

Krzysztof CZAJKOWSKI, Tomasz TRELA

Politechnika Krakowska, Instytut Teleinformatyki, Wydział Fizyki, Matematyki i Informatyki

SEMANTIC WEB – STANDARD, NARZĘDZIA, IMPLEMENTACJE

Streszczenie. Semantic Web udostępnia protokoły, które na podstawie standardów rekomendowanych przez W3C w jednoznaczny sposób pozwalają opisać zawartość dokumentów oraz dzielić te informacje pomiędzy różne aplikacje. Artykuł zawiera omówienie standardu, opis istniejących implementacji, przegląd narzędzi oraz wskazuje perspektywy zastosowania technologii semantycznych w przyszłości.

Słowa kluczowe: sieć semantyczna, ontologia, RDF

SEMANTIC WEB – STANDARD, TOOLS, IMPLEMENTATIONS

Summary. Semantic Web provides protocols and standards, which are already W3C recommendations and allow describe and share data among web or any applications. This paper covers overview of current standards, implementations an tools and tries to determine future adoption of the standard.

Keywords: semantic web, ontology, RDF

1. Wstęp

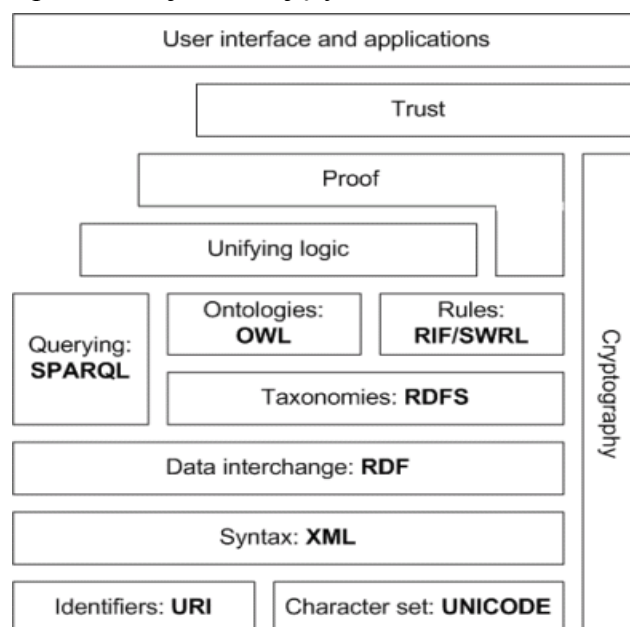
Powstanie sieci semantycznej miało na celu rozszerzenie funkcjonalności dotychczas istniejącej sieci WWW. Zaobserwowano, że treści dostępne w Internecie są niezrozumiałe dla programów komputerowych. Mimo że programy potrafią rozpoznać strukturę stron hipertekstowych i ich podstawowe elementy (nagłówki, łącza do innych stron), nie rozumieją znaczenia odnalezionych danych i pojęć z nimi związanych. Jediną formą przeszukiwania Internetu stało się wyszukiwanie fraz, które okazały się niewystarczająco skuteczne. Zauważono, że możliwości sieci WWW powoli się wyczerpują.

W 2001 r. opublikowany został artykuł Tima Berners-Lee [1], w którym przedstawiona została wizja utworzenia i rozpowszechnienia standardów opisywania treści w Internecie

w sposób, który umożliwi maszynom i programom (np. tzw. agentom) przetwarzanie informacji odpowiednio do ich znaczenia. Semantic Web (Sieć Semantyczna) pozwoli przypisać znaczenie poszczególnym elementom stron WWW, dzięki czemu odczytujące je programy będą mogły wykonywać skomplikowane zadania zlecone przez użytkownika, korzystając z dostępnych ontologii. Główną ideą Semantic Web jest zatem zwiększenie możliwości programów komputerowych w ten sposób, aby przetwarzały i rozumiały dane. Semantic Web zakłada, że programy komputerowe mogą pracować z danymi, które wcześniej nie zostały zdefiniowane. Możemy stworzyć nową ontologię opisującą dane, a inteligentny agent powinien pobrać ontologię i na tej podstawie mieć pełną wiedzę, co te dane reprezentują i co można z nimi zrobić.

2. Podstawy Semantic Web – „Semantic Web Stack”

Wizja inteligentnych agentów rozumiejących informacje zawarte w zasobach elektronicznych wymagała opracowania wielu standardów umożliwiających zdefiniowanie i przetwarzanie metadanych dołączonych do dokumentów elektronicznych. Szczególnie istotne jest to, że metadane mają nie tylko opisywać dokumenty, ale także rzeczywiste koncepty (idee i przedmioty) niemające reprezentacji w formie jakiegokolwiek zasobu. Dlatego język opisujący metadane musiał być niezwykle ekspresyjny oraz elastyczny – musiał pozwalać na definiowanie nowych pojęć na podstawie już istniejących.



Rys. 1. Stos sieci semantycznej

Fig. 1. Semantic Web Stack

Organizacja World Wide Web Consortium pomaga stworzyć zestaw technologii i narzędzi umożliwiający implementację agentów semantycznych. Ich trzon to tzw. stos sieci semantycznej.

tycznej – „Semantic Web Stack”, zwany również „Semantic Web Layer Cake”. Warstwy stosu tworzą hierarchię (rys. 1), która wyróżnia niezbędne elementy aplikacji semantycznej:

- unikalne adresowanie zasobów,
- tworzenie metadanych,
- tworzenie słowników pojęć,
- tworzenie języka opisu pojęć,
- mechanizmy wnioskowania,
- mechanizmy uwierzytelniania.

2.1. Stos semantyczny

U podstaw stosu semantycznego znajdują się standardy pozwalające reprezentować zasoby w jednoznaczny sposób oraz w każdym języku.

2.1.1. Unikod

Unikod jest zestawem znaków pozwalającym reprezentować tekst we wszystkich językach świata. Unikod definiują dwa standardy: „Unicode” oraz ISO 10646. Kodowanie znaków może być zaimplementowane na różne sposoby. Najpopularniejszym jest kodowanie UTF-8, które zachowuje te same znaki kodowe dla znaków ASCII.

2.1.2. URI – Uniform Resource Identifier

URI jest standardem internetowym, który umożliwia łatwą identyfikację zasobów w sieci. Zdefiniowany jest w dokumencie RFC 2396 [2].

W technologiach semantycznych istnieje wiele konwencji tworzenia identyfikatorów dla zasobów, wybrane techniki opisał Sandro Hawke [3]. W rozważaniach nad najlepszą konwencją najlepszym rozwiązaniem okazało się stosowanie URL z fragmentami (częścią URL po #). Rozwiązuje to problem rozróżniania stron internetowych od zasobów, jakie one opisują.

2.1.3. XML i XML Schema

XML (ang. *Extensible Markup Language*) [4] jest językiem służącym do przechowywania danych. XML jest podzbiorem języka SGML. Dokumenty XML przechowują dane w strukturalnej formie z wykorzystaniem znaczników. XML stanowi nie tylko format, w jakim dane są przechowywane, lecz również zbiór reguł opisujących dane. XML został wybrany przez W3C jako podstawa do zaimplementowania RDF z powodu ekspresyjności języka w opisywaniu danych, dużych możliwości rozszerzenia języka oraz licznych narzędzi ułatwiających pracę z XML, m.in. XSLT, XPath, XQuery.

XML Schema [5] jest najpopularniejszym językiem służącym do definiowania gramatyki dokumentu. XML Schema opisuje typ dokumentu XML, w szczególności typy danych, jakie przechowuje.

2.2. RDF

RDF, czyli „Resource Description Framework” [6], jest specjalnym językiem stworzonym do opisu danych, zrozumiałym dla komputera. Bardziej precyzyjnie, język ten ułatwia programom komputerowym przetwarzanie danych. Dzięki RDF można modelować dane odzwierciedlające istniejące pojęcia-koncepty. Model danych w RDF opiera się na tzw. trójkach (ang. *triple*). Trójki te złożone są z podmiotu (ang. *subject*), predykatu (ang. *predicate*) i obiektu (ang. *object*). Można powiedzieć, że podmiot <subject> ma właściwość <predicate> o wartości obiektu <object>.

Dalej przedstawiony przykład prezentuje znaczenie zdania „Politechnika Krakowska jest uczelnią wyższą”. W zdaniu można wyróżnić następujące elementy:

- Politechnika Krakowska – podmiot, identyfikowany przez adres internetowy,
- jest – predykat, w tym przypadku znaczenie słowa „jest” dotyczy określenia „czym jest” podmiot. Taki predykat został już zdefiniowany pod adresem <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>,
- uczelnia wyższa – obiekt, w tym przypadku skorzystano z definicji „uczelni” zamieszczonej pod adresem <http://swrc.ontoware.org/ontology#University>.

Każdy z tych elementów został jednoznacznie określony. Zatem można wskazane zdanie zapisać za pomocą następującej trójki RDF:

```
<http://www.pk.edu.pl#PK>  
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
  <http://swrc.ontoware.org/ontology#University>
```

2.2.1. Model RDF

Model RDF został zaprojektowany, aby spełnić następujące wymagania:

- umożliwić łatwą reprezentację danych,
- udostępnić formalną semantykę i udowodnialne wnioskowanie,
- wykorzystywać bazujące na URI słowniki,
- wykorzystywać składnię XML,
- wspierać typy danych dostępne w XML Schema,
- umożliwić tworzenie twierdzeń na temat wszystkich możliwych zasobów.

RDF opiera się na modelowaniu danych w postaci grafu skierowanego. Wierzchołkami grafu są podmioty i obiekty, predykaty natomiast są krawędziami grafu, ale mogą także występować jako wierzchołki. Kierunek krawędzi jest znaczący. Zbiór trójek RDF nazywamy

grafem RDF. RDF opiera się na logice asercji. Każda trójka reprezentuje pojedyncze twierdzenie. Wszystkie twierdzenia (trójki RDF) znajdujące się w danym grafie są w logicznej koniunkcji (logiczne AND) w obrębie grafu, do którego należą.

Wierzchołki grafu mogą być reprezentowane za pomocą literalów, URI lub mogą być „puste” (ang. *blank nodes*, tzn. nie mają odrębnej formy identyfikacji).

Najważniejszym pojęciem, jakie wprowadza teoria modelu RDF, to pojęcie relacji. Semantyka modelu RDF jest dokładnie opisana w rekomendacji W3C [7].

2.2.2. *RDF Schema*

RDF Schema (RDFS) [8] jest podstawowym rozszerzeniem języka RDF. Wprowadza podstawowe pojęcia pozwalające na strukturalne uporządkowanie grafu zdefiniowanego za pomocą RDF. RDFS znacząco rozszerza możliwości samego RDF. Wprowadza pojęcie klasy (`rdfs:Class`) i jej podtypów (`rdfs:subClassOf`), np. zasób (`rdfs:Resource`), oraz literalu (`rdfs:Literal`). Pozwala określać zakres właściwości.

2.2.3. *Tworzenie słownika pojęć (Vocabulary)*

Za pomocą RDFS tworzy się tzw. słowniki RDF (ang. *vocabularies*). Siłą RDF jest możliwość definiowania własnych pojęć na podstawie istniejących, już znanych, pojęć. Tworzone mogą być zarówno nowe klasy, jak i predykaty (właściwości).

Przykład wykorzystania istniejących pojęć do zdefiniowania nowego konceptu:

```
@prefix gadgets: <http://purl.org/net/vocab/2004/10/gadget>
:Tomofon a rdfs:Class.
:Tomofon rdfs:subClassOf gadgets:Gadget.
:Tomofon rdfs:comment ""Wyspecjalizowany telefon z opcją domofonu""
```

2.2.4. *Przestrzeń nazw*

Pojęcia definiowane w dokumentach RDF tworzą słownik (ang. *vocabulary*). Aby ułatwić posługiwanie się zdefiniowanymi słownikami, wprowadzono przestrzeń nazw (ang. *namespaces*).

Podstawowe przestrzenie nazw to:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

2.3. Notacje RDF

RDF umożliwia zapisanie trójek w wielu notacjach. Rekomendowanym przez W3C formatem dla RDF jest RDF/XML, a dla danych nieserializowalnych do XML – N3 (Notation 3). Format N3, opisywany przez Tim Berners-Lee [9], został zaprojektowany z myślą o czytelności dla użytkowników. Zapis w notacji N3 jest znacznie krótszy niż w formacie

XML. N3 jest bardzo wygodną notacją i na przykład, aby zapisać, że Politechnika Krakowska jest uczelnią, wystarczy:

```
@prefix swrc: <http://swrc.ontoware.org/ontology>.  
<http://www.pk.edu.pl/#PK> a swrc:University.
```

RDF/XML [10] jest zalecanym językiem do prezentacji modelu RDF, serializowanym za pomocą XML. Składnia języka jest zdefiniowana w rekomendacji W3C, która jest normatywną specyfikacją RDF/XML.

Aby reprezentować graf w języku XML, podmioty, predykaty i obiekty należy przedstawić w pojęciach XML, czyli odpowiednio jako elementy, atrybuty oraz zawartości elementów. XML limituje wyrażanie pewnych grafów RDF z powodu ograniczeń kodowania Unicode. Niektóre URI określające predykaty są niedozwolone w unikodzie. Te restrykcje nie dotyczą notacji N3 i Turtle.

2.3.1. Inne notacje RDF

Warto wspomnieć także o innych notacjach RDF:

- Notacja N-Triples [11] została stworzona z przeznaczeniem dla skryptów procesujących. Jest podzbiorem N3. Składnia tej notacji jest prostsza, ale ma mniejszą funkcjonalność. Opiera się o linie – każda linia to jedno wyrażenie.
- Turtle [12] (ang. *Terse RDF Triple Language*), czyli „skondensowany” opis trójek RDF, jest podzbiorem notacji N3 i nadzbiorem notacji N-Triples. Wszystkie wyrażenia zapisane w notacji Turtle powinny być prawidłowe w języku zapytań SPARQL.
- TriG jest formatem zapisu nazwanych grafów. Bazuje na Turtle, ale poszerza go o grupowanie trójek w nazwane grafy.
- TriX [13] jest eksperymentalną alternatywą zapisu trójek RDF w języku XML.
- RDFa jest standardem rozszerzającym zbiór atrybutów XHTML, umożliwiając umieszczanie metadanych bezpośrednio wewnątrz dokumentów w sieci.

2.4. Ontologie

W informatyce „ontologia” jest formalną reprezentacją wiedzy. Ontologie są wykorzystywane w wielu dziedzinach informatycznych. Najkrótszą definicję ontologii zaprezentował Tom Grubner [14] w 1992 r.: „An ontology is a specification of a conceptualization.”

Jest to najogólniejsza definicja, którą można przetłumaczyć jako formalny opis (specyfikację) abstrakcyjnych pojęć-konceptów oraz relacji pomiędzy tymi konceptami. Ontologie są tworzone po to, aby umożliwić formalny zapis wiedzy dotyczącej wybranej dziedziny. Do tworzenia ontologii można posłużyć się językami DAML+OIL lub omawianym w następnych podrozdziałach OWL 2. Jednak zanim zostaną one omówione, warto zaznaczyć, jakie zależ-

ności istnieją pomiędzy dotychczas wymienionymi standardami. Abstrakcyjny model danych RDF może być reprezentowany w wielu formatach, języki RDF Schema, DAML+OIL, OWL 2 ułatwiają natomiast modelowanie zaawansowanych zależności wiedzy dziedzinowej i tworzenie ontologii.

2.5. OWL 2

Pierwsza wersja OWL [15] została zaprojektowana dla aplikacji wykorzystywanych do procesowania danych, a nie tylko do ich prezentacji. OWL zwiększa możliwości opisu w stosunku do standardów XML, XML Schema, RDF oraz RDF Schema. OWL ułatwia opisywanie właściwości, relacji oraz typów.

Obecnie obowiązuje wersja OWL 2 stanowiąca rozwinięcie OWL, a wszystkie poprawne ontologie OWL wersji 1 mogą być traktowane jako prawidłowe ontologie OWL 2 (zachowano zgodność wstecz). Do różnic pomiędzy OWL 2 w stosunku do OWL można zaliczyć [27]:

- opracowanie nowej składni OWL 2 Manchester Syntax,
- wprowadzenie nowych konstrukcji mających zwiększyć siłę wyrazu języka,
- rozszerzenie liczby akceptowalnych typów danych,
- rozszerzenie możliwości adnotacji,
- zdefiniowanie dotychczas istniejących aksjomatów w celach ułatwienia ich zrozumienia, wydajniejszej implementacji oraz prostszego tworzenia ontologii.

Za szczególnie istotne w OWL 2 można uznać wprowadzenie trzech profili [40], określanych zwykle jako podjęzyki (podzbiory syntaktyczne), wykorzystywanych w różnych zastosowaniach. Profile te to:

- OWL 2 EL – znajdujący zastosowanie w aplikacjach pracujących na ontologiach z dużą liczbą własności i/lub klas.
- OWL 2 QL – ukierunkowany na aplikacje wykorzystujące wielkie liczby instancji danych i gdzie odpowiedź na zapytanie jest najważniejszym zadaniem wnioskowania. W OWL 2 QL łączone zapytania mogą być zaimplementowane z wykorzystaniem tradycyjnych relacyjnych baz danych.
- OWL 2 RL – ukierunkowany na aplikacje, które wymagają skalowalnego wnioskowania. Systemy wnioskujące z OWL 2 RL mogą być zaimplementowane przy użyciu silników regułowych.

2.6. Podział i taksonomia ontologii

Ontologie można klasyfikować według różnych kryteriów, z których najtrafniejszym wydaje się podział według ich zastosowania, podany przez Jos de Bruijn [16]:

- Ontologie najwyższego poziomu opisujące ogólne koncepty, niezależne od jakiejkolwiek dziedziny. Jednakże takie ontologie trudno jest utworzyć, ponieważ zawsze wkradają się elementy zależne od domeny zainteresowania.
- Ontologie dziedzinowe i zadaniowe opisujące ogólne koncepty dla poszczególnych dziedzin i zadań. Przykładem takiej ontologii może być ontologia medyczna lub bardziej szczególna, np. dermatologiczna.
- Ontologie aplikacyjne opisujące koncepty w danej aplikacji, zależne od innych ontologii domenowych.

Z powodu rozpowszechnienia się ontologii razem z wykorzystaniem metadanych istniało zapotrzebowanie na unifikację ontologii. Powstała inicjatywa „Dublin Core”, która doprowadziła do stworzenia podstawowego zbioru właściwości [17] ratyfikowanych jako standardy IEFT [18], ANSI/NISO [19] oraz ISO [20], rozpowszechnianych pod nazwą Dublin Core Metadata Element Set (DCMES) (w skrócie „Dublin Core”). Dublin Core definiuje następujące właściwości: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights. Powstały także inne ontologie opisujące szczegółowo wybrane dziedziny, takie jak biologia czy nawet bibliotekarstwo, jednak Dublin Core jest najbardziej znaczącą i popularną ontologią, na podstawie której budowane są kolejne rozszerzenia.

2.7. Mechanizmy wnioskowania

Mechanizmy wnioskowania są częścią Semantic Web, która wciąż jest w trakcie tworzenia. Obecnie istnieją pewne założenia, które mechanizmy te muszą spełniać. Niezwykle przydatnymi narzędziami byłyby mechanizmy pozwalające systemom komputerowym nie tylko rozpoznawać relacje pomiędzy konceptami, ale przede wszystkim wnioskować inne relacje wynikające z istniejących. Prostą zasadą wnioskowania jest zasada wnioskowania inferencyjnego. Polega ona na wnioskowaniu, że jeśli:

$$A \Rightarrow B \text{ i } B \Rightarrow C, \text{ to } A \Rightarrow C.$$

Istnieje kilka problemów związanych z wnioskowaniem opartym na ontologiach. Przykładowo ontologie odwołujące się do innych ontologii są od nich zależne, co ogranicza ich własną niezależność. Jeff Pan [24] proponuje usprawnienie (za pomocą algorytmów tablicowych) wnioskowania na logice opisowej, odpowiadającej konstrukcjom z OWL, przez wprowadzenie dwóch mechanizmów wnioskowania, pierwszego: tzw. subiektywnego, opartego na lokalnej interpretacji ontologii zapewniającej jej autonomiczność, oraz drugiego, obcego (ang. *foreign*), opartego na semantyce odwołującej się do innych ontologii.

Innym zagadnieniem związanym z wnioskowaniem jest Ontology Alignment. Może pojawić się sytuacja, że jeden koncept zostanie zdefiniowany w wielu niezależnych ontologiach.

Inaczej mówiąc, w kilku ontologiach będą znajdowały się równorzędne pojęcia, jednakże inaczej zdefiniowane. Problem ten nazwany jest problemem ewolucyjności (ang. *evolvability problem*). Jednak można go rozwiązać, dodając regułę stwierdzającą równorzędność danych pojęć.

2.8. SWRL i RIF

Mimo że język OWL oferuje bogaty zestaw konstrukcji do opisu relacji pomiędzy obiektami RDF, to jednak nie ma on możliwości, aby zobrazować relacje (mapping) pomiędzy różnymi ontologiami. Zauważono, że takie relacje lepiej jest reprezentować za pomocą języków opartych na regułach. SWRL [25] czy RIF [26] są językami uzupełniającymi zbiór aksjomatów OWL o zestaw rozszerzalnych reguł. Reguły te formułuje się w ten sposób, że implikuje się prawdziwość następnika, jeśli poprzednik jest prawdziwy. Przykładem takiej reguły może być implikacja, że prawdziwość zdań `hasParent` oraz `hasBrother` implikuje prawdziwość `isUncle`:

```
hasParent(?x1,?x2) ∧ hasBrother(?x2,?x3) → hasUncle(?x1,?x3)
```

Abstrakcyjna składnia tych reguł może być zapisana w następujący sposób:

```
Implies(Antecedent(hasParent(I-variable(x1) I-variable(x2))
                    hasBrother(I-variable(x2) I-variable(x3)))
        Consequent(hasUncle(I-variable(x1) I-variable(x3))))
```

2.9. Mechanizmy certyfikacji i zaufania

Model zabezpieczeń w głównej mierze opiera się na istniejących standardach, przede wszystkim infrastrukturze klucza publicznego i certyfikacji. Dokumenty RDF mogą być podpisywane przez autorów, dzięki temu aplikacje mogą uwierzytelniać posiadane informacje. Obecnie trwają prace nad bardziej zaawansowanymi technikami zabezpieczeń i integracji ich w językach definiowania ontologii.

2.10. Wybrane zastosowania

Zastosowanie technologii semantycznych jest bardzo szerokie. Interesujące zestawienie tych zastosowań przedstawione zostało przez Jarosława Bąka i Czesława Jędrzejka [27], którzy jako jedne z najciekawszych wskazują wyszukiwarki oraz bazy wiedzy. Przykładem wyszukiwarki opartej na technologii semantycznej jest PowerSet, umożliwiająca zadawanie pytań w języku naturalnym. Wyszukiwarka PowerSet została zintegrowana w wyszukiwarce Bing [28] firmy Microsoft. Przykładem bazy wiedzy jest DBPedia, opisana w dalszej części tego artykułu lub Freebase [29] – otwarta baza danych, gromadząca informacje o osobach,

miejscach czy rzeczach. Przechowuje ona informacje w grafie, tworząc relacje między wszystkimi encjami.

3. Języki i narzędzia

3.1. SPARQL

RDF umożliwia modelowanie relacji pomiędzy zasobami. Jednakże aby w pełni korzystać z możliwości RDF, niezbędne było stworzenie narzędzi pozwalających w efektywny sposób wydobywać dane z grafów RDF. SPARQL jest językiem, dzięki któremu można tworzyć zapytania RDF. Język ten jest zdefiniowany przez rekomendację W3C [30], która jest jego normatywną specyfikacją.

Składnia SPARQL przypomina składnię notacji Turtle, poszerzoną o słowa kluczowe znane z SQL, takie jak „SELECT”, „WHERE”. Zapytania w SPARQL tworzymy, konstruując relacje w notacji trójek Turtle. Wynikami zapytania są wszystkie trójki, które spełniają warunek, że relacje w zapytaniu są prawdziwe.

Przykładowym zapytaniem SPARQL, wyszukującym w tym grafie zadaną trójkę, może być zapytanie o tytuły prac magisterskich studentów danego wydziału Politechniki Krakowskiej:

```
PREFIX dc: <http://purl.org/dc/elements/1.1>
PREFIX swrc: <http://swrc.ontoware.org/ontology/>
PREFIX pk: <pk.edu.pl>

SELECT ?name ?title
WHERE
  { ?x swrc:studiesAt pk:FMI.
    ?x swrc:MasterThesis ?y.
    ?y dc:title ?title . }
```

3.2. Narzędzia semantyczne

Obecnie istnieją już narzędzia i biblioteki wykorzystujące technologie semantyczne w bardzo szerokim zakresie. Narzędzia programistyczne takie jak RDF Triple Store (bazy danych RDF) wspierające języki zapytań takie jak SPARQL są dostępne dla większości popularnych platform programistycznych. Open source’owe biblioteki udostępniające algorytmy wnioskujące, wyrównujące dane (ang. *data alignment*) również są ogólnie dostępne. W tym podrozdziale przedstawiono kilka wybranych narzędzi, które prezentują możliwości aplikacji semantycznych. Bardzo obszerna lista narzędzi oraz bibliotek umieszczona jest na stronie W3C poświęconej sieciom semantycznym [31].

3.2.1. Swoogle [32] – przeszukiwarka ontologii

Obecnie każdy może publikować własne ontologie. Problem, który może się pojawić, to wielokrotne definiowanie tych samych pojęć i niewykorzystywanie istniejących ontologii. Taka praktyka może spowodować, że procesowanie danych będzie utrudnione, ponieważ programy będą musiały obsługiwać ogromną ilość danych. Pomocnym narzędziem jest Swoogle. Jest to wyszukiwarka ontologii. Dzięki Swoogle można sprawdzić, czy szukane pojęcie jest już zdefiniowane w istniejących ontologiach.

3.2.2. OpenLink Data Explorer [33]

OpenLink Data Explorer jest rozszerzeniem dla przeglądarek internetowych: Mozilla Firefox, Safari oraz Google Chrome, pozwalającym na wyświetlenie „surowych” danych semantycznych zamieszczonych na stronie. Użytkownik może „przełączać” się pomiędzy danymi hipertekstowymi i danymi w postaci RDF. Jest to niezwykle użyteczne przy przeglądaniu sieci danych, ponieważ z łatwością można odnaleźć źródła informacji wyświetlanych na stronach.

3.2.3. OwlSight [34]

Bardzo przydatnym narzędziem jest przeglądarka ontologii. Przykładową przeglądarką jest OwlSight. Pozwala ona wczytać ontologię z podanego adresu URI i przeglądać jej strukturę, hierarchię klas i właściwości. Przeglądarka ta jest dostępna w wersji webowej, dlatego umożliwia przeglądanie ontologii bez potrzeby instalowania jakichkolwiek narzędzi.

3.2.4. DBpedia [35]

DBpedia jest open source’owym projektem mającym na celu udostępnienie bazy wiedzy z Wikipedii poprzez RDF. DBpedia umożliwia tworzenie semantycznych zapytań, dzięki czemu można oczekiwać bardzo precyzyjnych rezultatów. Wyniki wyszukiwania semantycznego z założenia są prawidłowe – rezultat zawsze spełnia warunek, że po podstawieniu do szukanego zapytania SPARQL trójki są prawdziwe. Dokładność można zwiększać, podając szczegółowe informacje, tj. dodając kolejne trójki RDF.

3.2.5. GoNTogle [36]

GoNTogle jest narzędziem do semantycznego oznaczania i wyszukiwania dokumentów, zbudowanym na bazie biblioteki Lucene i systemu Protege. Wspiera manualne i automatyczne tagowanie wybranych typów dokumentów (doc, pdf, rtf, txt, odt), korzystając z konceptów RDF.

3.2.6. *KiWi* [37]

KiWi (ang. *Knowledge In a Wiki*) jest przykładem zaawansowanego systemu Wiki opartego na technologii semantycznej. KiWi – semantyczne Wiki – przechowuje informacje w postaci RDF, umożliwiając dostęp do nich nie tylko z poziomu systemu Wiki, ale także z poziomu API RDF. Projekt jest fundowany przez Unię Europejską z programu European Community's Seventh Framework Programme (FP7/2007-2013).

4. Podsumowanie

Wyniki prowadzonych badań nad technologiami semantycznymi pozwalają sformułować wniosek, że technologie te, pomimo iż są już stosunkowo silnie ukształtowane i mają rozbudowane zaplecze, tj. specyfikacje, standardy, biblioteki oraz narzędzia zaimplementowane w większości popularnych językach programowania, to jednak nie są jeszcze powszechnie używane, co nie pozwala w pełni wykorzystać ich możliwości.

Obecnie duży wkład w tej dziedzinie odnotowują badania nad aplikacjami świadomymi kontekstu, wyszukiwarkami semantycznymi [41], a także narzędziami do zarządzania wiedzą [42]. Rosnąca popularność sieci semantycznych będzie powodować, iż sieć danych będzie bogatsza, a zatem możliwości sieci także będą większe.

Problemem, z jakim muszą zmierzyć się twórcy oprogramowania wykorzystującego sieci semantyczne, jest skalowalność tych sieci. Jak wskazują Bąk i Jędrzejek, problem ten może zostać rozwiązany w projekcie LarKC [38]. Jest to projekt fundowany z 7 Ramowego Programu Unii Europejskiej, mający na celu stworzenie platformy do masowego, rozproszonego, niepełnego wnioskowania, aby usunąć ograniczenia obecnych systemów wnioskowania dostępnych dla sieci semantycznych.

Warto wspomnieć o diagramie prezentującym tzw. chmurę LOD [39] (The Linking Open Data diagram). Wizualnie prezentuje ona połączenia pomiędzy istniejącymi zbiorami danych dostępnymi w sieci. Od czasu jej opublikowania w 2007 r. systematycznie rośnie zbiór otwartych baz danych. Obecnie zbiór ten zawiera około 300 elementów.

Technologie semantyczne wymagają jeszcze wielu lat badań, aby stały się dojrzałym rozwiązaniem, na którym będą budowane programy zastępujące użytkownika w wykonywaniu różnych czynności. Jednakże ciągle rosnące zainteresowanie sieciami semantycznymi, a w szczególności bardzo szerokie wsparcie Unii Europejskiej dla projektów dotyczących badań nad sieciami semantycznymi, pozwala przewidywać, że technologie semantyczne znajdą w przyszłości o wiele szersze zastosowania, niż ma to miejsce obecnie.

BIBLIOGRAFIA

1. 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.
2. Dokument RFC Uniform Resource Identifier, <http://tools.ietf.org/html/rfc2396>.
3. Hawke S.: How We Identify Things? <http://www.w3.org/2001/03/identification-problem/>, 2001.
4. Rekomendacja W3C, Extensible markup language, <http://www.w3.org/TR/2008/REC-xml-20081126/>, 2008.
5. Rekomendacja W3C, XML Schema Part 1: Structures Second Edition, <http://www.w3.org/TR/xmlschema-0/>, 2004.
6. Rekomendacja W3C, Resource Description Framework (RDF): Concepts and Abstract Syntax, <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>.
7. Rekomendacja W3C RDF Semantics, <http://www.w3.org/TR/rdf-mt/>.
8. Rekomendacja W3C dla języka RDF Schema, <http://www.w3.org/TR/rdf-schema>.
9. Berners-Lee T.: An RDF language for the Semantic Web, <http://www.w3.org/DesignIssues/Notation3.html>.
10. Rekomendacja W3C, RDF/XML Syntax, <http://www.w3.org/TR/rdf-syntax-grammar/>.
11. Rekomendacja W3C, RDF Test Cases, N-Tripples, <http://www.w3.org/TR/rdf-testcases/#ntriples>.
12. W3C Team Submission, Turtle – Terse RDF Triple Language, <http://www.w3.org/TeamSubmission/turtle/>, 28 March 2011.
13. Carroll J. J., Stickler P.: TriX : RDF Triples in XML. 2004.
14. Gruber T.: A translation approach to portable ontologies. Knowledge Acquisition, 1993.
15. Rekomendacja W3, OWL Web Ontology Language Overview, <http://www.w3.org/TR/owl-features/>.
16. De Bruijn J.: Using Ontologies. Technical Report DERI-2003-10-29.
17. Dublin Core Metadata Element Set, <http://dublincore.org/documents/dces/>.
18. IETF RFC 5013 of August 2007, <http://www.ietf.org/rfc/rfc5013.txt>.
19. The Dublin Core Metadata Element Set, ANSI/NISO Standard Z39.85-2007, 2007.
20. ISO Standard 15836-2003, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=52142.
21. Sauermann L., Van Elst L., Möller K.: Personal Information Model (PIMO).
22. Nepomuk Annotation Ontology, <http://www.semanticdesktop.org/ontologies/nao/>.

23. Ontologie NIE, <http://www.semanticdesktop.org/ontologies>.
24. Pan J. Z.: Description Logics: reasoning support for Semantic Web. PhD thesis of Manchester University, 2004.
25. Semantic Web Rule Language, <http://www.w3.org/Submission/SWRL/>.
26. Rekomendacja W3C, The Rule Interchange Format, <http://www.w3.org/TR/2010/REC-rif-core-20100622/>.
27. Bąk J., Jędrzejek C.: Semantic Web – technologie, zastosowania, rozwój. XV Konferencja PLOUG, 2009.
28. Wyszukiwarka Bing, www.bing.com.
29. Strona bazy danych Freebase, <http://www.freebase.com/>.
30. Rekomendacja W3C, SPARQL, <http://www.w3.org/TR/rdf-sparql-query/>.
31. Zestawienie narzędzi semantycznych, <http://www.w3.org/2001/sw/wiki/Tools>.
32. Strona narzędzia Swoogle, <http://swoogle.umbc.edu/>.
33. Strona projektu Open Link Data Explorer, <http://ode.openlinksw.com/>.
34. Strona projektu OWLBrowser, <http://pellet.owldl.com/ontology-browser/>.
35. Strona informacyjna projektu DBPedia, <http://dbpedia.org/About>.
36. <http://web.imis.athena-innovation.gr/projects/gontogle/>.
37. Strona projektu KIWI, <http://www.kiwi-project.eu/>.
38. Strona projektu LarKC, <http://www.larkc.eu/>.
39. Strona diagramu LOD, <http://richard.cyganiak.de/2007/10/lod/>.
40. Omówienie profilu OWL 2, <http://www.w3.org/TR/owl2-profiles/>.
41. Skulimowski M.: Analizator Internetu Semantycznego. Studia Informatica, Vol. 32, No. 2A(96), Wydawnictwo Politechniki Śląskiej, Gliwice 2011.
42. Goczyła K., Piotrowski P.: Application of Knowledge Views. Studia Informatica, Vol. 31, No. 2A(89), Wydawnictwo Politechniki Śląskiej, Gliwice 2010.

Wpłynęło do Redakcji 17 stycznia 2012 r.

Abstract

The emergence of semantic web was to extend the functionality of already existing World Wide Web. It was observed that the content available on the Internet is incomprehensible to computer programs. In 2001, an article was published by Tim Berners-Lee which presented a vision for the creation and spread of standards for describing content on the Internet in a way that allows machines and programs (called agents) processing of information according

to their importance. Since 2001, the Semantic Web has become a mature technology that provides protocols based on standards recommended by the W3C and allows describing clearly the content of these documents and share information between different applications. Semantic Web Stack is a set of basic standards allowing representation of information in semantic networks. It consists of well-known technologies such as URLs and XML, but also defines the RDF, the language to describe resources. The RDF model is based on the so-called “triples”: three-element statements with which it is possible to present any concept. With the rapid development of these standards a number of semantic web applications, for example semantic search engine and knowledge base, have been found. One of the very important areas of semantic web is the research on ontologies. For this purpose, a number of languages emerged that allow defining ontologies such as OWL and tools to view and work with ontologies. Present state of the art in the field of semantic web technology is mature and allows predicting major popularity increase in the future.

Adresy

Krzysztof CZAJKOWSKI: Politechnika Krakowska, Wydział Fizyki, Matematyki i Informatyki, Instytut Teleinformatyki, ul. Warszawska 24, 31-155 Kraków, Polska, kc@pk.edu.pl.

Tomasz TRELA: Politechnika Krakowska, Wydział Fizyki, Matematyki i Informatyki, Instytut Teleinformatyki, ul. Warszawska 24, 31-155 Kraków, Polska, tomas.trela@gmail.com.