczną inteligencję. Semantyka występująca w sieci semantycznej może być wyrażana w sposób formalny za pomocą ontologii.

Bardzo precyzyjny opis reprezentacji wiedzy zaproponował J. Sowa uznając, że: „Reprezentacja wiedzy jest interdyscyplinarną dziedziną badań, w której zastosowania znajdują:

1. Logika, która dostarcza formalne struktury oraz reguły wnioskowania.
2. Ontologia, która definiuje terminy występujące w dziedzinie.
3. Wsparcie w obliczeniach, które rozróżnia reprezentację wiedzy od czystej filozofii.

... Bez logiki, reprezentacja wiedzy jest niejasna, nie posiada mechanizmów pozwalających zdecydować czy stwierdzenie jest zbędne, czy sprzeczne. Bez ontologii, pojęcia i symbole są nieprecyzyjnie opisane, zagmatwane, mylące. I w końcu bez modeli obliczeniowych, logika i ontologia nie mogą być użyte w programach komputerowych [5].”

Uważana obecnie za klasyczną i jednocześnie najczęściej przytaczaną przez wielu znanych badaczy definicja ontologii w odniesieniu do informatyki została zaproponowana przez Grubera w 1993 roku: ontologia jest formalną, jednoznacznie specyfikacją współdzielonej konceptualizacji [6]. Konceptualizacja zaś jest abstrakcyjnym, uproszczonym obrazem świata, który chcemy opisać dla pewnych celów. Formalizm konceptualizacji, który charakteryzuje ontologię, potrzebny jest dla umożliwienia maszynom jednoznacznej identyfikacji i interpretacji wiedzy. Współdzielenie zaś oznacza, że ontologia nie ogranicza się do jednej aplikacji, ale może być wykorzystywana przez większą grupę.

Znaczenie ontologii w sieci semantycznej

Pomysł na inteligentne dane uwzględniające semantykę opiera się na opakowaniu dowolnych danych opisem utworzonym zgodnie z XML, który jest łatwy do przetworzenia przez program komputerowy. Ontologie zaś dostarczają terminologii na potrzeby takiego opisu. Może być ona zaczerpnięta ze sklasyfikowanych dziedzin, niekoniecznie zależnych. Dodatkowo ontologie muszą określać relacje logiczne między pojęciami, po to by zapewnić głęboki poziom analizy, a także usprawnić metody wyszukiwania i inteligentnego wnioskowania przez programy komputerowe.

Stąd też nasuwa się porównanie sieci semantycznej do opartej na ontologiach systemu zarządzania wiedzą, w którym to systemie ontologie dostarczają narzędzi do formalnego opisu wiedzy. Na bazie takiego opisu programy komputerowe mogą dokonywać wyszukiwania i przetwarzania danych. To właśnie dzięki ontologiom możliwa jest integracja ogromnej ilości istniejących obecnie w sieci Internet heterogenicznych dokumentów [7].

Ontologie określane są mianem technologicznego rdzenia sieci semantycznej. W takim podejściu ontologie mogą zapewnić współdzieloną interpretację pewnych dziedzin i umożliwić komunikację zarówno między ludźmi jak również i systemami aplikacji [8].

Połączenie wielu dziedzin, zaproponowanie odpowiednich ontologii na tyle ogólnych by można było odnosić się przez nie do jak największej liczby ontologii szczegółowych jest jedną z podstawowych idei sieci semantycznej [9].

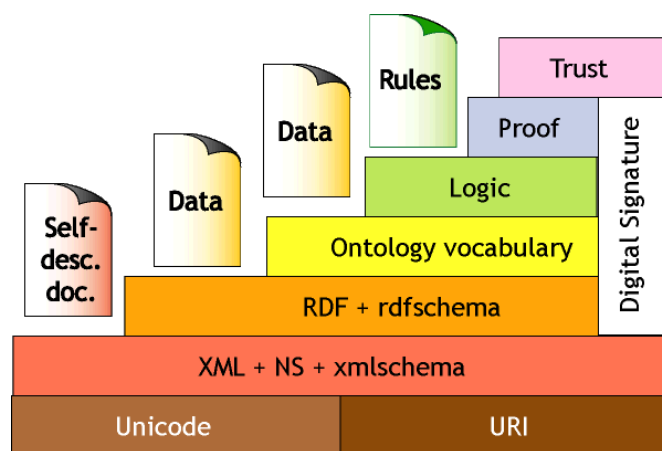
1.3. Architektura sieci semantycznej

Na architekturę sieci semantycznej składa się wielowarstwowa hierarchia wykorzystująca pokaźną liczbę standardów i technologii informatycznych zarówno tych już dostępnych jak i tych będących w trakcie opracowywania.

Sama sieć semantyczna obejmuje kilka zagadnień, do których zaliczyć można:

- definicje pojęć i obiektów oraz opis relacji między nimi,
- sposób reprezentacji wiedzy dziedzinowej przy wykorzystaniu ontologii,
- mechanizmy wnioskowania.

Logiczną strukturę technologii sieci semantycznej najczęściej przedstawia się za pomocą stosu semantycznego. Proces jego tworzenia polega na warstwowym dodawaniu standardów i technologii umożliwiających opisywanie i analizowanie zasobów poprzez tworzenie metadanych – danych o danych.



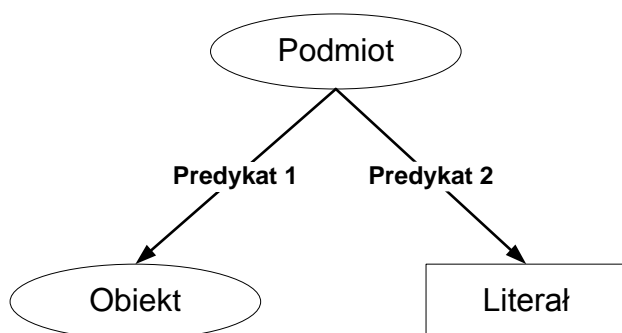
Rys. 2. Stos semantyczny – źródło <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html> (data pobrania: 21.02.2017).

Od czasu publikacji pierwszego dokumentu o Semantic Web minęło kilkanaście lat. W tym czasie semantyczna sieć WWW poddana została testom, badaniom prowadzącym do rozwoju i implementacji narzędzi i ich zastosowań.

2. TECHNOLOGIE SIECI SEMANTYCZNEJ

2.1. Język RDF i RDFS

Modelem danych używanym przez sieć semantyczną jest język RDF (*Resource Description Framework*) służący do opisywania zasobów sieciowych identyfikowanych za pomocą URI. W języku tym dane zapisywane są jako zbiór wyrażen zbudowanych z trzech elementów: podmiotu (*subject*), orzeczenia (*predicate*) i dopełnienia (*object/value*) tworzących uporządkowane trójki. Takie trójki RDF służą do opisu rzeczy poprzez tworzenie wyrażen opisujących ich właściwości. Podmiot to zasób, jaki chcemy opisać, orzeczenie określa związek między podmiotem a dopełnieniem, gdzie tym ostatnim może być zarówno obiekt jak i wartość określana mianem literalu.



Rys. 3. Graf ilustrujący dwa zdania RDF – opracowanie własne.

Opracowany został również język RDF Schema (RDFS), który zapewnia słownictwo do opisu zasobów, ich właściwości, klas i podklas. Wyrażenia w języku RDFS również są zapisywane przy postaci trójek RDF.

Podstawowymi klasami (dopełnieniami) są:

- Resource – klasa wszystkich podmiotów,
- Property – klasa wszystkich orzeczeń,
- Class – klasa wszystkich dopełnień,
- Literal – klasa wartości liczbowych bądź tekstowych,
- Datatype – klasa typów danych, podklasa Literal,
- XMLLiteral – klasa wartości dosłownych określonych w specyfikacji XML Schema.

Do podstawowych atrybutów (orzeczeń) zaliczyć można:

- subClassOf – relacja bycia podklasą,
- subPropertyOf – relacja bycia podatrybutem,
- range – relacja definiująca ograniczenie zbioru dopełnień do określonych klas dla słowa zdefiniowanego jako orzeczenie,
- domain – relacja definiująca ograniczenie zbioru podmiotów do określonych klas dla słowa zdefiniowanego jako orzeczenie,
- type – relacja określająca klasę danego podmiotu.

2.2. Język opisu ontologii OWL

Język OWL (*Web Ontology Language*) jest standardem konsorcjum W3C. Jest on przeznaczony do definiowania semantyki dokumentów poprzez specyfikowanie wiedzy dziedzinowej. Semantyka języka korzysta z badań w ramach sztucznej inteligencji w zakresie reprezentacji wiedzy. Język OWL umożliwia tworzenie ontologii, jako zbioru definicji pojęć, właściwości (relacji, atrybutów) a także obiektów będących przykładami pojęć (instancji) i relacji.

Zgodnie z konwencją wielowarstwowości w językach zapisu ontologii, OWL również posiada strukturę wielopoziomową. Wyróżnia się następujące warstwy języka:

- OWL Lite – to najniższy poziom, operuje na prostym podzbiorze słownictwa wystarczającym do modelowania prostych klasyfikacji. Warstwa ta nie pozwala na formułowanie rozszerzonych definicji pojęć.
- OWL DL (*OWL Description Logics*) – stanowi semantyczny odpowiednik logik deskrypcyjnych. Pozwala na tworzenie rozszerzonych definicji klas poprzez nakładanie kilku rodzajów ograniczeń. Stosowanie ograniczeń w znaczny sposób wpływa na efektywność i rozstrzygalność systemów realizujących tę warstwę języka. Warstwa ta jest najpowszechniej wykorzystywana w sieci semantycznej.
- OWL Full – stanowi zbiór konstruktorów, które nie posiadają żadnych ograniczeń, co przedkłada się na niską efektywność i brak rozstrzygalności. Warstwa ta nie posiada również formalnie zdefiniowanej semantyki. Klasa może być tu traktowana równocześnie jako zbiór jednostek i jako jednostka jako taka [10, 11].

Język OWL 2, w porównaniu z OWL, umożliwił tworzenie nowych konstrukcji, takich jak: klucze, nowe typy danych, kwalifikowane przez typ danych ograniczenia licznosci, nowe właściwości. Szczegółowo różnice między językiem OWL a OWL 2 zostały przedstawione w [12]. Poza pełnym językiem OWL 2 (który jest nierozstrzygalny) wyróżnia się także profile językowe: OWL 2 EL, OWL 2 QL i OWL 2 RL.

Ontologie OWL mogą mieć strukturę modułarną - można tworzyć ontologie na różnych poziomach szczegółowości, a następnie składać je w jedną całość. Język OWL wykorzystuje wszelkie dostępne obecnie metody i języki, które mają zastosowanie w semantyce dokumentów, czyli XML, XSD, RDF oraz RDFS. Taka integracja z językami znaczników, umożliwia rozszerzanie ontologii o inne zastosowania i języki.

2.3. Język reprezentacji reguł – SWRL

Język SWRL (*Semantic Web Rule Language*) - języka reprezentacji reguł, zaproponowany przez W3C. Wykorzystuje się go do zapisu reguł semantycznych. SWRL w swojej budowie rozszerza aksjomaty języka OWL. Dodaje reguły przedstawiane w postaci klauzul Horna, które mogą operować na pojęciach ontologicznych. Budowa reguły przedstawia się następująco:

- Ciało, które reprezentuje warunki
- Głowa, która określa konsekwencje wystąpienia warunków.

Zarówno głowa, jak i ciało nie muszą zawierać żadnych części atomowych. Atom to najmniejsza możliwa część, zapisana w danym formalizmie. Zbiór wielu atomów jest traktowany jak koniunkcja warunków, zakładając, że wszystkie one muszą być spełnione, aby wynik reguły był określony, jako prawdziwy. Interpretacja wyniku reguły przedstawia się następująco: zazwyczaj tożsama jest z dodaniem nowego faktu do bazy wiedzy. Pozwala to na zastosowanie nowych reguł.

Język SWRL posiada większą siłę wyrazu niż OWL DL, ale jego reguły, podobnie jak w przypadku OWL Full, nie zawsze muszą być rozstrzygalne. Propozycja W3C stała się nieformalnym językiem zapisu reguł. Podobnie jak OWL składnia języka SWRL opiera się na języku XML. Możliwy jest również zapis w postaci składni języka RDF.

2.4. Język zapytań SPARQL

Język SPARQL (*SPARQL Protocol And RDF Query Language*) to język zapytań do danych w formacie RDF, którego składnia zbliżona jest do języka SQL. Język SPARQL umożliwia zadawanie

zapytań w postaci grafów RDF uwzględniających zawartą w danych wiedzę (opisaną poprzez ontologię OWL). Wzorce występujące w zapytaniach mają formę trójek RDF wraz ze słowami kluczowymi języka. Zapytanie SPARQL składa się z trzech części:

- część prefiksów, w której definiowane są adresy URI danych i ontologii lub innych bytów.
- część opisująca rodzaj zapytania – analogicznie do SQL (SELECT, CONSTRUCT, ASK, DESCRIBE),
- Część wzorca (trójki RDF) wraz z dodatkowymi elementami (FILTER, ORDER BY, OPTIONAL).

W języku SPARQL są wyspecyfikowane cztery różne warianty zapytania w zależności od potrzeb:

- zapytanie SELECT - używane do wyciągnięcia surowych danych z SPARQL endpoint, wyniki są zwracane w formie tabeli. Po klauzuli SELECT podaje się listę zmiennych o które pytamy;
- zapytanie CONSTRUCT - służy do pobierania informacji z SPARQL endpoint, transformuje wyniki na graf RDF;
- zapytanie ASK - służy do sprawdzenia czy podane zapytanie zwróci jakiś wynik – wyrażenie zwraca wynik w postaci Prawda / Fałsz;
- zapytanie DESCRIBE - używane do wyodrębnienia grafu RDF z SPARQL endpoint, wynikiem będzie najmniejszy możliwy graf opisujący dany zasób.

Każde z opisanych rodzajów zapytań musi posiadać blok WHERE który może ograniczać zapytanie. W sekcji WHERE pomiędzy nawiasami sześcinnymi wpisujemy szablon grafu RDF, gdzie pewne elementy grafu można zastąpić zmiennymi zaczynającymi się od znaku zapytania. Tylko w przypadku zapytania DESCRIBE zapytanie WHERE jest opcjonalne.

Język SPARQL posiada podobnie jak SQL dodatkowe warunki typu ORDER BY, UNION, OPTIONAL i inne.

2.5. Wnioskowanie przy wykorzystaniu sieci semantycznej

Wnioskowanie jest konieczne wówczas, gdy nie można wprost znaleźć rozwiązania danego problemu. Wnioskowanie jest procesem polegającym na wyprowadzeniu, zgodnie z prawami logiki określonymi przez reguły wnioskowania, ze zdań uznanych za prawdziwe nowych twierdzeń. Proces wnioskowania można zdefiniować następująco: w bazie wiedzy znajdują się fakty oraz reguły, na ich podstawie należy udowodnić określoną hipotezę stanowiącą cel wnioskowania. W literaturze dotyczącej tematu wnioskowania spotkać się można z różnymi typami wnioskowania. Wśród nich wyróżnić można wnioskowanie: formalne, proceduralne, przez analogię, przez uogólnienie. Język naturalny nie jest precyzyjny, dlatego niewłaściwe używanie logiki może skutkować fałszywymi stwierdzeniami, lub uznaniem za prawdziwe stwierdzeń, które nie wynikają z przesłanek. Pomocne w rozwiązywaniu tego typu problemów okazują się sieci semantyczne, które zaliczane są do deklaratywnych metod reprezentacji wiedzy. W podejściu deklaratywnym wiedza z danej dziedziny jest zbiorem specyficznych faktów, zaś korzystanie z niej polega na stosowaniu do tego zbioru ogólnych procedur manipulacji faktami. Występuje wyraźne oddzielenie wiedzy z danej dziedziny od sposobu wykorzystania tej wiedzy w procesie wnioskowania.

Sieć semantyczna gromadzi całą wiedzę o semantyce pojęć. Stąd możliwość jej wykorzystania w systemach przetwarzających język naturalny. Znaczenie każdego pojęcia wynika ze zbioru relacji semantycznych wiążących to pojęcie z innymi pojęciami w sieci semantycznej. Wnioskowanie z wykorzystaniem sieci semantycznej odbywa się po krawędziach, które mogą posiadać wagi określające ich ważność. Jest to przeszukiwanie grafu, w którym rozpoczynając z jednego węzła grafu (pojęcia) i poruszając się po krawędziach (relacje między pojęciami), wychodząc z węzła, docieramy do

kolejnych węzłów, co odpowiada wnioskowaniu o właściwościach pojęć.

Mechanizmy wnioskowania są częścią sieci semantycznej, która wciąż się rozwija. Na chwilę obecną istnieją pewne założenia, które muszą być spełnione przez mechanizmy wnioskowania. Bardzo przydatne okazałyby się mechanizmy pozwalające systemom komputerowym nie tylko rozpoznawać relacje między conceptami, ale także i przede wszystkim wnioskować inne relacje wynikające z istniejących.

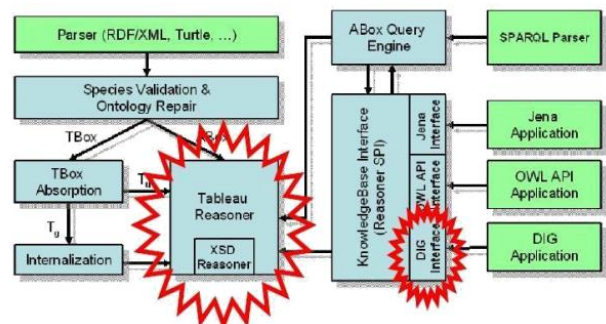
Systemy wnioskujące

Z danych, opisanych za pomocą języków RDF, RDFS, OWL i innych, można uzyskać nowe dane za pomocą wnioskowania. Taki rodzaj danych wymagał powstania nowych algorytmów oraz silników wnioskujących. Powstało szereg narzędzi wnioskujących umożliwiających uwzględnienie w procesie wnioskowania wiedzy opisanej w postaci ontologii. Podczas udzielania odpowiedzi na zapytania owa wiedza jest również brana pod uwagę przed zwróceniem ostatecznego wyniku.

Do najpopularniejszych narzędzi umożliwiających wnioskowanie oraz zadawanie zapytań w języku SPARQL zaliczyć można:

- Pellet – rozwiązanie typu Open Source umożliwiające wnioskowanie wykorzystujące RDFS, OWL, SWRL. Dodatkowo posiada wsparcie dla OWL 2 [13]. Pellet napisany jest w języku programowania Java.

Pellet architecture



Rys. 4. Architektura Pellet na podstawie [13].

- KAON2 – rozwiązanie niekomercyjne umożliwiające wnioskowanie RDFS, OWL, SWRL oraz połączenie z relacyjną bazą danych.
- HermiT – stworzony przez grupę pracowników Uniwersytetu Oksfordzkiego. Został napisany w języku Java i oferowany jest w postaci wolnego oprogramowania na licencji GNU na oficjalnej swojej stronie internetowej.
- RacerPro - rozwiązanie komercyjne, lecz możliwe jest uzyskanie licencji akademickiej.
- Oracle 11g RDFS/OWL – komercyjna baza danych wraz z systemem wnioskującym zgodnie z semantyką języków RDF i OWL.
- AllegroGraph – komercyjna baza danych RDF umożliwiająca wnioskowanie (głównie RDFS i parę aksjomatów OWL) oraz zadawanie zapytań.
- FaCT++ – rozwiązanie open Source dostępne na platformie Google Code. Powstał w języku C++. Wykorzystywany głównie do wnioskowania w kontekście warstwy OWL-DL.
- Jena2 – framework służący do budowy aplikacji sieci semantycznej rozwijany w laboratoriach firmy HP. Licencja typu Open Source. Na narzędzie składają się aplikacje umożliwiające składowanie danych (RDF), wnioskowanie oraz realizację zapytań.

3. WYBRANE PRAKTYCZNE ZASTOSOWANIA SIECI SEMANTYCZNYCH

3.1. Dbpedia

DBpedia (<http://wiki.dbpedia.org/>) jest projektem mającym na celu wyselekcjonowanie ustrukturyzowanej informacji z największej interaktywnej encyklopedii – Wikipedii. Rezultatem tego działania jest utworzenie wielkiego grafu – reprezentacji wiedzy. Zasada funkcjonowania DBpedii polega na stosowaniu złożonych zapytań zwracających konkretne informacje używając języka SPARQL. W przypadku klasycznej Wikipedii stosuje się jedynie wyszukiwanie na podstawie słów-kluczy.

3.2. GeoNames

GeoNames jest bazą danych dotyczących POI (miejsc i punktów zainteresowań) na Ziemi. Baza zawiera ponad 10 milionów miejsc. Dane te zebrane są z ponad 73 źródeł, ujednoczone i przechowane. Można przeglądać je na mapie, posiadają przypisanie:

- Kategorie
- Współrzędne geograficzne
- Nazwy w wielu językach

Baza ta jest otwarta i edytowalna. Twórcy udostępniają również darmowe API.

3.3. Wyszukiwarki

Rozwijane są semantyczne wyszukiwarki, przy pomocy których zapytania można zadawać w języku naturalnym, a zwracane wyniki posiadają uzasadnienie ich otrzymania oraz mogą zawierać dodatkowe linki do stron zawierających więcej informacji.

Przykładowe wyszukiwarki semantyczne:

- PowerSet (<http://www.powerset.com/>) – to wydajna wyszukiwarka semantyczna, której właścicielem jest firma Microsoft. Narzędzie umożliwia definiowanie zapytań w języku naturalnym (tylko angielski). Odpowiedzi udzielne są na podstawie danych pochodzących ze stron WWW, Wikipedii oraz bazy Freebase (<http://www.freebase.com/>). Dzięki „rozumieniu” treści zapytania wyszukiwarka pozwala na uzyskanie najbardziej adekwatnych wyników.
- Yahoo! SearchMonkey - wyświetla dodatkowe linki dołączane do już otrzymanych wyników. Dodatkowo istnieje możliwość tworzenia własnych aplikacji opartych o mechanizmy wyszukiwania SearchMonkey.
- Rich Snippets - narzędzie firmy Google.
- Hakia (<http://hakia.com/>) - początkowo wyszukiwarka była rozwijana przy udziale polskiej firmy PROKOM INVEST-MENTS S.A. Dzięki temu oferowany jest język Polski. Narzędzie znajduje się w fazie beta. Odpowiedzi na zapytania mogą być: konkretne fakty, powiązane linki oraz zdjęcia.

3.4. Zemanta

Zemanta (<http://www.zemanta.com/>) – system wsparcia dla autorów blogów. Pozwala na szybkie oznaczenie wpisywanych słów wraz z odnośnikami do stron i zdjęć. Sugeruje również znaczniki, których najlepiej użyć do oznaczenia treści bloga. Pozwala tworzyć blogi semantyczne opisane przy pomocy języka RDF, bez wymogu jego znajomości przez autora. Dzięki szybkiemu działaniu znacznie skraca czas pisania wiadomości. W sugerowaniu linków, zdjęć, muzyki, filmów i znaczników korzysta między innymi z serwisów takich jak: Wikipedia, Flickr, Youtube, IMDb, Musicbrainz, Amazon i innych.

PODSUMOWANIE

Wizja systemów „rozumiejących” i zdolnych do podejmowania za nas decyzji wymaga wciąż wielu lat badań i eksperymentów, jednak solidne podwaliny takich systemów zostały już ugruntowane. Obecnie prowadzonych jest szereg projektów związanych głównie z wydajnością i skalowalnością systemów semantycznych.

Obecny rozwój sieci semantycznej ukierunkowany jest na stworzenie narzędzi bazujących na wiedzy, umożliwiających integrację danych, wyszukiwanie, publikowanie ich w sieci oraz prezentowanie ich w formie ustrukturalizowanej informacji przydatnej dla użytkownika.

Dla serwisów, narzędzi komercyjnych a także przedsiębiorstw wszelkiej branży ważne jest, aby nowe technologie przynosiły nie tylko powszechne uznanie, ale również zyski. Będą one możliwe do osiągnięcia poprzez:

- lepsze zarządzanie treścią dokumentów, przez co czas uzyskania pożądanej treści ulegnie znacznemu skróceniu,
- przeprowadzanie analizy semantycznej danych w celu dostosowania się do potrzeb klienta,
- przejęcie przez komputery wielu prostych zadań, które obecnie musi jednak wykonywać człowiek (przygotowywanie raportów, wynajdywanie podobnych i rzetelnych informacji).

BIBLIOGRAFIA

1. Szeptuch A., Skuteczność metod informatycznych stosowanych w procesie zarządzania wiedzą na uczelniach, e-mentor, 3 (50), 2013.
2. Berners-Lee T., The Semantic Web – A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, Scientific American Magazine, 2001.
3. Feigenbaum L., Herman I., Hongsermeier T., Neumann E., Stephens S., *The Semantic Web in action*, Scientific American nr 297, 2007, s. 64-71.
4. Berners-Lee T., Hendler J., Lassila O., *The Semantic Web*, Scientific American nr 284, 2001, s. 34-43.
5. Sowa J. F., *Ontology, metadata, and semiotics*, In Conceptual Structures: Logical, Linguistic, and Computational Issues, Springer-Verlag, 2000, s. 55-81.
6. Gruber T. R., *A translation approach to portable ontologies*, Knowledge Acquisition, 5(2), 1993, s. 199-220.
7. Davies J., Fensel D., Harmelen F., *Towards the Semantic Web: Ontology-Driven Knowledge Management*, John Wiley&Sons, 2003.
8. Fensel D., *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlag Berlin Heidelberg, 2004
9. Daconta M. C., Obrst L. J., Smith K. T., *The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management*, Indianapolis, Indiana, USA, Wiley Publishing, 2003.
10. Roussey C., Pinet F., Kang M.A., Corcho O., *An introduction to ontologies and ontology engineering*, [w:] Falquet G., Metral C., Teller J., Tweed C. (red.), *Ontologies in Urban Development Projects*, Springer, 2011, s. 9-38.
11. Antoniou G., van Harmelen F., 2009, *Web Ontology Language: OWL*, [w:] Staab S., Studer R. (red.), *Handbook On Ontologies*, second edition, International Handbooks on Information Systems, Springer, 2009, s. 91-110.
12. Grau B.C., Horrocks I., Motik B., Parsia B., Patel-Schneider P., Sattler U., *OWL 2: The next step for OWL*, Web Semantics: Science, Services and Agents on the World Wide Web, vol. 6, no. 4, 2008, s. 309-322.

13. Evren S., Bijan P., Grau B., *Pellet: A Practical OWL-DL Reasoner*, Valencia 2004.

Semantic Web - new knowledge management in business aspects

The article presents the current state and development of the semantic web in terms of application, showing selected practical examples. A review of the most commonly used technologies used in the semantic web: RDF and RDFS triangles, OWL ontology description language, SWRL SWL writing language, SPARQL semantics query language, and application systems. A synthesis of selected semantic applications, both commercial and open source, has been made. Selected practical examples of semantic web applications are presented.

Autorzy:

dr inż. **Anna Gładysz** – Politechnika Rzeszowska