

Semantic Web

dr inż. Aleksander Smywiński-Pohl

Elektroniczne Przetwarzanie Informacji
Konsultacje: czw. 14.00-15.30, pokój 3.211

Plan prezentacji

Semantic Web

RDF

RDFS i OWL

Linked Data

Plan prezentacji

Semantic Web

RDF

RDFS i OWL

Linked Data

Motywacja

- ▶ Chcemy dojechać do odległej miejscowości.
- ▶ Jako środek transportu preferujemy autobus.
- ▶ Chcemy aby system komputerowy zaplanował dla nas całą trasę.
- ▶ Chcemy zapłacić jeden raz za całą usługę, niezależnie od przewoźników i wykorzystanych środków transportu.
- ▶ Wszystkie informacje są w Internecie, ale czy można je **automatycznie** przetworzyć?

HTML

```

<table>
  <tr>
    <td>
      106
      <br/>
      LINIA MIEJSKA
    </td>
    <td>
      <b> Łagiewniki</b>
      <br/>
      <b> TRASA: </b>
      ŁAGIEWNIKI - Brożka,
      Grota-Roweckiego,
      Kobierzyńska,
      Babińskiego,
      Skotnicka - SKOTNIKI SZKOŁA
    </td>
  </tr>
</table>

```

106 Łagiewniki		
LINIA MIEJSKA TRASA: ŁAGIEWNIKI - Brożka, Grota-Roweckiego, Kobierzyńska, Babińskiego, Skotnicka - SKOTNIKI SZKOŁA		
Dzień powszedni	Soboty	Święta
5 01 21 43	5 -	5 -
6 05 21 36 50	6 07 42	6 -
7 00 12 21 32 44 56	7 22 47	7 04 39
8 15 29 42 54	8 22 49	8 19 49
9 06 28 52	9 24 48	9 24 49
10 14 43	10 20 50	10 24 54
11 07 34 55	11 24 49	11 24 49
12 17 42	12 21 51	12 24 49
13 06 32 50	13 24 48	13 29 49
14 06 19 30 43 55	14 23 53	14 29 54
15 07 19 31 44 57	15 18 44	15 29 49
16 09 21 31 45 57	16 11 48	16 24 54
17 09 21 33 54	17 18 48	17 29 59
18 16 41	18 18 53	18 29 59
19 08 31 51	19 18 59	19 34
20 21 43	20 29	20 01 50
21 08 31 55	21 24	21 -
22 43	22 -	22 -
Zakłócenia w ruchu powodują zmiany czasów odjazdów. Rozkład jazdy ważny od 31.08.2013 do odwołania MPK Kraków		

HTML

```

<h2>
  <span>Miejscowość</span>
  <span>Odjazdy</span>
</h2>
<div>
  <p>Andrychów</p>
</div>
<div>
  <span> 20:50</span>
</div>

```

Miejscowość	Odjazdy
Andrychów	20:50  
Bełchatów	przez: Olszusz, Częstochowa, Radomsko, 16:00   21:00  
Będzin	przez: Dąbrowa Górnicza, 17:50  
Białystok	przez: Sandomierz, Lublin, 20:00  
Bielsko-Biała	05:35   06:10   07:40   08:40   09:45   10:20   10:55   11:40   12:20   12:25   13:20   13:35   14:40   16:00   16:35   17:15   17:20   18:15   19:20   19:20   19:55   21:45   21:50   przez: Katowice, 09:30  
Biłgoraj	przez: Tamobrzeg, Stalowa Wola, 14:05   przez: Pacanów, Tamobrzeg, Stalowa Wola, 13:30   15:30  

XML

MPK

```
<?xml version="1.0" ?>
<busInfo>
  <id>106</id>
  <route>
    <start>Łagiewniki</start>
    <end>Skotniki</end>
  </route>
  <timetable>
    <day>
      <name>poniedziałek</name>
      <hours>
        <hour>5:01</hour>
        <hour>5:21</hour>
        <hour>5:43</hour>
      </hours>
    </day>
  </timetable>
</busInfo>
```

PKS

```
<?xml version="1.0" ?>
<buses>
  <bus>
    <from>Kraków</from>
    <to>Andrychów</to>
    <departures>
      <days>all</days>
      <time hour="20" minutes="50">
    </departures>
    <arrivals>
      <days>all</days>
      <time hour="22" minutes="10">
    </arrivals>
  </bus>
</buses>
```

Semantic Web - Internet dla ludzi i maszyn

- ▶ Wykorzystanie zalet HTML – adresy URL oraz odnośniki
 - ▶ „Kraków”
`http://dbpedia.org/page/Kraków`
- ▶ Odnośniki posiadają znaczenie → RDF
 - ▶ „jest”
`http://www.w3.org/1999/02/22-rdf-syntax-ns#type`
- ▶ Znaczenie jest wspólne → RDFS, OWL
 - ▶ „przystankiem”
`http://sw.opencyc.org/concept/Mx4rwQBS...`

Plan prezentacji

Semantic Web

RDF

RDFS i OWL

Linked Data

RDF – Resource Description Framework

- ▶ Każdy **obiekt**, który chcemy opisać posiada adres URL
 - ▶ jako zasób
`http://dbpedia.org/resource/Cher`
 - ▶ lub jego fragment
`http://dbpedia.org/resource/Cher#name`
- ▶ Każdy **fakt** zamieniamy na trójkę:
 - ▶ podmiot
 - ▶ orzeczenie (predykat)
 - ▶ dopełnienie

Obiekt – kotek



Rysunek : Flickr/CC/jillhudsons

- ▶ Kilka faktów na temat kotka
 - ▶ imię Mruczuś
 - ▶ urodzony 20-01-2010
 - ▶ oczka szare

Jak przenieść kotka do Semantic Web?

- ▶ `http://mojadres.pl/mojkotek` – URL dla kotka
- ▶ `<http://mojadres.pl/mojkotek> rdfs:label "Mruczuś".`

Jak przenieść kotka do Semantic Web?

- ▶ `http://mojadres.pl/mojkotek` – URL dla kotka
- ▶ `<http://mojadres.pl/mojkotek> rdfs:label "Mruczuś".`

Jak przenieść kotka do Semantic Web?

- ▶ `http://mojadres.pl/mojkotek` – URL dla kotka
- ▶ `<http://mojadres.pl/mojkotek> rdfs:label`
`"Mruczuś".`

Jak przenieść kotka do Semantic Web?

- ▶ `http://mojadres.pl/mojkotek` – URL dla kotka
- ▶ `<http://mojadres.pl/mojkotek> rdfs:label "Mruczuś".`

Typowy zapis

Format N-triples:

- ▶ `<http://mojadres.pl/mojkotek> rdfs:label "Mruczuś".`
- ▶ `<http://mojadres.pl/mojkotek> rdf:type <http://dbpedia.org/resource/Cat> .`
- ▶ `<http://mojadres.pl/mojkotek> dbpprop:dateOfBirth 2010-01-20 .`
- ▶ `<http://mojadres.pl/mojkotek> myontology:eyesColor <http://dbpedia.org/resource/Green> .`

Można zwięźlej

Format turtle:

```
@base <http://mojadres.pl/> .  
<mojkotek> rdfs:label "Mruczuś",  
    rdf:type  
        <http://dbpedia.org/resource/Cat>,  
    dbpprop:dateOfBirth 2010-01-20,  
    myontology:eyesColor  
        <http://dbpedia.org/resource/Green>.
```

Można rozwleklej

RDF+XML

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-...">
  <rdf:Description rdf:about="http://mojadres.pl/mojkotek">
    <rdfs:label xml:lang="pl">Mruczuś</rdfs:label>
    <dbpprop:dateOfBirth>2010-01-20</dbpprop:dateOfBirth>
    <myontology:eyesColor
      rdf:resource="http://dbpedia.org/page/Green"/>
  </rdf:Description>
</rdf:RDF>
```


Plan prezentacji

Semantic Web

RDF

RDFS i OWL

Linked Data

Jak to wszystko uporządkować?

- ▶ `rdf:type`
- ▶ `rdfs:label`
- ▶ `dbpprop:dateOfBirth`
- ▶ `myontology:eyesColor`
- ▶ ...

RDF Schema (RDFS)

- ▶ Definiuje pojęcia:
 - ▶ zasobu `rdfs:Resource`
 - ▶ klasy `rdfs:Class`
 - ▶ literału `rdfs:Literal`
 - ▶ typu danej `rdfs:Datatype`
 - ▶ własności `rdf:Property`
- ▶ oraz własności:
 - ▶ domeny `rdfs:domain`
 - ▶ zakresu `rdfs:range`
 - ▶ typu `rdf:type`
 - ▶ pod-klasy `rdfs:subClassOf`
 - ▶ pod-własności `rdfs:subPropertyOf`
 - ▶ etykiety `rdfs:label`
 - ▶ komentarza `rdfs:comment`

OWL

- ▶ Znacznie bogatszy niż RDFS
- ▶ Pozwala na zaawansowane wnioskowanie
- ▶ Trzy wersje:
 - ▶ **Lite** – pomyślana głównie do konwersji wcześniej utworzonych taksonomii
 - ▶ **Description Logic (DL)** – gwarantuje obliczalność, znacznie bardziej ekspresywny niż Lite
 - ▶ **Full** – najbardziej ekspresywny, nie gwarantuje obliczalności

OWL – pojęcia i własności

- ▶ OWL definiuje podstawowe pojęcia:

- ▶ klasa `owl:Class`
- ▶ instancja `owl:Instance`

oraz własności

- ▶ synonimiczna klasa `owl:equivalentClass`
- ▶ synonimiczna własność `owl:equivalentProperty`
- ▶ to samo `owl:sameAs`
- ▶ coś innego `owl:differentFrom`
- ▶ rozłączność klas (DL) `disjointWith`
- ▶ ...

OWL – klasy i instancje

► klasa

- pojęcie, abstrakcyjny zbiór obiektów określonego typu, np. *kot domowy* jako gatunek biologiczny
- pojęcia mogą tworzyć taksonomię wykorzystując predykat `subClassOf`,
np. `<http://example.pl/kot> owl:subClassOf
<http://example.pl/zwierzę>`

► instancja

- obiekt, konkretny przedmiot, zjawisko, coś co możemy wskazać, np. kot, który był na zdjęciu wcześniej
- `rdf:type` wiąże instancję z jej typem (klasą),
np. `<http://mojadres/mojkot> rdf:type
<http://example.pl/kot>`

Plan prezentacji

Semantic Web

RDF

RDFS i OWL

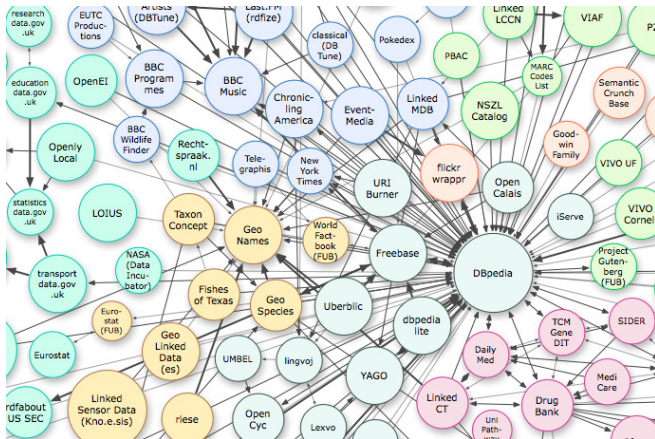
Linked Data

Czy ktoś z tego korzysta?

Semantic Web



Linked Data



Rysunek : linedata.org

Kto konkretnie?

- ▶ NewYork Times
- ▶ BBC
- ▶ Citeseer
- ▶ flickr
- ▶ Project Gutenberg
- ▶ ACM
- ▶ IEEE
- ▶ IBM
- ▶ MySpace

Ciekawsze ontologie

- ▶ FOAF – opisuje znajomości międzyludzkie
- ▶ Geonames – zależności geograficzne
- ▶ Music Ontology – przemysł muzyczny
- ▶ BIBO – dane związane z nauką
- ▶ DBpedia – zawiera dane wyekstrahowane z Wikipedii
- ▶ YAGO – jw. ale powiązane z WrodNetem
- ▶ Cyc – ontologia w całości konstruowana ręcznie, pozwala na faktyczne prowadzenie wnioskowań

Jak z tego korzystać?

- ▶ SPARQL – język zapytań dla danych grafowych
- ▶ np. <http://dbpedia.org/sparql>
- ▶ podobny do SQL:
 - ▶ `select distinct ?x where {?x rdf:type <http://xmlns.com/foaf/0.1/Person> .}`
`limit 50`
 - ▶ `select distinct ?x, ?l where {?x rdf:type <http://xmlns.com/foaf/0.1/Person> . ?x rdfs:label ?l. FILTER (lang(?l) = "pl")}` `limit 50`
- ▶ operujemy jednak na grafie a nie na tabelach!

DZIĘKUJĘ!