



Technologie Sieci Semantycznych

Andrzej Majczak

Uniwersytet Zielonogórski
Wydział Matematyki Informatyki i Ekonometrii
Data aktualizacji 2007.06.11

Zawartość prezentacji

1. Co to jest Sieć Semantyczna?
2. Co zostało zrobione?
3. Podsumowanie





Technologie Sieci Semantycznych

CO TO JEST SIEĆ SEMANTYCZNA?

Dwa główne cele ...

- Po pierwsze chodzi o to, aby powszechne formaty scalały i łączyły dane uzyskiwane z różnych źródeł, podczas gdy istniejący protokół komunikacyjny, na którym bazuje dzisiejszy Internet koncentruje się głównie na wymianie dokumentów.
- Chodzi także o to, aby język rejestrował sposób w jaki owe dane odnoszą się do przedmiotów użytku codziennego. Umożliwi to użytkownikowi lub też maszynie rozpoczęcie pracy od jednej bazy danych i progresywnie przejście do trybu pracy na nieograniczonej liczbie baz danych niezwiązanych łączami, lecz nawiązujących tematycznie do jednej określonej kwestii.

Źródło <http://www.w3.org/2001/sw/>

Sieć Semantyczna

Semantic Web jest rozszerzeniem dzisiejszej sieci Web, w której **informacja** ma **dobrze zdefiniowane znaczenie**, przez co umożliwi lepszą współpracę komputerów i ludzi.

Informacja będzie przedstawiona w postaci nie tylko do wyświetlania ale także w postaci **zrozumiałej dla maszyn**.

Sieć Semantyczna jest rodzajem siatki połączonych informacji w taki sposób aby informacja była **wygodna do przetwarzania** przez maszyny.

Zrozumienie przez maszyny przekazu informacji w postaci ludzkiej mowy, zawartości dokumentów, polegać ma nie na jakiejś magicznej sztucznej inteligencji, a raczej na zdolności maszyn to tego aby **rozwiązać dobrze zdefiniowane problemy** przez **wykonywanie dobrze zdefiniowanych operacji na dobrze zdefiniowanych danych**.

Jest to pomysł, który może być skuteczną drogą do reprezentowania danych w świecie World Wide Web lub jako globalna baza danych.

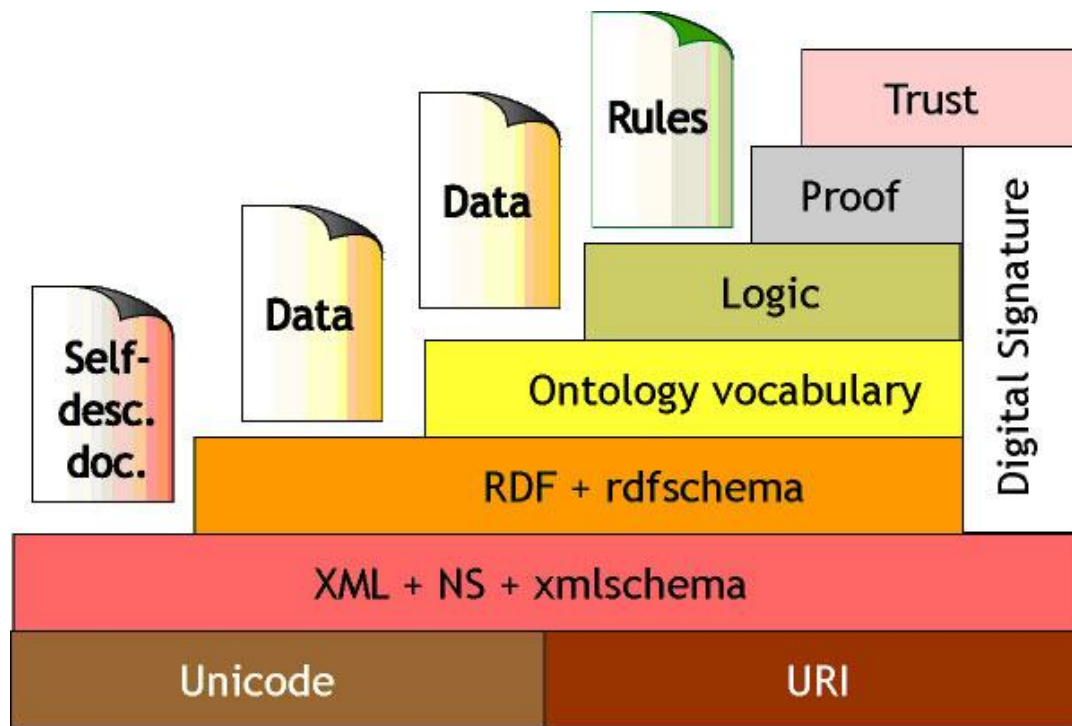
“The **Semantic Web** is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

— Tim Berners-Lee, James Hendler, Ora Lassila, “The Semantic Web”, Scientific American, May 2001.

Model Sieci Semantycznej

Model Sieci Semantycznej można podzielić na kilka warstw uporządkowanych w następujący sposób:

- **Unicode, URI** (*standard RFC*)
- **XML, przestrzenie nazw XML, schematy XML** (*standard W3C*)
- **RDF, Schematy RDF** (*standard W3C*)
- **Ontologie** (*standard W3C*)
- **Logika** (*brak standardów*)
- **Dowód** (*brak standardów*)
- **Zaufanie** (*brak standardów*)
- **Podpis cyfrowy** (*standard X.509, PGP*)



Źródło <http://www.w3.org/2000/Talks/1206-xml2k-tbl/>

Elementy Sieci Semantycznej

XML uniwersalny język znaczników przeznaczony do reprezentowania różnych danych w strukturalny sposób. Pozwala użytkownikom definiować i używać własnych znaczników, nie ma wbudowanych mechanizmów umożliwiających zrozumienie nowych znaczników przez innych użytkowników.

XML Schema to język który wprowadza ograniczenia dotyczące typu i struktury danych w dokumentach XML w celu gwarancji, że dane XML są poprawne w sensie syntaktycznym.

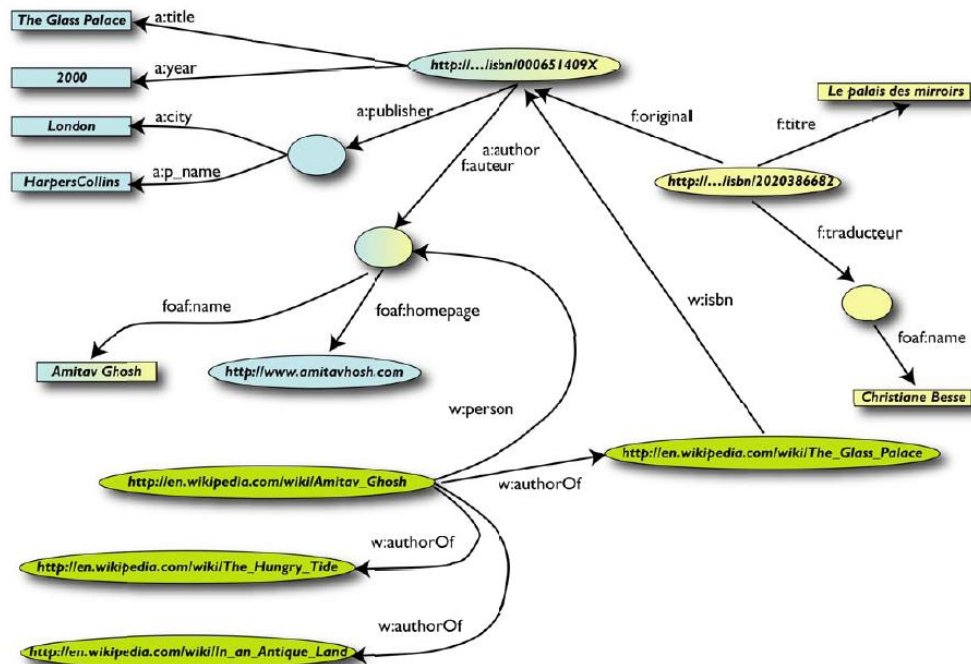
RDF jest standardem, który pozwala na zapis danych w postaci grafu skierowanego. W grafie tym dane zawarte są w wierzchołkach a relacje pomiędzy nimi i własności tychże znajdują się w krawędziach.

RDF Schema wprowadzają do grafów takie pojęcia jak klasy i podklasy, pozwalające na wspólne grupowanie danych mających cechy wspólne. Dowolna dana może znajdować się w wielu klasach.

OWL jest standardem pozwalającym na definiowanie klas na podstawie własności danych, a także na definiowanie logicznych charakterystyk relacji. OWL jest więc standardem formalnie zapisującym ontologie.

Ontologie to zbiory stwierdzeń zapisanych na przykład w RDF, definiujące relacje pomiędzy pojęciami oraz wyznaczające reguły wnioskowania. Komputery będą w stanie zrozumieć semantyczną zawartość dokumentów WWW, odwołując się do ontologii, do których odnoszą się występujące w nich pojęcia.

Reprezentacja danych w postaci grafu



Źródło <http://www.w3.org/2007/Talks/0419Paderborn-KB-IH/>

RDF Resource Description Framework

- **RDF** jest aplikacją języka XML, używanym do zapisu informacji w Sieci Semantycznej, przeznaczonym do przetwarzania maszynowego. W dosłownym tłumaczeniu RDF jest środowiskiem do opisu zasobów.
- **RDF** jest standardem, który pozwala na zapis danych w postaci grafu skierowanego. W grafie tym dane zawarte są w wierzchołkach a relacje pomiędzy nimi i własności znajdują się w krawędziach.
- **RDF** może być zakodowany w postaci języka XML, przez co daje możliwość jego wizualizacji.

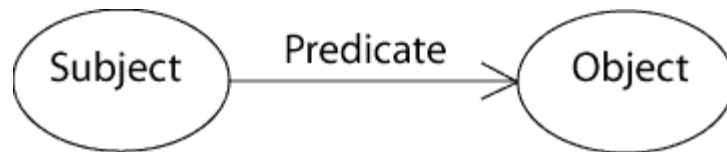
RDF Resource Description Framework

RDF używa następujących koncepcji:

- Graficznego modelu danych
- Słownictwa opartego na URI
- Typów Danych
- Literałów
- Serializacji składni XML
- Wyrażenia prostych faktów
- Wynikania

Graficzny model danych

- Zasadniczą strukturą każdego wyrażenia w RDF stanowi zbiór trójek, który nazywa się grafem RDF.
- Każda trójka reprezentuje oświadczenia związku pomiędzy rzeczami, które są określone przez węzły które je łączą. Kierunek ścieżki jest ważny, zawsze prowadzi w kierunku obiektu.
- Każdy taki zbiór składa się z następujących części:
 - **Subject** (temat lub podmiot)
 - **Predicate** (predykat) lub **Property** (właściwość), który określa związek.
 - **Object** (obiekt)



Trójki RDF (Triples)

Trójka RDF zawiera trzy komponenty:

- **Subject** (temat lub podmiot), który jest znacznikiem **URI** lub **pustym węzłem**
- **Predicate** (predykat) lub **Property** (własność), który jest znacznikiem **URI**
- **Object** (obiekt), który jest znacznikiem **URI**, **literałem** lub **pustym węzłem**

Elementy RDF

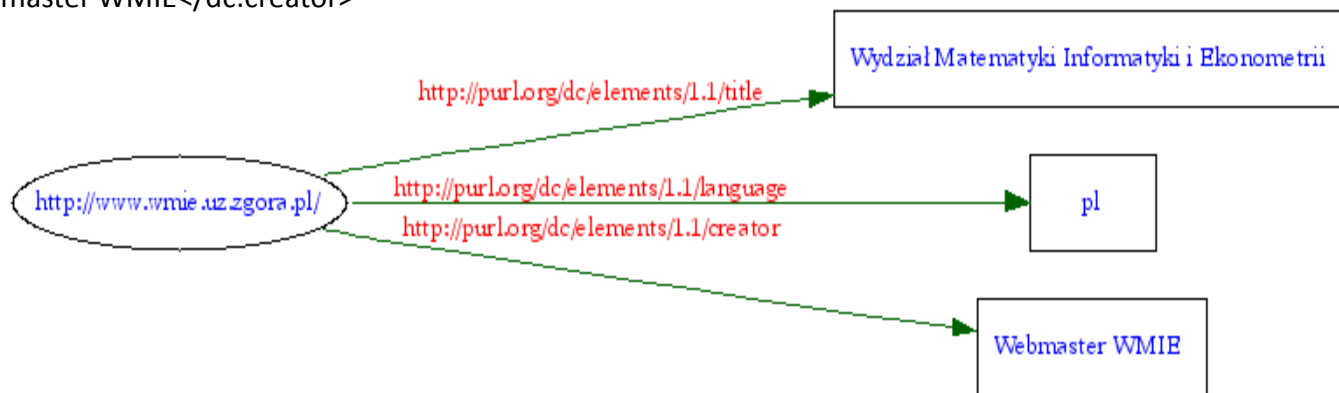
- **<rdf:RDF>** jest elementem root documentu RDF. Definiuje dokument XML jako dokument RDF. Zawiera odniesienia do przestrzeni nazw RDF
- **<rdf:Description>** element identyfikuje zasoby i ich atrybuty. Zawiera elementy opisujące zasoby
- **Properties as Attributes** (własności jako atrybuty) własności mogą być definiowane jako atrybuty zamiast elementów
- **Properties as Resources** (własności jako zasoby) własności mogą być zdefiniowane jako zasoby

Przykład RDF/XML „Description”

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://www.wmie.uz.zgora.pl/">
    <dc:title>Wydział Matematyki Informatyki i Ekonometrii</dc:title>
    <dc:language>pl</dc:language>
    <dc:creator>Webmaster WMIE</dc:creator>
  </rdf:Description>

</rdf:RDF>
```



Przykład RDF/XML „Contact”

```
<?xml version="1.0"?>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://wmie.uz.zgora.pl/rdf/1.1/">
```

```
<rdf:Description rdf:about="http://www.wmie.uz.zgora.pl/pracownicy/amajczak/">
```

```
<contact:name>Andrzej Majczak</contact:name>
```

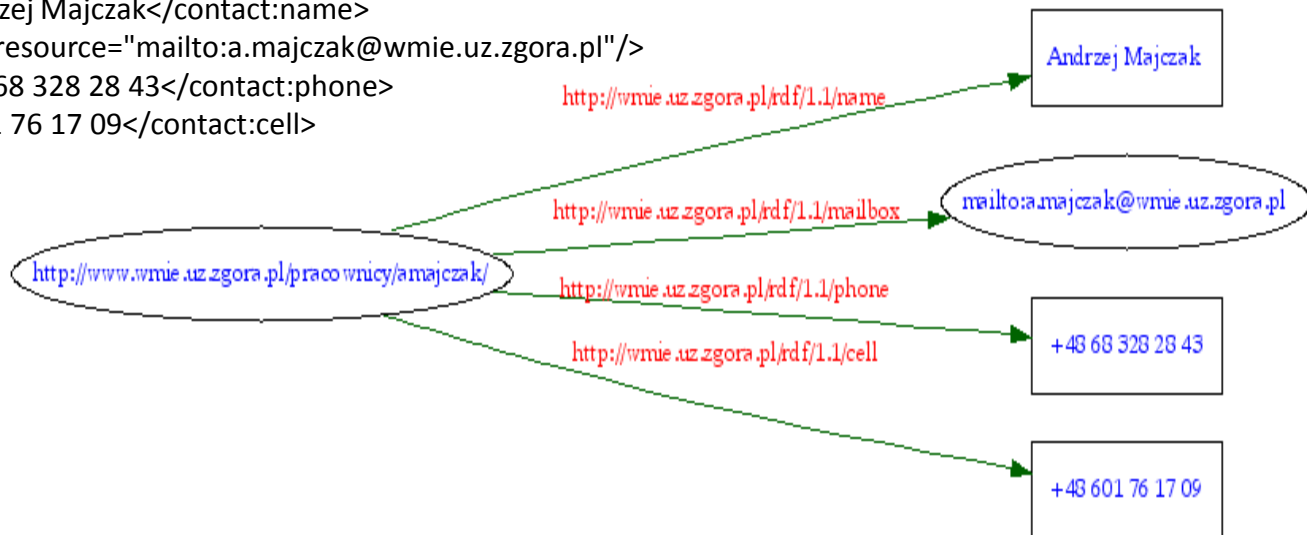
```
<contact:mailbox rdf:resource="mailto:a.majczak@wmie.uz.zgora.pl"/>
```

```
<contact:phone>+48 68 328 28 43</contact:phone>
```

```
<contact:cell>+48 601 76 17 09</contact:cell>
```

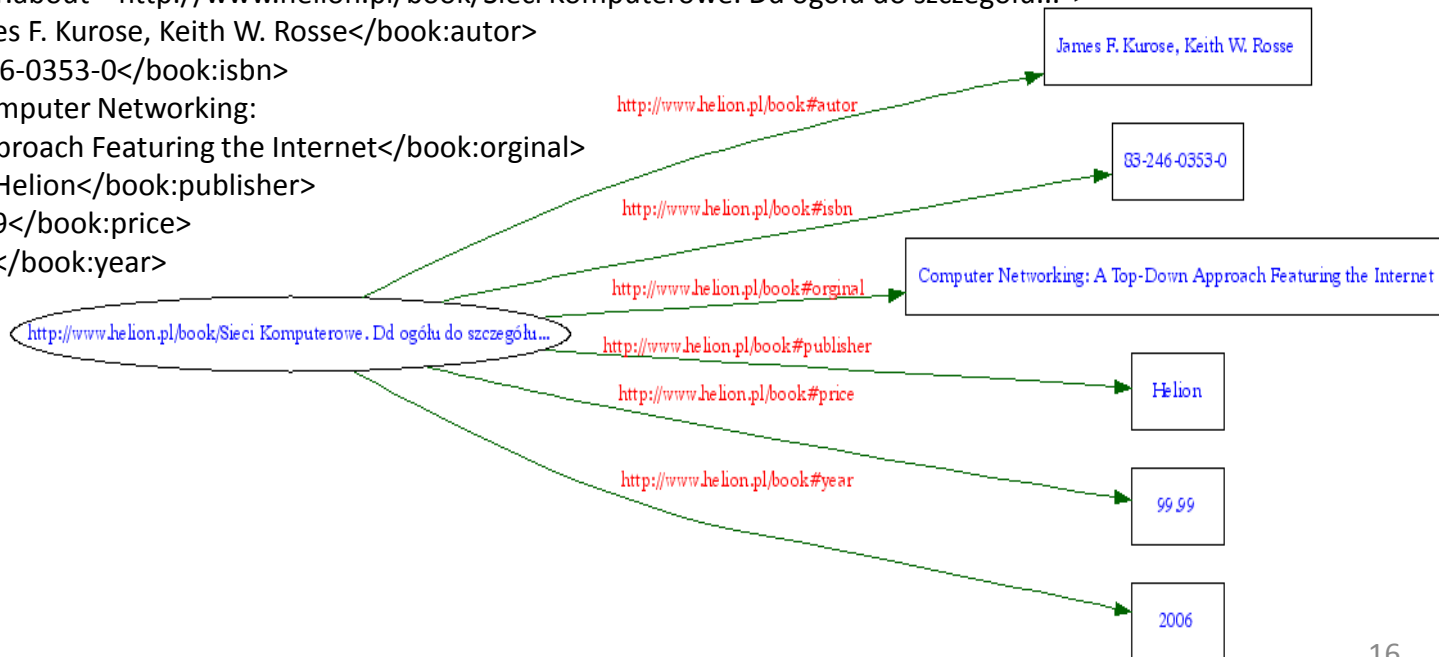
```
</rdf:Description>
```

```
</rdf:RDF>
```



Przykład RDF/XML „Book”

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:book="http://www.helion.pl/book#">
  <rdf:Description rdf:about="http://www.helion.pl/book/Sieci Komputerowe. Dd ogółu do szczególu...">
    <book:autor>James F. Kurose, Keith W. Rosse</book:autor>
    <book:isbn>83-246-0353-0</book:isbn>
    <book:original>Computer Networking:
      A Top-Down Approach Featuring the Internet</book:original>
    <book:publisher>Helion</book:publisher>
    <book:price>99.99</book:price>
    <book:year>2006</book:year>
  </rdf:Description>
</rdf:RDF>
```



(RDQL) RDF Data Query Language

RDQL (RDF Data Query Language) jest językiem zapytań do plików RDF

Przykładowy kod,

który wyciąga z bazy tytuł autora o nazwisku "Kurose"

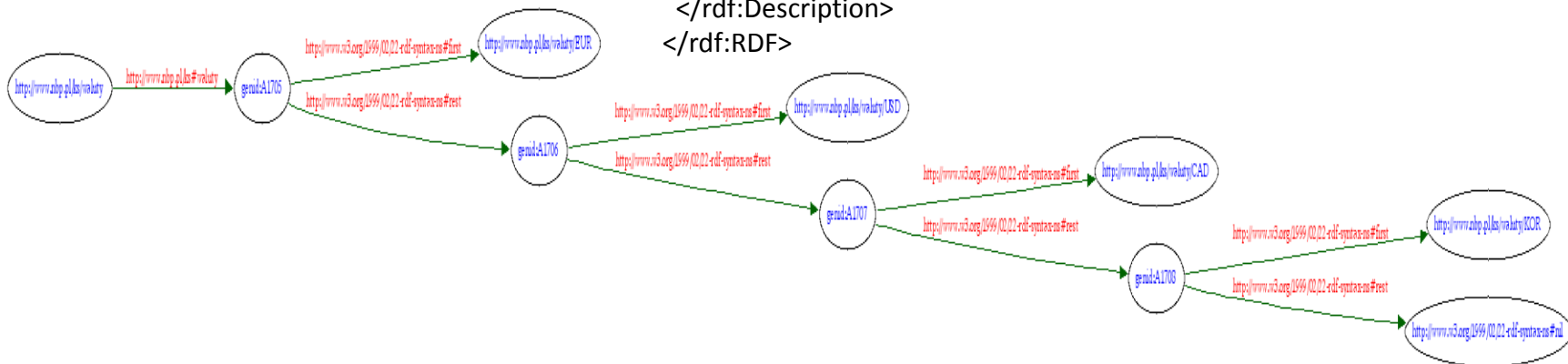
```
SELECT ?givenTitle
```

```
WHERE (?y, <http://www.helion.pl/books-rdf/3.0#Autor>, "Kurose") ,
```

```
    (?y, <http://www.helion.pl/books-rdf/3.0#Title>, ?givenTitle)
```

Przykład RDF Collection

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ks="http://www.nbp.pl/ks#">
  <rdf:Description rdf:about="http://www.nbp.pl/ks/waluty">
    <ks:waluty rdf:parseType="Collection">
      <rdf:Description rdf:about="http://www.nbp.pl/ks/waluty/EUR"/>
      <rdf:Description rdf:about="http://www.nbp.pl/ks/waluty/USD"/>
      <rdf:Description rdf:about="http://www.nbp.pl/ks/waluty/CAD"/>
      <rdf:Description rdf:about="http://www.nbp.pl/ks/waluty/KOR"/>
    </ks:waluty>
  </rdf:Description>
</rdf:RDF>
```

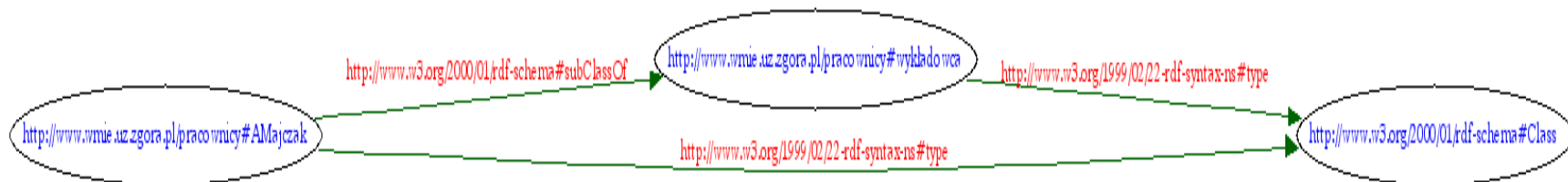


(RDFS) Schematy RDF

Schematy RDF (RDF Schema, RDFS) pełnią wobec RDF podobną rolę jak schematy XML wobec czystego XML. Sam RDF dostarcza jedynie podstawowych struktur do opisu abstrakcyjnych elementów zwanych zasobami. Brakuje jednak bardziej złożonych struktur służących do „specyfikowania konceptualizacji”. Nie można chociażby określić, że predykat powinien być używany jedynie do pewnego typu zasobów. RDFS umożliwia modelowanie pojęć bardzo zbliżone do używanego przez programistów modelowania obiektowego. Wprowadza bowiem takie pojęcia jak klasa, podklasa, właściwość.

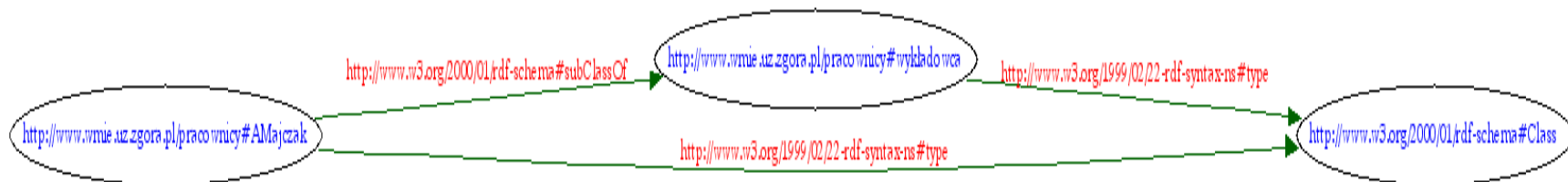
Przykład RDFS

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://www.wmie.uz.zgora.pl/pracownicy#">
  <rdf:Description rdf:ID="wykadowca">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/> </rdf:Description>
  <rdf:Description rdf:ID="AMajczak">
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <rdfs:subClassOf rdf:resource="#wykadowca"/> </rdf:Description>
</rdf:RDF>
```



Przykład RDFS (skrótowy)

```
<?xml version="1.0"?>  
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  
  xml:base="http://www.wmie.uz.zgora.pl/pracownicy#">  
  <rdfs:Class rdf:ID="wykladowca" />  
  <rdfs:Class rdf:ID="AMajczak">  
    <rdfs:subClassOf rdf:resource="#wykladowca"/> </rdfs:Class>  
</rdf:RDF>
```



Co to jest Ontologia?

- Ontologia jest pojęciem pochodzącym z filozofii, oznacza analizę pojęć i idei w celu ustalenia co istnieje oraz jakie są związki pomiędzy istniejącymi elementami.
- W informatyce ontologia oznacza formalną specyfikację pewnego obszaru wiedzy czy opisu elementów rzeczywistości w formie łatwej do wykorzystania przez oprogramowanie.
- Przy projektowaniu ontologii wykorzystywane są metody kategoryzacji i hierarchizacji. Pewnym pojęciom abstrakcyjnym i grupom obiektów posiadającym wspólne cechy przyporządkowywane są nazwy, tworzone są więc klasy obiektów (kategoryzacja). Tak utworzone klasy są następnie umieszczane w strukturze hierarchicznej.
- Przy tworzeniu ontologii można użyć różnych języków modelowania. Najczęściej wykorzystywany jest RDF/RDFS lub bardziej wyrafinowane języki zbudowane w oparciu o te technologie.

Ontologie według DAML+OIL

Ontologie są to abstrakcyjne kategorie odniesień, zdarzeń, które są zdefiniowane za pomocą terminologii klas i ich własności. Taka klasa wraz z jej własnościami nazywana jest ontologia. Ponieważ możliwe jest dziedziczenie klas oraz własności, ontologie mogą mieć strukturę hierarchiczną.

Ontologie: DAML+OIL

DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer) jest językiem, nad którym prace rozpoczęły się w roku 2000 w amerykańskim Departamencie Obrony. Jest zbudowany w oparciu o RDF ale ma znacznie więcej możliwości, z których najważniejsze to:

- definiowanie ograniczeń na wartości, jakie mogą przyjąć właściwości
- wprowadzenie jawnego odróżnienia właściwości, których obiektami mają być zasoby i literały
- możliwość definiowania właściwości o przeciwnym znaczeniu
- właściwości mogą być określone jako przechodnie
- dopuszczenie cykli w hierarchii klas
- możliwość określenia krotności właściwości
- wprowadzenie możliwości zdefiniowania jednoznacznej właściwości, identyfikującej zasób, pełniącej podobną rolę jak klucz pierwotny w relacyjnych bazach danych.

Ontologia według W3C

(Uważana za najbardziej praktyczną definicję)

Ontologia definiuje pojęcia, używane do opisanie i reprezentacji gałęzi wiedzy. Ontologie są używane przez ludzi, bazy danych i aplikacje, które potrzebują informacji pochodzących z danej dziedziny. Dziedzina jest rozumiana jako specyficzna gałąź wiedzy taka, jak np. medycyna, wytwarzanie narzędzi, handel nieruchomościami, naprawa samochodów, zarządzanie finansami itp. Ontologie zawierają definicje pojęć z danej dziedziny i relacji zachodzących pomiędzy tymi pojęciami, które są czytelne dla komputera. Definicje te nie muszą być ścisłe w sensie rozumianym przez logikę, ale raczej powinny być zrozumiałe dla aplikacji. Ontologie klasyfikują wiedzę w postaci gałęzi wiedzy, ale także wiedzę pochodzącą z różnych dziedzin. W ten sposób czynią tę wiedzę dostępną dla człowieka i aplikacji.

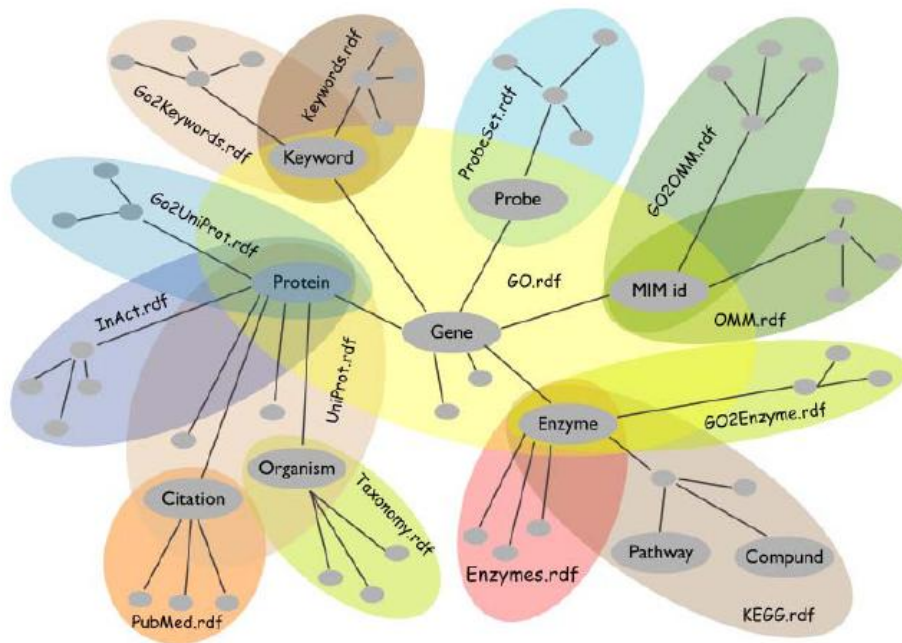
Ontologia jest to słownik, który łączy zbiór pojęć pochodzących z pewnej dziedziny wiedzy oraz relacje zachodzące pomiędzy nimi, z opisami tekstowymi czytelnymi dla człowieka, a także regułami poprawnego użycia tych pojęć czytelnymi dla "inteligentnej" aplikacji. Konceptualizacja z kolei, jest uproszczonym obrazem świata, jaki ma inteligentna aplikacja. Może ona być utożsamiana z wiedzą o zbiorze pojęć dotyczących pewnej dziedziny wiedzy i sposobie ich użycia.

Ontologia: OWL

OWL (Web Ontology Language) jest językiem opracowanym przez W3C w dużym stopniu bazowanym na DAML+OIL. Istnieją trzy dialekty OWL, istotnie różniące się pod względem możliwości:

- **OWL Lite** pozwala na tworzenie hierarchii klas z prostymi ograniczeniami, narzędzia przeznaczone dla OWL Lite są łatwe w implementacji. OWL Lite można traktować jak rozszerzenie RDF.
- **OWL DL** (OWL Description Logic) powstał z myślą o bardziej zaawansowanych zastosowaniach. Szczególny nacisk położono w nim na wsparcie dla wnioskowania. Dowolne wnioskowanie jest obliczalne w skończonym czasie. Na użycie konstrukcji językowych nałożono bardzo dużo ograniczeń. OWL DL można traktować jak rozszerzenia RDF z restrykcjami.
- **OWL Full** można traktować jak mniej restrykcyjną wersję OWL pod względem ograniczeń w użyciu konstrukcji językowych niż OWL DL, których można używać dużo swobodniej. Ale w związku z tym nie ma gwarancji związanych z obliczalnością wnioskowania. OWL Full można traktować jak rozszerzenia RDF z restrykcjami.

Przykład słowników/ontologi



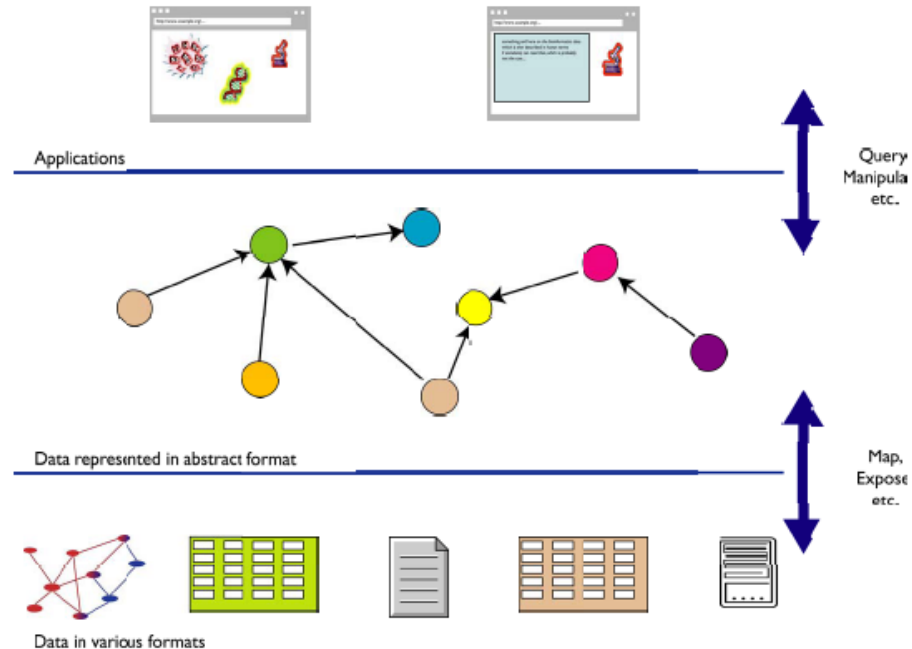
Źródło <http://www.w3.org/2007/Talks/0403-Tampere-IH/>

Technologie Sieci Semantycznych



CO ZOSTAŁO ZROBIONE?

Co zostało zrobione?



Źródło <http://www.w3.org/2007/Talks/0419Paderborn-KB-IH/>

Gdzie jest Semantic Web?

- Semantic Web dostarcza technologie, które umożliwiają taką integrację!

Na przykład:

- *abstrakcyjny model dla relacyjnych grafów: **RDF***
 - *możliwość wyciągnięcia informacji w postaci RDF z dokumentów XML np. XHTML: **GRDDL***
 - *możliwość dodania informacyjnej struktury (RDF) do XHTML stron: **RDFa***
 - *SQL adoptowany do relacyjnych grafów: **SPARQL***
 - *różne technologie do określenia relacji i kategoryzacji zasobów: **RDFS** (RDF Schemas), **OWL** (Web Ontology Language), **SKOS**, **Rule Interchange Format***
- Niektóre z tych technologii już istnieją, inne są w fazie rozwoju

(SPARQL) Protokół i język zapytań

SPARQL (SPARQL Protocol And RDF Query Language) jest językiem zapytań i protokołem dla plików RDF. Pozwala wyciągać z nich dane zawężone według kryteriów określonych poprzez predykaty RDF. Jest opisany przez kilka specyfikacji W3C:

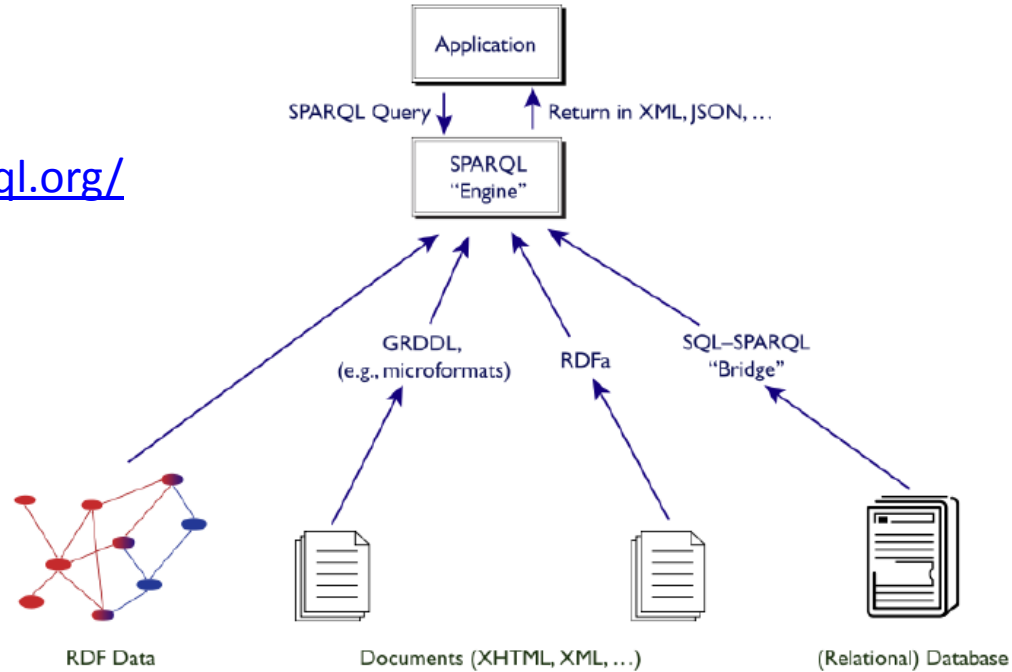
- [SPARQL Protocol for RDF](#)
- [SPARQL Query Language for RDF](#)
- [SPARQL Query Results XML Format](#)

Przykładowy kod:

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>  
SELECT ?title  
WHERE { <http://example.org/book/book1> dc:title ?title }
```

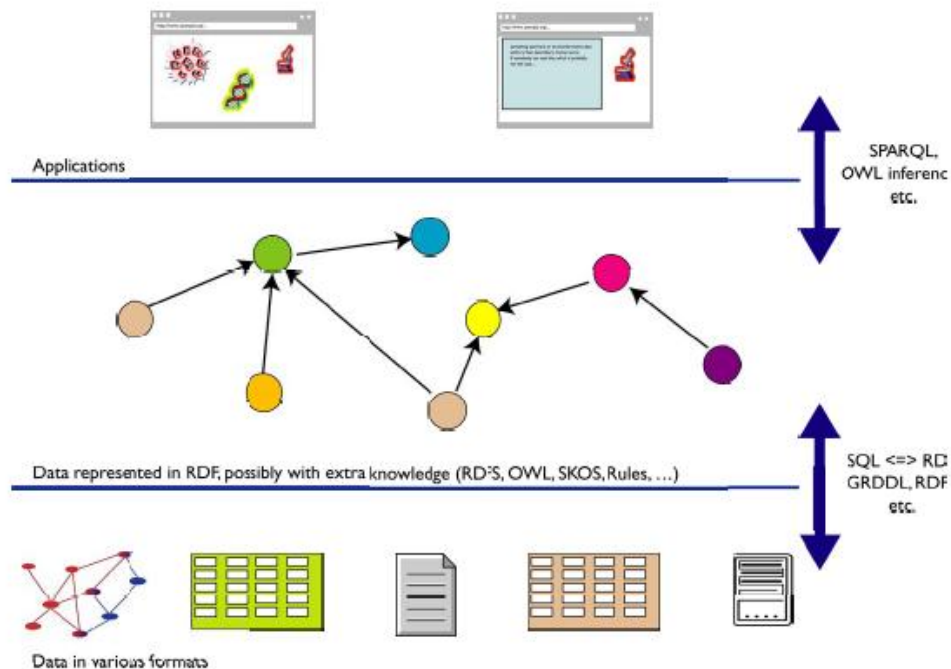
SPARQL cd.

<http://www.sparql.org/>



Źródło <http://www.w3.org/2007/Talks/0419Paderborn-KB-IH/>

Co zostało zrobione?



Źródło <http://www.w3.org/2007/Talks/0419Paderborn-KB-IH/>

Trzy Warstwy Semantycznych Technologii

Złożone aplikacje będą łączyć i wykorzystywać semantyczne modele poprzez trzy warstwy:

- 1) Złożony interfejs użytkownika
- 2) Złożony schemat biznesowych ontologii
- 3) Złożone zapytania

Złożone aplikacje połączą informacje w kontekst, umożliwiając narzędziom wnioskowania na: praktyczne załatwianie różnych spraw, śledzenie zagrożeń, zarządzanie ryzykiem, ocenę sytuacji, sterowanie i kontrolę.

Anatomy of a composite application — Semantic technologies at 3 levels:

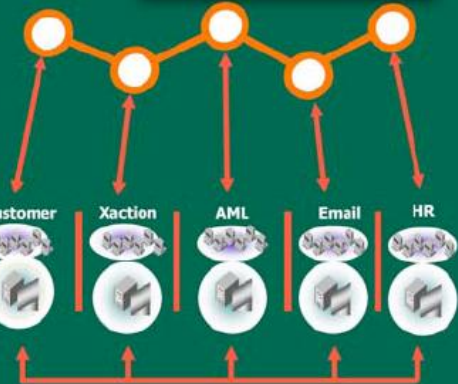
(1) Composite User Interface

UI must persist and expose semantics such that users can interact with meaningful objects



(2) Composite schema (business ontology)

A business ontology describes the semantics of data relationships, workflow, events, and business rules



(3) Composite query (EII)

Logically map multiple databases, applications, and web services as if they came from a single source

Source: Digital Harbor

Źródło <http://www.digitalharbor.com/>

Narzędzia

- Istnieje już cała grupa narzędzi typu Edytory Ontologi, Walidatory, OWL Reasoners (Ontologie wnioskujących), narzędzia typu Program Development Tools,...
- Aktualna lista narzędzi dostępna jest na stronach W3C Wiki <http://esw.w3.org/topic/SemanticWebTools>
- Warto zwrócić uwagę na znane firmy, które już oferują lub wkrótce będą oferować narzędzia dla Sieci Semantycznych: Adobe, Oracle, IBM, HP, Software AG, webMethods, Northrop Gruman, Altova, ...

Narzędzia cd..

Altova **SemanticWorks™ 2007** pozwala tworzyć w postaci graficznej i redagować dokumenty RDF, słowniki RDFS i ontologie OWL z pełnym sprawdzeniem składni i semantyki. Kontekstowy system z listą aktualnych wyborów opartych na RDF lub dialekcie OWL do łatwego i szybkiego tworzenia dokumentów.

http://www.altova.com/semantic_web.html

Technologie Sieci Semantycznych

PODSUMOWANIE

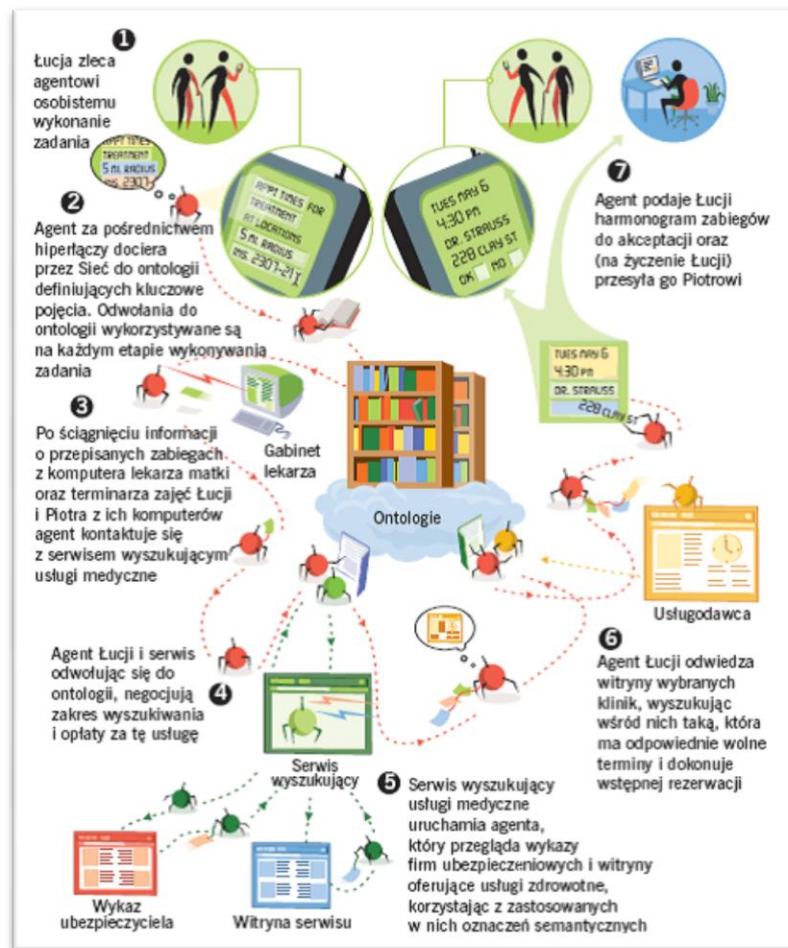


Ten scenariusz wkrótce będzie możliwy...

Agent Łucji wyszukuje dla jej matki klinikę rehabilitacji, która spełnia zadane kryteria i ma wolne terminy nie kolidujące z rozkładem zajęć Łucji i jej brata Piotra.

Program typu **agent** będzie potrafił wykonać stawiane jemu zadania dzięki umieszczonym na stronach WWW oznaczeniom semantycznym.

Ontologie, w których zdefiniowane są niezbędne pojęcia i ich powiązania semantyczne, umożliwiają agentom rozpoznawanie znaczenia występujących w Sieci Semantycznej elementów oraz współpracę z agentami witryn i automatycznymi serwisami.



Podsumowanie

- Sieć Semantyczna to nowy format dokumentów hipertekstowych, który umożliwi komputerom rozpoznanie ich treści.
- Zrozumienie przez maszyny przekazu informacji będzie polegać nie na jakiejś magicznej sztucznej inteligencji, a raczej na zdolności maszyn do tego aby rozwiązać dobrze zdefiniowane problemy przez wykonywanie dobrze zdefiniowanych operacji na dobrze zdefiniowanych danych.

Podsumowanie cd..

- Sieć Semantyczna to narzędzie do wykonywania konkretnych zadań.
- Właściwie zaprojektowana może przyczynić się do dalszego rozwoju wiedzy całej ludzkości. Otrzymamy w ten sposób nowe, skuteczne narzędzia ułatwiające nam życie, pracę i naukę.

Technologie Sieci Semantycznych

DZIĘKUJĘ ZA UWAGĘ

