

Michał ŚWIĄTKIEWICZ
Politechnika Śląska, Instytut Informatyki

AKADEMICKA BAZA INFORMACJI JAKO PRZYKŁAD WYKORZYSTANIA TECHNOLOGII SIECI SEMANTYCZNEJ

Streszczenie: Współczesna sieć WWW została stworzona w taki sposób, że jest dostępna praktycznie wyłącznie dla człowieka. Rozbudowanie jej, zgodnie z koncepcją Sieci Semantycznej, umożliwi przekształcenie Internetu w rozproszoną globalną bazę danych, opisanych ontologicznie i możliwych do zautomatyzowanego przetwarzania programowego. Artykuł przedstawia sposób stworzenia takiej bazy informacji dla środowiska akademickiego.

Słowa kluczowe: sieć semantyczna, ontologie, RDF

ACADEMIC INFORMATION BASE AS AN APPLICATION OF SEMANTIC WEB TECHNOLOGIES

Summary: WWW we know today is full of information accessible only for human readers. Changing WWW according to Semantic Web vision will allow us to treat Internet as an machine-processable distributed database of ontologically described information. The article presents some way of creating such database for academic community.

Keywords: semantic web, ontology, RDF

1. Wprowadzenie

W czasach zaawansowanego rozwoju technologii internetowych sieć WWW stanowi jedno z najważniejszych źródeł informacji i dróg komunikacji międzyludzkiej. W środowiskach rozproszonych terytorialnie, lecz powiązanych wspólnymi zainteresowaniami Internet stanowi podstawę tworzenia wspólnej bazy wiedzy i forum wymiany informacji, a także odgrywa znaczną rolę socjologiczną, ułatwiając kontakty osób zaangażowanych w podobne przedsięwzięcia. Przykładowo, w środowisku akademickim za pośrednictwem Internetu znacznie

zwiększa się możliwość odnalezienia informacji na temat prowadzonych w danej dziedzinie badań, istniejących publikacji, zaangażowanych ośrodków naukowych, zainteresowanych osób lub planowanych konferencji tematycznych. Wszystkie tego typu informacje publikowane na stronach uczelni czy ośrodków naukowych, na stronach prywatnych pracowników naukowych, a także na stronach konferencji lub wydawnictw, umożliwiają zorientowanie się w aktualnym stanie zaawansowania badań i kierunkach rozwoju wybranej dziedziny.

Niestety, odnalezienie i powiązanie tych informacji wymaga obecnie bardzo dużego nakładu pracy i zaangażowania. Co więcej, nawet jeżeli pewne informacje zostały opublikowane, nie ma pewności, że zostaną odnalezione albo prawidłowo skojarzone z innymi. Sytuacja taka wynika z aktualnej kondycji sieci WWW – informacje są publikowane na najróżniejszych stronach, w najróżniejszej formie, przy zastosowaniu wielu rozmaitych formatów i technik prezentacji, a czasem nawet odmiennego słownictwa do opisu tej samej dziedziny pojęciowej. Z drugiej jednak strony różnorodność Internetu, mnogość źródeł informacji i podmiotów, publikujących dane stanowią jego największą siłę. Stosowanie rozwiązań scentralizowanych, np. wprowadzenie jednej centralnej bazy danych do opisu działań wszelkich ośrodków naukowych, byłoby zaprzeczeniem tej idei, a ponadto rozwiązanie takie byłoby trudne w utrzymaniu i jako takie nie miałyby większych szans powodzenia. Rozwiązaniem tego typu problemów może być zastosowanie technologii rozwijanych w ramach koncepcji Sieci Semantycznej.

Idea Sieci Semantycznej (ang. *Semantic Web*)¹ stanowi rozwinięcie aktualnej koncepcji Internetu, w szczególności sieci WWW, w taki sposób, aby publikowane informacje były możliwe do automatycznego przetwarzania, aby ich analiza i integracja nie wymagały żmudnej pracy człowieka, lecz mogły być zlecone agentom programowym. Dzięki wprowadzeniu technologii Sieci Semantycznej Internet można potraktować z jednej strony jako rozproszoną globalną bazę danych, w której publikowane informacje są opisane semantycznie i przez wykorzystanie ontologii stanowią bazę wiedzy z najróżniejszych dziedzin, a z drugiej jako platformę wymiany kontaktów i informacji między jednostkami zainteresowanymi daną dziedziną problemową. Sieć Semantyczna jest koncepcją bardzo szeroką, prace nad jej urzeczywistnieniem trwają od kilku lat i nadal bardzo wiele problemów, związanych z jej funkcjonowaniem pozostaje nie rozwiązanych. Jednak mimo to istnieje już wystarczająco dużo standardowych rozwiązań, aby tworzyć praktyczne zastosowania.

¹ Rozwój koncepcji Semantic Web koncentruje się wokół W3C Semantic Web Activity, <http://www.w3.org/2001/sw/>.

2. Technologie Sieci Semantycznej

Spojrzenie na Sieć Semantyczną jak na globalną, rozproszoną bazę wiedzy wymaga sprawnego funkcjonowania trzech podstawowych mechanizmów:

- reprezentacji danych w sposób umożliwiający opisanie różnorodnych typów obiektów i relacji między nimi,
- reprezentacji wiedzy z możliwością wyciągania logicznych konsekwencji na podstawie przedstawionych faktów,
- łączenia i wyszukiwania informacji z wielu rozproszonych źródeł.

Dzięki pracom standaryzacyjnym, prowadzonym w ramach konsorcjum W3C², stopień realizacji idei *Semantic Web* pozwala już na spełnienie powyższych wymagań. Aktualnie obowiązującymi standardami w zakresie reprezentacji danych, wiedzy i mechanizmów wyszukiwania są specyfikacje języków RDF, OWL i SPARQL.

RDF (*Resource Description Framework*) [1] umożliwia przedstawienie informacji o dowolnych zasobach i relacjach między nimi. Mimo że RDF powstał na potrzeby reprezentacji informacji w Sieci, występujące w nim pojęcie zasobu nie jest ograniczone jedynie do dokumentów dostępnych w Internecie, ale równie dobrze może dotyczyć obiektów świata rzeczywistego, a także dowolnych zjawisk lub pojęć abstrakcyjnych. Model RDF składa się z wyrażeń trójkowych w postaci {podmiot, predykat, obiekt}, gdzie podmiot stanowi opisywany zasób, obiekt opisuje inny zasób lub wartość stałą, natomiast predykat stanowi relację między zasobami lub wyznacza opisywaną cechę podmiotu. Można na ten model spojrzeć jak na zbiór twierdzeń na temat zasobów, albo jak na graf, którego węzłami są podmioty i obiekty, a predykaty stanowią krawędzie skierowane od węzła podmiotu do węzła obiektu. Jako identyfikatory zasobów oraz słowniki predykatów w wyrażeniach modelu RDF stosuje się URI (*Uniform Resource Identifier*) [2], a do opisu typów danych specyfikację *XML Schema* [3], co powoduje, że RDF jest mocno osadzony we współczesnych technologiach, wykorzystywanych w Internecie. Mimo że podstawową formą modelu RDF jest graf, najczęściej przedstawia się go w formie zserializowanej do postaci XML, według specyfikacji RDF/XML [4]. W takiej też postaci będą przedstawiane przykłady w dalszej części.

Podstawowe możliwości opisu zasobów, oferowane przez RDF, są jednak zbyt ograniczone dla praktycznych zastosowań w systemach reprezentacji wiedzy. Dlatego język ten został dalej rozwinięty w specyfikacji *RDF Vocabulary Description Language* (zwanej też *RDF Schema*) [5], która z kolei stanowi podstawę do specyfikacji języka OWL (*Web Ontology Language*) [6]. OWL umożliwia nie tylko wyrażenie informacji o dowolnych zasobach, ich klasach i hierarchiach, lecz w praktyce pozwala również na pełne opisanie zaawansowa-

² World Wide Web Consortium, <http://www.w3.org/>.

nych ontologii z dowolnej dziedziny pojęciowej. Dzięki temu język OWL jest podstawowym sposobem opisu pojęć w bazie wiedzy Sieci Semantycznej. Mimo że Sieć Semantyczna wciąż stanowi ideę nie zrealizowaną w całości, to w Internecie dostępnych jest już dużo danych w formacie RDF i dużo najróżniejszych ontologii,³ opisanych w języku OWL lub jego poprzednikach (takich jak np. język DAML+OIL).

Wyszukiwanie informacji w modelach danych RDF jest możliwe za pomocą języków zapytań, podobnych składniowo do języka SQL. Istnieje kilka różnych języków tego typu, opracowanych przez dostawców narzędzi przetwarzania modeli RDF, jednak aktualne prace W3C wprowadzają w ich miejsce zestaw oficjalnych specyfikacji SPARQL (*SPARQL Protocol And RDF Query Language*). Specyfikacja ta określa zarówno standard języka zapytań dla danych RDF [7], jak i protokół komunikacji z serwerem, przetwarzającym takie zapytania [8] oraz format wyników, zwracanych w postaci XML [9].

Dzięki połączeniu możliwości, oferowanych przez technologie RDF, OWL i SPARQL przy dzisiejszym stanie zaawansowania realizacji idei Sieci Semantycznej możliwe jest stworzenie i sprawne przeszukiwanie rozproszonej bazy informacji, opisanej ontologicznie.

3. Tworzenie Akademickiej Bazy Informacji

Wśród danych, które wydają się istotne z punktu widzenia platformy wymiany wiedzy, informacji i kontaktów w środowisku akademickim, znajdują się dane, należące do trzech podstawowych kategorii:

- dane o osobach – najczęściej pracowników naukowych, zawierające oprócz danych osobowych i kontaktowych również opis dorobku naukowego, informacje o wykładach i uczestnictwie w konferencjach,
- dane o publikacjach – najczęściej uczelnianych, takich jak: prace naukowe, zeszyty naukowe, a także prace niepublikowane – magisterskie lub doktoraty,
- dane o wydarzeniach – czyli informacje o konferencjach, sympozjach itp.

Każda z tych kategorii informacji jest publikowana przez inne jednostki i najczęściej w innym miejscu, niezależnie od pozostałych.

RDF umożliwia wykorzystywanie w jednym modelu wielu niezależnych słowników lub ontologii. Ustalenie jednej ontologii, na potrzeby opisu wspólnej bazy informacji, może w pewnych sytuacjach ułatwić komunikację między użytkownikami takiej bazy, chociaż nie jest formalnie wymagane. Do dalszych rozważań zostanie przede wszystkim wykorzystana

³ Do jednych z liczniejszych wykazów ontologii, dostępnych w Internecie należy serwis SchemaWeb, <http://www.schemaweb.info/>.

ontologia SWRC (*Semantic Web for Research Communities*),⁴ która została opracowana na potrzeby opisu środowisk badawczych. Zawiera ona podstawowe klasy pojęć, takich jak: osoba, organizacja, projekt, publikacja, wydarzenie i wiele pojęć pochodnych. Konstrukcja tej ontologii z założenia ułatwia jej rozbudowę i dostosowywanie do specyficznych potrzeb konkretnych zastosowań. Aktualnie bazują na niej rozwiązania różnych portali semantycznych, dotyczących działalności społeczności lub instytucji naukowych. Opis tych rozwiązań można znaleźć w [10].

3.1. Dane o osobach

Dane o osobach, które leżą w zakresie zainteresowania Akademickiej Bazy Informacji, to oprócz danych teleadresowych, publikowanych przez uczelnie czy jednostki, zatrudniające daną osobę, także informacje pochodzące ze stron prywatnych członków społeczności naukowej. Na takich stronach pracownicy naukowcy często publikują informacje o swoim dorobku naukowym, wykładach lub uczestnictwie w konferencjach (byłych i przyszłych). Udostępnienie tych danych w formacie RDF umożliwia włączenie ich w skład bazy informacji Sieci Semantycznej. Najłatwiej można to uzyskać umieszczając na serwerze, oprócz strony przeznaczonej dla człowieka (w postaci plików HTML), również informacje RDF w postaci pliku, zgodnego z formatem RDF/XML. Taki dodatkowy plik może być wskazany na stronie podstawowej, za pomocą znacznika `<link rel="meta">`. Przykładową zawartość pliku RDF/XML z informacjami na temat osoby, przedstawiono poniżej. Przy wykorzystaniu ontologii SWRC zapisano w nim podstawowe dane osobowe oraz wskazano położenie zasobów, opisujących trzy publikacje oraz uczestnictwo w projekcie badawczym.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:swrc="http://swrc.ontoware.org/ontology#"
  xmlns="http://example.org/abi/osoba#">

  <swrc:PhDStudent rdf:about="http://example.org/abi/osoba#MK">
    <swrc:name>Mikołaj Kowalski</swrc:name>
    <swrc:firstName>Mikołaj</swrc:firstName>
    <swrc:lastName>Kowalski</swrc:lastName>
    <swrc:email>m.kowalski@student.uni.edu.pl</swrc:email>
    <swrc:phone>+48699112233</swrc:phone>
    <swrc:photo>http://kowalski.name/mikolaj/pictures/ja.jpg</swrc:photo>
    <swrc:homepage>http://kowalski.name/mikolaj/</swrc:homepage>
    <rdfs:seeAlso rdf:resource="http://kowalski.name/mikolaj/foaf.rdf"/>

    <swrc:publication rdf:resource="http://example.org/abi/pub#MK2005b"/>
    <swrc:publication rdf:resource="http://example.org/abi/pub#MK2006a"/>
    <swrc:publication rdf:resource="http://example.org/abi/pub#MK2006b"/>
```

⁴ *SWRC Ontology* to jeden z projektów nadzorowanych przez *OntoWare Group*, <http://ontoware.org/projects/swrc/>.

```
<swrc:supervisor rdf:resource="http://example.org/abi/osoba#AB"/>
<swrc:worksAtProject rdf:resource="http://www.uni.edu.pl/research/rdf/p7"/>
</swrc:PhDStudent>
</rdf:RDF>
```

Istnieje również możliwość umieszczenia informacji RDF wewnątrz dokumentu HTML lub XHTML, poprzez odpowiedni dobór znaczników i ich atrybutów. W tym celu można wykorzystać np. rozwiązanie RDF/A [11]. Publikowana jest wtedy tylko jedna wersja strony, przeznaczona zarówno do czytania przez człowieka, jak i do przetwarzania programowego, nie ma więc niebezpieczeństwa podania dwóch, sprzecznych wersji tej samej informacji. Niestety wadą takiego rozwiązania jest konieczność używania dodatkowego mechanizmu przetwarzającego (najczęściej opartego o transformacje XSLT, np. rozwiązania typu GRDDL [12]), który przekształca taką stronę na postać RDF, oczekiwaną przez oprogramowanie Sieci Semantycznej.

Ponieważ model RDF funkcjonuje w „świecie otwartym”, możliwe jest konstruowanie wyrażeń o tych samych zasobach w zupełnie innym kontekście. Na przykład, osoba opisana wg ontologii SWRC może równocześnie podlegać opisowi wg innych popularnych ontologii, takich jak np. FOAF (*Friend Of A Friend*)⁵. W powyższym przykładzie zastosowano jedynie odwołanie do pliku dodatkowego opisu tego samego zasobu, za pomocą wyrażenia `rdfs:seeAlso`.

3.2. Dane o publikacjach

Informacje o publikacjach mogą być umieszczane zarówno na stronach uczelni lub ośrodków naukowych (jako dorobek naukowy pracowników), jak i na stronach prywatnych, a także na stronach konferencji, na których te publikacje były ogłoszone. W przypadku publikacji danych RDF, równoległe ze stronami HTML można wykorzystywać mechanizmy opisane powyżej. Jednak dane o publikacjach, szczególnie przechowywane przez instytucje (np. uczelnie), w znacznie większym stopniu niż dane o osobach, posiadają charakter tabelaryczny i masowy. Z tego względu dane te najczęściej przechowywane są w bazach danych w postaci właściwej dla lokalnego systemu informatycznego. Włączenie takiej bazy do wspólnego repozytorium wiedzy w ramach Sieci Semantycznej nie wymaga zmiany tych systemów lub zaprzestania ich użytkowania – należy jedynie umożliwić eksport danych do postaci RDF. Konwersja taka może być wykonywana cyklicznie, a wyniki są umieszczane w plikach, dostępnych za pośrednictwem Internetu lub na żądanie, za pomocą stworzonego, specjalnie do tego celu, serwisu sieciowego. Z punktu widzenia Sieci Semantycznej mechanizm ten jest zupełnie nieistotny, ważne jest jedynie, aby odwołanie do określonego adresu

⁵ <http://www.foaf-project.org/>.

URL skutkowało otrzymaniem oczekiwanego grafu RDF w jednej z powszechnie akceptowalnych postaci (standardowo w formacie RDF/XML).

Poniższy listing przedstawia przykład pliku RDF/XML, zawierającego informacje o publikacjach, wyrażone przy pomocy ontologii SWRC. Identyfikatory tych publikacji były już użyte poprzednio, a dodatkowo jako autor wskazany jest zasób o identyfikatorze wykorzystywanym wcześniej do oznaczenia pracownika. W ten sposób te dwa fragmenty RDF zawierają relacje między zasobami, nawet jeśli są umieszczone w rozłącznych plikach, umieszczonych w różnych lokalizacjach w sieci internetowej.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:swrc="http://swrc.ontoware.org/ontology#"
  xmlns="http://example.org/abi/pub#">

  <swrc:ProjectReport rdf:about="http://example.org/abi/pub#MK2005b">
    <swrc:describesProject rdf:resource="http://www.uni.edu.pl/research/rdf/p7"/>
    <swrc:author rdf:resource="http://example.org/abi/osoba#MK"/>
    <swrc:author rdf:resource="http://example.org/abi/osoba#KP"/>
    <swrc:title>Akademicka Baza Informacji</swrc:title>
    <swrc:year>2006</swrc:year>
    <swrc:creationDate>2005-10-07</swrc:creationDate>
    <swrc:keywords>semantic web, rdf</swrc:keywords>
  </swrc:ProjectReport>

  <swrc:InBook rdf:about="http://example.org/abi/pub#MK2006a">
    <swrc:author rdf:resource="http://example.org/abi/osoba#MK"/>
    <swrc:title xml:lang="pl">Portal Semantyczny</swrc:title>
    <swrc:title xml:lang="en">Semantic Portal</swrc:title>
    <swrc:year>2006</swrc:year>
    <swrc:creationDate>2006-02-01</swrc:creationDate>
    <swrc:booktitle>Aplikacje Sieci Semantycznej</swrc:booktitle>
    <swrc:chapter>4</swrc:chapter>
    <swrc:pages>112-160</swrc:pages>
    <swrc:series>Nowe Technologie w Internecie</swrc:series>
    <swrc:publisher rdf:resource="http://example.org/abi/org#EasyPress"/>
    <swrc:keywords>rdf, owl</swrc:keywords>
  </swrc:InBook>

  <swrc:Article rdf:about="http://example.org/abi/pub#MK2006b">
    <swrc:author rdf:resource="http://example.org/abi/osoba#MK"/>
    <swrc:title>RDF w Relacyjnych Bazach Danych</swrc:title>
    <swrc:creationDate>2006-03-01</swrc:creationDate>
    <swrc:journal>DB Review</swrc:journal>
    <swrc:year>2006</swrc:year>
    <swrc:volume>6</swrc:volume>
    <swrc:number>1</swrc:number>
    <swrc:pages>12-19</swrc:pages>
    <swrc:publisher rdf:resource="http://example.org/abi/org#DBJ"/>
    <swrc:keywords>database, rdf</swrc:keywords>
  </swrc:Article>
</rdf:RDF>
```

W przypadku danych o publikacjach naturalne wydaje się wykorzystanie do ich opisu zestawu metadanych DC (*Dublin Core*). Jest to zestaw opracowany przez *Dublin Core Meta-*

data Initiative,⁶ który jest w ramach Sieci Semantycznej szeroko wykorzystywany do opisu wszelkiego rodzaju dokumentów i publikacji. Jednak w przypadku zastosowania ontologii SWRC dane, podawane w ontologii DC (autor, tytuł, słowa kluczowe, itp.), były w znacznym stopniu zdublowane. W praktycznych zastosowaniach ontologii SWRC (np. w [10]) proponuje się wręcz jawne opisanie równoważności niektórych wyrażeń (np. `dc:creator` i `swrc:author`), za pomocą wyrażenia `owl:equivalentProperty` języka OWL.

3.3. Dane o wydarzeniach

W rozważanym zakresie zastosowań leżą również różnego rodzaju wydarzenia, związane z działalnością grup badawczych lub organizacji naukowych. Należą do nich konferencje, wykłady, seminaria, kolokwia itp. Ontologia SWRC pozwala opisać wiele rodzajów wydarzeń i wskazać ich uczestników lub inne osoby i wydarzenia z nimi powiązane. Dane tego typu mogą być publikowane zarówno przez osoby odpowiedzialne za organizację danego wydarzenia, np. konferencji, jak i przez osoby w nich uczestniczące.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:swrc="http://swrc.ontoware.org/ontology#"
  xmlns="http://example.org/abi/osoba#">

  <swrc:Workshop rdf:about="http://example.org/abi/ws#WS2006">
    <swrc:eventTitle>Rozmowy O Sieci 2006</swrc:eventTitle>
    <swrc:startDate>2006-03-03</swrc:startDate>
    <swrc:endDate>2006-03-05</swrc:endDate>
    <swrc:atEvent rdf:resource="http://example.org/abi/conf#ISWC2006"/>
    <swrc:participant rdf:resource="http://example.org/abi/osoba#MK"/>
    <swrc:participant rdf:resource="http://example.org/abi/osoba#JL"/>
    <swrc:participant rdf:resource="http://example.org/abi/osoba#AB"/>
    <swrc:participant rdf:resource="http://example.org/abi/osoba#MP"/>
  </swrc:Workshop>

  <swrc:Conference rdf:about="http://example.org/abi/conf#SWC2006">
    <swrc:eventTitle>Semantic Web Conference 2006</swrc:eventTitle>
    <swrc:startDate>2006-03-01</swrc:startDate>
    <swrc:endDate>2006-03-05</swrc:endDate>
    <swrc:hasPartEvent rdf:resource="http://example.org/abi/ws#WS2006"/>
    <swrc:hasPartEvent rdf:resource="http://example.org/abi/ws#SA2006"/>
    <swrc:hasPartEvent rdf:resource="http://example.org/abi/ws#IT72006"/>
    <swrc:hasPartEvent rdf:resource="http://example.org/abi/event#Meeting2006"/>
  </swrc:Conference>

  <swrc:Lecture rdf:about="http://example.org/abi/wyklad#SW1">
    <swrc:eventTitle>Semantic Web (I)</swrc:eventTitle>
    <swrc:givenBy rdf:resource="http://example.org/abi/osoba#MK"/>
  </swrc:Lecture>

</rdf:RDF>
```

⁶ <http://www.dublincore.org/>.

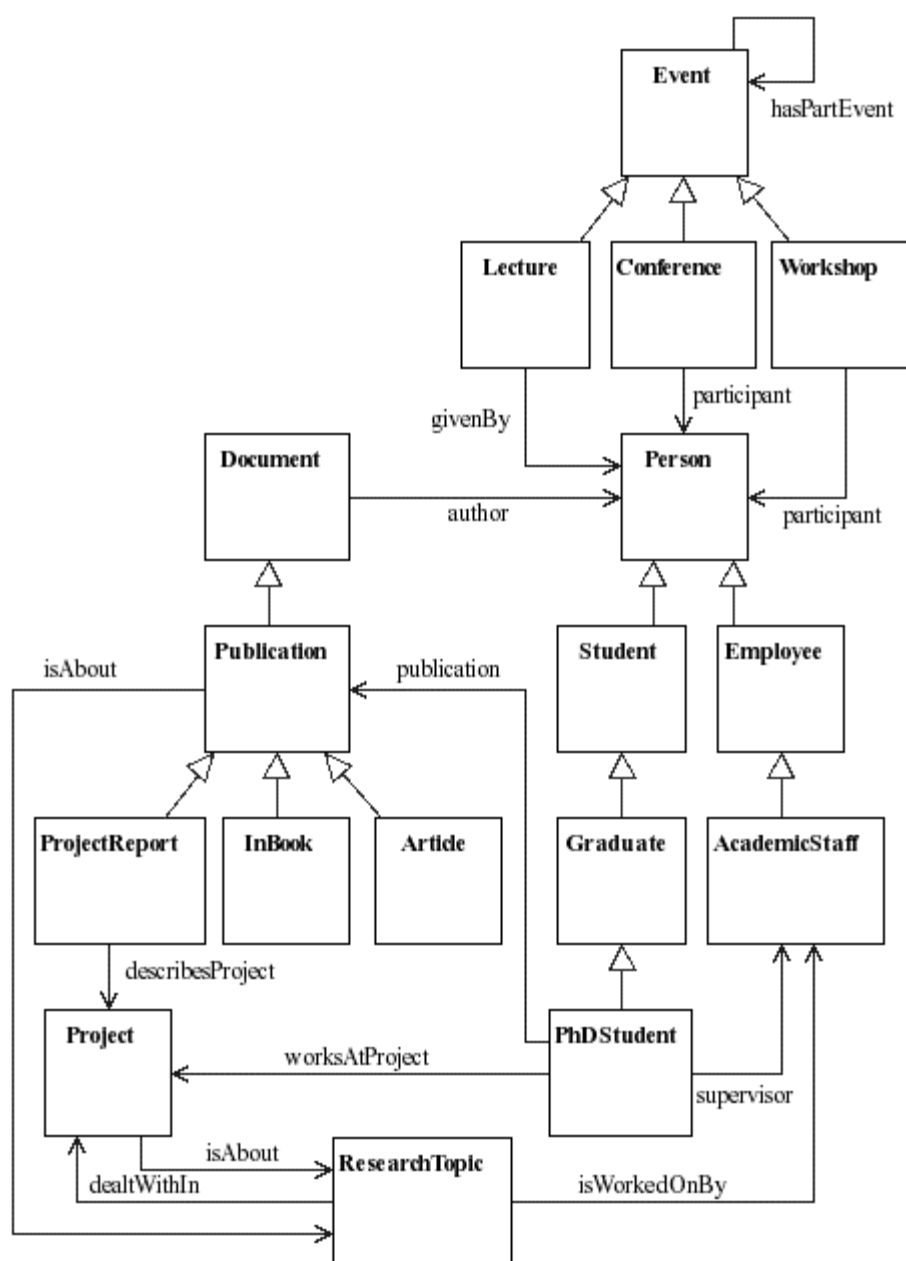
Analogicznie do innych rodzajów danych, dane o wydarzeniach opisane w ontologii SWRC mogą również posiadać dodatkowy opis, wykorzystujący inne ontologie, np. iCal. Ontologia ta opracowana w ramach *RDF Calendar Workspace*⁷ jest próbą przeniesienia do notacji RDF sposobu opisu danych kalendarza według standardu iCalendar (*Internet Calendaring and Scheduling Core Object Specification*, RFC2445) i jako taka stanowi popularny sposób opisu informacji o spotkaniach i terminach.

4. Łączenie i wyszukiwanie informacji

Przedstawione powyżej opisy zasobów w postaci RDF zawierają oprócz wartości poszczególnych cech danego zasobu również wskazania na inne zasoby (za pomocą wyrażenia `rdf:resource`), pozostające w relacjach, których semantyka wynika z zastosowanej ontologii. Opis łączonych zasobów może się znajdować w tym samym lub w różnych grafach RDF, przy czym grafy te mogą pozostać stale rozdzielone (wtedy odwołania do różnych grafów muszą być obsługane programowo) lub mogą podlegać połączeniu w jeden graf (za pomocą prostej procedury *RDF merge*, oficjalnie zdefiniowanej w [13]). W żadnym z tych przypadków nie następuje zaburzenie znaczenia wyrażenia łączącego oba zasoby. W ten sposób RDF umożliwia zapisanie relacji między dowolnymi obiektami, niezależnie od tego czy zostały one opisane w jednym modelu, czy w różnych rozproszonych w sieci internetowej. Oprócz połączeń między konkretnymi zasobami, definicje klas zawarte w ontologii zawierają opis relacji (w przypadku ontologii SWRC są to przede wszystkim hierarchie) między poszczególnymi klasami. Na tej podstawie można spojrzeć na dotychczas opisane przykłady wykorzystania ontologii SWRC jak na model obiektowy, którego schemat przedstawia rys. 1.

Model ten obrazuje relacje klas zasobów i pojęć, opisanych w ontologii w sposób podobny jak relacje klas obiektów w programowaniu obiektowym. Przykładowo zapis w ontologii SWRC mówi, iż predykat `swrc:publication` dla zasobu typu `swrc:PhDStudent` wskazuje zasób typu `swrc:Publication`. Zatem, w poprawnie skonstruowanym modelu RDF predykat ten może łączyć zasób, opisujący doktoranta z obiektem publikacji lub dowolnego typu pochodnego. Natomiast predykat `swrc:supervisor` wskazuje zasób klasy `swrc:AcademicStaff`, o którym wiadomo, że jest klasą pochodną od `swrc:Employee` i w dalszej kolejności od `swrc:Person`. Dzięki takim ograniczeniom oprogramowanie, wykorzystujące dane RDF, opisane według tej ontologii może z jednej strony rozpoznawać sytuacje występowania błędnych danych, a z drugiej wyciągać wnioski na temat odczytywanych zasobów. W prezentowanym przykładzie zasób wskazany przez `swrc:supervisor` z pewnością ma dane osobowe, a zasób wskazywany przez `swrc:publication` powinien mieć tytuł itd.

⁷ <http://www.w3.org/2002/12/cal/>.



Rys. 1. Ontologia SWRC jako model obiektowy (fragment)

Fig. 1. SWRC ontology as an object model (fragment)

Model taki może być również potraktowany jako baza danych i może podlegać przeszukiwaniu, za pomocą mechanizmów podobnych jak w przypadku klasycznych baz relacyjnych. Odpowiednikiem języka zapytań SQL, stosowanego w relacyjnych bazach danych jest, w środowisku Sieci Semantycznej, język zapytań SPARQL [7]. Wynikiem zapytania w tym języku jest zbiór wyrażeń RDF, występujących w przeszukiwanym modelu (zbiornie danych) i pasujących do podanego wzorca. Na przykład zapytanie wybierające projekt, datę, autora i tytuł publikacji opisujących dowolny z projektów, w których uczestniczy Mikołaj Kowalski, mogłoby wyglądać tak jak na listingu poniżej. Zakłada ono istnienie dwóch grafów RDF me-

rytorycznie ze sobą powiązanych – jednego z osobami i drugiego z publikacjami, ale oczywiście ilość tych plików, jak również podział ich zawartości oraz fizyczne umieszczenie w Internecie może być dowolne.

```
PREFIX swrc: <http://swrc.ontoware.org/ontology#>

SELECT ?projekt ?data ?autor ?tytul
FROM <http://example.org/abi/osoby.rdf>
FROM <http://example.org/abi/publikacje.rdf>
WHERE {
  [ swrc:name "Mikołaj Kowalski" ] swrc:worksAtProject ?projekt .
  ?pub swrc:describesProject ?projekt ;
    swrc:title ?tytul ;
    swrc:creationDate ?data ;
    swrc:author [ swrc:name ?autor ] .
}
ORDER BY ?projekt DESC(?data)
```

W bardziej praktycznym podejściu przeszukiwane grafy RDF zawsze muszą być wskazywane bezpośrednio w zapytaniu. Częściej występuje sytuacja, gdy grafy zawierające elementy składowe przeszukiwanego zbioru danych pobierane są na żądanie, na podstawie adresów przechowywanych w głównym modelu RDF. Na przykład, domyślny graf RDF utrzymywany przez oprogramowanie, wykonujące wyszukiwanie informacji może zawierać tylko adresy plików RDF, opisujących publikacje i przy takim założeniu zapytanie o tytuły i rok wydania publikacji Mikołaja Kowalskiego mogłoby wyglądać następująco:

```
PREFIX swrc: <http://swrc.ontoware.org/ontology#>

SELECT ?tytul ?rok
WHERE {
  [ swrc:name "Mikołaj Kowalski" ] swrc:publication ?src .
  GRAPH ?src {
    ?pub swrc:title ?tytul .
    ?pub swrc:year ?rok .
  }
}
ORDER BY DESC(?rok) ?tytul
```

W zapytaniu tym identyfikator URI grafu RDF, z którego pobierane są dane na temat publikacji, jest reprezentowany przez zmienną `?src`, czyli wskazywany przez predykat `swrc:publication` dla zasobu nazwanego Mikołaj Kowalski. W praktyce jest to adres URL pliku RDF, opisującego daną publikację. W ten sposób to proste zapytanie w języku SPARQL obejmuje informacje zapisane w źródłach rozproszonych w Internecie, w dowolny sposób.

Zapytania SPARQL mogą również zawierać elementy opcjonalne oraz dodatkowe warunki, np. filtry. Przykładowo: nazwiska i adresy (jeżeli zostały podane) osób, które w tym roku opublikowały cokolwiek w jednym z czasopism, w których publikował również Mikołaj Kowalski można uzyskać za pomocą poniższego zapytania.

```

PREFIX swrc: <http://swrc.ontoware.org/ontology#>

SELECT ?journal ?name ?email
WHERE {
  [ swrc:name "Mikołaj Kowalski" ] swrc:publication [ swrc:journal ?journal ] .
  ?publikacja swrc:journal ?journal ;
              swrc:year ?rok .
  ?osoba swrc:publication ?publikacja ;
         swrc:name ?name .
  OPTIONAL { ?osoba swrc:email ?email } .
  FILTER (?rok = "2006")
}
ORDER BY ?journal ?name

```

Z powyższych przykładów wynika, że język SPARQL bardzo dobrze nadaje się do wyszukiwania informacji zapisanej w bazie danych, reprezentowanej przez modele RDF, bez względu na ich rozproszenie i fizyczne rozmieszczenie w sieci internetowej. W kontekście rozważanego zastosowania Akademickiej Bazy Informacji oznacza to przede wszystkim, że programy działające w Sieci Semantycznej mają łatwą możliwość łączenia i przetwarzania wszelkich informacji, istotnych dla środowiska akademickiego, mimo że są one publikowane w różnych miejscach i przez niezależne od siebie podmioty.

5. Podsumowanie

Technologie Sieci Semantycznej umożliwiają łatwe skonstruowanie współdzielonej bazy informacji, rozproszonej w sieci internetowej. Centralne nadzorowanie takiej bazy nie jest wymagane, a tworzenie elementów składowych niezależne i stosunkowo proste. Dzięki temu nie ponosi się dodatkowych nakładów na aktualizację i utrzymanie, jakie byłyby konieczne w przypadku zastosowaniu centralnego repozytorium. Praktyczne wykorzystanie tej koncepcji wymaga co prawda ustalenia między użytkownikami i współautorami bazy wspólnego zestawu wykorzystywanych ontologii lub utworzenia nowej, jednolitej ontologii na potrzeby tego rozwiązania, jednak ze względu na to, że istnieje aktualnie w Internecie wiele dostępnych ontologii, a utworzenie własnej także nie jest szczególnym problemem, wymaganie to nie powinno być istotną przeszkodą.

Dzięki zastosowaniu ujednoliconych specyfikacji języka opisu informacji, oprogramowanie Sieci Semantycznej nie wymaga stosowania specyficznych formatów zapisu i dane mogą być wymieniane między programami, operującymi na danych RDF bez żadnych ograniczeń. Tym samym opisana powyżej koncepcja Akademickiej Bazy Informacji nie stanowi jednolitej aplikacji, lecz raczej środowisko do działania różnych programów. W przyszłości środowisko to będzie stanowić pole działania przede wszystkim dla agentów programowych, dokonujących wyszukiwania informacji na zlecenie użytkownika.

Istnienie takiej bazy informacji otwiera drogę dla wielu istotnych zastosowań. Z punktu widzenia edukacji akademickiej są to nie tylko zastosowania w zakresie sprawdzenia aktualnego stanu zaawansowania badań naukowych w danej dziedzinie, ale także większe możliwości kontroli, dotyczące zautomatyzowanego sprawdzania powtarzalności tematów prac, sprawdzenia zasadności podejmowania jakiegoś tematu pracy lub publikacji itp.

Jednak aby zastosowanie takie miało szansę powodzenia koniecznie jest rozpowszechnienie tej idei i publikowanie opisu semantycznego wszelkich zasobów, istotnych z punktu widzenia wymiany informacji w omawianej dziedzinie. Ponadto wiele zagadnień, związanych z funkcjonowaniem Sieci Semantycznej nie jest jeszcze dostatecznie opracowanych. Dotyczy to w szczególności zagadnień, związanych z weryfikacją i przekazywaniem zaufania do danych, które są pobierane z Internetu. Nieograniczona ilość różnorodnych źródeł informacji jest z jednej strony poważną zaletą takiego rozwiązania, jednak z drugiej wprowadza ryzyko wykorzystywania informacji błędnych lub wręcz świadomie dezinformujących. Zjawisko takie występuje oczywiście również w dzisiejszej sieci WWW, jednak nabiera dużo większej wagi w sytuacji, gdy dane pochodzące z Internetu miałyby być przetwarzane przez oprogramowanie automatycznie, bez udziału człowieka. Jest to aktualnie najbardziej intensywnie opracowywana grupa zagadnień, dotyczących *Semantic Web* i z pewnością niedługo zostaną opracowane oficjalne rozwiązania i standardy również w tej kwestii.

LITERATURA

1. Klyne G., Carroll J. (red.): Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, W3C 2004, <http://www.w3.org/TR/rdf-concepts/>.
2. Berners-Lee T., Fielding R., Masinter L.: Uniform Resource Identifier (URI): Generic Syntax, Request for Comments: 3986. The Internet Society 2005, <http://www.ietf.org/rfc/rfc3986.txt>.
3. Biron P. V., Permanente K., Malhotra A.: XML Schema Part 2: Datatypes Second Edition. W3C Recommendation, W3C 2004, <http://www.w3.org/TR/xmlschema-2/>.
4. Beckett D.: RDF/XML Syntax Specification (Revised). W3C Recommendation, W3C 2004, <http://www.w3.org/TR/rdf-syntax-grammar/>.
5. Brickley D., Guha R. V.: RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, W3C 2004, <http://www.w3.org/TR/rdf-schema/>.
6. McGuinness D., van Harmelen F.: OWL Web Ontology Language Overview. W3C Recommendation, W3C 2004, <http://www.w3.org/TR/owl-features/>.

7. Prud'hommeaux E., Seaborne A.: SPARQL Query Language for RDF. W3C Working Draft, W3C 2006, <http://www.w3.org/TR/rdf-sparql-query/>.
8. Clark K. G.: SPARQL Protocol for RDF. W3C Working Draft, W3C 2006, <http://www.w3.org/TR/rdf-sparql-protocol/>.
9. Beckett D., Broekstra J.: SPARQL Query Results XML Format. W3C Working Draft, W3C 2006, <http://www.w3.org/TR/rdf-sparql-XMLres/>.
10. Sure Y., Bloehdorn S., Haase P., Hartmann J., Oberle D.: The SWRC Ontology – Semantic Web for Research Communities. Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005), Covilha 2005.
11. Adida B., Birbeck M.: RDF/A Primer 1.0. Embedding RDF in XHTML. W3C Working Draft, W3C 2006, <http://www.w3.org/TR/xhtml-rdfa-primer/>.
12. Hazaël-Massieux D., Connolly D.: Gleaning Resource Descriptions from Dialects of Languages (GRDDL), W3C Team Submission, W3C 2005, <http://www.w3.org/Team-Submission/grddl/>.
13. Hayes P.: RDF Semantics. W3C Recommendation, W3C 2004, <http://www.w3.org/TR/rdf-mt/>.

Recenzent: Dr inż. Grzegorz Hryń

Wpłynęło do Redakcji 2 maja 2006 r.

Abstract

Accessing information available in WWW requires a lot of effort from human reader and is usually inaccessible for automated processing by software agents. The Semantic Web is a vision of changing this situation by providing infrastructure for expressing data and knowledge in machine-understandable format. To make this vision possible W3C Semantic Web Activity prepared set of specifications including RDF (*Resource Description Language*) for representing statements about resources, OWL (*Web Ontology Language*) for representing ontologies and SPARQL (*SPARQL Protocol And RDF Query Language*) for querying RDF data sets.

Although Semantic Web is still a work in progress, there is already a lot of ontologies available today. In the article the SWRC (*Semantic Web for Research Communities*) ontology was chosen as an example. This ontology models key entities in academic and research communities and can be used for describing information about members (academic staff, students), publications (books, articles, reports) and events (workshops, conferences, lectures)

as well as different relations between them. Knowledge about resources described in SWRC ontology can be represented as an object model, as shown in Fig.1.

Using this model Semantic Web application will be able to automatically collect, integrate and process information from many sources and RDF models provided by involved organizations and different community members. SPARQL query language provides an easy way of querying distributed RDF models even not using specialized and complicated software.

In conclusion, using Semantic Web technologies already available today, one could create an distributed machine-readable, self-describing database of related information. All data in such database are provided by independent sources and no central and cost-consuming maintenance is needed.

Adres

Michał ŚWIĄTKIEWICZ: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-101 Gliwice, Polska, michal@opole.home.pl .