

**Podstawy Inżynierii Wiedzy**

Informatyka Stosowana

Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki



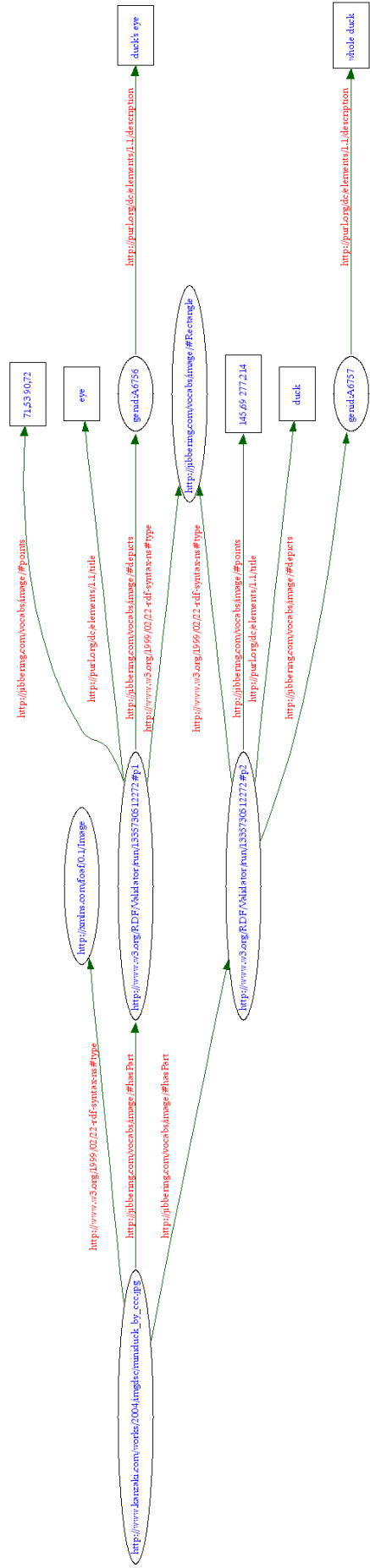
**AGH**

# **Sprawozdanie nr 1: Wprowadzenie i XML**

Jarosław Bodnar

Mikołaj Mesjasz

Grupa: środa 8.00



## Triples of the Data Model

Number	Subject	Predicate	Object
1	<a href="http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg">http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg</a>	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://xmlns.com/foaf/0.1/Image">http://xmlns.com/foaf/0.1/Image</a>
2	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p1">http://www.w3.org/RDF/Validator/run/1335730512272#p1</a>	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://jibbering.com/vocabs/image/#Rectangle">http://jibbering.com/vocabs/image/#Rectangle</a>
3	<a href="http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg">http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg</a>	<a href="http://jibbering.com/vocabs/image/#hasPart">http://jibbering.com/vocabs/image/#hasPart</a>	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p1">http://www.w3.org/RDF/Validator/run/1335730512272#p1</a>
4	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p1">http://www.w3.org/RDF/Validator/run/1335730512272#p1</a>	<a href="http://jibbering.com/vocabs/image/#points">http://jibbering.com/vocabs/image/#points</a>	"71,53 90,72"
5	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p1">http://www.w3.org/RDF/Validator/run/1335730512272#p1</a>	<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>	"eye"
6	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p1">http://www.w3.org/RDF/Validator/run/1335730512272#p1</a>	<a href="http://jibbering.com/vocabs/image/#depicts">http://jibbering.com/vocabs/image/#depicts</a>	genid:A6756
7	genid:A6756	<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	"duck's eye"
8	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p2">http://www.w3.org/RDF/Validator/run/1335730512272#p2</a>	<a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#type">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</a>	<a href="http://jibbering.com/vocabs/image/#Rectangle">http://jibbering.com/vocabs/image/#Rectangle</a>
9	<a href="http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg">http://www.kanzaki.com/works/2004/imgdsc/miniduck_by_ccc.jpg</a>	<a href="http://jibbering.com/vocabs/image/#hasPart">http://jibbering.com/vocabs/image/#hasPart</a>	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p2">http://www.w3.org/RDF/Validator/run/1335730512272#p2</a>
10	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p2">http://www.w3.org/RDF/Validator/run/1335730512272#p2</a>	<a href="http://jibbering.com/vocabs/image/#points">http://jibbering.com/vocabs/image/#points</a>	"145,69 277,214"
11	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p2">http://www.w3.org/RDF/Validator/run/1335730512272#p2</a>	<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>	"duck"
12	<a href="http://www.w3.org/RDF/Validator/run/1335730512272#p2">http://www.w3.org/RDF/Validator/run/1335730512272#p2</a>	<a href="http://jibbering.com/vocabs/image/#depicts">http://jibbering.com/vocabs/image/#depicts</a>	genid:A6757
13	genid:A6757	<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	"whole duck"

2.

**SKOS** – system organizacji wiedzy, pozwala na opis powiązań między pojęciami, łączenie ich z innymi danymi. SKOS jest wykorzystywany w tezaurusach, taksonomii oraz w innych typach struktur, organizujących słownictwo. Programy wykorzystujące SKOS:

- ontologia GeoName do reprezentacji danych geograficznych,
- tezaurus w SWED Environmental Directory.

**Dublin Core** – standard metadanych, który może służyć do opisu zasobów internetowych  
Wykorzystanie DC:

- specyfikacja OMF(Open Source Metadata Framework), wykorzystywana pośrednio w Gnome czy KDE.
- PBCore – rozszerza DC o elementy do opisu contentu audio-wideo.
- Zope Content Management Framework – framework do zarządzania contentem, pozwala na katalogowanie, personalizację contentu.

**FOAF** (Friend of a friend) – ontologia opisująca osoby, relacje pomiędzy nimi i ich aktywność/działania. Pozwala na opis relacji pomiędzy ludźmi bez użycia centralnej bazy danych.  
Wykorzystanie FOAF:

- Platformy blogowe: Live Journal, DeadJournal wspierają powiązania pomiędzy użytkownikami wykorzystując FOAF
- Społeczność MyOpera, związana z przeglądarka Opera
- Yandex – wyszukiwanie blogów

**DOAP** (Description of a Project) – schemat RDF wraz ze znacznikami xml, służące do opisu projektów informatycznych (głównie Open Source). Wykorzystywany w projektach Mozilla Foundation oraz na stronie Freecode (dawniej Freshmeat), umożliwiającej wyszukiwanie oprogramowania.

3.

Plik rdf dostępny pod adresem: <http://student.agh.edu.pl/~seelus/Other/foaf.rdf>

```
<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>Jan Kowalski</foaf:name>
<foaf:title>Mr</foaf:title>
<foaf:givenname>Jan</foaf:givenname>
```

```


<foaf:family_name>Kowalski</foaf:family_name>
<foaf:nick>kowal</foaf:nick>
<foaf:mbox_sha1sum>1258e13a2a755aeb1d959b557f96b9b8c2379834</foaf:mbox_sha1sum>
<foaf:phone rdf:resource="tel:123456789"/>
<foaf:schoolHomepage rdf:resource="AGH"/>
<foaf:knows>
<foaf:Person>
<foaf:name>Tomasz Nowak</foaf:name>
<foaf:mbox_sha1sum>27b523021db59a46d4bfe37905c8221f926c7891</foaf:mbox_sha1sum></foaf:
:Person></foaf:knows>
<foaf:knows>
<foaf:Person>
<foaf:name>Marek Kwiatkowski</foaf:name>
<foaf:mbox_sha1sum>9c0aafa56097d608e8828750651a9192d54f4e81</foaf:mbox_sha1sum></foaf
:Person></foaf:knows></foaf:Person>
</rdf:RDF>


```

Screen z wizualizacji:

## Jan Kowalski

### Knows:

Marek Kwiatkowski  *sha1sum of a personal mailbox URI name:*  
9c0aafa56097d608e8828750651a9192d54f4e81

Tomasz Nowak  *sha1sum of a personal mailbox URI name:*  
27b523021db59a46d4bfe37905c8221f926c7891

***schoolHomepage:***  
<http://student.agh.edu.pl/~seelus/Other/AGH>

***phone:***  
<tel:123456789>

***sha1sum of a personal mailbox URI name:***  
1258e13a2a755aeb1d959b557f96b9b8c2379834


***nickname:***  
kowal


***family\_name:***  
Kowalski


***Given name:***  
Jan

***title:***  
Mr

#### Referenced Vocabularies

 <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
The RDF Vocabulary (RDF)

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

 <http://webns.net/mvcb/>

4. Ontologie z katalogu powstały w latach 2000 - 2003. Ich wielkość jest różna, większość z nich zawiera kilkanaście - kilkadziesiąt rekordów.

5, 6. Strona programu owl:sight nie jest dostępna (sprawdzone codziennie przez ostatni tydzień), co uniemożliwia wykonanie ćwiczeń.

## 7. Linked Data

- Open Library – otwarty katalog książek, udostępniający użytkownikom możliwości dodawania, edycji informacji o książkach. Mottem twórców „biblioteki” jest: „One web page for every book ever published”. Obecnie w katalogu zgromadzonych jest ok. 20 mln rekordów.
- UK Postcodes – zbiór zawierający wszystkie kody pocztowe w Wielkiej Brytanii. Możliwy jest ich eksport do formatów: XML, HTML, JSON, RDF oraz CSV.
- ProductDB – celem twórców projektu jest zbudowanie największej na świecie bazy danych o produktach i powiązaniach pomiędzy nimi. Dane zgromadzone w ramach projektu pochodzą z przetwarzania i analizy informacji z serwisów: ProductWiki, MusicBrainz, dbpedia, FreeBase i OpenLibrary oraz wielu innych.
- Mortality EnAkTing – zawiera statystyki dotyczące śmiertelności w Wielkiej Brytanii w latach 2008-2009. Dane pochodzą z UK Home Office (Ministerstwo Spraw Wewnętrznych Wielkiej Brytanii)
- GeoSpecies – zawiera dane o gatunkach biologicznych, wraz z odnośnikami do google, Wikipedii i innych źródeł, pozwalających poszerzyć wiedzę o przedstawicielach różnych gatunków.

8.1. RDFizers – z punktu widzenia programisty Javy bardzo interesująca wydaje się możliwość wygenerowania RDF-a z javadoc’a. Wydaje się, że taki RDF pozwoliłby na lepszą orientację w projekcie (zwłaszcza złożonym) poprzez umożliwienie szybkiego sprawdzenia czy metoda realizująca pożądaną funkcjonalność już istnieje, czy też konieczne jest jej samodzielne napisanie.

W naszej opinii ciekawym i dość prostym do realizacji konwerterem byłby MP3->RDF. W MP3 (i innych plikach muzycznych) istnieją systemy tagów jak ID3, z których możliwe byłoby wydobycie informacji i zapisanie w postaci znaczników rdf. Poprzez system powiązań możliwe byłoby chociażby stworzenie osobistej płytoteki, a szerzej nawet bazy danych wykonawców, płyt, gatunków muzycznych itp.

## 8.2.

### Niedziałające(6):

- Razorbase – domena nie została przedłużona – na stronie znajdują się reklamy
- Tabulator – przy próbie uruchomienia najnowszej wersji błąd: Failed to get privileges: Error: Skrypt z -<http://dig.csail.mit.edu> nie otrzymał uprawnień UniversalXPConnect UniversalBrowserRead.
- Zitgist RDF Browser – za każdym razem komunikat o braku informacji: Zitgist has no information to display about that subject.
- Objectviewer – strona nie działa
- iLOD – strona nie działa
- zLinks – w pop-upach nie pojawiają się żadne linki

### Działające(5):

- Disco
- OpenLinkDataExplorer – dostępne pod adresem <http://uriburner.com/ode/>
- DBpedia Mobile
- Sigma
- Quick & Dirty RDF Browser

### Inne:

Dla domyślnych danych nie działa <http://wiki.dbpedia.org/Marbles?v=71e#h53-3> jednak już dla naszego pliku RDF zwracane są poprawne dane.

## 8.3.

W większości działających mechanizmów wyszukiwujących wystarczy wpisać nazwę firmy/przedmiotu/osobę, o których informacji poszukujemy. Wyszukiwanie przy pomocy „semantycznych” wyszukiwarek od strony użytkownika nie różni się niczym od obecnie wykorzystywanych, standardowych wyszukiwarek internetowych typu Google.

## Control Questions:

1. Możliwe jest dodanie takiej adnotacji ręcznie (korzystając z języka RDF) lub wykorzystanie narzędzi, które pozwalają na automatyczne jej wygenerowanie (np. flickr2rdf).
2. Semantic Web Stack – przedstawia hierarchię języków, wykorzystywanych przy tworzeniu i działaniu sieci semantycznych. Ponadto, demonstruje on sposób komunikacji pomiędzy poszczególnymi warstwami: każda warstwa korzysta z informacji udostępnionych przez warstwę „niższe” w hierarchii.
3. Składnia RDF oparta jest na „trójkach”: podmiot – orzeczenie – dopełnienie (funkcje jak w zdaniu) . Ponadto, należy wspomnieć, że RDF opiera się na xml-u, którego największymi zaletami są: niezależność od platformy sprzętowej i łatwość w tworzeniu opisu zasobów, a co za tym idzie łatwiejsze zrozumienie takiego opisu.
4. Ontologie reprezentują model dziedziny wiedzy. W skład modelu danej dziedziny wchodzi:
  - zbiór pojęć
  - relacje między pojęciami

W ten sposób w ontologii opisana jest dana dziedzina wiedzy, jednakże model może służyć również do wnioskowania o właściwościach opisywanych pojęć. Ponadto, ontologie są wykorzystywane do organizacji i nawigacji pomiędzy stronami.

5. 4 zasady Linked Data:
  - Do identyfikacji zasobów służą adresy URI,
  - Dzięki wykorzystaniu adresów URI w oparciu o protokół http dostęp do zasobów mają również inni ludzie i user agent'y,
  - Informacje powinny być udostępniane w formacie XML lub RDF.
  - Udostępniane dane powinny być powiązane z już istniejącymi zasobami w celu ułatwienia szukania powiązanych informacji.

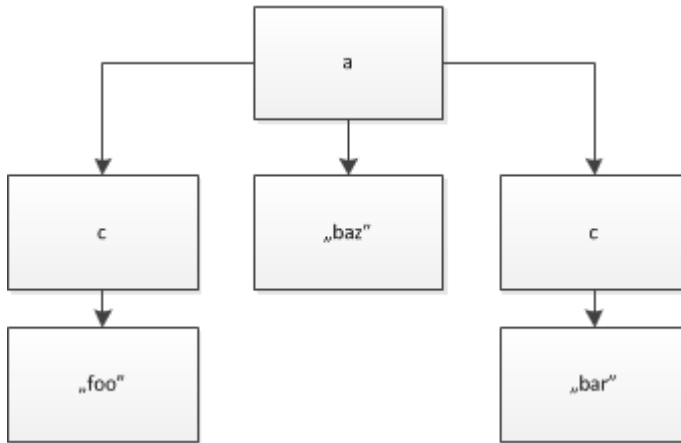


# XML

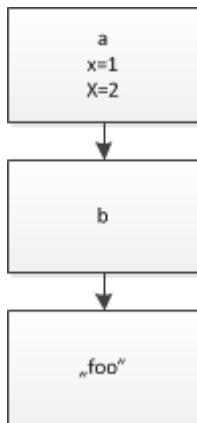
1.

Poprawne:

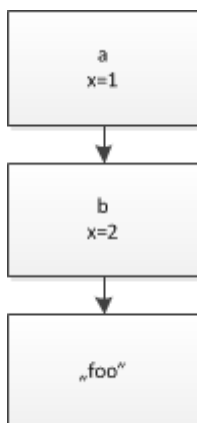
- `<a><b><c>foo</c>baz<c>bar</c></b></a>`



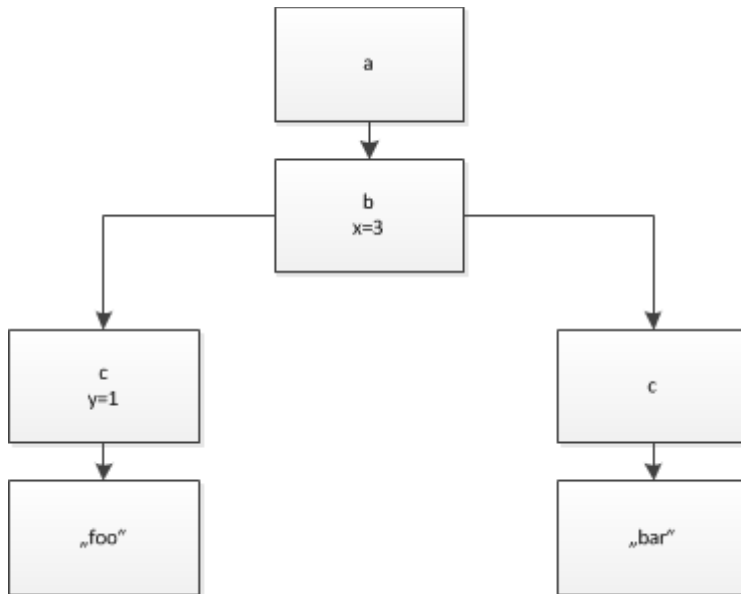
- `<a x="1" X="2"><b>foo</b></a>`



- `<a x="1"><b x="2">foo</b></a>`



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



Błędne:

- `<a><b/><a>` - zamiast pogrubionego znacznika otwierającego powinien znaleźć się znacznik zamykający
- `<a><b>foo</b></a><a>bar</a>` - pogrubiony znacznik a nie ma komplementarnego znacznika otwierającego
- `<a><b>foo</b><b>bar</a>` - pogrubiony znacznik nie jest zamknięty
- `<a><b><c>foo</b>bar</c></a>` - znacznik `<b>` jest zamknięty przed znacznikiem `<c>`
- `<a/><b>foo</b><b>bar<b>` - znacznik korzenia jest zamknięty przed deklaracją innych znaczników
- `<a x="1" y="2" x="3"><b>foo</b></a>` - zdublowany atrybut x

2.

`<book>`

`<author>Robert Ludlum</author>`

`<translator>Jan Krasko</translator>`

`<title>The Janson Directive</title>`

`<plTitle>The Janson Directive</plTitle>`

`<date>2008-10-09</date>`

`<publisher>Amber</publisher>`

`</book>`

3.

```
<?xml version="1.0" ?>
<!DOCTYPE book [
    <!ELEMENT book (author+, translator*, title, plTitle?, date,
publisher)>
    <!ELEMENT author (#PCDATA)>
    <!ELEMENT translator (#PCDATA)>
    <!ELEMENT title (#PCDATA)>
    <!ELEMENT plTitle (#PCDATA)>
    <!ELEMENT date (#PCDATA)>
    <!ELEMENT publisher (#PCDATA)>
]>
```

5.

```
<?xml version="1.0" encoding="ISO-8859-2" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="book">
        <xs:complexType>
            <xs:sequence>
                <xs:element name="author" type="xs:string"
maxOccurs="unbounded"/>
                <xs:element name="translator" type="xs:string"
maxOccurs="unbounded" minOccurs="0" />
                <xs:element name="title" type="xs:string"/>
                <xs:element name="plTitle" type="xs:string" minOccurs="0"/>
                <xs:element name="date" type="xs:date"/>
                <xs:element name="publisher" type="xs:string"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
</xs:schema>
```

## 7. Control Questions

7.1. prolog: `<?xml version="1.0" ?>`

7.2. `<znacznik/>` lub `<znacznik></znacznik>` - brak zawartości, ale może mieć atrybuty

7.3. Nie.

7.4. `<!--komentarz -->`

7.5. XML jest niezależny od platformy.

7.6. Dane tekstowe, które nie są parsowane przez parser xml-owy.

7.7. Przed ':' znajduje się przestrzeń nazw

7.8. DTD – definiuje składnię dokumentu (np. XML-owego). Zazwyczaj DTD definiuje każdy dopuszczalny element dokumentu, jego zbiór atrybutów i dopuszczalne wartości. DTD określa także zagnieżdżanie i wymagalność poszczególnych elementów w dokumencie. W praktyce DTD przeważnie składa się z definicji ELEMENT i definicji ATTLIST.

7.9. XSD (XML Schema) – standard definiujący strukturę dokumentów XML, w odróżnieniu od DTD nie wymaga stosowania dodatkowej, nie-xml'owej składni. Dokumenty zawierające definicje XML Schema zapisuje się zwykle w plikach z rozszerzeniem .xsd (od XML Schema Definition).

Specyfikacja XML Schema składa się z 3 części:

- Primer – Elementarz – nienormatywna część zawierająca podstawy tego języka
- Structures – Struktury
- Datatypes – Typy danych

7.10. XSD ma większe możliwości niż DTD.

7.11. XPath - język pozwalający na adresowanie składowych dokumentu XML.

7.12. Tak

7.13. XSLT jest to język, który dostarcza reguł, pozwalających na transformację dokumentów z jednego formatu XML na dowolny inny format zgodny ze składnią XML-a (czyli np. HTML).