

Lab 1 – Introduction to the Semantic Web

1. Extracting RDF metadata:

The original RDF/XML document

```
1: <?xml version="1.0"?>
2: <rdf:RDF
3:   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4:   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
5:   xmlns:dc="http://purl.org/dc/elements/1.1/"
6:   xmlns:foaf="http://xmlns.com/foaf/0.1/"
7:   xmlns:image="http://jibbering.com/vocabs/image/#"
8:   xmlns:an="http://www.w3.org/2000/10/annotation-ns#">
9:   <foaf:Image
rdf:about="http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg">
10:     <image:hasPart>
11:       <image:Rectangle rdf:ID="p1">
12:         <image:points>173,139 255,200</image:points>
13:         <dc:title>Kaczka 1</dc:title>
14:         <dc:description>Kaczka pierwsza</dc:description>
15:         <image:depicts rdf:parseType="Resource">
16:           <dc:description>Kaczka pierwsza</dc:description>
17:         </image:depicts>
18:       </image:Rectangle>
19:     </image:hasPart>
20:     <image:hasPart>
21:       <image:Rectangle rdf:ID="p2">
22:         <image:points>54,102 133,153</image:points>
23:         <dc:title>Kaczka 2</dc:title>
24:         <dc:description>Kaczka druga</dc:description>
25:         <image:depicts rdf:parseType="Resource">
26:           <dc:description>Kaczka druga</dc:description>
27:         </image:depicts>
28:       </image:Rectangle>
29:     </image:hasPart>
30:     <image:hasPart>
31:       <image:Rectangle rdf:ID="p3">
32:         <image:points>159,19 203,61</image:points>
33:         <dc:title>Kaczka 3</dc:title>
34:         <dc:description>Kaczka trzecia</dc:description>
35:         <image:depicts rdf:parseType="Resource">
36:           <dc:description>Kaczka trzecia</dc:description>
37:         </image:depicts>
38:       </image:Rectangle>
39:     </image:hasPart>
40:   </foaf:Image>
41: </rdf:RDF>
```

Triples of the Data Model

Nr	Subject	Predicate	Object
1	http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://xmlns.com/foaf/0.1/Image
2	http://www.w3.org/RDF/Validator/run/1336425588063#p1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://jibbering.com/vocabs/image/#Rectangle
3	http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg	http://jibbering.com/vocabs/image/#hasPart	http://www.w3.org/RDF/Validator/run/1336425588063#p1
4	http://www.w3.org/RDF/Validator/run/1336425588063#p1	http://jibbering.com/vocabs/image/#points	"173,139 255,200"
5	http://www.w3.org/RDF/Validator/run/1336425588063#p1	http://purl.org/dc/elements/1.1/title	"Kaczka 1"
6	http://www.w3.org/RDF/Validator/run/1336425588063#p1	http://purl.org/dc/elements/1.1/description	"Kaczka pierwsza"
7	http://www.w3.org/RDF/Validator/run/1336425588063#p1	http://jibbering.com/vocabs/image/#depicts	genid:A15826
8	genid:A15826	http://purl.org/dc/elements/1.1/description	"Kaczka pierwsza"
9	http://www.w3.org/RDF/Validator/run/1336425588063#p2	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://jibbering.com/vocabs/image/#Rectangle
10	http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg	http://jibbering.com/vocabs/image/#hasPart	http://www.w3.org/RDF/Validator/run/1336425588063#p2
11	http://www.w3.org/RDF/Validator/run/1336425588063#p2	http://jibbering.com/vocabs/image/#points	"54,102 133,153"
12	http://www.w3.org/RDF/Validator/run/1336425588063#p2	http://purl.org/dc/elements/1.1/title	"Kaczka 2"
13	http://www.w3.org/RDF/Validator/run/1336425588063#p2	http://purl.org/dc/elements/1.1/description	"Kaczka druga"
14	http://www.w3.org/RDF/Validator/run/1336425588063#p2	http://jibbering.com/vocabs/image/#depicts	genid:A15827
15	genid:A15827	http://purl.org/dc/elements/1.1/description	"Kaczka druga"

16	http://www.w3.org/RDF/Validator/run/1336425588063#p3	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://jibbering.com/vocabs/image/#Rectangle
17	http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg	http://jibbering.com/vocabs/image/#hasPart	http://www.w3.org/RDF/Validator/run/1336425588063#p3
18	http://www.w3.org/RDF/Validator/run/1336425588063#p3	http://jibbering.com/vocabs/image/#points	"159,19 203,61"
19	http://www.w3.org/RDF/Validator/run/1336425588063#p3	http://purl.org/dc/elements/1.1/title	"Kaczka 3"
20	http://www.w3.org/RDF/Validator/run/1336425588063#p3	http://purl.org/dc/elements/1.1/description	"Kaczka trzecia"
21	http://www.w3.org/RDF/Validator/run/1336425588063#p3	http://jibbering.com/vocabs/image/#depicts	genid:A15828
22	genid:A15828	http://purl.org/dc/elements/1.1/description	"Kaczka trzecia"

Graph of the data model



2. Common vocabularies:

- **SKOS** – Simple Knowledge Organisation System – służy m.in do reprezentacji słowników, schematów klasyfikacji, taksonomii.
Zastosowanie: internetowy katalog Biblioteki Kongresu.
- **Dublin Core** – składający się z 15 elementów ogólny standard opisu metadanych.
Zastosowanie: biblioteki cyfrowe systemu dLibra.
- **FOAF** – Friend of a Friend – schemat formatowania danych służących do opisu danej osoby oraz jej aktywności. Umożliwia odwołania do innych dokumentów FOAF tworząc w ten sposób swego rodzaju sieć społecznościową.
Zastosowanie: DeadJournal – serwis bloggerski,
Live_Journal – serwis bloggerski,
Indent.ca – sieć społecznościowa, mikroblogi,
FriendFeed – serwis społecznościowy,
MyOpera – sieć społecznościowa.

- **DOAP** – Description of a project – schemat formatowania danych służących do opisu projektów programistycznych.

Zastosowanie: Freecode – serwis katalogujący oprogramowanie open source.

3. FOAF:

```
<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:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic rdf:resource="#me"/>
  <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="me">
<foaf:name>Piotr Kamoda</foaf:name>
<foaf:title>Mr</foaf:title>
<foaf:givenname>Piotr</foaf:givenname>
<foaf:family_name>Kamoda</foaf:family_name>
<foaf:nick>kmo</foaf:nick>
<foaf:mbox_sha1sum>af8e6a871a932b7d959fe6d7da35b31852474324</foaf:mbox_sha1sum>
<foaf:homepage rdf:resource="http://www.agh.edu.pl"/>
<foaf:phone rdf:resource="tel:123789543"/>
<foaf:schoolHomepage rdf:resource="http://www.agh.edu.pl"/>
<foaf:knows>
<foaf:Person>
<foaf:name>Jan Kowalski</foaf:name>
<foaf:mbox_sha1sum>587f63f68c95c40a2bec6a06d2885d80c6dbfddd</foaf:mbox_sha1sum></foaf:
Person></foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>Piotr Nowak</foaf:name>
<foaf:mbox_sha1sum>587f63f68c95c40a2bec6a06d2885d80c6dbfddd</foaf:mbox_sha1sum></foaf:
Person></foaf:knows></foaf:Person>
</rdf:RDF>
```

URL: <http://dl.dropbox.com/u/11924760/file.rdf>

4. Schemas:

Wszystkie zawarte ontologie zostały utworzone w latach 2000-2003. Podstawową różnicą pomiędzy nimi jest ich rozmiar. Znajdziemy tam ontologie o bardzo rozbudowanej strukturze (kilkaset elementów), jak i składające się zaledwie z kilku (1-2) elementów.

5. Ontology visualization:

Aplikacja niedostępna.

6. Ontology reasoning:

Aplikacja niedostępna.

7. Linked Open Data:

- **DBpedia** – ustrukturyzowane informacje pochodzące z Wikipedii i powiązania pomiędzy nimi.
- **TWC LOGD** – projekt zawierający powiązane informacje rządowe.
- **YAGO 2** – ogromna baza wiedzy pochodzącej m.in. z Wikipedii, WordNet, GeoNames.
- **Dailymed** – publikowany przez National Library of Medicine serwis udostępniający informacje na temat rynku leków.
- **Last.fm RDFizer** – serwis wyszukujący informacje zawarte w serwisie Last.fm.

8. Semantic Web Tools:

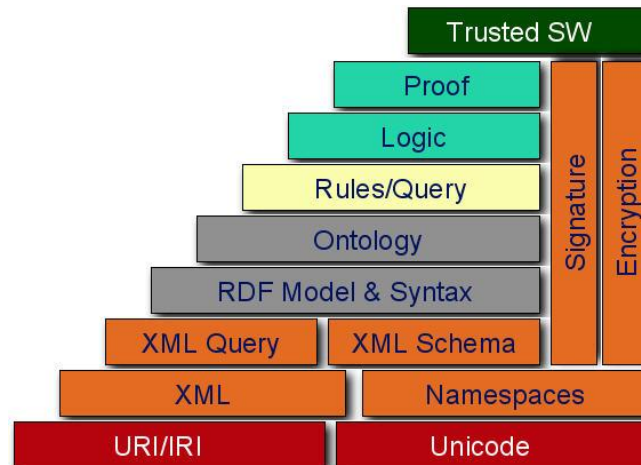
- **RDFizers** – najciekawszymi wydają się być Java->RDF, Javadoc->RDF, Jira->RDF, Subversion->RDF. Wykorzystują one technologie i narzędzia najczęściej przeze mnie używane w związku z czym zapewne ich wykorzystanie mogłoby usprawnić codzienną pracę. Ze względu na zainteresowania muzyczne najchętniej widziałbym RDFizery dla najbardziej powszechnych formatów plików muzycznych.
- **Semantic Web browsers** – na stronie dostępnych jest 13 przeglądarek, z których większość ciągle jest w użyciu.
- **Semantic Web Search Engines** – większość wyszukiwarek jest prosta w użyciu. Ich obsługa jest intuicyjna. W odpowiednie pole należy wprowadzić szukaną frazę, a mechanizm wyszukiwania wyświetli wyniki po przeszukaniu plików RDF.
- **Applications** – większość z zaprezentowanych aplikacji jest ciekawa jednak najciekawszą wydaje się być BBC Music. Przejrzysta i łatwa w nawigacji pozwala na wyszukiwanie informacji ze świata muzyki.

9. Control questions:

- **How one can add semantic annotation to a web page?**

Można to zrobić poprzez dodanie odpowiednich metadanych ręcznie, bądź wykorzystując aplikacje generujące dane w odpowiednim formacie np. RDF. Aby uniknąć błędów składniowych bezpieczniej jest wykorzystać gotowe generatory.

- **Explain the Semantic Web Stack of technologies.**



Semantic Web Stack ilustruje hierarchie technologii wchodzących w skład Semantic Web. Każda warstwa wykorzystuje zalety warstwy poniżej, tzn. że do opisu bazy wiedzy wykorzystywane są relacje pomiędzy odpowiednimi technologiami.

XML – jest standardem pozwalającym w ogólności na zapis danych;

XML Query/XML Schema – wprowadzają ograniczenia i restrykcje dotyczące typu i struktury danych;

RDF – jest językiem do opisu modeli danych, bazujący na XML-u;

RDF Schema – wprowadza ograniczenia i restrykcje dotyczące podziału modelu na klasy i podklasy;

OWL/Ontology – wprowadza dodatkowe słownictwo opisujące właściwości i klasy jak i relacje zachodzące między nimi.

- **What is the main syntax for RDF? What are its advantages over other syntaxes?**

Dokumenty RDF oparte są na składni XML. XML jest elastyczny i prosty w użyciu. Wspierany przez języki programowania.

- **What is the role of the ontologies in the Semantic Web?**

Ontologie służą do określania znaczenia oraz opisu zasobów informacyjnych. Ontologie budują Semantic Web.

- **What are the 4 principles of Linked Data?**

URI – identyfikacja zasobów;
HTTP URI – umożliwia dostęp do zasobów przez agentów i inne osoby;
Używanie ustandaryzowanych formatów RDF/XML do kodowania użytecznych informacji;
Łączenie powiązanych zasobów poprzez linki.

Lab 2 – XML

1. Warm up: Well formedness of the XML:

`<a><a>`

Niepoprawne. Brak zamknięcia znacznika `<a>`.

`<a>foo<a>bar`

Niepoprawne. Brak znacznika nadrzędnego.

`<a>foobar`

Niepoprawne. Błędna kolejność znaczników.

`<a><c>foobar</c>`

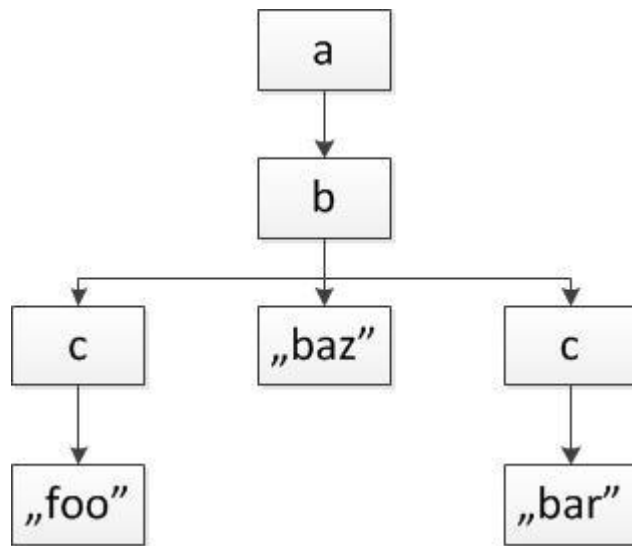
Niepoprawne. Błędna kolejności znaczników.

`<a/>foobar`

Niepoprawne. Brak zamknięcia nadrzędnego znacznika

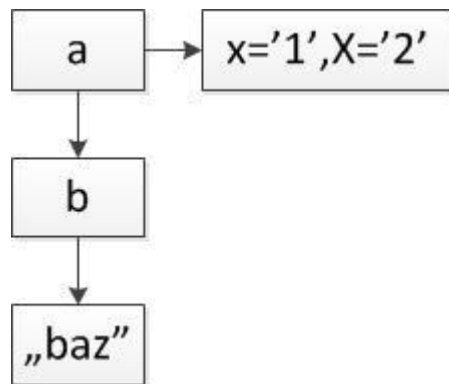
`<a><c>foo</c>baz<c>bar</c>`

Poprawne.



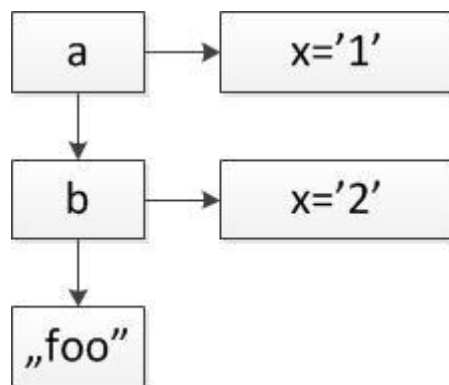
`foo`

Poprawne.



`<b x="2">foo`

Poprawne.

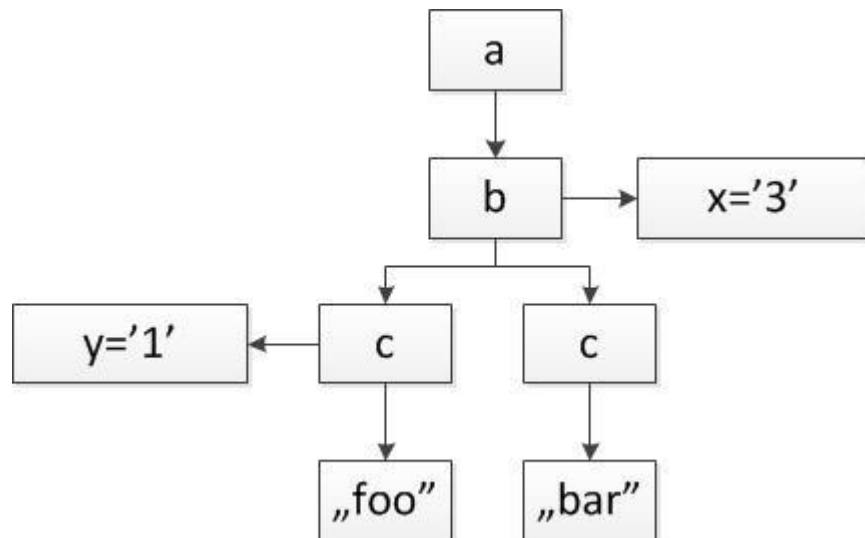


`foo`

Niepoprawne. Dwukrotnie zdefiniowany parameter x znacznika <a>.

`<a><b x="3"><c y="1">foo</c><c>bar</c>`

Poprawne.



2. Creating XML document:

```

<book>
  <author>Bret Easton Ellis</author>
  <title>American Psycho</title>
  <date>2011</date>
  <publisher>vis-a-vis</publisher>
</book>

```

3. DTD:

```

<!DOCTYPE book [
  <!ELEMENT book (author+, title, date, publisher)>
  <!ELEMENT author (#PCDATA)>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT date (#PCDATA)>
  <!ELEMENT publisher (#PCDATA)>
]>

```

4. XSD:

```

<?xml version="1.0" encoding="utf-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="book">

```

```
<xs:complexType>
<xs:sequence>
  <xs:element name="title" type="xs:string"/>
  <xs:element name="publisher" type="xs:string"/>
  <xs:element name="year" type="xs:string"/>
  <xs:element name="authors">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="author" type="xs:string"
minOccurs="1" maxOccurs="unbounded"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:sequence>
</xs:complexType>
</xs:element>
</xs:schema>
```