

XV Konferencja PLOUG  
Kościelisko  
Październik 2009

# Semantic Web – technologie, zastosowania, rozwój

Jarosław Bąk, Czesław Jędrzejek  
Politechnika Poznańska

*jaroslaw.bak@put.poznan.pl, czeslaw.jedrzejek@put.poznan.pl*

**Abstrakt.** Niniejszy artykuł przedstawia bieżący stan oraz rozwój Semantycznej Sieci WWW (Semantic Web) w ujęciu aplikacyjnym. Zaprezentowano najważniejsze zagadnienia historyczne dotyczące powstałych rozwiązań oraz dokonano zgrubnego podziału możliwych zastosowań. Praca zawiera krótką charakterystykę najpowszechniej używanych technologii wykorzystywanych w Semantycznej Sieci: języka opisu ontologii (OWL), języka opisu danych w postaci trójek (RDF i RDFS), języka zapisu reguł (SWRL) oraz języka służącego do definiowania zapytań (SPARQL) do struktur semantycznych. Przedstawione zostały różnego rodzaju aplikacje semantyczne, wśród których znajdują się narzędzia mające zastosowanie komercyjne oraz rozpowszechniane na licencji typu Open Source. Zarysowano również dalszy rozwój Sieci Semantycznej oraz zaprezentowano bieżące prace badawcze dotyczące skalowalności aplikacji semantycznych.



## 1. Wprowadzenie

Pojęcie semantyki występuje w kilku znaczeniach. Po pierwsze jest to dział językoznawstwa zajmujący się znaczeniem słów, zdań i tekstów. W odniesieniu do technik informatycznych ma związek z rozumieniem przez komputer języka naturalnego (technologia Natural Language Processing). Po drugie, w węższym znaczeniu, semantyka dotyczy opisu formalnego modelu (np. semantyka ontologii OWL), Semantyczna Sieć WWW (ang. Semantic Web) to przyszłościowa wizja Sieci zarysowana początkowo w dokumencie [Berner] przez „ojca” WWW Tima Bernersa-Lee. W dokumencie tym autor postulował potrzebę ewolucji Sieci, w której treść dokumentów w formie nie do końca zdefiniowanych struktur semantycznych mogłaby być przetwarzana przez programy oraz komputery. Jest to istotne ze względu na ogrom istniejących zasobów i ciągle oraz szybkie ich powiększanie się. Automatyczne przetwarzanie zawartości dokumentów wymaga jednak wiedzy nie tylko o składni, ale również o semantyce. Połączenie danych i metadanych semantycznych umożliwiłoby zautomatyzowanie przetwarzania zawartości znaczeniowej dokumentów, co z kolei pozwoliłoby na:

- wyszukiwanie informacji na podstawie znaczenia haseł wpisywanych w wyszukiwarce, a nie tylko po słowach kluczowych, co często skutkuje błędnymi odpowiedziami i uniemożliwia korzystanie z synonimów,
- rozróżnianie danych na podstawie kontekstu ich znaczenia,
- ekstrakowanie informacji z różnych źródeł i ich automatyczna integracja,
- prezentowanie tylko istotnych danych z punktu widzenia użytkownika wraz z możliwością wyjaśnienia podjętych przez komputer działań/wniosków,
- przeprowadzanie wnioskowania na danych opisanych semantycznie, co skutkuje uzyskaniem nowych informacji (które nie były dotychczas jawnie zapisane), a te z kolei poddane procesom wnioskowania mogą prowadzić do uzyskania kolejnych nowych faktów itd.,
- wprowadzanie wiedzy zawartej w dokumentach danego typu, przez co ułatwiłoby to zarządzanie dokumentami oraz informacją, a także prowadziło do utrzymywania niesprzeczności i aktualności gromadzonych danych,
- automatyczną generację dokumentów opisanych semantycznie,
- automatyczną generację stron WWW opisanych semantycznie bez udziału użytkownika (wiedza służąca do opisu strony czerpana jest z semantyki zawartej w ontologii).

Idea Semantycznej Sieci WWW zakłada, iż komputery będą w stanie prawidłowo przetwarzać informację w sposób adekwatny do jej znaczenia. Istotny jest tutaj fakt, iż taka Sieć korzystać ma z istniejącej infrastruktury technicznej obecnego Internetu. Zmianie ulega jedynie sposób przetwarzania przesyłanych danych.

Idea Semantycznej Sieci jest przedstawiana w sposób niesformalizowany i wizjonerski, ta jednak wymaga formalnych środków, dzięki którym komputery będą potrafiły przetwarzać dane zgodnie z ich znaczeniem i przy wykorzystaniu algorytmów sztucznej inteligencji. Semantyka występująca w Semantic Web może być wyrażana w sposób formalny za pomocą ontologii definiowanej w języku OWL [OWL]. Ontologię stanowią metadane zawierające bazę wiedzy dziedzinowej, której dotyczą opisywane dane.

Od czasu publikacji pierwszego dokumentu o Semantic Web minęło już ponad 10 lat. Tym samym Semantyczna Sieć WWW ma za sobą dziesięciolecie badań, rozwoju, implementacji narzędzi i zastosowań. Dzisiejsza Semantyczna Sieć WWW to:

- rozproszona sieć danych, w której każda informacja może być powiązana w sposób logiczny z dowolną inną informacją,
- największa formalna baza danych istniejąca na świecie a jednocześnie najbardziej „zabałaganiona”,
- zbiór standardów W3C takich, jak: RDF [RDF], RDFS [RDFS], OWL i inne,
- motywacja do rozwoju aplikacji sztucznej inteligencji, w których skalowalność będzie porównywalna do współczesnych możliwości relacyjnych baz danych i Internetu.

Niniejsza praca ma na celu podsumowanie dotychczasowych osiągnięć, przedstawienie powstałych i obecnie wykorzystywanych narzędzi do tworzenia Semantycznej Sieci, a także zaprezentowanie aplikacji mających zastosowanie komercyjne oraz rozpowszechnianych na licencjach typu Open Source.

Następny rozdział charakteryzuje najbardziej popularne i użyteczne technologie Semantycznej Sieci WWW.

## 2. Technologie Semantycznej Sieci WWW

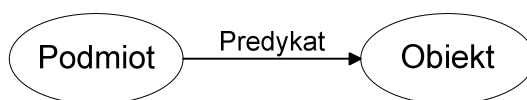
Rozwój Semantic Web wymusił niejako powstawanie nowych narzędzi umożliwiających tworzenie Sieci zgodnie z przyjętą ideą. Są one wykorzystywane do:

- opisu informacji pozyskiwanej z Internetu bądź relacyjnych baz danych,
- definiowania baz wiedzy w postaci ontologii dziedzinowych i fundacjonistycznych,
- przetwarzania, analizowania i wnioskowania na danych opisanych semantycznie,
- efektywnej realizacji zapytań zapisywanych w sposób nieformalny a jednocześnie „rozumiany” i interpretowany przez maszyny.

Poniżej znajdują się krótka charakterystyka najpowszechniej używanych narzędzi i technologii Semantycznej Sieci WWW.

### 2.1. 2.1. Języki RDF i RDFS do reprezentowania metadanych

RDF (Resource Description Framework) jest modelem służącym do opisywania zasobów sieciowych identyfikowanych za pomocą URI. Każdy opis zasobu (*resource*, *subject*) jest zbiorem stwierdzeń (*statements*), z których każde specyfikuje właściwość (*predicate*) zasobu oraz wartość tej właściwości (*object*). Pojedyncze stwierdzenie ma postać trójki: (subject, predicate, object), reprezentowanej na różne sposoby, np. w postaci elementarnego grafu przedstawionego na rys. 1, w zapisie wyrażonym w języku XML, w języku trójkowym N3 i innych. Stwierdzenia mogą podlegać reifikacji stając się zasobami, które mogą być również opisywane poprzez stwierdzenia wyższych rzędów.



Rys. 1. Struktura danych RDF

Język RDFS (RDF Schema) umożliwia tworzenie taksonomii klas za pomocą subsumpcji (właściwość *rdfs:subClassOf*) oraz taksonomii relacji (właściwość *rdfs:subPropertyOf*), a także określanie powiązań pomiędzy klasami za pomocą relacji (właściwości *rdfs:domain* i *rdfs:range*). Możliwe jest również posługiwanie się obiektami klas (typowanie zasobów za pomocą właściwości *rdf:type*).

## 2.2. Język reprezentacji ontologii – OWL

Język OWL (Web Ontology Language) jest standardem konsorcjum W3C. Jest on przeznaczony do definiowania semantyki dokumentów poprzez specyfikowanie wiedzy dziedzinowej. Stanowi integralną formę, która łączy zalety języków RDF, RDFS oraz DAML+OIL. W związku z jego pochodzeniem, forma składniowa oparta jest na języku XML, przy czym liczba form zapisu jest dość rozległa (RDF(S), N3, N-Triple i inne). Język OWL umożliwia tworzenie ontologii jako zbioru definicji pojęć, właściwości (relacji, atrybutów) a także obiektów będących przykładami pojęć (instancjami) i relacji. Zgodnie z konwencją wielowarstwowości w językach zapisu ontologii, OWL również posiada strukturę wielopoziomową. Warstwy języka to tzw. „gatunki sów” (ang. *species*):

- warstwa OWL Lite – to najniższy poziom, który pozwala tworzyć taksonomię pojęć ontologicznych opartą o relacje subsumcji. Pozwala również na nakładanie ograniczeń na relacje – jednak wystąpić mogą tylko więzy licznosciowe (ang. *cardinality constraints*). Warstwa ta nie pozwala na formułowanie rozszerzonych definicji pojęć (które są jednocześnie tzw. konstruktorami klas),
- warstwa 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 Semantycznej Sieci,
- warstwa 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.

Ontologie OWL mogą mieć strukturę modułarną. Dzięki zastosowaniu tzw. importom, 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 (np. SWRL).

## 2.3. Język reprezentacji reguł – SWRL

Język SWRL (Semantic Web Rule Language) [SWRL] jest propozycją języka reprezentacji reguł, zaproponowaną przez konsorcjum W3C. Służy on do zapisu reguł semantycznych. Język ten stanowi kombinację języków OWL-DL, OWL-Lite oraz elementów RuleML. SWRL rozszerza aksjomaty języka OWL o reguły w postaci klauzul Horna, które mogą operować na pojęciach ontologicznych. Budowa reguły jest oparta o standardowy schemat: ciało reguły, reprezentujące warunki oraz głowa reguły, określająca efekt zajścia 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. Wynik reguły oznacza zazwyczaj dodanie nowej danej (faktu) do bazy wiedzy, co umożliwia zastosowanie nowych reguł (w przypadku sterowania reguł danymi). Zmienne w regułach są zawsze poprzedzone znakiem „?”, w przypadku wartości należy je podawać w cudzysłowie.

Język SWRL można wzbogacić o tzw. funkcje wbudowane w postaci ograniczeń wyrażonych w języku SWRLB (SWRL Built-ins) [SWRLB]. Więzy można nakładać na wartości występujących w regułach zmiennych oraz wyrażać zależności pomiędzy nimi (np.  $?x > ?y$ ).

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 (nie jest to rekomendacja, ale propozycja – ang. *submission*) językiem zapisu reguł. Podobnie, jak

RuleML czy 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. Dzięki takiemu podejściu programiści zaznajomieni z relacyjnymi bazami danych w prosty sposób będą mogli przenieść się na technologie semantyczne. 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 oraz ontologii lub innych dokumentów,
- część opisująca rodzaj zapytania (SELECT, CONSTRUCT, ASK, DESCRIBE),
- część składająca się ze wzorca w postaci trójek RDF wraz z dodatkowymi elementami (FILTER, ORDER BY, OPTIONAL itd.).

SPARQL to również protokół przesyłu danych typu RDF poprzez technologie Web Service: WSDL 2.0 i SOAP 1.2. Ponadto rozwijany jest również SPARQL Query Results XML Format będący specyfikacją standardu dokumentów opisujących rezultaty zapytań typu SELECT i ASK.

## 2.5. 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. Wiedza ta jest również brana pod uwagę w procesie udzielania odpowiedzi na zapytania. Wiele systemów pozwala na realizację zapytań, podczas gdy w tle dokonywane jest wnioskowanie [BąJę07, BąJę08]. Stosowane są mechanizmy wnioskowania w przód, wstecz oraz mieszane (hybrydowe).

Do najpopularniejszych narzędzi umożliwiających wnioskowanie i/lub zadawanie zapytań w języku SPARQL należą:

- Pellet [Pellet] – narzędzie typu Open Source, możliwe wnioskowanie RDFS, OWL, SWRL; posiada również wsparcie dla OWL 2 [OWL2],
- KAON2 [KAON2] – niekomercyjna wersja narzędzia Ontobroker [OntBro] umożliwiająca wnioskowanie RDFS, OWL, SWRL oraz połączenie z relacyjną bazą danych,
- RacerPro [Racer] – komercyjny silnik wnioskujący, możliwe jest uzyskanie licencji akademickiej,
- Oracle 11g RDFS/OWL [Oracle] – komercyjna baza danych wraz z systemem wnioskującym zgodnym z semantyką języków RDF i OWL
- AllegroGraph – [Allegro] – komercyjna baza danych RDF umożliwiająca wnioskowanie (głównie RDFS i parę aksjomatów OWL) oraz zadawanie zapytań,
- FaCT++ [Fact] – narzędzie typu Open Source, służy do wnioskowania w warstwie OWL-DL,
- Jena2 [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. Zastosowania Semantycznej Sieci WWW

Pomimo wszelkiego rodzaju przeciwności, związanych głównie ze skalowalnością rozwiązań, Semantyczna Sieć WWW znajduje coraz więcej uwagi ze strony firm komercyjnych, których produkty wpasowują się w ideę Tima Bernersa-Lee. Poniżej zamieszczone zostały najciekawsze, według Autorów, zastosowania i narzędzia Semantic Web.

#### 3.1. Wyszukiwarki

Obecnie istniejące wyszukiwarki nie zawsze są w stanie prawidłowo odpowiedzieć na zadane zapytanie. Często odpowiedzi dotyczą zupełnie innych dokumentów niż te, których się spodziewano. Dlatego też rozwijane są semantyczne wyszukiwarki, przy pomocy których zapytania można zadawać np. w języku naturalnym, a zwracane wyniki posiadają uzasadnienie ich otrzymania oraz mogą zawierać dodatkowe linki do stron zawierających więcej informacji.

Do przykładowych wyszukiwarek semantycznych należą:

1. PowerSet [PowerSet] – 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 [Freebase]. Przykładowo, zadając zapytanie o zarząd firmy Apple: who is a member of apple's board – dostaniemy listę członków wraz ze zdjęciami oraz linkami do Wikipedii. Można zadawać bardziej precyzyjne zapytania, np. data założenia Uniwersytetu Harvarda: when was Harvard University founded – otrzymamy dokładną datę: September 8, 1636 (373 years ago). Dzięki „rozumieniu” treści zapytania wyszukiwarka pozwala na uzyskanie najbardziej adekwatnych wyników.
2. Yahoo! SearchMonkey – crawler tej wyszukiwarki pozwala zaindeksować dane RDF, dane z rodziny microformats [MicroFor], dane z serwisu del.icio.us i inne, które są opisane semantycznie. Wyświetla dodatkowe linki dołączane do już otrzymanych wyników. Dodatkowo istnieje możliwość tworzenia własnych aplikacji opartych o mechanizmy wyszukiwania SearchMonkey.
3. Rich Snippets [RiSn] – narzędzie firmy Google zbliżone do rozwiązania Yahoo!. Strony opisane przy pomocy RDF lub microformats mogą być brane pod uwagę w przypadku udzielania odpowiedzi na zapytanie dotyczące ludzi lub produktów. W przypadku produktów mamy możliwość wyświetlania opinii, ceny itd.
4. Hakia [Hakia] – wyszukiwarka pozwalająca na zadawanie zapytań w języku naturalnym. Początkowo wyszukiwarka była rozwijana przy udziale polskiej firmy PROKOM INVESTMENTS 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.

#### 3.2. Semantyczne Wiki

Idea Wikipedii i rozwiązań Wiki poszła w kierunku stworzenia semantycznych stron tego typu. Do najpopularniejszych rozwiązań należą:

1. Semantic MediaWiki [SMW] – to darmowe rozszerzenie platformy MediaWiki [Media], które pozwala na dołączanie do zwykłej treści Wiki semantycznych adnotacji. W tym przypadku semantyka upraszcza strukturę samej Wiki, pozwala użytkownikom na szybsze znalezienie informacji a także poprawia jakość i spójność zawartych informacji. Semantyka pozwala również na automatyczne generowanie list (lista miast kraju posegregowana wg ludności, lista ludzi urodzonych w danym mieście itd.), graficzne wyświetlanie informacji poprzez tzw. semantyczne mapy czy też format Semantic Result Format. Dodatkowo SMW posiada mechanizmy ułatwiające eksport danych do innych formatów, takich jak RDF, CSV czy JSON.

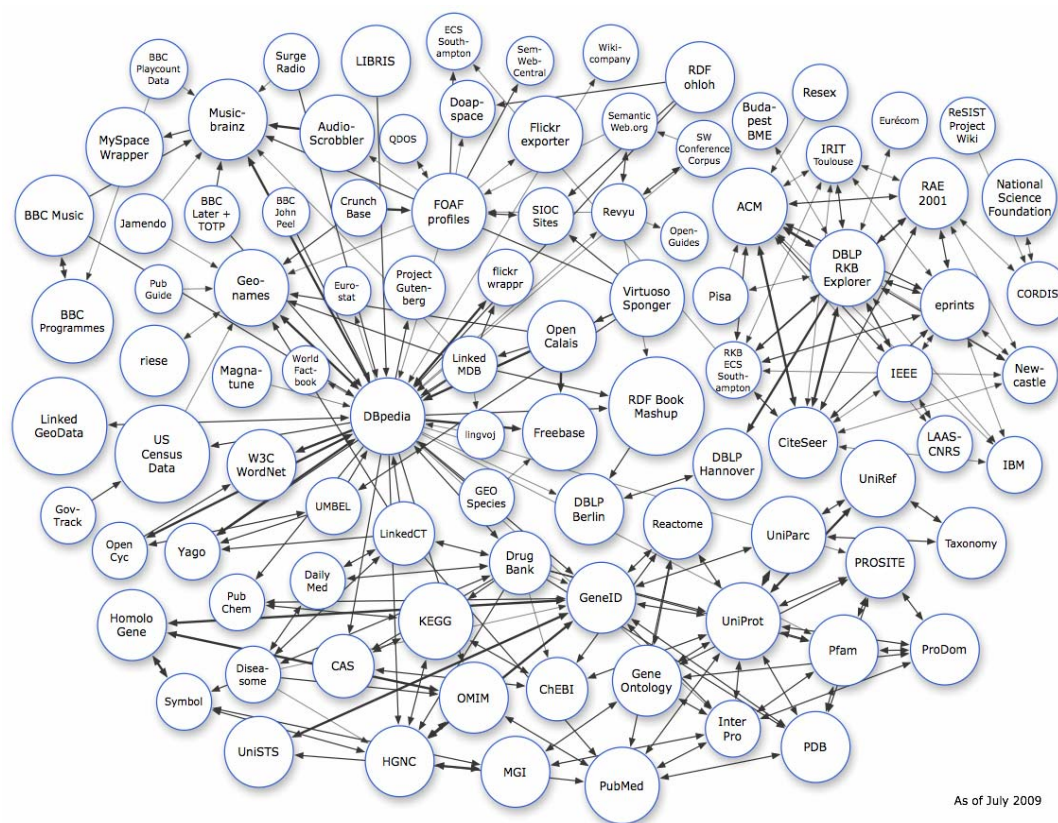
2. BiomedGT [Biomed] – semantyczna Wiki umożliwiająca wspólne tworzenie biomedycznej terminologii przez ekspertów z całego świata. Wiki dotyczy głównie kwestii związanych z chorobami nowotworowymi, ale nie tylko. Rozwijana jest przez National Cancer Institute Center for Bioinformatics oraz przez firmę Apelon, Inc.
3. VisWiki [VisWiki] – specyficzny rodzaj Wiki umożliwiający wyszukiwanie w zależności od dostępnych multimediów, które pobierane są ze stron serwisów: Wikipedia i Youtube. Umożliwia również wyszukiwanie w serwisach Yahoo!, Flickr i innych.

### 3.3. Bazy wiedzy

Poza wyszukiwarkami i systemami Wiki pojawiają się również inicjatywy, których celem jest gromadzenie danych i wiedzy z różnych heterogenicznych źródeł. Do serwisów realizujących takie podejście należą między innymi:

1. Freebase [FreeBase] – internetowa baza danych umożliwiająca integrację informacji pochodzących z różnych źródeł. Dane pochodzące z serwisów WWW są opisywane za pomocą znaczników dostarczanych przez Freebase. Można również tworzyć własne znaczniki. Baza tworzona jest przez społeczność użytkowników. System pozwala na wyszukiwanie zintegrowanych informacji na zadany temat. Odpowiedzi są przedstawiane w postaci ustrukturalizowanej informacji wraz z odwołaniami do innych serwisów, takich jak: Wikipedia, NNDB, IMDb, New York Times i innych. Dane mogą być reprezentowane w postaci grafów RDF.
2. Linking Open Data [LOD] – inicjatywa mająca na celu zebranie wszystkich dostępnych otwartych zbiorów danych RDF w jeden globalny graf. Graf (przedstawiony na Rysunku 2) zawiera powiązania RDF pomiędzy heterogenicznymi bazami danych opisanymi za pomocą trójek RDF. Do tak skonstruowanego grafu można zadawać zapytania w języku SPARQL, wykorzystywać ontologie oraz przeprowadzać wnioski. Projekt LOD jest w ciągłym rozwoju. Rysunek 2 przedstawia stan na lipiec 2009. Na LOD składa się obecnie 4,7 miliarda trójek RDF, które są powiązane 142 milionami interlinków. Baza ta ciągle się powiększa i możliwe, iż niedługo nie będzie można policzyć wszystkich powiązań pomiędzy dokumentami (ze względu na złożoność oraz ciągle dodawanie nowych powiązań). Celem projektu jest również sformułowanie ogólnie akceptowalnych reguł i standardów tworzenia dokumentów RDF i wzajemnych powiązań pomiędzy nimi.
3. DBpedia [DBped] – projekt mający na celu wyekstrahowanie ustrukturalizowanej informacji z Wikipedii. Wydobyta informacja jest odpowiednio opisana za pomocą ontologii i reprezentowana w postaci trójek RDF. Dzięki takiemu podejściu mamy możliwość definiowania wyrafinowanych zapytań w języku SPARQL do Wikipedii. Niestety projekt dotyczy tylko języka angielskiego (tylko hasła i krótkie streszczenia są opisywane w 30 językach), a przekształcanie danych z formatu Wikipedii do DBpedii nie jest dokonywane automatycznie (ze względu na złożoność tego procesu). Obecnie baza DBpedii zawiera 274 miliony trójek RDF, co odpowiada opisowi ponad 2,6 miliona haseł: osób, miejsc, albumów muzycznych, filmów, firm czy też linków do innych stron WWW i dokumentów RDF. DBpedia jest częścią projektu LOD.

Każda z powyższych baz udostępnia interfejs sieciowy, który może być wykorzystywany przez deweloperów innych stron WWW. Serwis Freebase udostępnia usługi typu Web Service, natomiast dostęp do serwisów LOD i DBpedia możliwy jest poprzez protokół SPARQL.



As of July 2009

Rys. 2. Graf RDF projektu Linking Open Data, tzw. chmura danych LOD (ang. *LOD data cloud*)

### 3.4. Inne zastosowania

W Internecie pojawiają się również innego rodzaju zastosowania Semantic Web, niż wymienione powyżej. Wynika to z faktu, iż jej ciągły rozwój i nowe technologie stwarzają zupełnie nowe możliwości przetwarzania danych, stąd ciągle pojawiające się nowe pomysły i zastosowania. Należą do nich między innymi:

1. Faviki [Faviki] – system opisu ulubionych stron WWW (zakładek) opisanych semantycznie za pomocą znaczników standardu Common Tag [ComTag] oraz RDF. Podczas opisywania zakładek mamy możliwość skorzystania z wiedzy zawartej w Wikipedii oraz DBpedii. Można również współdzielić zakładki w celu dokładniejszego scharakteryzowania opisu danej strony WWW lub też pozostawić zakładkę i jej opis tylko dla siebie. System pozwala również na tworzenie własnych znaczników wraz z opisem semantycznym (np. czy jaguar to samochód, zwierzę czy też system operacyjny). W ten sposób powstaje największa baza semantycznie opisanych najczęściej odwiedzanych (ulubionych) stron WWW.
2. Zemanta [Zema] – system wsparcia dla autorów blogów. Pozwala na szybkie oznaczenie wpi-sywanych 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.

## 4. Przyszłość Semantic Web

Realizacja Semantycznej Sieci WWW powoli zbliża się do wizji przedstawionej pod koniec lat 90-tych. O ile do systemów „rozumiejących” i zdolnych do podejmowania za nas decyzji pozostało wiele lat badań i eksperymentów, o tyle solidne podwaliny takich systemów zostały już ugruntowane. Obecnie prowadzonych jest szereg projektów związanych głównie z wydajnością i skalowalnością systemów semantycznych.

Do takich należy narzędzie LarKC [LarKC]. Large Knowledge Collider rozwijany jest w ramach projektu 7 Programu Ramowego Unii Europejskiej poświęconemu skalowalności systemów semantycznych. Jest to najbardziej obiecująca inicjatywa, w której Semantyczna Sieć WWW ma być skalowalna do rozmiaru obecnego Internetu. Głównymi celami projektu są:

- stworzenie platformy dla rozproszonego niekompletnego wnioskowania,
- usunięcie barier dotyczących skalowalności obecnej Semantycznej Sieci,
- połączenie wnioskowania z wyszukiwaniem i ekstrakcją informacji,
- ustalenie kompromisu pomiędzy jakością odpowiedzi a czasem jej uzyskania (w obecnych systemach semantyczne przetworzenie danych często wymaga zbyt dużo czasu),
- opracowanie algorytmów wnioskujących z określonym prawdopodobieństwem,
- zapewnienie modularnej architektury opartej na wtyczkach.

Wszystkie zastosowania Semantic Web wymienione w Rozdziale 3 są w ciągłym rozwoju. Ponadto powstają nowe inicjatywy oraz rozwijane są nowe meta-języki. Do najważniejszych należą:

3. OWL 2 – nowa wersja języka opisu ontologii. OWL 2 będzie wykorzystywany przez algorytmy projektu LarKC. Najważniejsze różnice pomiędzy OWL a OWL 2 to:
  - opracowanie nowej składni – OWL 2 Manchester Syntax,
  - nowe konstrukcje zwiększające siłę wyrazu języka,
  - poszerzona liczba akceptowalnych typów danych,
  - rozszerzone możliwości adnotacji,
  - zdefiniowanie dotychczas istniejących aksjomatów w celu ułatwienia ich zrozumienia, wydajnej implementacji oraz prostszego tworzenia ontologii,
  - wprowadzenie profili w zależności od skali problemu oraz efektywności procesu wnioskowania,
  - zachowanie kompatybilności w stosunku do pierwszej wersji języka OWL.
4. Rule Interchange Format [RIF] – inicjatywa mająca na celu opracowanie jednolitego formatu reguł wnioskowania.
5. SPARQL – dopracowywane są wszystkie elementy tej rodziny (SPARQL nadal nie jest standardem a jedynie rekomendacją konsorcjum W3C).

Obecny rozwój Semantycznej Sieci 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. Zbyt małe praktyczne zastosowanie ontologii powinno ulec zmianie ze względu na szereg pojawiających się ustępstw dotyczących złożoności przetwarzania oraz rozwoju OWL 2.

Dla serwisów i narzędzi komercyjnych ważne jest, aby nowe technologie przynosiły nie tylko powszechne uznanie, ale również zyski. Osiągnięte to może być na kilka sposobów, między innymi poprzez:

- wiarygodną integrację danych sterowaną semantyką,
- lepsze zarządzanie treścią dokumentów, przez co czas uzyskania požądanej 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 itd.),
- inne, o których dziś nie mamy zupełnie pojęcia.

## 5. Podsumowanie

W niniejszej pracy przedstawiono krótką charakterystykę Semantycznej Sieci WWW, obecny jej stan, technologie, zastosowania oraz dalsze plany rozwoju. W artykule zamieszczono tylko najważniejsze, zdaniem Autorów, kwestie. Ciągły rozwój Semantic Web skutkuje tym, iż przedstawienie wszystkich aplikacji w jednym artykule jest praktycznie niemożliwe. Ponadto powszechnie uznaje się, iż zastosowania Semantic Web są trudne do przewidzenia (abstrahując od obecnych wyobrażeń). Wizja komputerowych agentów komunikujących się w imieniu ludzi, potrafiących samodzielnie np. zarejestrować wizytę u lekarza [BHL01] wydaje się być nieco odległa. Jednak obecnie prowadzone prace badawcze oraz wysiłki entuzjastów z całego świata skupiają coraz większą uwagę wielkich firm, takich jak: Oracle, Google, Yahoo!, IBM czy Microsoft. Szybko też rozwijają się firmy, które oferują aplikacje wymagające wyjście poza relacyjne struktury danych, np. przeszukiwanie struktur grafowych [Detica, StHo09].

Obecnie dają się zauważyć dwie tendencje rozwoju Sieci: w wizji sieci semantycznej (nacisk położony na semantykę) użyte są złożone struktury danych (np. OWL i wnioskowanie). Aplikacje przetwarzania tekstu i aplikacje używane przez firmy farmaceutyczne są bliskie automatycznemu przetwarzaniu wiedzy, w więc sztucznej inteligencji. Sieć semantyczna zakłada formalnie uboższe, ale bardziej różnorodne struktury danych. Dla współczesnej Sieci WWW, określanej jako Web 2.0 treść serwisów jest generowana przez użytkowników (sieci społeczne). Dla Sieci Web 3.0, poza dotychczas wykorzystywanymi mechanizmami (XML, AJAX, Web Service, RSS, inne) powszechne będą aplikacje wykorzystujące sztuczną inteligencję, rozwiązania semantyczne oraz systemy przetwarzania języka naturalnego.

Praca ta została sfinansowana ze środków na naukę MNiSzW w latach 2006-2010 jako projekt badawczy rozwojowy "Narzędzie wspomagające procedury śledcze wykorzystujące automatyczne wnioskowanie" oraz przez grant Politechniki Poznańskiej 45-083/09/DS.

## Bibliografia

- [Allegro] AllegroGraph, <http://www.franz.com/agraph/allegrograph/>
- [BaJe07] Bał J., Jędrzejek C.: Wnioskowanie hybrydowe w relacyjnej bazie danych, wykorzystujące ontologię OWL wzbogaconą regułami języka SWRL, Materiały konferencyjne PLOUG2007, Kościelisko, , ISSN 1641-211, str. 285-300
- [BaJe08] Bał J., Jędrzejek C., Wnioskowanie hybrydowe w relacyjnej bazie danych wykorzystujące podejście semantyczne, Konferencja BDAS'08 (Bazy Danych: Aplikacje i Systemy), Ustroń 27-30 maja 2008, str. 335-350
- [Berner] Berners-Lee T. Semantic Web road map, <http://www.w3.org/DesignIssues/semantic.html>.
- [BHL01] Berners-Lee T., Hendler J., Lassila O., The Semantic Web, Scientific American 284(5):34-43 (May 2001)

---

[Biomed]	Baza BiomedGT, <a href="http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page">http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page</a>
[ComTag]	Standardy Common Tag, <a href="http://www.commontag.org/Home">http://www.commontag.org/Home</a>
[Detica]	<a href="http://www.detica.com/">http://www.detica.com/</a>
[DBped]	Baza DBPedia, <a href="http://dbpedia.org/">http://dbpedia.org/</a>
[Fact]	Narzędzie wnioskujące Fact++, <a href="http://owl.man.ac.uk/factplusplus/">http://owl.man.ac.uk/factplusplus/</a>
[Faviki]	Serwis Faviki, <a href="http://www.faviki.com/pages/welcome/">http://www.faviki.com/pages/welcome/</a>
[FreeBase]	Serwis Freebase, <a href="http://www.freebase.com/">http://www.freebase.com/</a>
[Hakia]	Wyszukiwarka Hakia, <a href="http://hakia.com/">http://hakia.com/</a>
[Jena2]	Biblioteka Jena2, <a href="http://jena.sourceforge.net/">http://jena.sourceforge.net/</a>
[KAON2]	Narzędzie KAON2, <a href="http://kaon2.semanticweb.org/">http://kaon2.semanticweb.org/</a>
[LarKC]	Projekt LarKC, <a href="http://www.larkc.eu/">http://www.larkc.eu/</a>
[LOD]	Projekt Linking Open Data, <a href="http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/">http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/</a>
[Media]	Narzędzie MediaWiki, <a href="http://www.mediawiki.org/wiki/MediaWiki">http://www.mediawiki.org/wiki/MediaWiki</a>
[MicroFor]	Microformats, <a href="http://microformats.org/">http://microformats.org/</a>
[OntBro]	Narzędzie OntoBroker OWL, <a href="http://www.ontoprise.de/de/en/home/products/ontobroker.html">http://www.ontoprise.de/de/en/home/products/ontobroker.html</a>
[Oracle]	Technologie Semantyczne Oracle <a href="http://www.oracle.com/technology/tech/semantic_technologies/index.html">http://www.oracle.com/technology/tech/semantic_technologies/index.html</a>
[OWL]	Web Ontology Language (OWL) Guide Version 1.0, <a href="http://www.w3.org/TR/owl-features/">http://www.w3.org/TR/owl-features/</a>
[OWL2]	OWL 2 Web Ontology Language, <a href="http://www.w3.org/2007/OWL/wiki/Document_Overview">http://www.w3.org/2007/OWL/wiki/Document_Overview</a>
[Pellet]	Narzędzie wnioskujące Pellet, <a href="http://clarkparsia.com/pellet">http://clarkparsia.com/pellet</a>
[PowerSet]	Serwis PowerSet, <a href="http://www.powerset.com/">http://www.powerset.com/</a>
[Racer]	Narzędzie RacerPro, <a href="http://www.racer-systems.com/">http://www.racer-systems.com/</a>
[RDF]	RDF, <a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a>
[RDFS]	<a href="http://www.w3.org/TR/rdf-schema/">http://www.w3.org/TR/rdf-schema/</a>
[RIF]	Standard Rule Interchange Format, <a href="http://www.w3.org/2005/rules/wiki/RIF_Working_Group">http://www.w3.org/2005/rules/wiki/RIF_Working_Group</a>
[RiSn]	Rich Snippets, <a href="http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html">http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html</a>
[SMW]	Semantic Media Wiki, <a href="http://semantic-mediawiki.org/wiki/Semantic_MediaWiki">http://semantic-mediawiki.org/wiki/Semantic_MediaWiki</a>
[SPARQL]	Prud'hommeaux E., Seaborne A.: SPARQL Query Language for RDF, W3C Candidate Recommendation, <a href="http://www.w3.org/TR/rdf-sparql-query/">http://www.w3.org/TR/rdf-sparql-query/</a> , 2008
[StHo09]	Stonebraker, M. and Hong, J. 2009. Saying good-bye to DBMSs, designing effective interfaces. <i>Commun. ACM</i> 52, 9 (Sep. 2009),
[SWRL]	SWRL: A Semantic Web Rule Language Combining OWL and RuleML, <a href="http://www.w3.org/Submission/SWRL/">http://www.w3.org/Submission/SWRL/</a>
[SWRLB]	SWRLB, <a href="http://www.w3.org/2003/11/swrlb">http://www.w3.org/2003/11/swrlb</a>
[VisWiki]	Visual Wiki, <a href="http://www.viswiki.com/en/">http://www.viswiki.com/en/</a>
[Zema]	Serwis Zemanta, <a href="http://www.zemanta.com/">http://www.zemanta.com/</a>