

Modelowanie wiedzy

Topic Maps – geneza

- „W dzisiejszych czasach większość ludzi nie potrzebuje *więcej* informacji. Jeśli już, to potrzebują jej *mniej*, ponieważ już toną w ogromnych jej ilościach.”
Pepper, S. Euler, *Topic Maps and Revolution*,
<http://www.infoloom.com/tmsample/pep4.htm>
- Oryginalna motywacja (1991):
 - jednolity standard do reprezentacji indeksów książkowych,
 - scalanie indeksów.
- Pomysł:
 - utworzenie nad warstwą zasobów warstwy abstrakcyjnych *pojęć*,
 - powiązanie obu warstw poprzez *wystąpienia* pojęć w zasobach.

Pojęcia

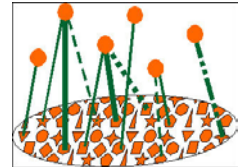
- Pojęcie (topic):
 - abstrakcyjny byt,
 - "co autor miał na myśli".
- Typ pojęcia:
 - także jest pojęciem.
- Pojęcie posiada:
 - nazwy,
 - wystąpienia,
 - role pełnione w powiązaniach.



Źródło: Pepper, S. Euler, *Topic Maps and Revolution*,
<http://www.infoloom.com/tmsample/pep4.htm>

Wystąpienia

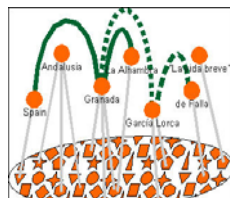
- Wiążą pojęcia z warstwą zasobów:
 - zasoby nie są częścią mapy pojęć.
- Nadają sens pojęciom.
- Role wystąpień.



Źródło: Pepper, S. Euler, *Topic Maps and Revolution*,
<http://www.infoloom.com/tmsample/pep4.htm>

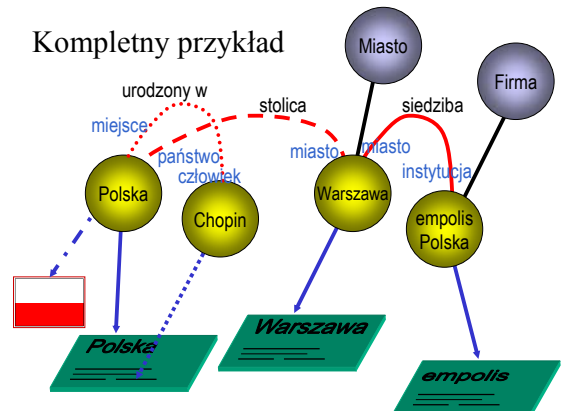
Powiązania

- Tworzą sieć zależności między pojęciami:
 - typy powiązań,
 - role pojęć w powiązaniu,
 - powiązania łączące więcej niż dwa pojęcia.



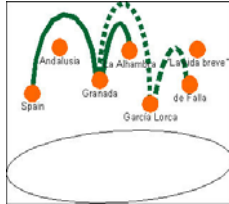
Źródło: Pepper, S. Euler, *Topic Maps and Revolution*,
<http://www.infoloom.com/tmsample/pep4.htm>

Kompletny przykład



Co z tego mamy?

- Mapa pojęć jako samodzielny, niezależny byt (dokument):
 - oddzielona od warstwy zasobów,
 - linki (powiązania) niezależne od warstwy zasobów.
- Więcej:
 - nad danym zbiorem zasobów można skonstruować wiele różnych map,
 - jedna mapa może być użyta do nawigacji po wielu zbiorach zasobów.
- Reprezentacja **wiedzy!**
- „The GPS of the information universe”.



Zródło: Pepper, S. Euler, *Topic Maps and Revolution*, <http://www.infoloom.com/tmsample/pep4.htm>

Zastosowania

- Wydawnictwa encyklopedyczne:
 - wartość dodana do informacji encyklopedycznej (poprzez możliwość łatwego znalezienia informacji),
 - *Mother Encyclopaedia*.
- Zarządzanie witryną internetową:
 - struktura witryny jako mapa pojęć,
 - interfejs nawigacyjny – dzięki odpowiednim przekształceniom XSLT.
- Zarządzanie wiedzą w organizacjach.
- Wymiana/przesyłanie zakodowanej wiedzy.

Status

- Topic Maps:
 - pierwotnie zwany Topic Navigation Maps,
 - ISO/IEC 13250:2000, Second Edition, 19 May 2002,
 - twórcy: Michel Biezunski, Martin Bryan, Steven R. Newcomb,
 - oparty na SGML-u i Hy-Time.
- XML Topic Maps (XTM):
 - rozwijany przez TopicMaps.org – niezależne konsorcjum specjalistów,
 - aktualnie dostępna wersja: 1.0 z 6.08.2001,
 - twórcy: Steve Pepper i Graham Moore,
 - oparty na XML-u i XLink.

XML Topic Maps – przykład

```
<topicMap>
  <topic id="kompozytor">
    <baseName><baseNameString>kompozytor</baseNameString</baseName>
  </topic>
  ...
  <topic id="chopin">
    <instanceOf><topicRef xlink:href="#kompozytor"/></instanceOf>
    <baseName><baseNameString>Fryderyk Chopin</baseNameString</baseName>
    <baseName><baseNameString>Chopin, Fryderyk</baseNameString</baseName>
    <occurrence><resourceRef
      xlink:href="http://www.encyklopedia.pl/chopin.htm"/></occurrence>
  </topic>

  <topic id="polska">
    <instanceOf><topicRef xlink:href="#kraj"/></instanceOf>
    <baseName><baseNameString>Polska</baseNameString</baseName>
    <occurrence><resourceRef
      xlink:href="http://www.encyklopedia.pl/polska.htm"/></occurrence>
  </topic>






  <association>
    <instanceOf><topicRef xlink:href="#urodzony-w"/></instanceOf>
    <member><roleSpec><topicRef xlink:href="#osoba"/></roleSpec>
      <topicRef xlink:href="chopin"/> </member>
    <member><roleSpec><topicRef xlink:href="#kraj"/></roleSpec>
      <topicRef xlink:href="polska"/> </member>
  </assoc>
</topicMap>
```

TMQL – Topic Maps Query Language

- Trwają prace rozwojowe.
- Dostępne różne propozycje i prototypy ich implementacji.
- Przykład (składnia proponowana przez empolis):

```
SELECT topic x WHERE
x instance_of topic named "Job seeker"
AND
assoc template y named "Person is skilled in"
AND
x in (assoc template_is y) has topic named "Java programming"
AND
x in (assoc template_is y) has topic named "German - fluent"
```

Narzędzia

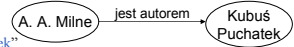
- Topic Maps Loom, InfoLoom  www.infoloom.com
- Ontopia Knowledge Suite, Ontopia  www.ontopia.net
- ITM Intelligent Topic Manager, Mondeca  www.mondeca.com
- TM4J, Topic Maps for Java (open source)  tm4j.org
-  www.techquila.com

Resource Description Framework

- Resource Description Framework – metodologia opisu zasobów (w Internecie).
- Zastosowania:
 - metainformacje zasobów,
 - modelowanie zależności pomiędzy zasobami,
 - wnioskowanie na podstawie modelu wiedzy (RDQL – Resource Description Query Language, język zapytań),
 - Semantic Web.

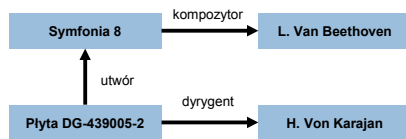
Reprezentacja wiedzy w RDF

- Zdania logiczne (trójki):
 - podmiot orzeczenie przedmiot
 - np. A. A. Milne jest autorem „Kubusia Puchatka”
- Pojęcia:
 - zasoby:
 - „A. A. Milne”, „Kubus Puchatek”
 - typy właściwości:
 - „jest autorem”, „słowo kluczowe”
 - wartości właściwości:
 - zasób „Kubus Puchatek”, literal „miód”
- Typy właściwości są zasobami.
- Właściwości mogą być zasobami.
- Elastyczność:
 - nie ogranicza się repertuaru właściwości.



Interpretacja trójek RDF

- Zbiór zdań RDF reprezentuje graf skierowany.
 - węzły z których wychodzą łuki reprezentują zasoby,
 - łuki reprezentują własności.



- Specyfikacja RDF definiuje sposób serializacji grafu do XML-a.
- Grafy RDF można w prosty sposób łączyć.

Identyfikowanie zasobów

- URI – Universal Resource Identifier
- URI nie musi być URL, tzn. nie musi oznaczać lokalizacji internetowej

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

Indentyfikator przestrzeni nazw

Nazwa lokalna

- 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`

Budowanie ontologii w RDF

- 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).

Dziedziczenie w RDF – przykład

- A. A. Milne jest autorem „Kubusia Puchatka”
- Klasa: człowiek
 - Podklasa: pisarz
 - Instancja: A. A. Milne
- Klasa: dzieło
 - Podklasa: książka
 - Instancja: „Kubus Puchatek”
- Własność: jest autorem
 - Dziedzina (domain): człowiek
 - Zasięg (range): dzieło

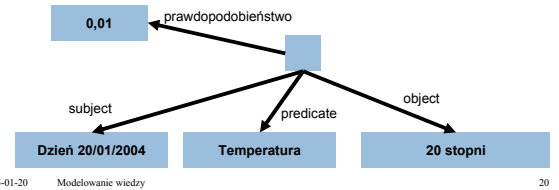


Inferencje w RDF

- Semantyka predykatów służących do budowania ontologii.
- Przykłady reguł inferencji (wnioskowania):
 - jeżeli A jest podklasą B i a jest obiektem typu A, to a jest też obiektem typu B
(A, rdfs:subClassOf, B) (a, rdf:type, A)
=> (a, rdf:type B)
 - jeżeli P jest podwłasnością B i zachodzi (a, P, b), to zachodzi też (a, R, b)
(P, rdfs:subPropertyOf, R) (a, P, b)
=> (a, R, b)
 - jeżeli dziedzina P jest klasa A oraz a ma własność P o dowolnej wartości, to a jest typu A
(P, rdfs:domain, A) (a, P, x)
=> (a, 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 *reifikowane* tzn. zamienione na zestaw zdań (metazdań)
- Przykład:
 - Prawdopodobieństwo faktu, że 20 stycznia będzie temperatura 20 stopni celsjusza jest bliskie zero



RDQL – język zapytań

- Resource Description Query Language:
 - łatwy język zapytań wzorowany na SQL,
 - dostępny dla popularnych języków programowania np. Java.
- Przykład:

```
select ?user where
(<http://strona.com>,
 <http://property/created-by>, ?user)
```

Standardy oparte na RDF

- Problem:
 - w RDF można wyrazić *dowolne* własności,
 - komunikacja przy pomocy RDF ma sens jeśli partnerzy posługują się tym samym słownikiem,
 - RDF nie definiuje słownika, jedynie sposób zapisu metadanych!
- Standardy oparte na RDF (słowniki pojęć, ontologie):
 - Dublin Core,
 - RSS – RDF Site Summary,
 - FOAF – Friend-Of-A-Friend,
 - OWL – Web Ontology Language.

Dublin Core

- Dublin Core Metadata Initiative (DCMI) – organizacja promująca stosowanie standardów metadanych i rozwój specjalizowanych słowników pojęć do opisu zasobów.
- Dublin Core:
 - definiuje podstawowe i rozszerzone słowniki pojęć,
 - określa standardy użycia np. w HTML-owych tagach META,
 - określa schematy zapisu metadanych w postaci RDF.
- Przestrzeń nazw:
<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 na potrzeby portalu my.netscape.com do syndykacji informacji.
- Kanał RSS:
 - lista hiperłączy do zasobów,
 - metainformacje.
- Typowe wykorzystanie kanału RSS:
 - pobierany przez programy agregujące,
 - przetwarzany i wyświetlany jako linki do oryginalnych zasobów.
- Przykłady:
 - wiadomości z ostatniej chwili,
 - lista ostatnich wpisów w blogu lub najnowszych artykułów w serwisie,
 - lista ofert pracy.
- Przestrzeń nazw:
<http://purl.org/rss/1.0>
- Słownik pojęć
 - Channel
 - Item
 - Title
 - Description

Przykład: RSS i Dublin Core

```
<?xml version="1.0" encoding="iso-8859-1" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://purl.org/rss/1.0/"
xmlns:mbox="http://usefulinc.com/rss/manifold/" xmlns:dc="http://purl.org/dc/elements/1.1/">
<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:ed@xmlhack.com)</dc:creator>
<dc:rights>Copyright 1999-2001, xmlhack team</dc:rights>
<image rdf:resource="http://xmlhack.com/images/myntscapet08.gif" />
<items>
<rdf:Seq>
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2131" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2129" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2125" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2124" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2123" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2120" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2119" />
<rdf:li rdf:resource="http://xmlhack.com/read.php?item=2118" />
</rdf:Seq>
</items>
</channel>
<image rdf:about="http://xmlhack.com/images/myntscapet08.gif">
<item rdf:about="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 Montréal,
Canada.</dc:description>
<dc:subject>Community</dc:subject>
</item>
</rdf:RDF>
```

25

Friend-Of-A-Friend (FOAF)

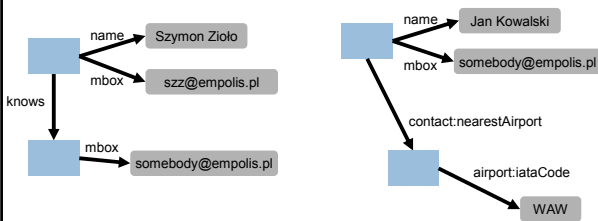
- Scenariusz wykorzystania:
 - osoby publikują swoje pliki FOAF,
 - roboty lub aplikacje agregują informacje z wielu źródeł,
 - zagregowane informacje mogą zostać wykorzystane np. do znalezienia znajomych, których mogą spotkać w mieście w którym odbywa się konferencja na którą jadą.
- Identyfikacja osób przy pomocy e-maili.
- Przykładowa zawartość pliku FOAF:
 - miejsce pobytu,
 - projekty w których pracuje,
 - osoby które znam.
- Przestrzeń nazw:
 - <http://xmlns.com/foaf/0.1/>

- Słownictwo
 - agent
 - person
 - name
 - nick
 - mbox
 - knows
 - depiction
 - publications
 - homepage
 - organization
 - group
 - project

2005-01-20 Modelowanie wiedzy

26

Łączenie grafów RDF

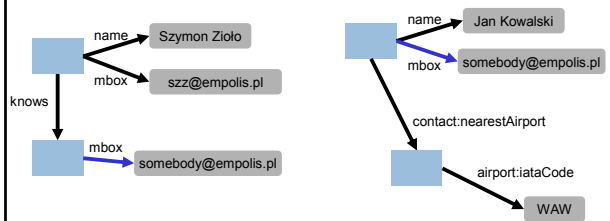


2005-01-20 Modelowanie wiedzy

27

Łączenie grafów RDF

Wartość własności mbox identyfikuje jednoznacznie osobę!

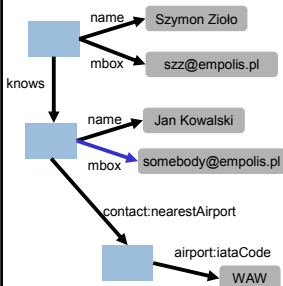


2005-01-20 Modelowanie wiedzy

28

Łączenie grafów RDF

Wartość własności mbox identyfikuje jednoznacznie osobę!



2005-01-20 Modelowanie wiedzy

29

OWL Web Ontology Language

- Sformalizowany język budowania ontologii:
 - ogranicza wolność opisu RDF, dając w zamian dobrze zdefiniowaną interpretację zdań,
 - słownik formalnego opisu klas i własności,
 - korzysta z dorobku AI w zakresie reprezentacji wiedzy i wnioskowania.
- Rekomendacja W3C z 10 lutego 2004.

- | | | |
|--|--|---|
| <ul style="list-style-type: none"> Podstawowe obiekty: <ul style="list-style-type: none"> Class Property Individual | <ul style="list-style-type: none"> Definiowanie klas: <ul style="list-style-type: none"> oneOf intersectionOf unionOf własności instancji: <ul style="list-style-type: none"> minCardinality maxCardinality | <ul style="list-style-type: none"> Definiowanie własności: <ul style="list-style-type: none"> TransitiveProperty SymmetricProperty FunctionalProperty inverseOf |
|--|--|---|

2005-01-20 Modelowanie wiedzy

30

Semantyka OWL

- Przykłady:
 - **TransitiveProperty:**
If a property, P, is specified as transitive then for any x, y, and z:
 $P(x, y)$ and $P(y, z)$ implies $P(x, z)$
 - **FunctionalProperty:**
If a property, P, is tagged as functional then for all x, y, and z:
 $P(x, y)$ and $P(x, z)$ implies $y = z$
- Źródło: *OWL Web Ontology Language Guide*, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-guide>
- Możliwość wykorzystania formalnej semantyki w silnikach inferencyjnych (wnioskujących).

RDF – narzędzia

- W3C RDF Validation Service
www.w3.org/RDF/Validator
- RedFoot – An RDF-Centric Application Server (Python)
redfoot.net
- Jena – S Semantic Web Framework for Java
jena.sourceforge.net/
- Sesame – repozytorium, silnik wnioskowania i zapytań (Java)
www.openrdf.org/
- IsaViz – A Visual Authoring Tool for RDF
www.w3.org/2001/11/IsaViz/

RDF ↔ Topic Maps

- Bardzo prosty aparat pojęciowy
- Wsparcie dla ontologii
- W RDF można zakodować dowolną mapę Topic Map
- Skomplikowany aparat pojęciowy
- Brak wsparcia dla ontologii
- Przy pomocy Topic Map można zakodować graf RDF

Równoważna siła wyrazu:

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.*
Tim Berners-Lee, Scientific American, May 2001
- Internetowa infrastruktura publikacji danych:
 - neutralna (niezależna od aplikacji),
 - umożliwiająca przetwarzanie informacji przez programy w celu:
 - automatyzacji,
 - agregacji,
 - wielokrotnego użycia.
- To jest ciągle wizja:
 - zdania nie oznaczają faktów,
 - URI nie identyfikują bytów rzeczywistych,
 - potrzebna jest kodyfikacja *wiedzy codziennej*,
 - pojawiają się problemy związane z etyką oraz bezpieczeństwem.
- Pojawiają się pierwsze technologie.

Gdzie szukać dalej

- Resource Description Framework
www.w3.org/RDF
- Bray, T., What Is RDF?
www.xml.com/pub/a/2001/01/24/rdf.html
- Dublin Core Metadata Initiative
dublincore.org
- RDF Site Summary (RSS) 1.0
web.resource.org/rss/1.0
- The FOAF Project
www.foaf-project.org
- W3C Web Ontology (WebOnt) Working Group
www.w3.org/2001/sw/WebOnt



Gdzie szukać dalej

- Berners-Lee, T., Lassila, O., Hendler, J., *Semantic Web*
 Scientific American, May 2001
www.kbs.uni-hannover.de/Lehre/K11/WS02/trails/skript/modul4/rawdata/Article.html
- The Semantic Web Community Portal
www.semanticweb.org
- W3C Semantic Web Activity
www.w3.org/2001/sw
- topicmap.com – Hand-crafted Machine-generated Knowledge Interchange
www.topicmap.com
- Learn more about Topic Maps
www.ontopia.net/topicmaps/learn_more.html

