


8 stycznia 2003

Mariusz Olko

Mariusz.Olko@empolis.pl


Od metadanych do inteligentnego Internetu



Metadane


Informacja o informacji (sic!)

- Różnorakie informacje o dokumencie
 - Data utworzenia
 - Słowa kluczowe
 - Autor
- Schemat bazy danych
 - Powiązania klucz główny/klucz obcy
 - Typy i rozmiary pól
 - Więzy integralności





Skąd się biorą metadane?

- Zdania wypowiedzianie (zapisane) niezależnie od dokumentu
- Wyróżnione elementy w dokumencie



```

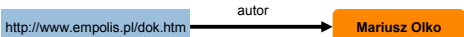

<HTML>
<HEAD>
<TITLE>Pan Tadeusz</TITLE>
<META NAME="Author" CONTENT="Adam Mickiewicz">
<META http-equiv="Content-Type"
  content="text/html; charset=iso-8859-2">
</HEAD>
  
```

Zdania o dokumencie


`http://www.empolis.pl/dok.htm` ma `autora` o nazwie `Mariusz Olko`

Podmiot *subject* Orzeczenie *predicate* Dopelnienie *object*

Resource Description Framework


- Zdania są zapisane w postaci trójek
(**subject, predicate, object**)
- Podmiotem (subject) w trójce są zasoby (resources), które
 - Mogą być adresowalne w internecie np. dokument HTML
 - Mogą też oznaczać koncepty abstrakcyjne np. osobę lub dziedzinę sztuki
- Predykat w trójce jest zawsze konceptem abstrakcyjnym określającym własność
- Dopelnieniem trójki może być
 - Inny zasób abstrakcyjny lub rzeczywisty
 - Bądź też literał czyli napis nie podlegający interpretacji w ramach RDF




Universal Resource Identifier

- Zasoby są identyfikowane przez URI
- URI nie musi być URL, tzn. nie musi oznaczać lokalizacji internetowej

`http://www.w3.org/1999/02/22-rdf-syntax-ns#label`

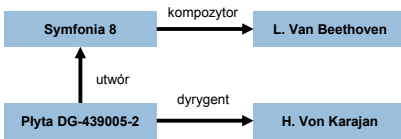


- Używając przestrzeni nazw w serializacji XML można
 - zdefiniować prefix `rdf` dla URI przestrzeni nazw
`http://www.w3.org/1999/02/22-rdf-syntax-ns#`
 - Zapisać pełne URI zasobu w postaci skróconej
`rdf:label`



Interpretacja trójek RDF

- Zbiór zdań RDFowych reprezentuje graf skierowany
- Węzły z których wychodzą łuki reprezentują zasoby
- Łuki reprezentują własności



- Specyfikacja RDF definiuje sposób serializacji grafu do XMLa
- Grafy RDFowe można w sposób prosty łączyć

Narzędzia do modelowania

- RDF definiuje podstawowy zestaw pojęć, który może być użyty do modelowania informacji i budowy ontologii
- Własności dotyczące zasobów
 - label (krótka etykieta nadana zasobowi)
 - description (dłuższy opis zasobu)
 - type (określenie typu zasobu)
- Własności dotyczące klas
 - subclassOf (podział klasy na podklasy)
- Własności dotyczące własności
 - subPropertyOf
 - domain (klasa obiektów, które mogą być podmiotem zdań z danym predykatem)
 - range (klasa obiektów, które mogą być dopełnieniem zdań z danym predykatem)

Inferencje w RDFach

Nadanie interpretacji językowi oraz predykatom służącym do budowania modeli pozwala na skonstruowanie reguł zwanych regułami inferencji.

- Reguła dotycząca podklas
 - Jeżeli A jest podklasą B i a jest obiektem typu A
 - To a jest też obiektem typu B
 - $(A, \text{rdfs:subclassOf}, B) (a, \text{rdf:type}, A) \Rightarrow (a, \text{rdf:type}, B)$

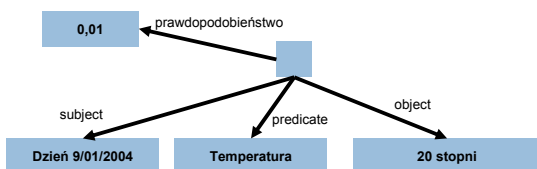
Reguły inferencyjne

- Reguła dotycząca podwłasności
 - Jeżeli P jest podwłasnością B i zachodzi (a, P, b)
 - To zachodzi też (a, R, b)
 - $(P, \text{rdfs:subPropertyOf}, R) (a, P, b) \Rightarrow (a, R, b)$
- Reguła dotycząca dziedziny własności
 - Jeżeli dziedziną P jest klasa A oraz a ma własność P o dowolnej wartości
 - To a jest typu A
 - $(P, \text{rdfs:domain}, A) (a, P, x) \Rightarrow (a, \text{rdf:type}, A)$

Reifikacja

- Nie można w prosty sposób wypowiadać zdań na temat innych zdań!
- Aby można było się wypowiedzieć na temat jakiegoś zdania, musi ono zostać **zreifikowane** tzn. zamienione na zestaw zdań (metazdań)

Prawdopodobieństwo faktu, że 9 stycznia będzie temperatura 20 stopni celsjusza jest bliskie zeru



Dublin Core

- *Dublin Core Metadata Initiative (DCMI)* jest organizacją której celem jest promowanie stosowania standardów metadanych i rozwój specjalizowanych słowników pojęć do opisu zasobów
- Zdefiniowane zostały podstawowe i rozszerzone słowniki pojęć
- Określono standardy użycia np. w HTMLowych tagach META
- Określono schematy zapisu metadanych w postaci grafów RDF
- Namespace URI <http://purl.org/dc/elements/1.1/>

Słownik pojęć

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Source
- Language
- Rights

RDF Site Summary (RSS)

- Powstał dla potrzeb portalu my.netscape.com syndykacji informacji
- XMLowa lista hiperłączy do zasobów internetowych razem z metainformacjami dotyczącymi tych zasobów
- Zazwyczaj pliki RSS są pobierane przez programy agregujące przetwarzane i wyświetlane jako hiperlinki do oryginalnych zasobów
- Listy mogą zawierać:
 - Wiadomości z ostatniej chwili
 - Listę ostatnich wpisów w blogu lub najnowszych artykułów w serwisie
 - List ofert pracy
 - Wyniki poszukiwań
- Namespace URI <http://purl.org/rss/1.0>

Słownik pojęć

- Channel
- Item
- Title
- Description

RSS example

```

<?xml version="1.0" encoding="UTF-8" ?>
<rss xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rss="http://xmlns.com/foaf/0.1/"
xmlns:rdfs="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:atom="http://purl.org/rss/1.0/"
channel rdf:about="http://xmlhack.com/rss10.php">
<title>XMLHack</title>
<link>http://xmlhack.com/</link>
<dc:source rdf:resource="http://xmlhack.com/" />
<description>Developer news from the XML community</description>
<dc:language>en-us</dc:language>
<dc:publisher>Edd Dumbill (mailto:webmaster@xmlhack.com)</dc:publisher>
<dc:creator>Edd Dumbill (mailto:editor@xmlhack.com)</dc:creator>
<dc:rights>Copyright 1999-2001, xmlhack team.</dc:rights>
<image rdf:resource="http://xmlhack.com/images/mynetscape88.gif" />
</channel>
<rss:Seq>
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2131" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2129" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2128" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2127" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2126" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2119" />
<rss:item rdf:resource="http://xmlhack.com/read.php?item=2118" />
</rss:Seq>
</rss>
</channel>
<image rdf:about="http://xmlhack.com/images/mynetscape88.gif">
<atom:link href="http://xmlhack.com/read.php?item=2131">
<title>Extreme Markup Languages 2004</title>
<link>http://xmlhack.com/read.php?item=2131</link>
<dc:description>Originally announced at XML 2003, the Call for Participation for Extreme Markup
2004 is now open. The conference will be held from 3-6 August in Montreal,
Canada.</dc:description>
<dc:subject>Community</dc:subject>
</atom:link>
</channel rdf:about="http://xmlhack.com/read.php?item=2129">

```

Friend-Of-A-Friend (FOAF)

**W jaki sposób jednoznacznie zidentyfikować osobę?
Użyć jej adres email!**

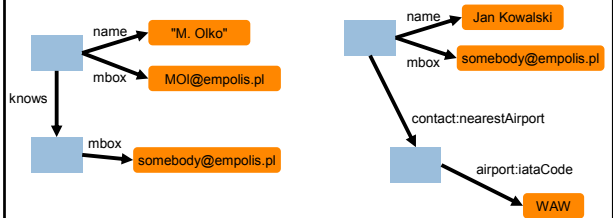
- Osoby publikują swoje pliki FOAF
- Pliki mogą zawierać informacje np.
 - nt. miejsca pobytu
 - projektów na których pracują
 - osób które znają
- Roboty lub aplikacje agregują informacje z wielu źródeł
- Zagregowane informacje mogą zostać wykorzystane do prezentacji np.
 - Kogo ze znajomych mogę spotkać w mieście w którym odbywa się konferencja na którą jadę?

Słownictwo

- Agent
- Person
- name
- nick
- Mbox
- knows
- depiction
- publications
- homepage
- organization
- group
- project

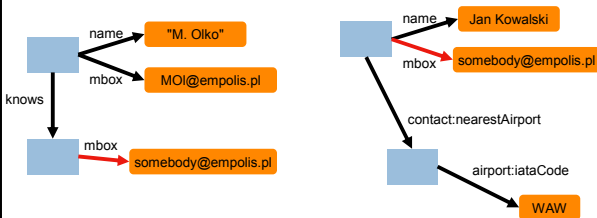
<http://xmlns.com/foaf/0.1/>

Łączenie grafów RDFowych



Łączenie grafów RDFowych

Wartość własności mbox identyfikuje jednoznacznie osobę!



Łączenie grafów RDFowych

Wartość własności mbox identyfikuje jednoznacznie osobę!



OWL Web Ontology Language

OWL

- Obecnie Proposed Recommendation w W3C
- Bardziej sformalizowany język budowania ontologii
- Dostosowany do możliwości implementacji silników inferencyjnych
- Ogranicza wolność opisu RDFowego dając w zamian dobrze zdefiniowaną interpretację zdań
- Składnia oparta o dokumenty RDF, "internetowy" charakter
- Semantyka korzysta z dorobku AI w zakresie reprezentacji wiedzy

Słownictwo OWL

Definicja klasy

- `owl:oneOf` przez wyliczenie przynależnych instancji
- `owl:restriction` i `owl:onProperty` przez podanie ograniczenia na własności instancji
 - `owl:allValuesFrom` wszystkie wartości własności są w podanej klasie
 - `owl:someValuesFrom` przynajmniej jedna w własności jest w podanej klasie
 - `owl:minCardinality` i `owl:maxCardinality`
- Przez wykonanie operacji teorii mnogościowej na innych klasach
 - `owl:intersectionOf`
 - `owl:unionOf`
 - `owl:complementOf`

OWL słownictwo

- `owl:sameAs`
 - stwierdza tożsamość dwóch obiektów
- `owl:equivalentProperty`
 - ta sama dziedzina i ten sam zakres
 - mogą mieć całkowicie inne znaczenie
- `Owl:TransitiveProperty`
 - przechodność własności
- `Owl:FunctionalProperty`
 - Własność P, która może mieć tylko jedną *unikalną* wartość dla każdej instancji obiektu
 - Inaczej jeżeli dwa obiekty mają jako tę samą wartość P to oznacza, że obiekty są tożsame
- `owl:differentFrom`
 - Dwa URI odnoszą się do dwóch różnych bytów

OWL - trzy poziomy języka

OWL full

- Pozwala mieszać słownictwo RDFSchem oraz OWL
- Pozwala na dowolność w opisywaniu obiektów, klas i własności

OWL DL

- Stawia ograniczenia na używanie słownictwa RDFSchem
- Wymaga rozdzielenia klas, własności, instancji
- Pozwala na zastosowanie istniejących już silnych narzędzi do wnioskowania opartych na Description Logic

OWL Lite

- Stawia te same ograniczenia co OWL DL na użycie słownictwa RDFSchem oraz na rozdzielenie klas, własności i instancji
- Wybiera prosty podzbiór słownictwa umożliwiający łatwy start użytkownikom oraz zapewniający prostotę implementacji narzędzi

Semantic Web

"The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

[Scientific American, May 2001](#)

Internetowa infrastruktura publikacji danych

- neutralna (niezależna od aplikacji)
- umożliwiająca przetwarzanie informacji przez programy w celu
 - automatyzacji
 - agregacji
 - wielokrotnego użycia

Semantic Web - aplikacje

- Poprawienie jakości wyszukiwania informacji
- Umożliwienie wyszukiwania usług sieciowych
- Automatyzacja czynności codziennych
 - Rezerwacja wizyty u lekarza, w urzędzie
 - Przygotowanie podróży

Uwaga

- To jest ciągle wizja
- Technologie się pojawiają ale
- Zdania nie oznaczają faktów
- URI nie identyfikują bytów rzeczywistych
- Potrzebna jest kodyfikacja *wiedzy codziennej*
- Pojawiają się problemy związane z etyką oraz bezpieczeństwem

Projekt SWAP

- Kombinacja dwóch technologii
 - Peer-to-Peer
 - Semantic Web
- Zarządzanie informacją i wiedzą w środowisku rozproszonym

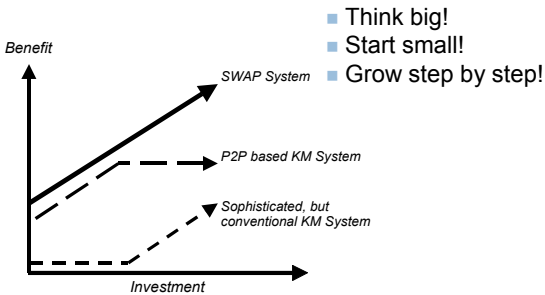
Konwencjonalny systemy zarządzania wiedzą

- Trudne w obsłudze
- Narzucają własne widzenie świata (ontologie)
- Wymagają "oddania" dokumentów

Systemy peer-to-peer

- Dużo efektywniejsze w wymianie informacji
- Wyszukiwanie informacji jest zbyt proste - słowa kluczowe
- Nie umożliwiają dzielenie się "wiedzą" (ontologiami)

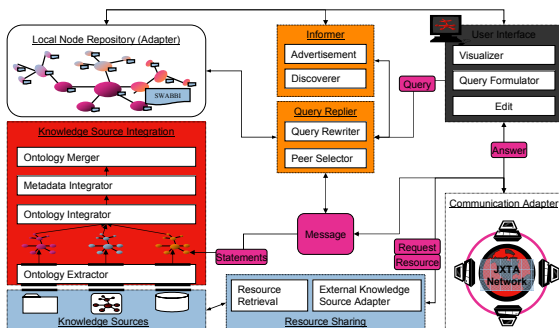
Korzyści płynące ze SWAPa



Scenariusz

- Ontologia jest budowana ze struktury informacji przechowywanych na komputerze takich jak:
 - Dokumenty
 - Bookmarki
 - Poczta
- Zostają automatycznie sklasyfikowane i udostępnione innym użytkownikom
- Użytkownicy mogą podawać zapytania dotyczące ontologii przechowywanej w węzle
- Komputer może gromadzić "wiedzę" przechowywaną na innych węzłach poprzez integrację wyników zapytań
- Użytkownicy mogą pobierać dokumenty zgromadzone na odległych węzłach

SWAP Architecture



Referencje

Resource Description Framework

- [Resource Description Framework na serwerze W3C](#)
- [OWL Web Ontology Language](#)

Strony inicjatyw opartych o RDF

- [Inicjatywa Dublin Core](#)
- [Projekt Friend-Of-A-Friend](#)

Artykuły

- [Scientific American nt. Semantic Web](#)
- [Tutorial RSS](#)