

# Bazy wiedzy

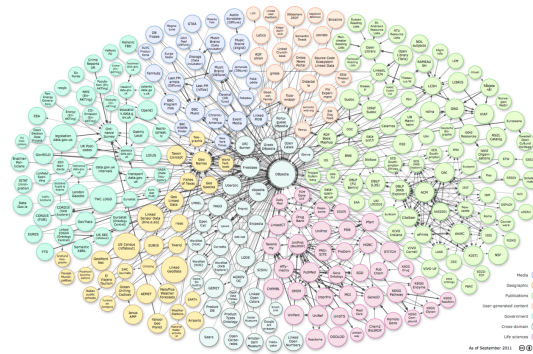
Agnieszka Ławrynowicz

Poznań, rok akademicki 2014/2015

# Baza wiedzy – pragmatyczna definicja

**Baza wiedzy**  $\approx$  kolekcja encji, klas i faktów o postaci „subject-predicate-object” (atrybuty, relacje), istotnych ogólnie lub w danej dziedzinie, która jest kompleksowa, zorganizowana semantycznie i zdatna do odczytu maszynowego

# Grafy wiedzy



**Facebook Entity Graph**



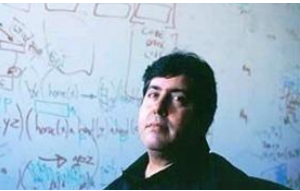
**Microsoft Bing Satori**



# Historia baz wiedzy

## ludzie dla ludzi

Cyc



WordNet



WIKIPEDIA  
The Free Encyclopedia

Wikipedia



## algorytmy dla maszyn



1985

1990

2000

2005

2010

# Cyc

- **Baza wiedzy zawiera mikro-teorie – kolekcje pojęć i faktów dotyczących jakiejś dziedziny zainteresowania.**
- **Mikro-teoria musi być wolna od sprzeczności, ale cała baza wiedzy nie musi.**
- **Reguły FOL**
- **CycL (notacja w Lisp)**
  - (#\$implies
  - (#\$and
  - (#\$isa ?OBJ ?SUBSET)
  - (#\$genls ?SUBSET ?SUPERSET))
  - (#\$isa ?OBJ ?SUPERSET))
- **Dostępny wariant OpenCyc**



# WordNet

- leksykalne kategorie rzeczowników, czasowników, przymiotników, przysłówków
- słowa z tej samej kategorii, z grubsza synonimy, są pogrupowane w **synsety** (pojęcia)
- hierarchie, zdefiniowane by hiperonimy (IS-A)
- Polska Słowosieć (rozwijana od 2005 roku na Politechnice Wrocławskiej)

```
dog, domestic dog, Canis familiaris
=> canine, canid
=> carnivore
=> placental, placental mammal,
eutherian, eutherian mammal
=> mammal
=> vertebrate, craniate
=> chordate
=> animal, animate being,
beast, brute, creature, fauna
=> ...
```

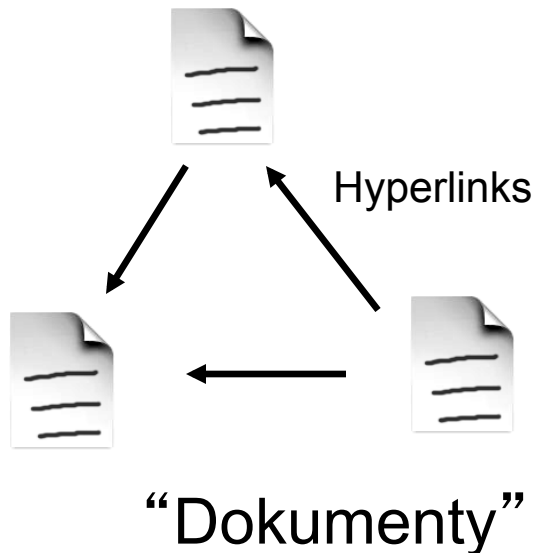
## „Powiązane Dane” (ang. Linked Data): definicja

***“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. “ (Tim Berners-Lee)***

**Powiązane Dane** – wykorzystanie technologii Sieci Semantycznej do publikowania ustrukturalizowanych danych w Sieci i do ustanawiania powiązań między źródłami danych

# Od „Sieci Dokumentów” do „Sieci Danych”

## Sieć dokumentów

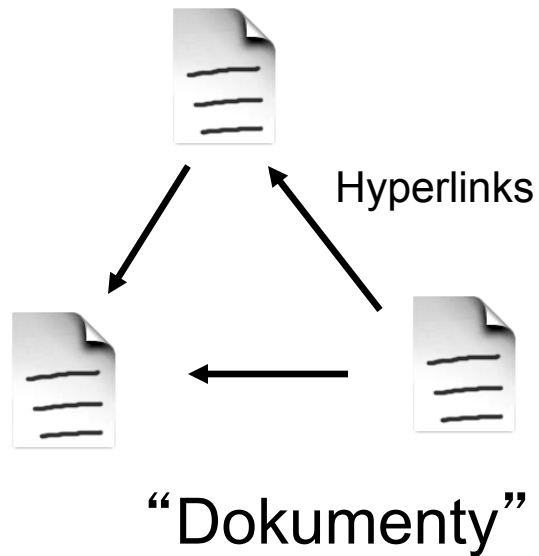


- Podstawowe elementy:
  1. Nazwy (URI)
  2. Dokumenty (Zasoby) opisane w HTML, XML, itp.
  3. Interakcja poprzez HTTP
  4. (Hiper)linki pomiędzy dokumentami lub anchors w dokumentach
- Wady:
  - Nietypowane linki
  - Wyszukiwarki nie potrafią obsłużyć skomplikowanych zapytań

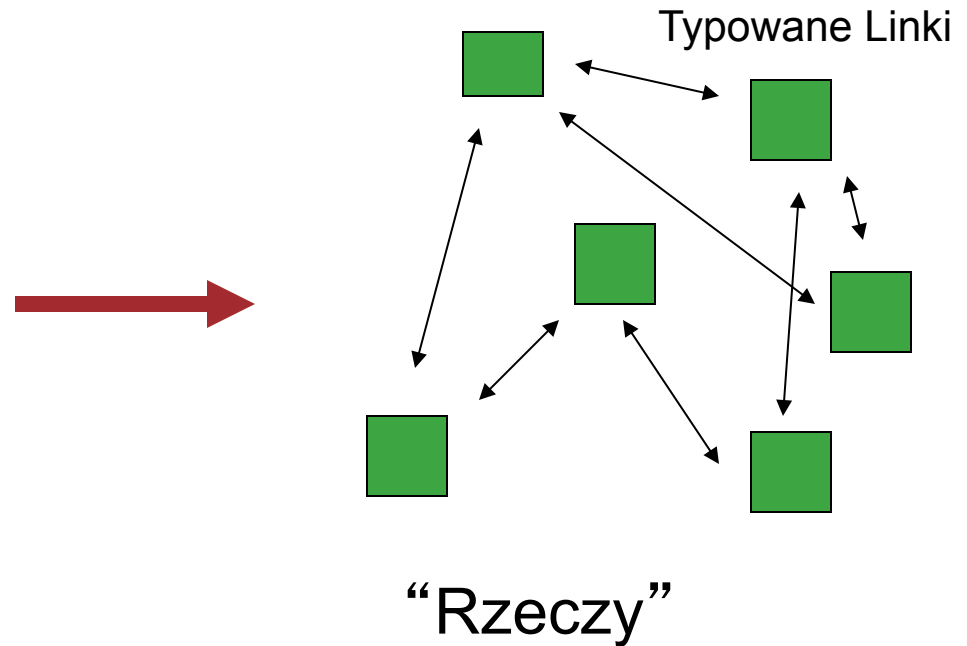


# Od „Sieci Dokumentów” do „Sieci Danych”

## •Sieć Dokumentów



## •Sieć Danych



# Od „Sieci Dokumentów” do „Sieci Danych”

- Cechy:

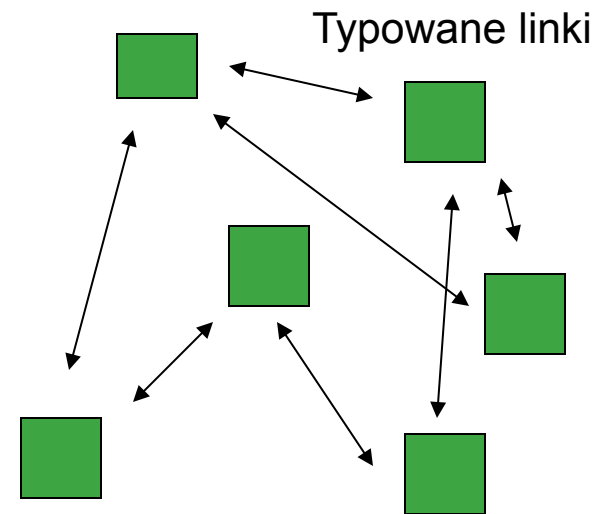
- Linki pomiędzy dowolnymi rzeczami (np. osobami, lokalizacjami, zdarzeniami, budynkami)

- Struktura danych na stronach WWW jest jawna

- Rzeczy opisane na stronach mają nazwę i URI

- Linki pomiędzy rzeczami są jawne i typowane

## Sieć danych



“Rzeczy”

## Wizja Sieci Danych 1/2

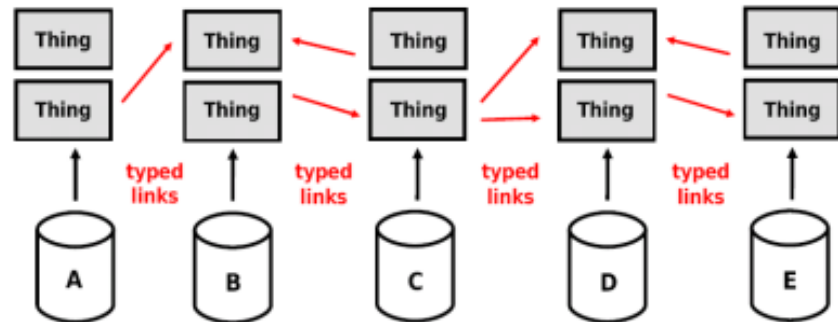
- Sieć dzisiaj

- składa się z odizolowanych silosów danych, które są dostępne poprzez wyspecjalizowane wyszukiwarki
- jedna strona (silos danych) przechowuje filmy, inne recenzje, jeszcze inne informacje o aktorach
- wiele popularnych rzeczy jest reprezentowanych w wielu różnych zbiorach danych
- „linkowanie” identyfikatorów łączy te zbiory danych

## Wizja Sieci Danych 2/2

- Sieć Danych - **globalna baza danych**

- składa się z **obiektów i ich opisów**
- obiekty są ze sobą powiązane linkami**
- z **wysokim stopniem ustrukturalizowania obiektów**
- z **jawną semantyką linków i treści**
- zaprojektowana dla ludzi i maszyn



## „Powiązane Dane” - zasady

**Używaj URI jako nazwy dla rzeczy.**

**Używaj HTTP URI tak aby ludzie mogli wyszukiwać tych nazw.**

**Kiedy użytkownik wyszukuje URI, dostarcz użytecznej informacji w RDF.**

**Zawrzyj wyrażenia RDF, które są powiązane linkami do innych identyfikatorów URI tak aby mogły one pomóc w wykryciu powiązanych rzeczy.**

# Projekt „Linking Open Data” (Otwarte Powiązane Dane)



**Projekt społecznościowy ze wsparciem W3C**

**Cel: Pomoc w utworzeniu Sieci Semantycznej poprzez publikowanie zbiorów danych z wykorzystaniem RDF.**

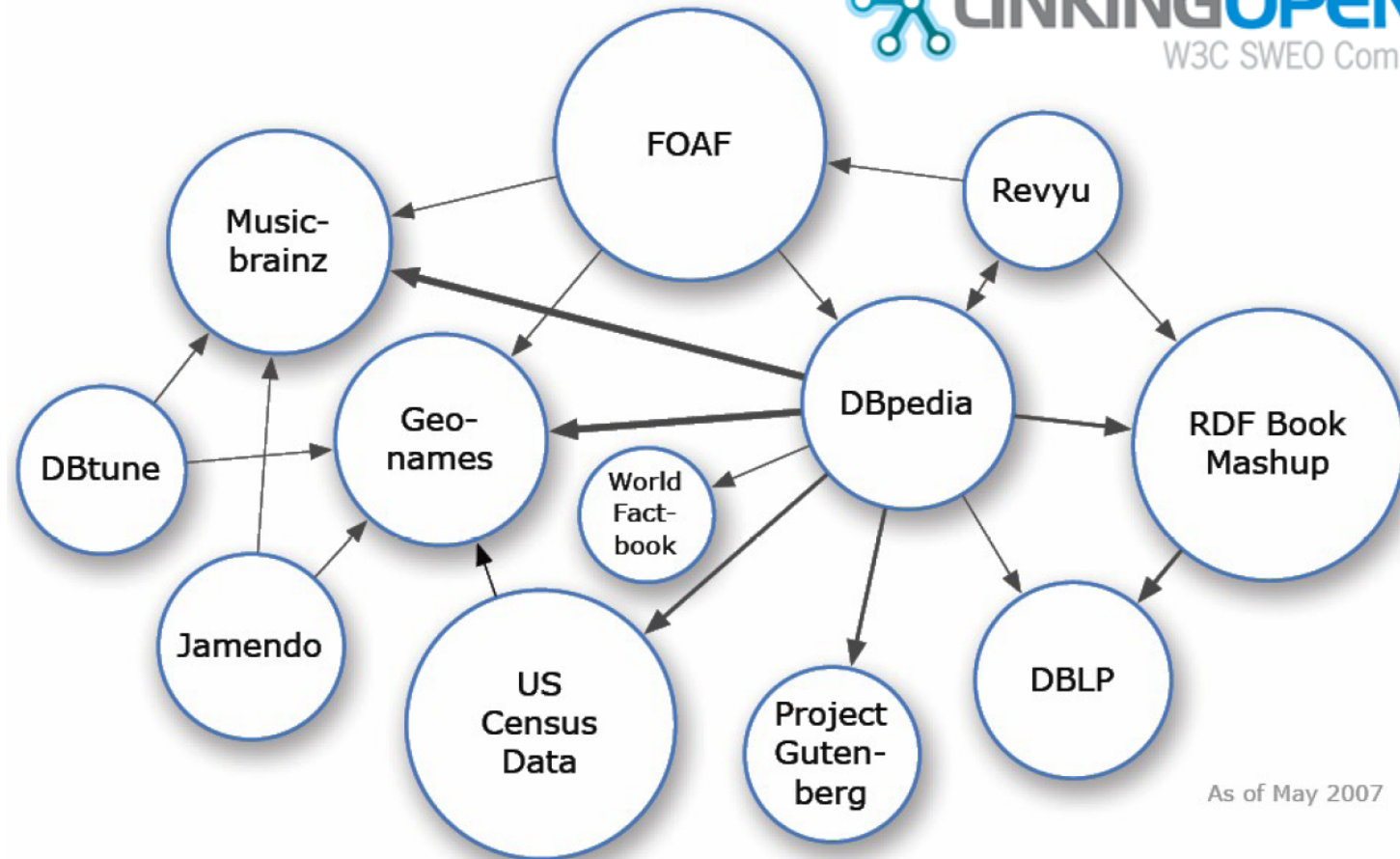
**Spełnia zasady „połączonych danych” (Linked Data principles)**

**Główna idea: wziąć istniejące (otwarte) zbiory danych i uczynić je dostępnymi w Sieci w formacie RDF**

**Raz opublikowane w RDF, połączyć je linkami z innymi zbiorami danych**

**Przykładowy link RDF: <http://dbpedia.org/resource/Berlin>  
[Identyfikator Berlina w DBPedia] owl:sameAs  
<http://sws.geonames.org/2950159> [Identyfikator Berlina w Geonames].**

# Chmura LOD - Maj 2007



As of May 2007

# Chmura LOD - Maj 2007



## Ogólnie:

Chmura Powiązanych Otwartych Danych (Linked Open Data) jest zbiorem powiązanych między sobą zbiorów danych, które zostały opublikowane i powiązane linkami zgodnie z zasadami „powiązanych danych”.

## Fakty:

Punkty „ogniskujące”:

- DBPedia: wersja Wikipedii w formacie RDF; wiele przychodzących i wychodzących linków

- Zbiory danych dotyczące muzyki

Duże zbiory danych zawierają: FOAF, US Census data

Rozmiar w przybliżeniu 1 bilion trójek, 250k linków





# Chmura LOD - Wrzesień 2008

## **Fakty:**

Więcej niż 35 powiązanych zbiorów danych

Gracze komercyjni dołączyli do chmury (np. BBC)

Firmy zaczęły publikować i przechowywać zbiory danych  
(OpenLink, Talis, Garlik)

Rozmiar w przybliżeniu 2 bilion'y trójek, 3 miliony linków



# Chmura LOD - Marzec 2009

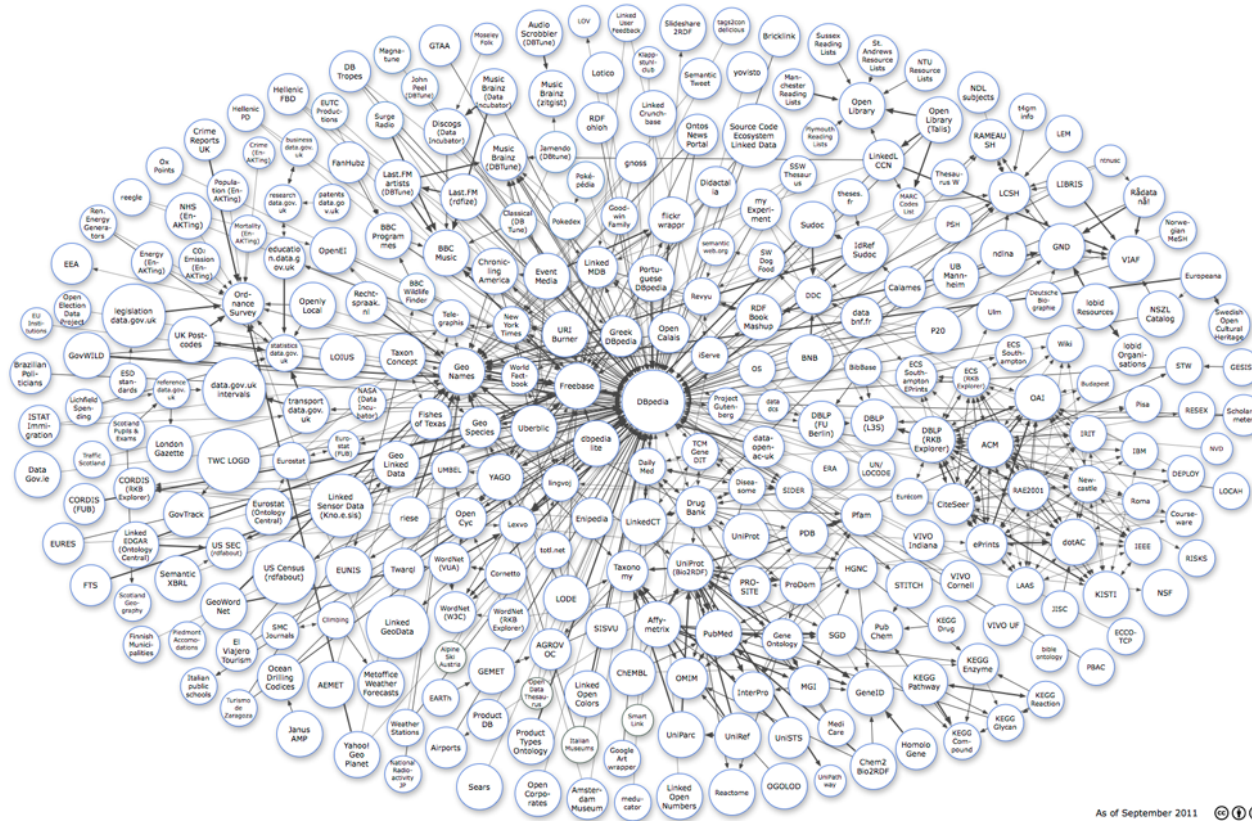
## **Fakty:**

Wielka część z chmury 'Drug' i projektu BIO2RDF

Znaczące nowe zbiory danych: Freebase, OpenCalais, ACM/  
IEEE

Rozmiar > 10 bilionów trójek

# Chmura LOD - Wrzesień 2011



As of September 2011 © 1 1 1

Liczba zbiorów danych: 295  
Liczba trójek: 31 634 213 770



# Publikowanie Powiązanych danych w 7 krokach

**Wybór słowników – ważne ponowne wykorzystanie istniejących słowników - interoperacyjność**

**Partycjonowanie grafu RDF do “stron danych”**

**Przyznanie URI każdej stronie danych**

**Stworzenie wariantów HTML każdej strony danych - do renderowania stron w przeglądarkach**

**Przyznanie URI każdej encji**

**Dodanie metadanych stron i linków np. publisher, license, topics**

**Dodanie semantycznej mapy strony (semantic sitemap) – ważna dla pajaków w celu znalezienia zbioru danych lub końcówki SPARQL z dostępem do danych**

## Tworzenie powiązań (linków)

• **Popularne predykaty:** owl:sameAs,  
foaf:homepage, foaf:topic, foaf:based\_near,  
foaf:maker/foaf:made, foaf:depiction, foaf:page,  
foaf:primaryTopic, rdfs:seeAlso



# Przykładowe zbiory danych

**DBpedia**

**BBC Music**

**Open government (UK), Data.gov (US)**

**Freebase**

**Zbiory danych biologicznych i medycznych**

# DBpedia



## **Inicjatywa społeczna:**

Ekstrakcja strukturalnej informacji z Wikipedii

Udostępnienie informacji w Sieci na otwartej licencji

Powiązanie linkami zbioru danych DBpedii z innymi zbiorami danych w Sieci

**DBpedia to jeden z najbardziej centralnych "hubów" w tworzącej się Sieci Danych**

# Data.gov



## 1. Zgromadź dane

– z wielu miejsc, udostępnił je za darmo developerom, naukowcom, obywatelom

## 2. Połącz społeczność

– w znajdowaniu rozwiązań pozwalających na współpracę poprzez media społecznościowe, wydarzenia, platformy

## 3. Dostarcz infrastrukturę

– w oparciu o standardy i interoperacyjność

## 4. Zachęć twórców technologii

– do tworzenia aplikacji, map, wizualizacji danych, które wzmocnią wybory dokonywane przez ludzi

## 5. Zgromadź więcej danych

– i połącz więcej ludzi



*“A Strategy for American Innovation”  
wrzesień 2009*

## Powiązane Dane – Narzędzia i Aplikacje

**Narzędzia do przenoszenia danych z innych formatów i z funkcjonujących wewnątrz systemów do Sieci Danych**

**Narzędzia do wykorzystywania Powiązanych Danych: przeszukiwanie, przeglądanie, tworzenie *mashups*, inne**

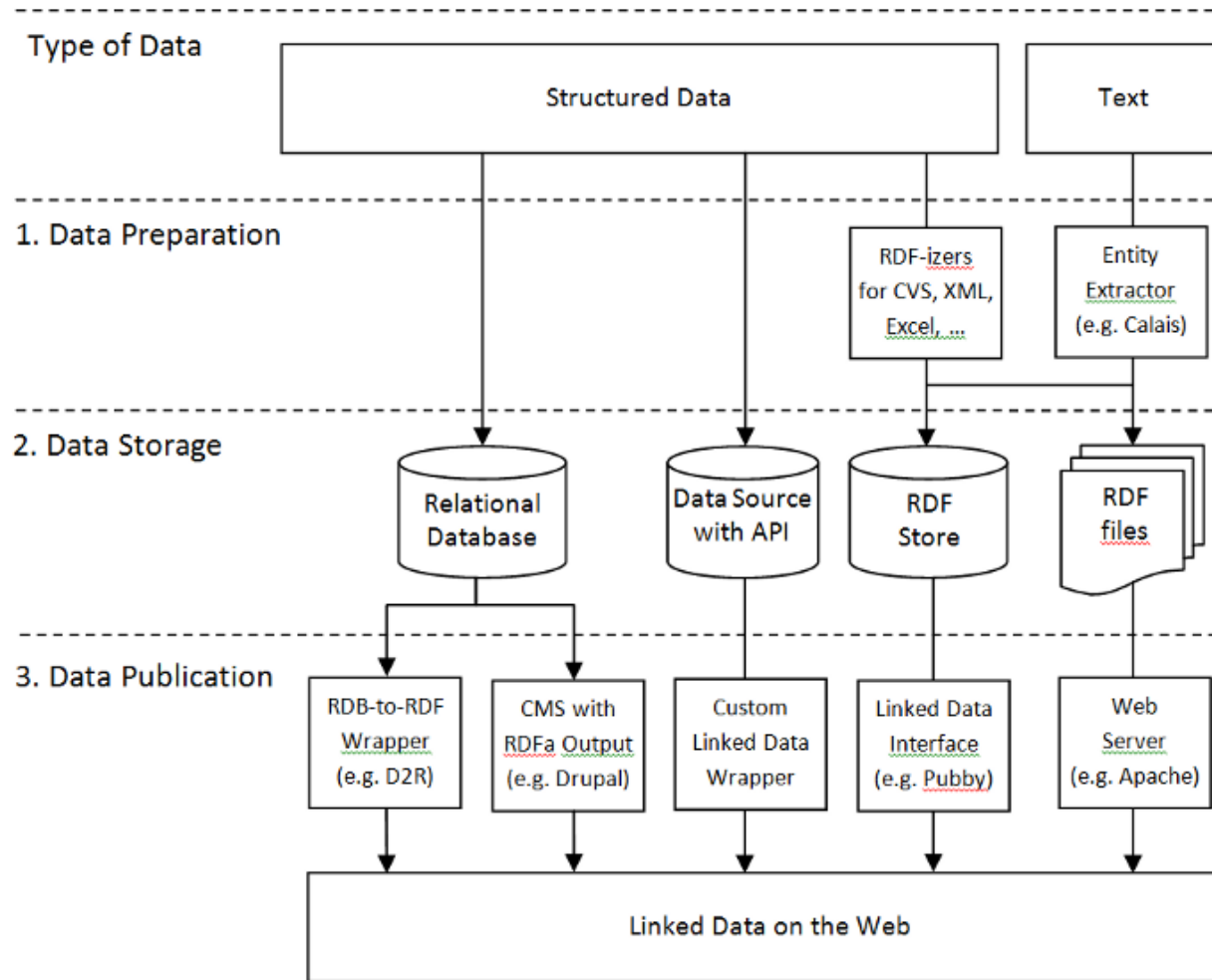
# Przenoszenia danych z innych formatów do Sieci Danych

- Dostarczenie danych przechowywanych w relacyjnych bazach danych do Sieci Danych:
  - **Triplify**: pozwala na specyfikację zapytań SQL i zrenderowanie wyników jako RDF
  - **D2RQ, atop**: odwzorowanie relacyjnych baz danych do RDF; dostarczają końcówkę SPARQL z dostępem do danych
  - **Virtuoso RDF Views**: oferuje deklaratywny język do tworzenia odwzorowań pomiędzy danymi SQL i RDF
- Ekstrakcja danych z Sieci WWW (np. **DBPedia**: dane z Wikipedii)
- Konwersja istniejących danych i ekstrakcja z nich RDF: z JPEG, Email, BibTex, Java bytecode, Javadoc, Excel

# Repozytoria trójek RDF

- **OWLIM**: natywne, wykorzystuje mechanizm wnioskowania wprzód (forward chaining) i materializację
- **AllegroGraph**: natywne
- **Jena TDB**: natywne
- **Open Link Virtuoso**: hybrydowe, hostuje zbiór Dbpedia, Virtuoso 7 - Virtuoso Column Store
- **BigData**: hybrydowe

# Publikowanie powiązanych - typowe wzorce



# Konsumowanie Powiązanych Danych

- **Przeglądarki Powiązanych Danych: eksplorowanie rzeczy i zbiorów danych i nawigacja pomiędzy nimi**
  - Tabulator Browser, Marbles, OpenLink RDF Browser, Zitgist RDF Browser, Disco Hyperdata Browser, Fenfire
- ***Mashup'y* Powiązanych Danych: strony, które łączą („mieszają”) powiązane dane**
  - Revyu.com, DBtune Slashfacet, DBPedia Mobile, Semantic Web Pipes
- **Wyszukiwarki powiązanych danych**
  - Falcons, Sindice, MicroSearch, Watson, SWSE, Swoogle



## Przykładowy *Mashup*: [Revyu.com](http://Revyu.com) 1/2

**Revyu.com** - strona do oceniania wszystkiego.

**Powiązane Dane** wykorzystywane do wzbogacania ocen.

**Oceny** zawierają linki do ocenianej “rzeczy” i linki „seeAlso” do Wikipedii i innych zbiorów danych.

# Przykładowy Mashup: Revyu.com 2/2

The screenshot shows a Windows Internet Explorer browser window. The address bar displays the URL <http://revyu.com/things/volver-movie-almodovar/about/html>. The browser's toolbar includes various icons for search, sharing, and navigation. The page content features the Revyu.com logo with the tagline "Review ANYTHING". Navigation links include Home, Browse Things, Search Things, Browse People, Login/Register, and New Review. The main content area is titled "Volver" and includes a "Links" section pointing to the Sony Classics homepage, a "Tags" section with "almodovar" and "movie", and a "Reviews (2)" section. The first review is a 5-star rating by "esfandia" dated "14 May 2007". The review text is partially obscured by a yellow box containing the URL <http://revyu.com>. The second review is also a 5-star rating by "esfandia" dated "14 May 2007".

**REVIU.COM**  
Review ANYTHING

[Home](#) | [Browse Things](#) | [Search Things](#) | [Browse People](#)  
[Login/Register](#) | [New Review](#)

## Volver

**Links**  
Homepage: <http://www.sonyclassics.com/volver/>

**Tags**  
[almodovar](#) [movie](#)

**Reviews (2)**

☆☆☆☆☆ by [esfandia](#) on 14 May 2007

\* synopsis: A mother needs to cover up a crime committed by her own mother... <http://revyu.com> th appearances of the ghost of her

\* comment: A touching tribute to women's resilience to hardships. Ever since the passing of her own mother, Almodovar has seemed more interested in female characters and the exploration of their personalities, rather than in the funny and kitsch dramas that made him famous in the 80s. That this maturation has not come at the expense of tedious intellectualism is remarkable.

☆☆☆☆☆ by [esfandia](#) on 14 May 2007

# Przykładowa wyszukiwarka: Sindice 1/2

## Wyszukiwarka Powiązanych Danych.

**Pozwala na wyszukiwanie treści Sieci Semantycznej na bazie:**

- słów kluczowych
- URI (identyfikujących obiekty, pojęcia, lub dokumenty).

# Przykładowa wyszukiwarka: Sindice 2/2



The screenshot shows a Windows Internet Explorer browser window displaying search results for the term "poznan" on the Sindice website. The browser's address bar shows the URL <http://sindice.com/search?q=poznan&qt=term>. The search bar contains the text "poznan" and a "SEARCH" button. The search results are displayed below the search bar, showing the title "Search results for term 'poznan', found 28.76 thousand". The results list includes:

- [Poznan](#) (RDF)  
2010-08-06 - 17 triples in 2.1 kb  
<http://dbpedia.org/resource/Poznan> (Search) Inspect: (Cached) (Live)
- [Poznan](#) (LICENSE)  
2010-10-12 - 46 triples in 7.4 kb  
<http://plumbot.com/Poznan.html> (Search) Inspect: (Cached) (Live)
- [Lech Poznan](#) (RDFa)  
2010-11-06 - 2 triples in 233 bytes  
<http://www.ligafutbol.net/tag/lech-poznan/> (Search) Inspect: (Cached) (Live)
- [Poznan'](#) (RDF)  
2009-03-19 - 10 triples in 1.1 kb  
<http://sws.geonames.org/505656/> (Search) Inspect: (Cached) (Live)
- [Poznan](#) (RDF)

## Inne inicjatywy Open\*

**Open Source**

**Open Content**

**Open Science (Open Notebook Science)**

**Open Access**

**Open CourseWare**

**Open Society Foundations**

**Open Health**

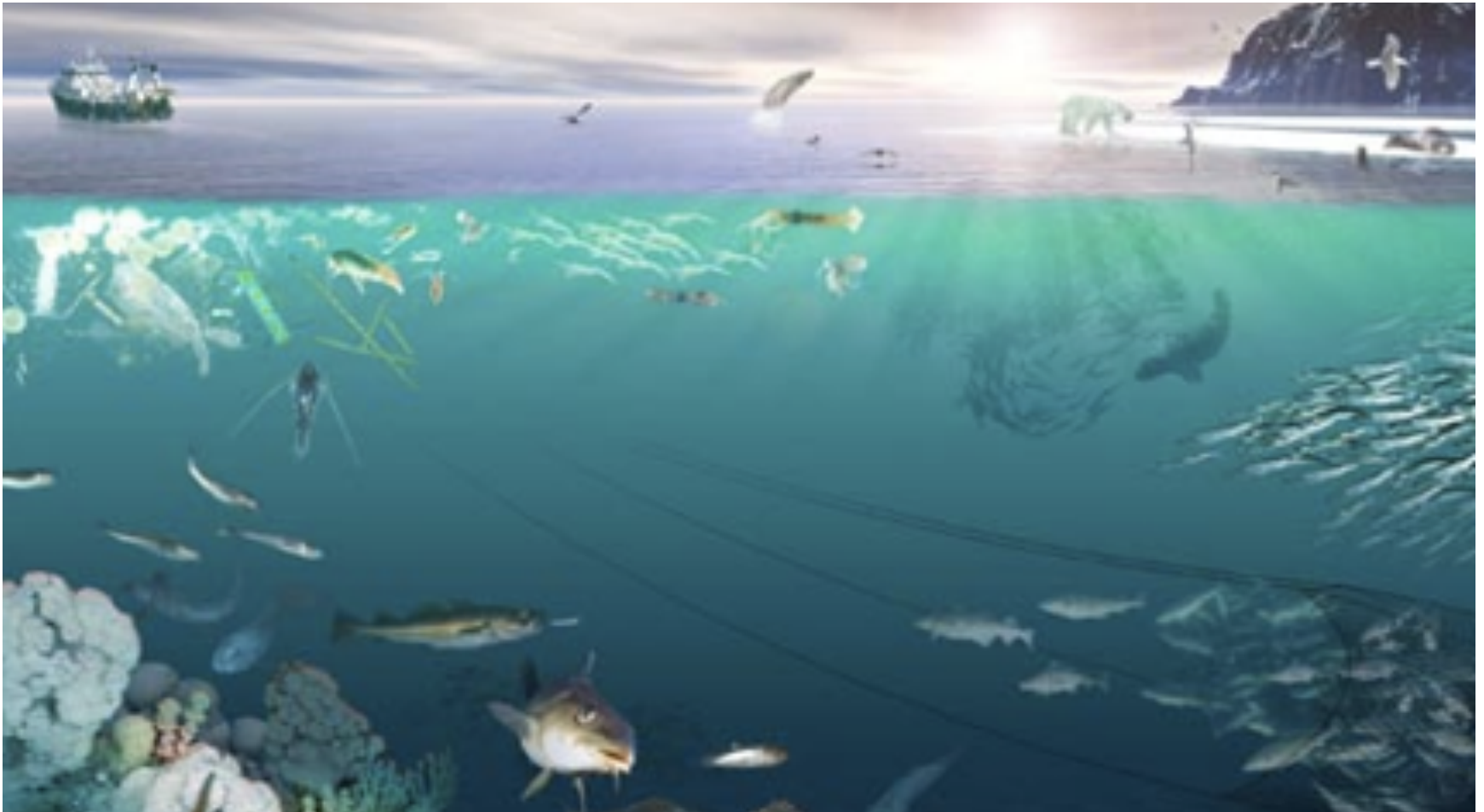
# Otwarte dane – przykład aplikacji

## Green Button

- 21 milionów amerykańskich gospodarstw domowych może ściągnąć dane dot. zużycia energii w ich domu
- Następnie wykorzystać aplikacje, które zarządzają ich zużyciem energii i zaoszczędzić pieniądze (i być bardziej ekologicznym)
- Więcej: [Energy.Data.gov](http://Energy.Data.gov)



# Otwarte dane to ekosystem



# Wspólna wizja

1. Wizja: Co będzie łączyć społeczność, jak współpraca będzie wyglądać w przyszłości?
2. Liderzy: Kto będzie przewodzić społeczności?
3. Uczestnicy: Kto będzie uczestniczyć?
4. Wyniki: Jakie są oczekiwane wyniki, miary ich osiągnięcia?
5. Funkcjonalność: Jakie typy aktywności będą funkcjonować (fora, blogi, wiki, rankingi konkursy, aplikacje)?
6. Treść: Jaka treść będzie pokazywana?
7. Interakcyjność: Jak społeczność będzie komunikować się z liderami i z zewnętrznymi osobami, jednostkami?



# Google: Graf Wiedzy

## semantyczne wyszukiwanie

**maj 2012: baza wiedzy wykorzystywana przez Google do rozszerzenia wyników wyszukiwania**

wiele źródeł wiedzy: *CIA World Factbook, Freebase, Wikipedia*

sieć semantyczna zawiera ponad 570 mln obiektów i ponad 18 mld faktów

**maj 2013: polska wersja językowa; zadawanie pytań raczej niż wyszukiwanie, informacje i powiązania między nimi raczej niż zestaw linków**

system poszukujący nie fraz kluczowych, lecz "bytów" stojących za wpisanymi w wyszukiwarkę słowami

## Dwa główne sposoby działania Grafu Wiedzy

**dopasowywanie odpowiedzi do kontekstu; w przypadku dwuznacznych haseł prezentacja różnych wersji odpowiedzi**

**podsumowania najbardziej istotnych informacji: - biogramy, wyróżnione najważniejsze elementy, powiązania między kluczowymi hasłami, odnośniki do kolejnych informacji**

# Graf Wiedzy: przykład

Google

[Internet](#) [Grafika](#) [Mapy](#) [Więcej ▾](#) [Narzędzia wyszukiwania](#)

Okolo 824,000 wyników (0,22 s)

[Maria Skłodowska-Curie – Wikipedia, wolna encyklopedia](#)  
pl.wikipedia.org/wiki/Maria\_Skłodowska-Curie ▾  
**Maria** Salomea Skłodowska-Curie (ur. 7 listopada 1867 w Warszawie, zm. 4 lipca 1934 w Passy) – uczona polsko-francuska, fizyczka, chemiczka, dwukrotna ...  
[Życiorys](#) - [Ciekawostki](#) - [Wyróżnienia i nagrody](#) - [Upamiętnienie](#)

[Marie Curie - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/Marie\_Curie ▾ [Tłumaczenie strony](#)  
**Marie Curie** (7 November 1867 – 4 July 1934), née Maria Salomea Skłodowska was a Polish physicist and chemist, working mainly in France, who is famous for ...

[Uniwersytet Marii Curie-Skłodowskiej w Lublinie](#)  
www.umcs.lublin.pl/ ▾  
Serwis Uniwersytetu Marii Curie - Skłodowskiej; informacje o studiach pierwszego i drugiego stopnia, jednolitych magisterskich, międzywydziałowych, ...


[Marie Curie Actions - Research Fellowship Programme - European ...](#)  
ec.europa.eu > ... > [Research & Innovation](#) ▾ [Tłumaczenie strony](#)  
**Marie Curie** Actions are European research grants for scientists in Europe and abroad.

[Maria Curie-Skłodowska \(1867-1834\) - Historia Kobiet](#)  
historia\_kobiet.w.interia.pl/teksty/skłodowska.html ▾  
**Maria Curie**-Skłodowska (1867-1834) wybitna fizyczka i chemiczka polska, żyjąca i pracująca we Francji, pierwsza kobieta będąca profesorem Sorbony, ...

[Maria Skłodowska-Curie – Wikicytaty](#)  
pl.wikiquote.org/wiki/Maria\_Skłodowska-Curie ▾  
**Maria** Salomea Skłodowska-Curie (1867–1934) – polsko-francuska fizyczka i chemiczka, z wykształcenia także matematyczka, dwukrotna laureatka Nagrody ...

[Marie Curie - Biography - Nobelprize.org](#)  
www.nobelprize.org > ... > [Nobel Prize in Physics](#) ▾ [Tłumaczenie strony](#)  
Short profile from the foundation that awards the Nobel Prize.

[Ludzie \(People\)](#)



[Więcej obrazów](#)

## Maria Skłodowska-Curie

Maria Salomea Skłodowska-Curie – uczona polsko-francuska, fizyczka, chemiczka, dwukrotna noblistka. Większość życia spędziła we Francji, tam też studiowała, a następnie rozwinęła swoją karierę naukową. Prekursorka nowej gałęzi chemii – radiochemii. Wikipedia

**Data i miejsce urodzenia:** 7 listopada 1867, [Warszawa](#)

**Data i miejsce śmierci:** 4 lipca 1934, [Passy, Francja](#)


**Odkrycia:** [Polon](#), [Rad](#)

**Wykształcenie:** [Sorbona](#) (1903), [Sorbona](#) (1894), [Sorbona](#) (1891–1893)

**Dzieci:** [Irène Joliot-Curie](#), [Ève Curie](#)

**Nagrody:** Nagroda Nobla w dziedzinie fizyki, Nagroda Nobla w dziedzinie chemii

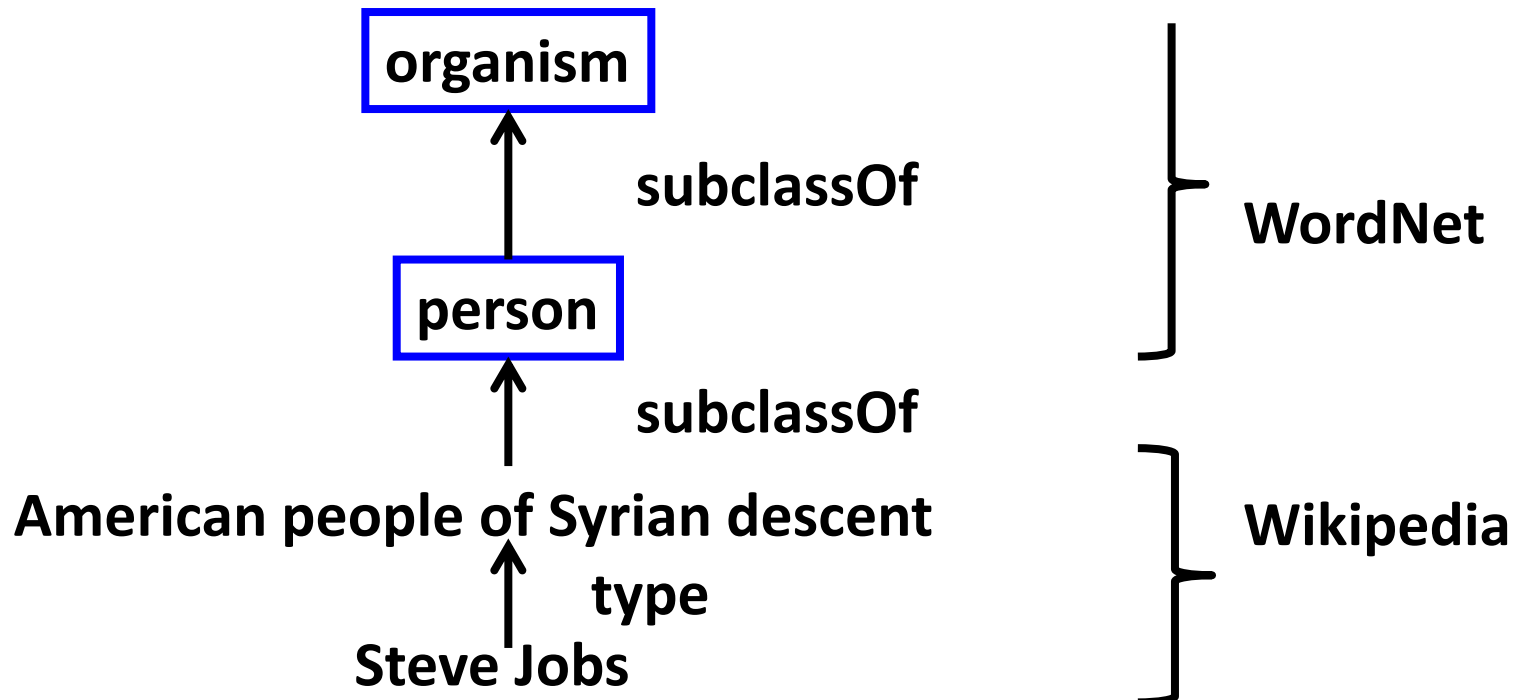
### Podobne wyszukiwania



[Pierre Curie](#) [Henri Becquerel](#) [Albert Einstein](#) [Ernest Rutherford](#) [Irène Joliot-Curie](#)

# Yago

- **Wordnet + Wikipedia**



## Ekstrakcja wiedzy (fakty)

- **dane semi-strukturalne**  
Wikipedia: infoboksy i kategorie  
tabelki i listy HTML
- **tekst**  
wzorce Hearsta  
„płytkie” NLP
- **tabele Web**

# Wzorce Hearsta

- **Cel: znajdź instancje klas**
- **Hearst zdefiniował leksykalno-syntaktyczne wzorce:**
  - X such as Y; X like Y;
  - X and other Y; X including Y;
  - X, especially Y;
- **Znajdywanie takich wzorców w tekście:**
  - companies such as Apple
  - Google, Microsoft and other companies
  - Internet companies like Amazon and Facebook
- **Wywiedzenie typu (Y,X)**
  - type(Apple, company), type(Google, company), ...

# Ekstrakcja z Infoboksów

- język znaczników

**Andrzej Sapkowski**



Andrzej Sapkowski at Book World Fair in Prague in 2010

<b>Born</b>	21 June 1948 (age 66) Łódź, Poland
<b>Occupation</b>	Novelist
<b>Citizenship</b>	<span><span><span></span></span><span> </span></span> Poland
<b>Period</b>	1986–present
<b>Genre</b>	Fantasy, history
<b>Notable works</b>	The Witcher Saga The Hussite Trilogy
<b>Notable awards</b>	Janusz A. Zajdel Award Paszport Polityki

```
{{Infobox writer <!-- for more information see [[:Template:Infobox writer/doc]] -->
| name      = Andrzej Sapkowski
| image     = Andrzej Sapkowski - Book World Fair 2010 Prague.jpg
| caption   = Andrzej Sapkowski at Book World Fair in [[Prague]] in 2010
| alt       = Andrzej Sapkowski at Book World Fair in Prague in 2010
| birth_date = {{birth date and age|1948|6|21|df=y}}
| birth_place = [[Łódź]], [[Poland]]
| occupation = [[Novelist]]
| citizenship = {{POL}}
| period    = 1986–present
| genre     = [[Fantasy]], [[history]]
| movement  =
| notableworks = [[The Witcher]] Saga<br />The Hussite Trilogy
| awards    = [[Janusz A. Zajdel Award]]<br />[[Paszport Polityki]]
| signature = Andrzej_Sapkowski_signature.JPG
| website   = {{URL|http://www.sapkowski.pl/}}
}}
```

## Ekstrakcja z Infoboksów

- ekstrakcja za pomocą wyrażeń regularnych

Przykład – ekstrakcja daty urodzin:

```
\{\{birth date \|(\d+)\|(\d+)\|(\d+)\}\}
```

```
| birth_date = {\{birth date |1948|6|21\}}
```



# Tabele Web

- dostarczają strukturalnej/relacyjnej informacji

## List of laureates

Year ↕	Physics ↕	Chemistry ↕	Physiology or Medicine ↕	Literature ↕	Peace ↕	Economics ↕
1901	Wilhelm Röntgen	Jacobus Henricus van 't Hoff	Emil Adolf von Behring	Sully Prudhomme	Henry Dunant; Frédéric Passy	—
1902	Hendrik Lorentz; Pieter Zeeman	Hermann Emil Fischer	Ronald Ross	Theodor Mommsen	Élie Ducommun; Charles Albert Gobat	—
1903	Henri Becquerel; Pierre Curie; Marie Curie	Svante Arrhenius	Niels Ryberg Finsen	Bjørnstjerne Bjørnson	Randal Cremer	—
1904	Lord Rayleigh	William Ramsay	Ivan Pavlov	Frédéric Mistral; José Echegaray	Institut de Droit International	—
1905	Philipp Lenard	Adolf von Baeyer	Robert Koch	Henryk Sienkiewicz	Bertha von Suttner	—
1906	J. J. Thomson	Henri Moissan	Camillo Golgi; Santiago Ramón y Cajal	Giosuè Carducci	Theodore Roosevelt	—
1907	Albert Abraham Michelson	Eduard Buchner	Charles Louis Alphonse Laveran	Rudyard Kipling	Ernesto Teodoro Moneta; Louis Renault	—
1908	Gabriel Lippmann	Ernest Rutherford	Élie Metchnikoff; Paul Ehrlich	Rudolf Christoph Eucken	Klas Pontus Arnoldson; Fredrik Bajer	—
1909	Karl Ferdinand Braun; Guglielmo Marconi	Wilhelm Ostwald	Emil Theodor Kocher	Selma Lagerlöf	Auguste Marie François Beernaert; Paul-Henri-Benjamin d'Estournelles de Constant	—
1910	Johannes Diderik van der Waals	Otto Wallach	Albrecht Kossel	Paul Heyse	International Peace Bureau	—

# Pozyskiwanie wiedzy poprzez CrowdSourcing

- **wiedza powszechna (nie-ekspercka)**
- **gry z celem**
- **platformy Amazon Mechanical Turk, CrowdFlower**

## Gry z celem

Games with a purpose (GWAP) =  
Technika oparta na obliczeniach  
wykonywanych przez ludzi (human-based  
computation)

**Proces obliczeniowy wykonywany jest poprzez zlecenie niektórych czynności ludziom do wykonania w zabawny, zajmujący sposób**

**GWAP wykorzystuje różnice w umiejętnościach i kosztach pracy ludzi i metod informatycznych w celu osiągnięcia symbiotycznej interakcji człowiek-komputer**

# Gry z celem

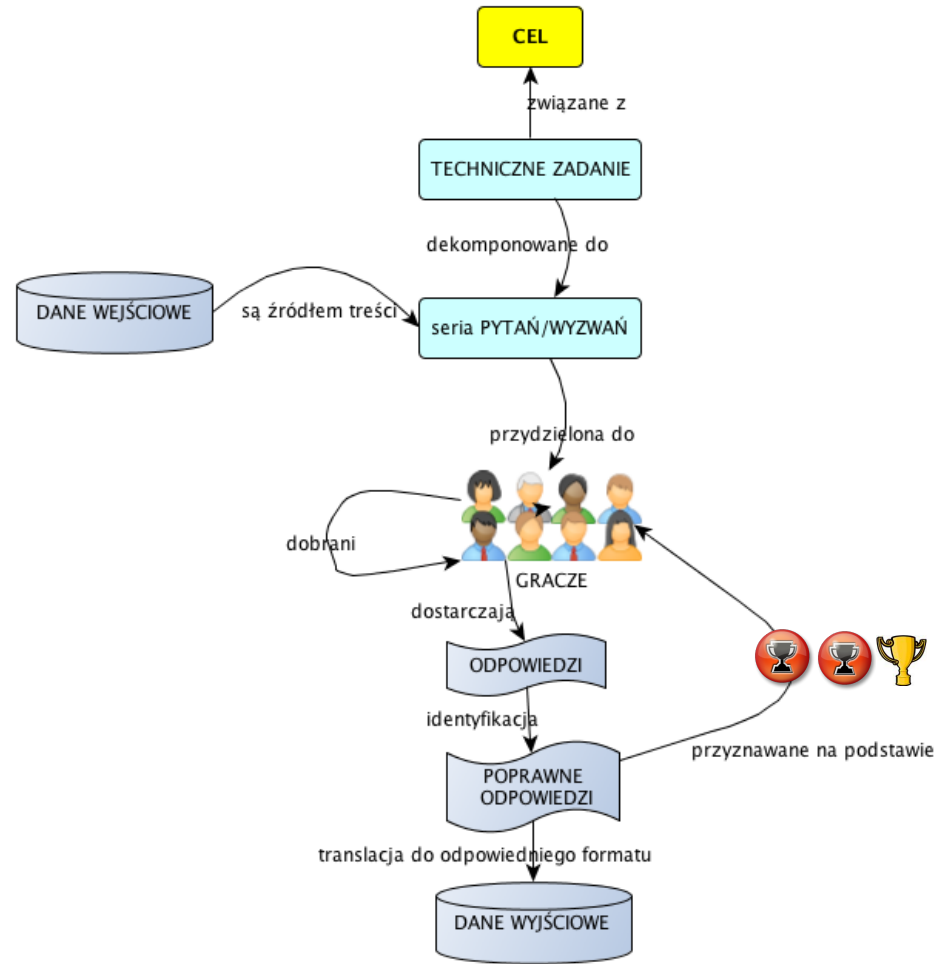
Luis Von Ahn (2006)

Główna motywacja: nie leży w rozwiązaniu instancji problemu obliczeniowego, jest to ludzkie pragnienie **zabawy**

W GWAP ludzie wykonują **pożyteczne obliczenia** jako **efekt uboczny przyjemnej rozrywki**

Miarą **użyteczności** GWAP jest kombinacja **wygenerowanych wyników** i **przyjemności rozgrywki**

# Kluczowe elementy GWAP



## **Gry z celem tworzenia treści, wiedzy**

**Adnotacja tekstu/audio/obrazów/video**

**Konstrukcja ontologii**

**Mapowanie ontologii**

**Tworzenie linków między zasobami**

**„Wyścigi Wiki”**

# **Konstrukcja ontologii: OntoPronto (Ontogame)**

**Dwuosobowa gra quizowa**

**Cel: budowa ontologii dziedzinowej będącej  
rozszerzeniem ontologii Proton**

**Zasady: Gracze czytają streszczenie losowo wybranego  
artykułu z Wikipedii i odpowiadają na zapytania o  
relacji tego artykułu w stosunku do ontologii  
Proton.**

**Dane wyjściowe: Ontologia dziedzinowa ufundowana  
na ontologii Proton**

**Walidacja: konsensus, większość**

# OntoPronto (Ontogame)

time left  
1:45

score  
20

## ontogame

all data taken from wikipedia.org

### Deviled crab

A deviled crab (croqueta de jaiba) is a simple croquette made from seasoned and cooked crab meat which is breaded, rolled into the approximate shape of a football or small potato, and fried.[1]The dish originated in the immigrant community of Ybor City, Tampa, Florida during the Great Depression and uses ingredients that were cheap and readily available at the time: crab and stale brea ...

**your choice**

Is this wikipedia page about a:

single object or happening  
= instance →

or rather describing a:

set/type of objects  
←  = class

**hint** → „Dog“ is a class (as several entities of this class exist)  
„Lassie“ is an Instance

Starting Singleplayer mode ...

SKIP OK



# Mapowanie ontologii: SpotTheLink

**Dwuosobowa gra quizowa**

**Cel: uzgadnianie ontologii, np. Dbpedia i Proton**

**Zasady: Graczom prezentowane jest pojęcie z jednej ontologii. Pierwszy krok: zgadzają się co do odpowiadającego mu pojęcia w drugiej ontologii. Krok drugi: zgadzają się co do relacji wiążącej te dwa pojęcia.**

**Dane wyjściowe: Odwzorowanie (w języku SKOS) pomiędzy pojęciami w ontologiach**

**Walidacja: konsensus, większość**

# SpotTheLink

time left  
01:28

score  
20

ontogame

Choose a related concept!

Film Festival,  
e.g.: Sarajevo Film Festival, Istanbul  
International Film Festival or  
Melbourne International Film Festival

- Happening
  - Event
  - Situation
  - TimeInterval
- Abstract
- Object

This Concept  is more specific  This Concept  
 is the same

Select the most fitting concept in the tree.

OK SKIP

# **Tworzenie linków między zasobami:** **VeriLinks**

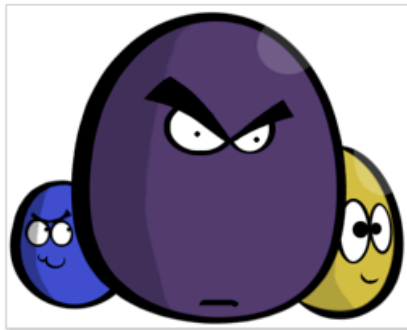
**Cel: walidacja linków w arbitralnym zbiorze danych**

**Zasady: Zgoda graczy co do poprawności linku jest nagradzana monetami, które są następnie wykorzystywane do zwalczania najeźdźców w grze polegającej na obronie wieży.**

**Dane wyjściowe: zwalidowane linki**

# VeriLinks

## Welcome to Pea-Invasion



**Pea-Invasion** belongs to the *Games With A Purpose (GWAP)* genre. Playing Pea-Invasion means turning this world in to a better place.. well.. kind of..

While playing this game, you'll generate useful data as a byproduct - which will be used to improve the semantic web. The data is collected when you answer the questions in the **quiz**. You can also skip difficult questions, so please answer them only if you're certain :)

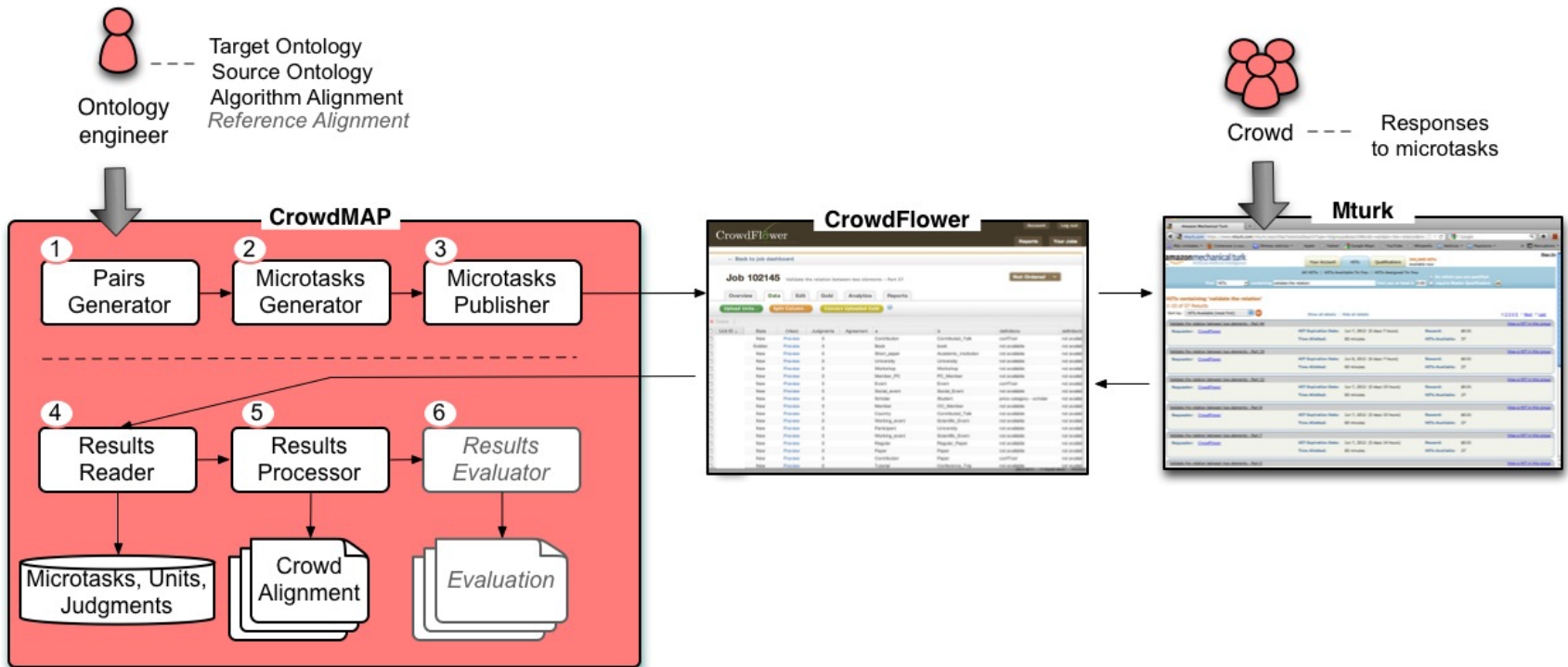
### Goal

- [1] Fight the enemy troops and increase your score!
- [2] Answer the quiz to gain coins!

### Credits

CONTINUE

# CrowdMap



Cristina Sarasua, Elena Simperl, Natalya Fridman Noy:

CrowdMap: Crowdsourcing Ontology Alignment with Microtasks. International Semantic Web Conference (1) 2012: 525-541

## **Dalsze uwagi**

**Nie każde zadanie da się łatwo przerobić na GWAP  
(wymóg dekompozycji na mikro-zadania)**

**Tworzenie niektórych ontologii wymaga bardzo  
specjalistycznej wiedzy**

**To co powstaje w wyniku GWAP jest raczej „płytkim”  
modelem**

**GWAP wymaga strategii zapobiegania oszustwom**

# Scenariusze użycia baz wiedzy

- **Wyszukiwanie w Internecie**
  - encje a nie słowa
- **Odpowiadanie na zapytania**
  - np. IBM Watson, nawigatory wiedzy
- **Analityka tekstów**
  - np. teksty medyczne
- **Analityka *Big Data***
  - efekty uboczne kombinacji leków
- **Semantyczne wyszukiwanie**
  - np. wiadomości

# Bibliografia

- [1] C. Bizer, T. Heath, and T. Berners-lee “Linked Data – The Story So Far” International Journal on Semantic Web and Information Systems (IJSWIS) (2009)
- [2] T. Heath, and C. Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
- [3] RDFa Primer, <http://www.w3.org/TR/xhtml-rdfa-primer/> (last accessed on 18.03.2009)