Semantic Data Analytics – The key to challenging Big Data

Pascal Hitzler
Kno.e.sis Center
Wright State University, Dayton, OH
http://www.pascal-hitzler.de/

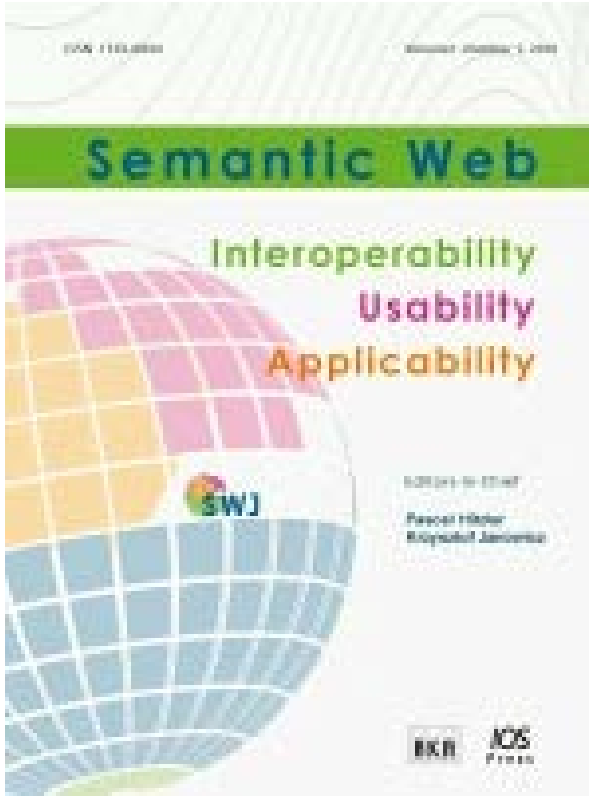
Semantic Web journal

- **EiCs:** **Pascal Hitzler**
  Krzysztof Janowicz

- **New journal with significant initial uptake.**

- **We very much welcome contributions at the "rim" of traditional Semantic Web research – e.g., work which is strongly inspired by a different field.**

- **Non-standard (open & transparent) review process.**



- **http://www.semantic-web-journal.net/**

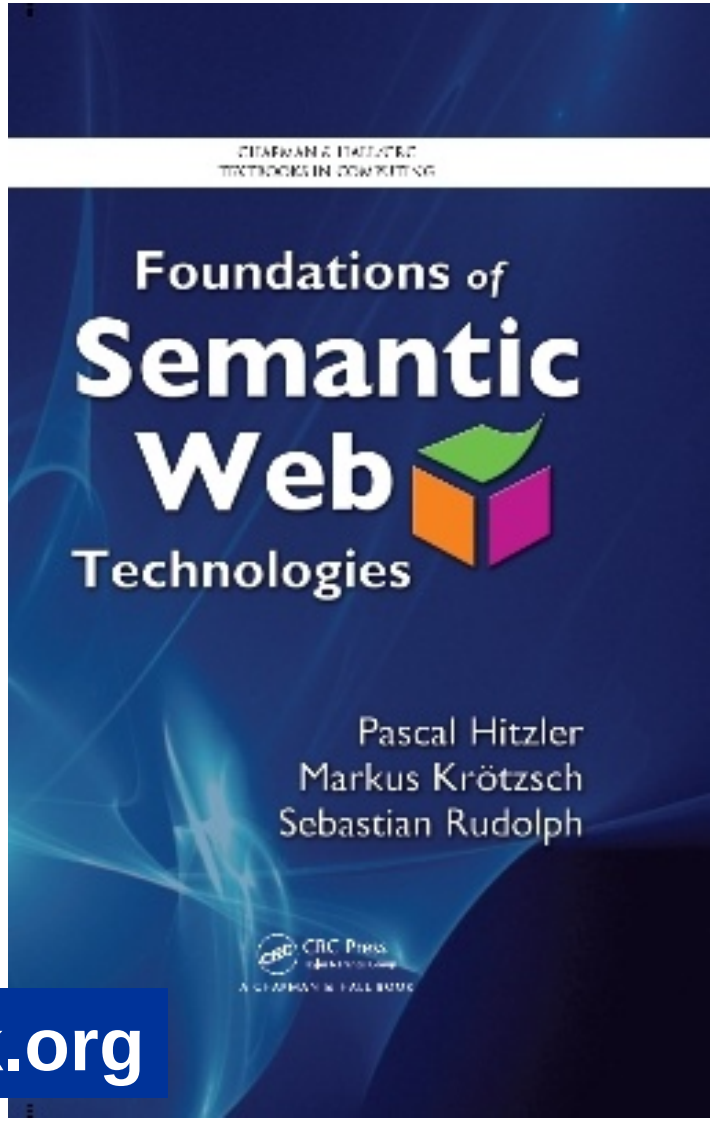
Textbook

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Choice Magazine Outstanding Academic Title 2010 (one out of seven in Information & Computer Science)



http://www.semantic-web-book.org

Contents

- Big Data, Linked Data, Semantic Web
- An Example: Linked Data Querying
- The Big Data Added Value Pipeline

Big Data

KNO.E.SIS

ITBUSINESSEDGE | An IT Business Edge Site

ITBUSINESSEDGE

Welcome, Guest
Log in | Register

Home | IT Projects | Blogs | IT Downloads | White Papers | Newsletters | Google™ Custom Search | Search

Business Alignment & Management | Business Integration | Governance | Infrastructure | Mobile Technology | Security | Sourcing | Vendors & Markets | More

Home → Blogs → Charting Your IT Career → Gartner: Big Data Will Generate 6 Million U.S. Jobs By 2015

Like us:

Related Content

Topic : Data Management
Data management involves a variety of tasks involved with the full data lifecycle

Blog : Six Reasons IT Can Be Thankful on Thursday

Article : The Fast Evolution - and Transformation - of Mobile Device Management

Gartner: Big Data Will Generate 6 Million U.S. Jobs by 2015

Susan Hall | CHARTING YOUR IT CAREER | 23 OCT, 2012

Share

Gartner predicts that 4.4 million IT **jobs will be created to support Big Data** by 2015, with 1.9 million of them to be in the United States.

In addition, every Big Data-related role in the United States will create employment for three people outside of IT, pushing the total to 6 million U.S. jobs, Peter Sondergaard, senior vice president at Gartner and global head of research, told those attending the Gartner Symposium/ITxpo. He said:

> But there is a challenge. There is **not enough talent** in the industry. Our public and private education systems are failing us. Therefore, only one-third of the IT jobs will be filled. Data experts will be a scarce, valuable commodity," he said. "IT leaders will need immediate focus on how their organization develops and attracts the skills required. These jobs will be needed to grow your business. These jobs are the future of the new information economy.

Though I don't follow Sondergaard's math, we know there's a **shortage of analytics talent for Big Data** and for **engineering talent as well**

Subscribe to our Newsletters

Sign up now and get the best business technology insights

Big Data

Big Data is characterized not only by the enormous volume or the velocity of its generation but also by the heterogeneity, diversity and complexity of the data.

Suzi Iacono, source: http://community.topcoder.com/coeci/nitrd/

- **volume: the sheer size of the data**
- **velocity: new data is added at breathtaking speed**
- **variety: different formats and different perspectives**
- **(value: how useful is the data?)**
- **(veracity: how good/reliable is the data?)**

Contents

- Big Data, Linked Data, Semantic Web
- An Example: Linked Data Querying
- The Big Data Added Value Pipeline

Linked Data: Volume

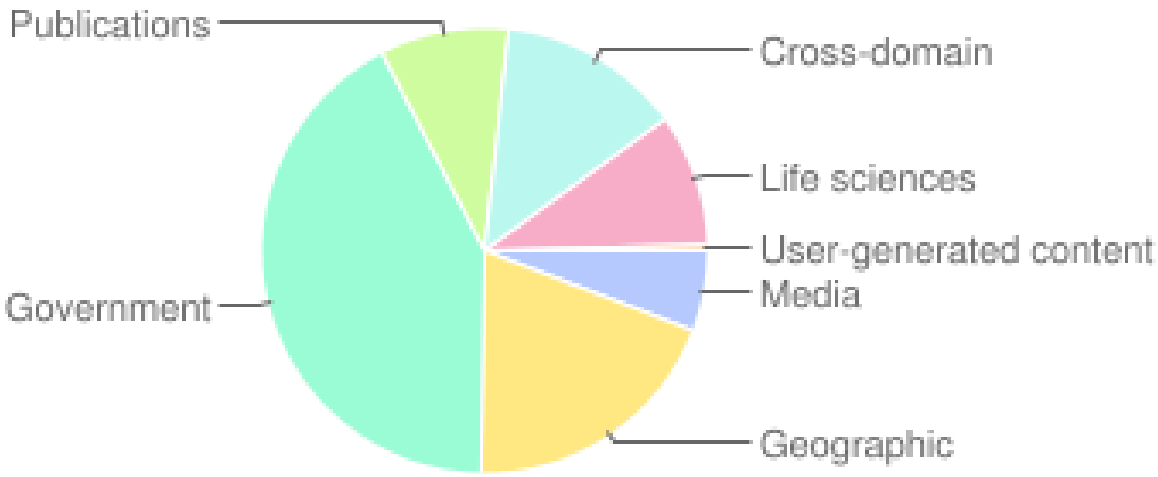
Number of Datasets

2011-09-19	**295**
2010-09-22	**203**
2009-07-14	**95**
2008-09-18	**45**
2007-10-08	**25**
2007-05-01	**12**

Number of triples (Sept 2011)

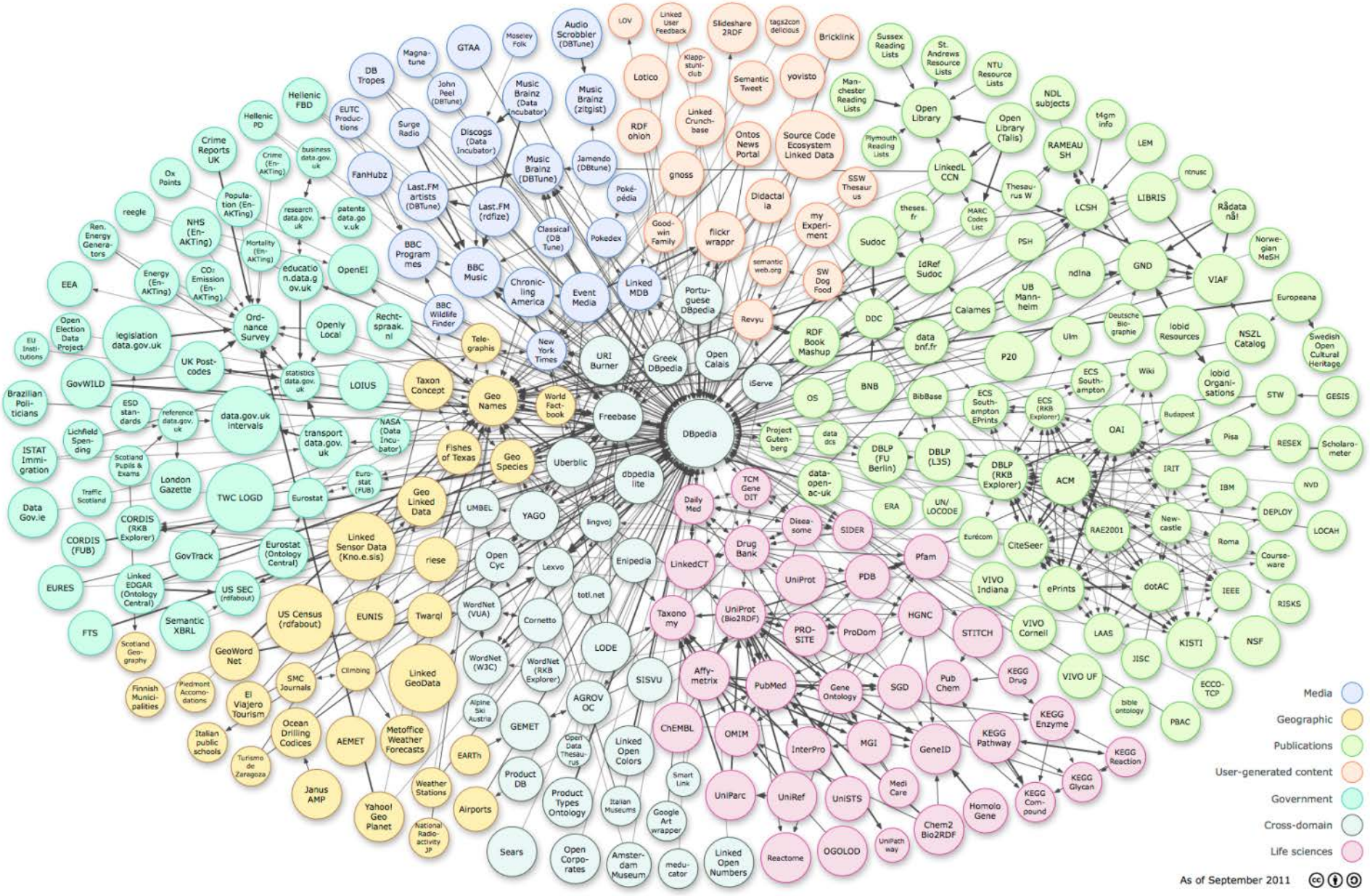
31,634,213,770

with 503,998,829 out-links



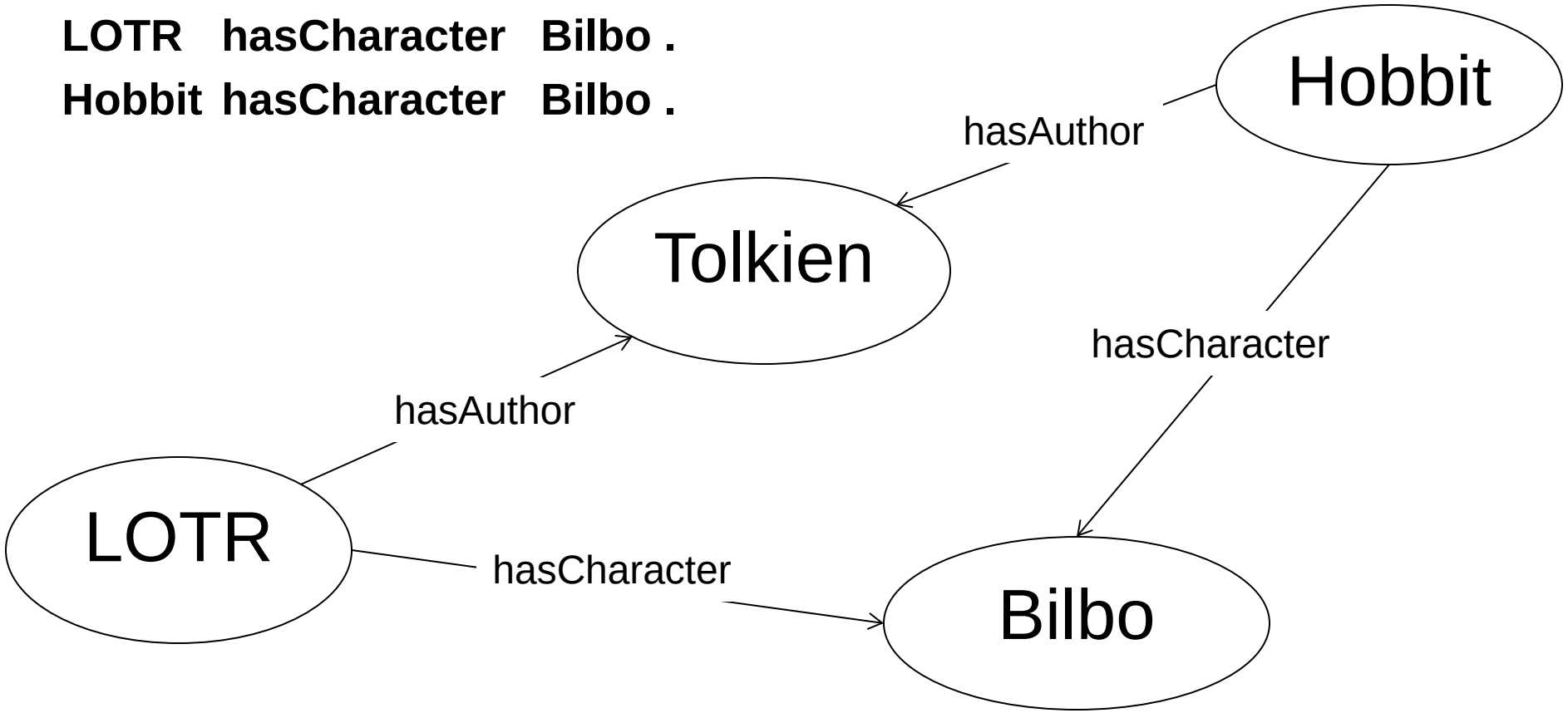
From http://www4.wiwiss.fu-berlin.de/lodcloud/state/

Linked Data 2011



Information as RDF triples / graph

LOTR hasAuthor Tolkien .
Hobbit hasAuthor Tolkien .
LOTR hasCharacter Bilbo .
Hobbit hasCharacter Bilbo .



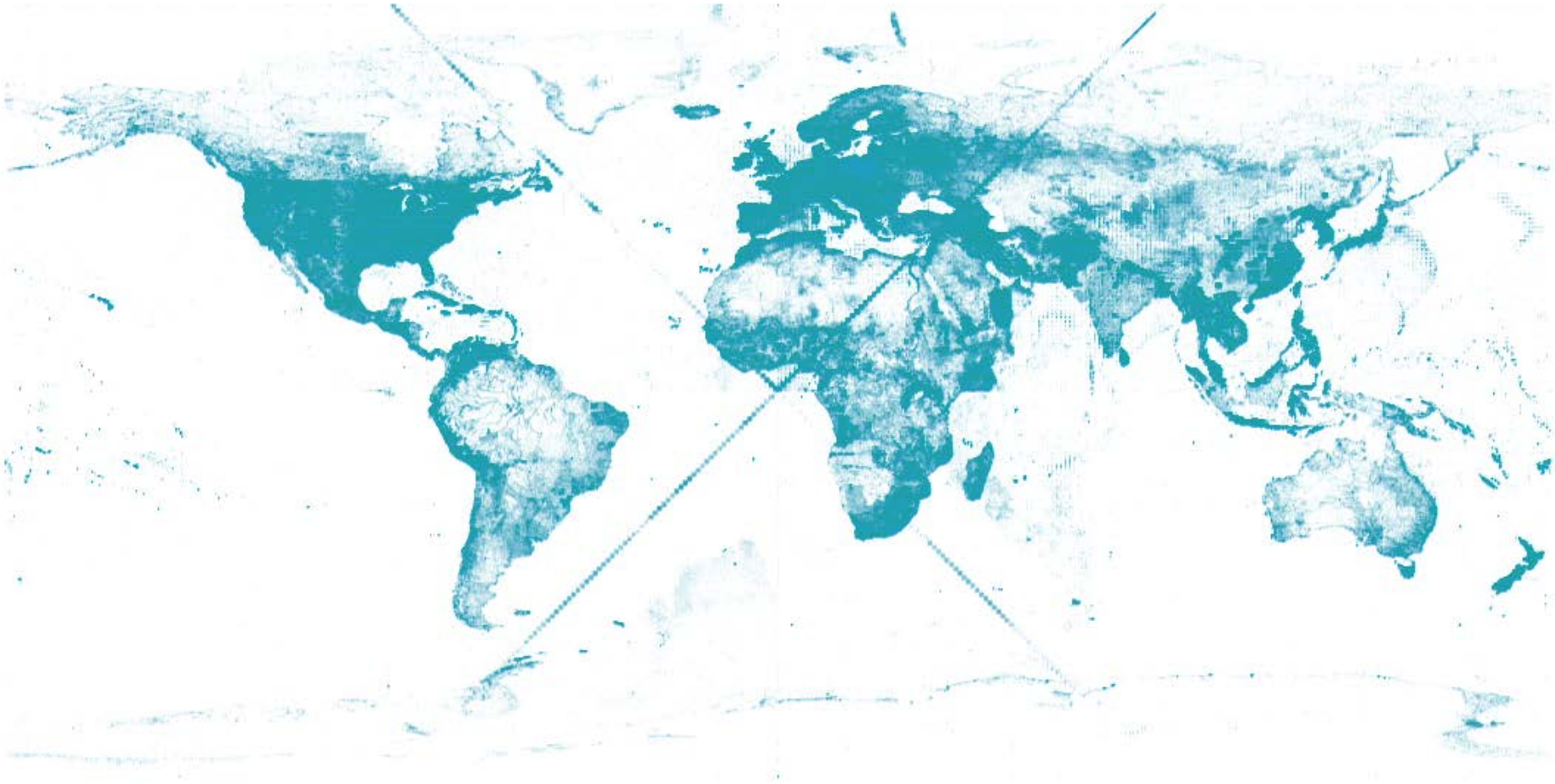
DBpedia: LOTR page

dbpedia-owl:thumbnail	▪ http://upload.wikimedia.org/wikipedia/commons/thumb/6/62/Jrrt_lotr_cover_design.jpg/200px-Jrrt_lotr_cover_design.jpg
dbpedia-owl:wikiPageExternalLink	▪ http://lotr.wikia.com ▪ http://www.glyphweb.com/arda/ ▪ http://www.tolkienlibrary.com/ ▪ http://www.tolkien.co.uk/ ▪ http://www.houghtonmifflinbooks.com/features/lordoftheringstrilogy/
dbpprop:author	▪ dbpedia:J._R._R._Tolkien
dbpprop:books	▪ dbpedia:The_Two_Towers ▪ dbpedia:The_Return_of_the_King ▪ dbpedia:The_Fellowship_of_the_Ring ▪ "Volumes:"
dbpprop:country	▪ England
dbpprop:expiry	▪ 20 (xsd:integer)
dbpprop:genre	▪ dbpedia:Adventure_novel ▪ dbpedia:High_fantasy
dbpprop:hasPhotoCollection	▪ http://www4.wiwiss.fu-berlin.de/flickrwrappr/photos/The_Lord_of_the_Rings
dbpprop:imageCaption	▪ Tolkien's own cover designs for the three volumes
dbpprop:language	▪ English
dbpprop:mediaType	▪ Print
dbpprop:name	▪ The Lord of the Rings
dbpprop:pages	▪ 1216 (xsd:integer)
dbpprop:precededBy	▪ dbpedia:The_Hobbit
dbpprop:pubDate	▪ 21 (xsd:integer)
dbpprop:publisher	▪ dbpedia:Allen_&_Unwin
dbpprop:small	▪ yes
dbpprop:wikiPageUsesTemplate	▪ dbpedia:Template:Infobox_book_series ▪ dbpedia:Template:Pp-vandalism
dcterms:subject	▪ category:Monomyths ▪ category:High_fantasy_novels ▪ category:Middle-earth_books ▪ category:British_fantasy_novels ▪ category:Fantasy_books_by_series ▪ category:1950s_fantasy_novels ▪ category:Sequel_novels ▪ category:The_Lord_of_the_Rings ▪ category:English_novels

Linked Data: Volume

Geoindexed Linked Data – courtesy of Krzysztof Janowicz

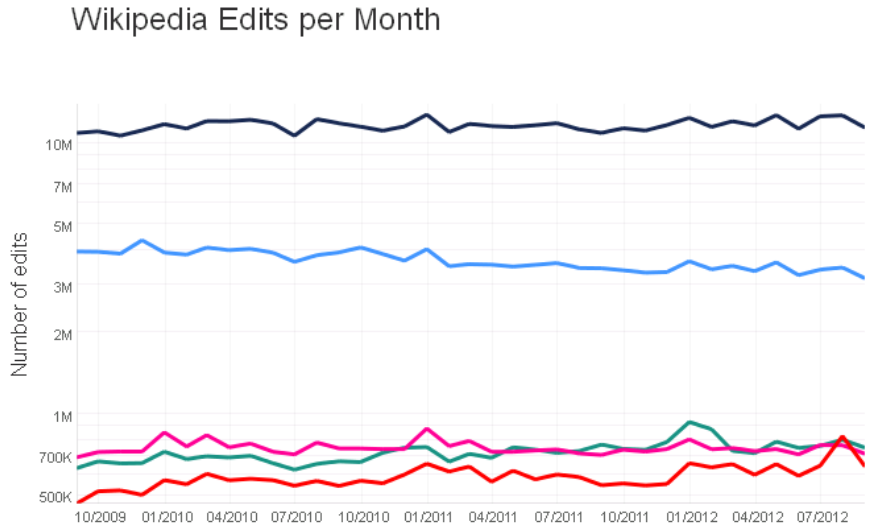
http://stko.geog.ucsb.edu/location_linked_data



Data Velocity

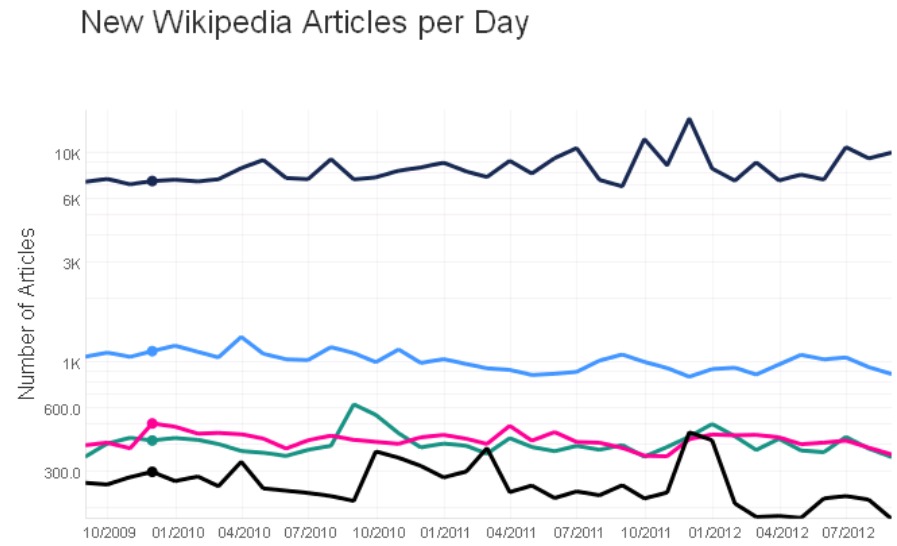
11.39 Million

Sep 11 — Sep 12 4.64%
Aug 12 — Sep 12 -9.85%



10,028.00

Sep 11 — Sep 12 45.19%
Aug 12 — Sep 12 6.68%



- **Weather sensors**
- **Tweets**
- **Satellite images**
- **…**

Linked Data: Variety

Copernicus lunar crater located on earth – courtesy of Krzysztof Janowicz http://stko.geog.ucsb.edu/location_linked_data (missing reference coordinate system)

Information Workbench

Print Help Login

Copernicus (lunar crater)

You do not have permission to edit this page. View Revisions

Copernicus is a lunar impact crater named after the astronomer Nicolaus Copernicus, located in eastern Oceanus Procellarum. It is estimated to be about 800 million years old, and typifies craters that formed during the Copernican period in that it has a prominent ray system.

Contents

Characteristics

Copernicus is visible using binoculars, and is located slightly northwest of the center of the Moon's Earth-facing hemisphere. South of the crater is the Mare Insularum, and to the south-south west is the crater Reinhold. North of Copernicus are the Montes Carpatus, which lie at the south edge of Mare Imbrium. West of Copernicus is a group of dispersed lunar hills. Due to its relative youth, the crater has remained in a relatively pristine shape since it formed.

The circular rim has a discernible hexagonal form, with a terraced inner wall and a 30 km wide, sloping rampart that descends nearly a kilometer to the surrounding mare. There are three distinct terraces visible, and arc-shaped landslides due to slumping of the inner wall as the crater debris subsided.

Most likely due to its recent formation, the crater floor has not been flooded

Location of Copernicus.

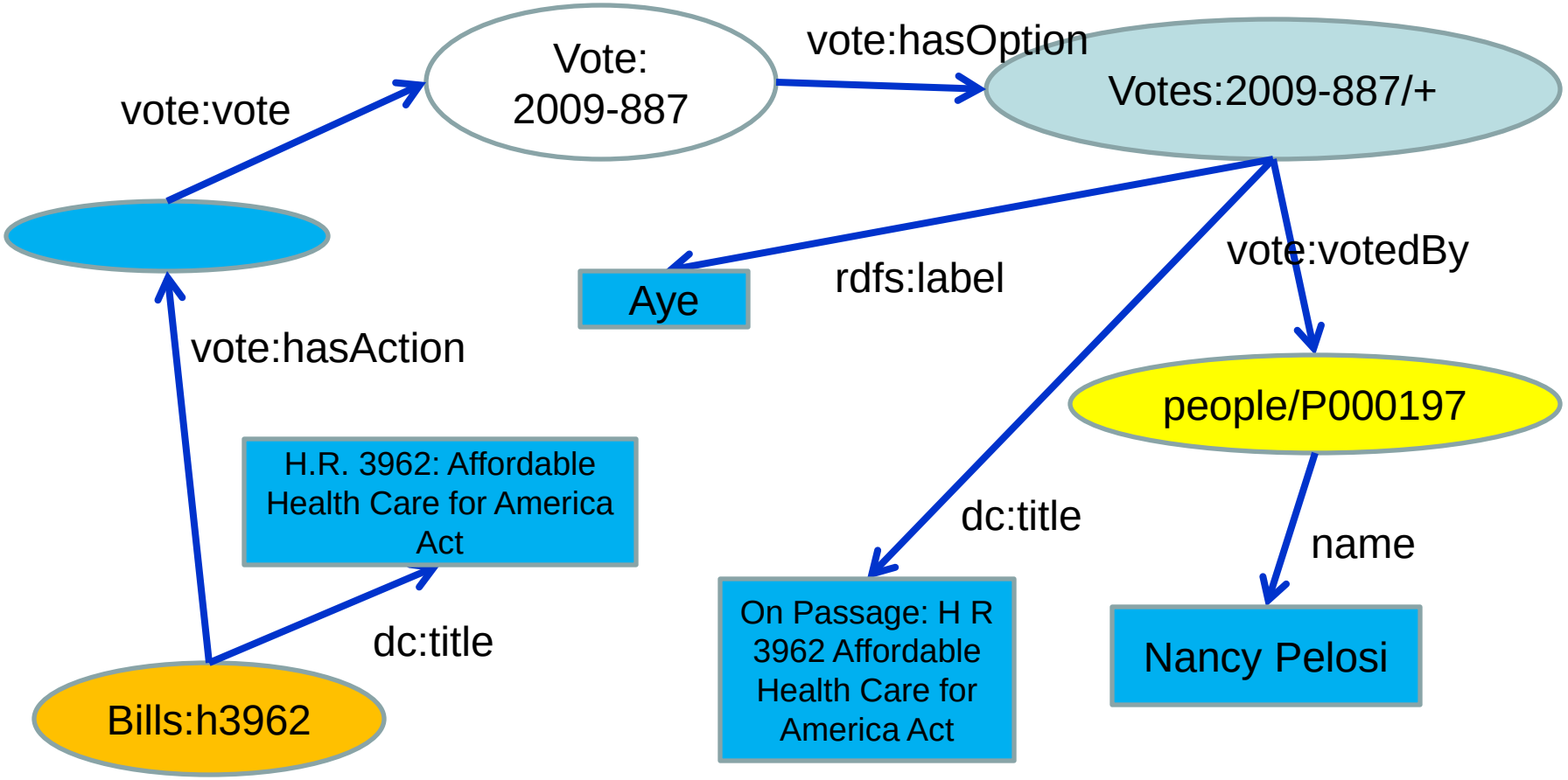
Location of Copernicus.

Image

Google Map

Linked Data: Value (GovTrack)

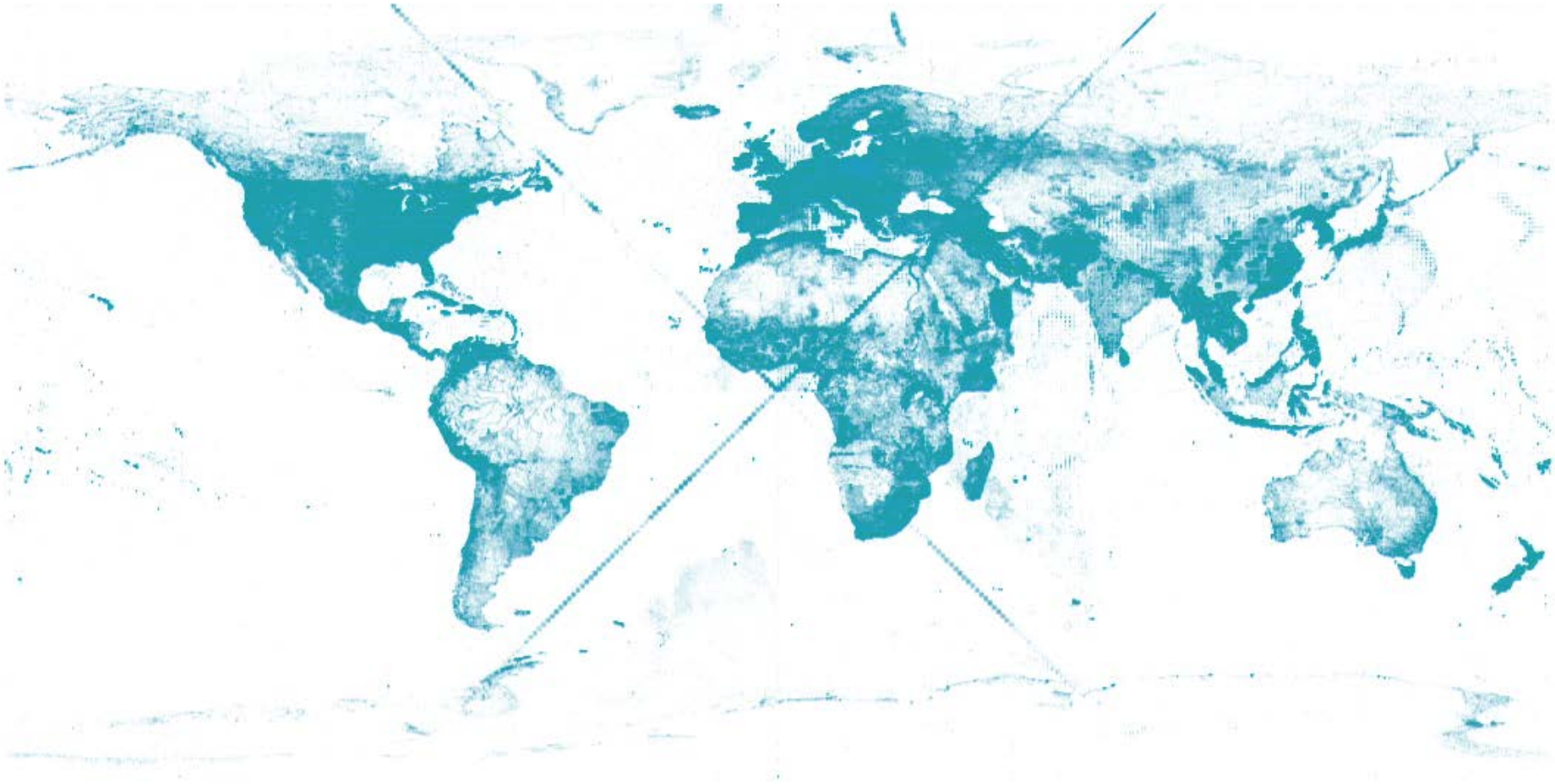
“Nancy Pelosi voted in favor of the Health Care Bill.”



Linked Data: Veracity

Geoindexed Linked Data – courtesy of Krzysztof Janowicz

http://stko.geog.ucsb.edu/location_linked_data



Linked Data: Veracity

Courtesy of Krzysztof Janowicz

http://stko.geog.ucsb.edu/location_linked_data

FactForge

RDF Search and Explore | SPARQL | RelFinder | About | Contact |

SPARQL Query

Results for your query (6) - Edit query

View as Exhibit Download SPARQL Results in: JSON | XML

place	populationCount
dbpedia:Keta	18077
http://sws.geonames.org/2304548/	29748
w-flick:Aneho	47579
http://sws.geonames.org/6295630/	6814400000
dbpedia:Lomé	749700
http://sws.geonames.org/2393947/	9847

Linked Data: Veracity

Courtesy of Krzysztof Janowicz

http://stko.geog.ucsb.edu/location_linked_data

FactForge

RDF Search and Explore | SPARQL | RelFinder | About | Contact

Welt

RDF Rank

Source: http://sws.geonames.org/6295630/

RDF Search and Explore

Subject (100 of 8935127) | Predicate | Object | All

Download in: JSON | RDF | N3/Turtle | N-Triples

Statements in which the resource exists as a subject. Named Gr

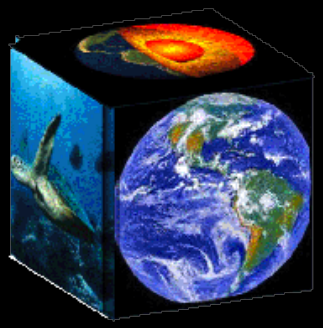
Predicate	Object
rdf:type	http://schema.org/Place, geo-ont:Feature,
rdfs:seeAlso	dbpedia:Earth, http://sws.geonames.org/6
rdfs:isDefinedBy	http://sws.geonames.org/6295630/about.r
rdfs:label	Earth@en, Globe@en, World@en
skos:altLabel	Earth@en, Globe@en, World@en
dc:type	geo-ont:L, geo-ont:L.AREA
dc-term:type	geo-ont:L, geo-ont:L.AREA
geo-pos:long	0
geo-pos:lat	0

Coordinates: 1°0′N 4°0′E

Map of the Gulf of Guinea, showing the chain of islands formed by the Cameroon line of volcanoes.

EarthCube requires

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support



Semantic Web studies

- information integration
- interoperability
- conceptual modeling
- intelligent search
- data-model intercomparison
- data publishing support

Pascal Hitzler, WSU; Krzysztof Janowicz, UCSB

Linked Data and Big Data

- **Linked Data is a kind of structured Big Data**
- **Linked Data is Big Data in a nutshell**

Many of the same problems

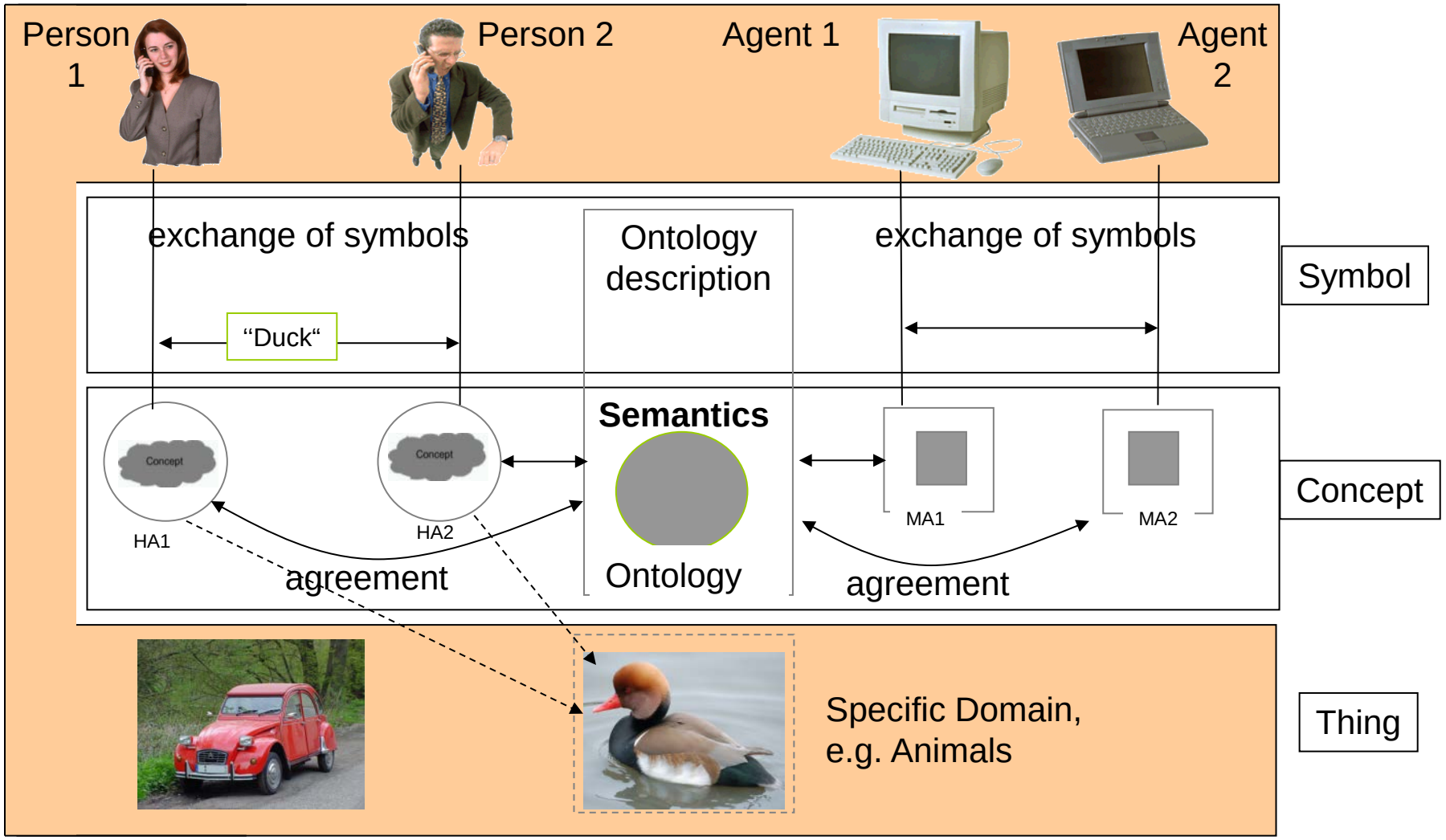
Testbed for Big Data solutions

Intermediate stage for getting *semantics* into Big Data

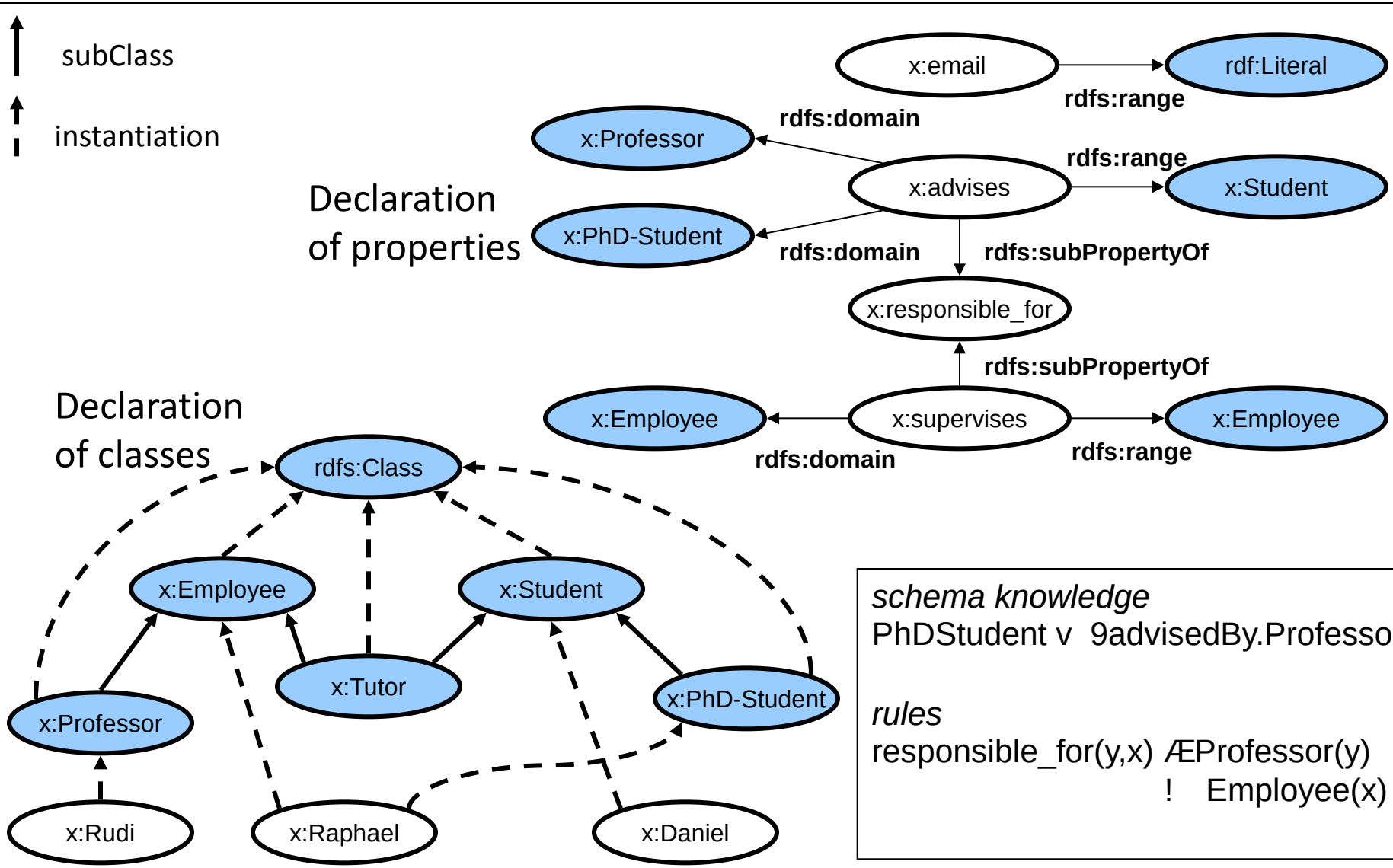
Contents

- **Big Data, Linked Data, Semantic Web**
- **An Example: Linked Data Querying**
- **The Big Data Added Value Pipeline**

Basic Idea of the Semantic Web



Ontology Example



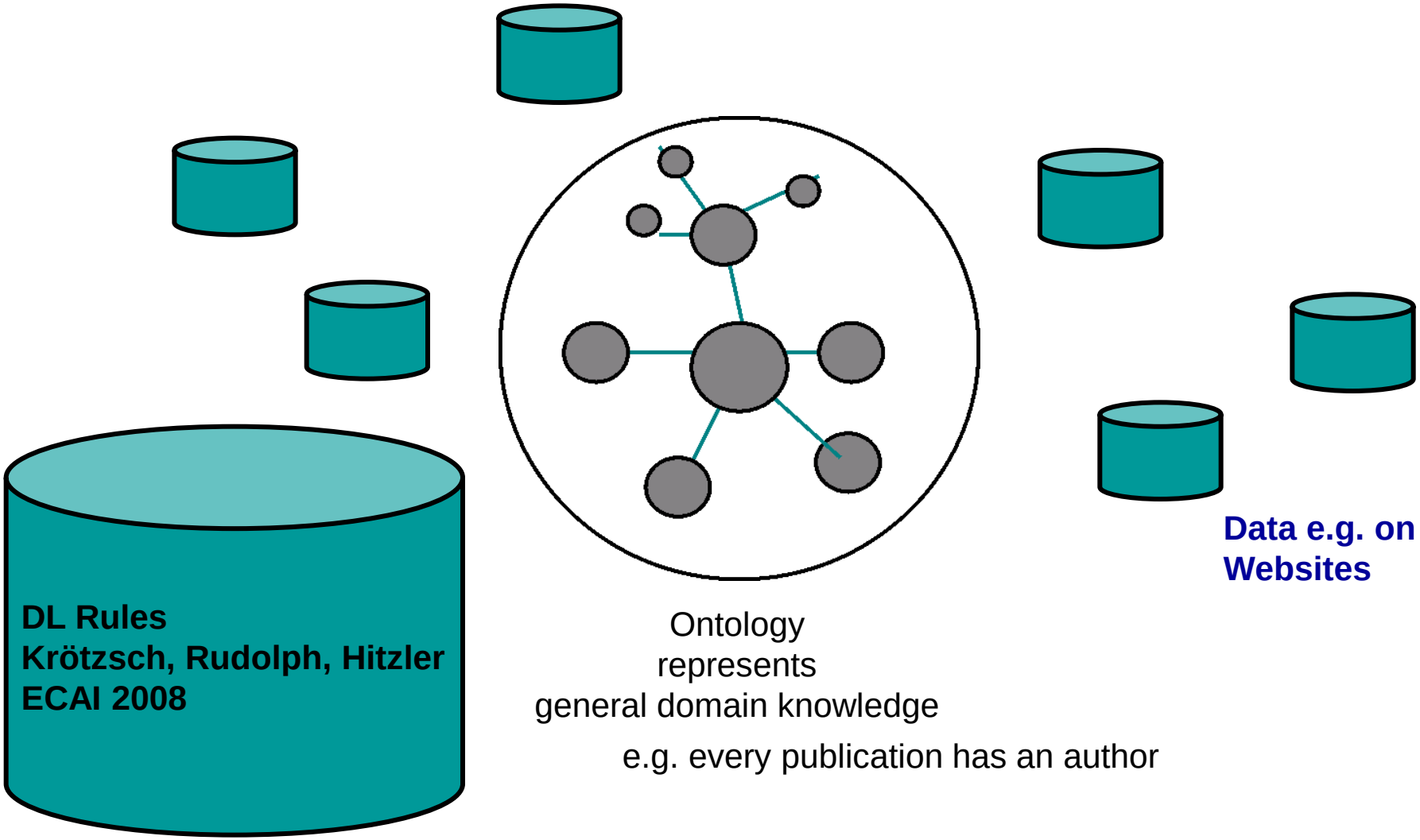
schema knowledge

PhDStudent v 9advisedBy.Professor

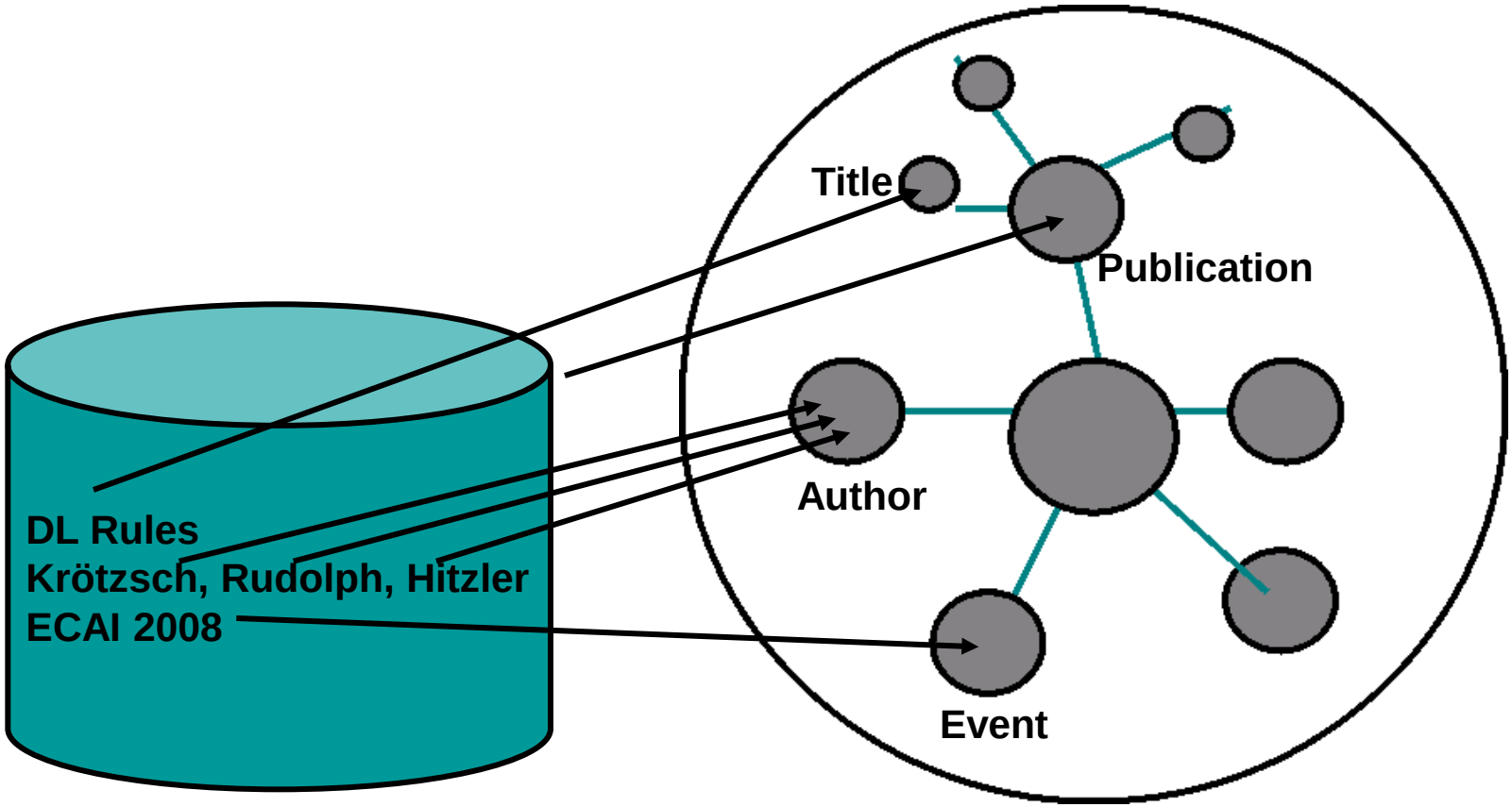
rules

responsible_for(y,x) ÆProfessor(y)
! Employee(x)

Basic Idea of the Semantic Web

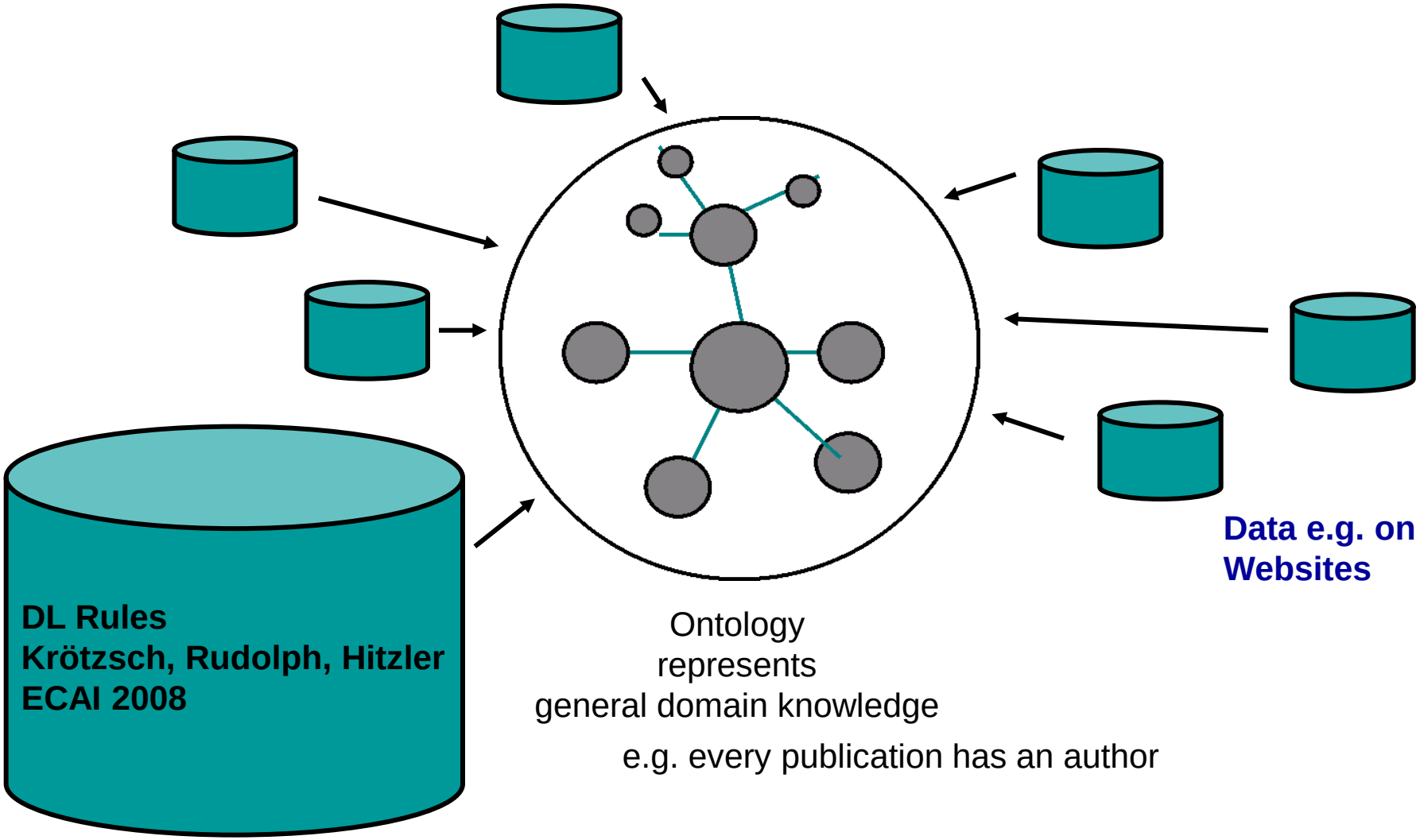


Basic Idea of the Semantic Web



e.g. every publication has an author

Basic Idea of the Semantic Web



IBM

IBM WATSON

The Science Behind an Answer

Watson performs so fast that it can rival the greatest human contestants in understanding a Jeopardy! clue and arriving at a single, precise answer. The significance of this accomplishment can be difficult to comprehend.

Watch the video to see how the computing system designed to play Jeopardy! works.

Possible Answers

bake
balance
ban
bang
bathe
battle
be
beam
bear

The first person mentioned by name
in 'The Man in th
is this hero of a
by the same aut

IBM Watson Solutions and WellPoint, America's largest health benefits company with

The DeepQA hypothesis is that by complementing classic knowledge-based approaches with recent advances in NLP, Information Retrieval, and Machine Learning to interpret and reason over huge volumes of widely accessible naturally encoded knowledge (or "unstructured knowledge") we can build effective and adaptable open-domain QA systems. While they may not be able to formally prove an answer is correct in purely logical terms, they can build confidence based on a combination of reasoning methods that operate directly on a combination of the raw natural language, automatically extracted entities, relations and available structured and semi-structured knowledge available from for example the Semantic Web.

What is Watson?

Implications for analytics, system design and industry transformation ›

Watson for a Smarter Planet™

Join the conversation on IBM Watson Connect

Watson - A System Designed for Answers

Optimize Your Growing

Store | Mac | iPod | iPhone | iPad | iTunes | Support

iPhone

Features | Built-in Apps | From the App Store | iOS | iCloud | Tech Specs | Buy iPhone

Siri. Beta

Your wish is its command.

Siri on iPhone 4S lets you use yo
voice to send messages, schedu
meetings, place phone calls, an
more. Ask Siri to do things just
talking the way you talk. Siri

mediabistro | semanticweb.com | SemTechBiz SF | more >>

semanticweb.com™ The Voice of Semantic Web Technology and Linked Data Business

Home | Events | Community | Learning | Industry Verticals | Answers | Search semanticweb.com | SEARCH

Apple Buys Siri: Once Again The Back Story Is About Semantic Web

By Bernard Lunn on April 28, 2010 2:33 PM

According to Robert Scoble who got it from tracking FTC, Apple is buying Siri. (This has yet to be confirmed by Siri or Apple). The front story is mobile, specifically a bruising battle between Apple and Google. But once again the back story is semantic technology. Siri is not some cute iPhone app banged together in a garage over a Red Bull fueled long weekend. Siri has hard core semantic tech that originated from Darpa (just like that little system called the Internet).

Like the Facebook OpenGraph story, this is another example of semantic web going mainstream. The Open Graph front story was all about social media, but the back story was their adoption of RDFa. That has been a big boost to the semantic web community.

Siri looks like a good exit for investors and will give them confidence to invest more in companies

Semanticweb.com Newsletter
Enter your email for updates | and your ZIP | SIGN UP

Semanticweb.com Event Updates
Enter your email for updates | and your ZIP | SIGN UP

Send an anonymous tip | SEND

SemanticWeb.com on Facebook
Like 1,467

Twitter
Mobile
RSS

SEMANTIC TECH & BUSINESS CONFERENCE
June 3-7, 2012 San Francisco
presented by SemanticWeb.com and WebMediaBrands

WRIGHT STATE UNIVERSITY

Februar

- **Thing**: additionalType, description, image, name, url
  - **CreativeWork**: about, accountablePerson, aggregateRating, alternativeHeadline, associatedMedia, audience, audio, author, award, awards, comment, contentLocation, contentRating, contributor, copyrightHolder, copyrightYear, creator, dateCreated, dateModified, datePublished, discussionUrl, editor, encoding, encodings, genre, headline, inLanguage, interactionCount, isFamilyFriendly, keywords, mentions, offers, provider, publisher, publishingPrinciples, review, reviews, sourceOrganization, text, thumbnailUrl, version, video
    - **Article**: articleBody, articleSection, wordCount
      - **BlogPosting**
      - **NewsArticle**: dateline, printColumn, printEdition, printPage, printSection
      - **ScholarlyArticle**
        - **MedicalScholarlyArticle**: citation, publicationType
    - **Blog**: blogPost, blogPosts
    - **Book**: bookEdition, bookFormat, illustrator, isbn, numberOfPages
    - **Comment**
    - **Diet**: dietFeatures, endorsers, expertConsiderations, overview, physiologicalBenefits, proprietaryName, risks
    - **ExercisePlan**: activityDuration, activityFrequency, additionalVariable, exerciseType, intensity, repetitions, restPeriods, workload
    - **ItemList**: itemListElement, itemListOrder
    - **Map**
    - **MediaObject**: associatedArticle, bitrate, contentSize, contentUrl, duration, embedUrl, encodesCreativeWork, encodingFormat, expires, height, interactionCount, offers, playerType, regionsAllowed, requiresSubscription, uploadDate, width
      - **AudioObject**: transcript
      - **ImageObject**: caption, exifData, representativeOfPage, thumbnail
      - **MusicVideoObject**
      - **VideoObject**: caption, productionCompany, thumbnail, transcript, videoFrameSize, videoQuality
    - **Movie**: actor, actors, director, duration, musicBy, producer, productionCompany, trailer
    - **MusicPlaylist**: numTracks, track, tracks
      - **MusicAlbum**: byArtist
    - **MusicRecording**: byArtist, duration, inAlbum, inPlaylist
    - **Painting**
    - **Photograph**
    - **Recipe**: cookingMethod, cookTime, ingredients, nutrition,
    - **Review**: itemReviewed, reviewBody, reviewRating
    - **Sculpture**
    - **SoftwareApplication**: applicationCategory, applicationSub
      downloadUrl, featureList, fileFormat, fileSize, installUr
      releaseNotes, requirements, screenshot, softwareVersion, storageRequirements
      - **MobileApplication**: carrierRequirements
      - **WebApplication**: browserRequirements
    - **TVEpisode**: actor, actors, director, episodeNumber, musicBy, partOfSeason, partOfTVSeries, producer, productionCompany, trailer
    - **TVSeason**: endDate, episode, episodes, numberOfEpisodes, partOfTVSeries, seasonNumber, startDate, trailer
    - **TVSeries**: actor, actors, director, endDate, episode, episodes, musicBy, numberOfEpisodes, producer, productionCompany, season, seasons, startDate, trailer

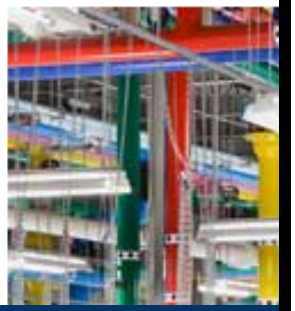
schema.org for enhancing web search
joint effort including Bing, Google, Yahoo, Yandex

Google and the future of search: Amit Singhal and the Knowledge Graph

Google has revolutionised the way we holiday, shop, work and play. Now, with Knowledge Graph, it plans to radically transform the way we search the internet... again. But some voice qualms about the company's ambitions

Tim Adams
The Obs

Jump



senses, that attribute is in the process of changing. This year, Google will roll out what it calls its Knowledge Graph, the closest any system has yet come to creating what Tim Berners-Lee, originator of the web itself, called "the semantic web", the version that had understanding as well as data, that could itself provide answers, not links to answers.

The Knowledge Graph is a database of the 500 million most searched for people, places and

Wright State University

Sign i

Web Images Maps Shopping More Search tools

About 32,400,000 results (0.23 seconds)

Wright State University

www.wright.edu/

Expanded scholarship program offered. **Wright State University** has unveiled a new scholarship program designed to make a college education more affordable.

Score: **21** / 30 - 11 Google reviews - Write a review

3640 Colonel Glenn Hwy Dayton, OH 45435
(937) 775-3333

Wings
Get a WINGS username and password. Alternative Login for ...

Graduate School
Programs - Admissions - Apply - Check Application Status - ...

Undergraduate Admissions
... your new home away from home. Visit Undergraduate ...

The Official Wright State ...
Playing in the new and updated facilities, the Raiders have won ...

Academics
We think Wright State is pretty amazing—and national ...

Libraries
Research Databases - Hours - Contact Us - Staff - ...

Search wright.edu

Wright State University - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Wright_State_University

Wright State University is a public research university in Fairborn, Ohio just outside of Dayton. The school offers degrees at the associate, bachelor's, master's, ...

Wright State University

Directions

Wright State University is a public research university in Fairborn, Ohio just outside of Dayton. The school offers degrees at the associate, bachelor's, master's, and doctoral level. Wikipedia

Nickname: Raiders

Address: 3640 Colonel Glenn Hwy, Dayton, OH 45435

Phone: (937) 775-3333

Mascot: Rowdy Raider

Founded: 1967

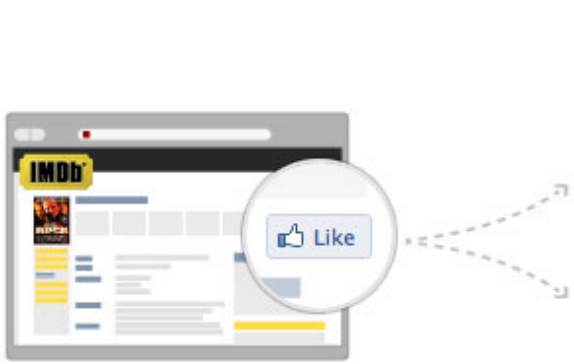
Colors: Gold, Green

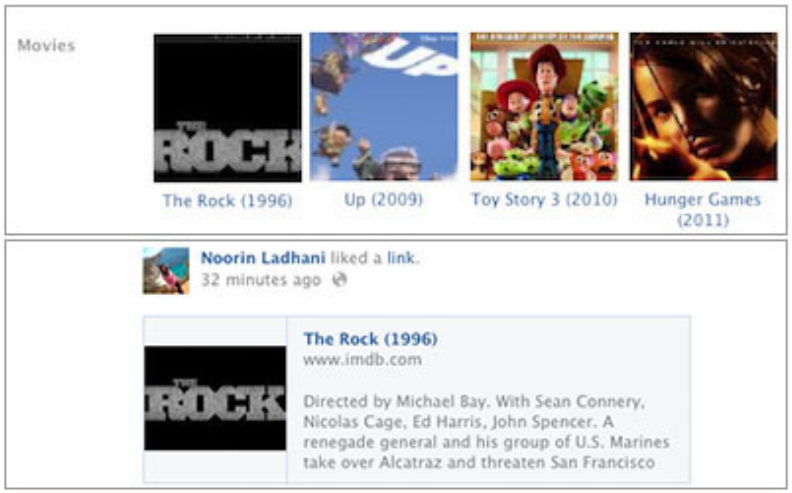
Open Graph Protocol

We announced the next version of the Open Graph at f8 2011. It is currently released to the public. You can view the Open Graph documentation here. The below documentation refers to the first version of the Open Graph that shipped with the Like Button at f8 2010.

We have also updated how the Like Button will function with respect to the next version of Open Graph. Please read the following developer doc about the Like Button Migration.

The Open Graph Protocol enables you to integrate your Web pages into the social graph. It is currently designed for Web pages representing profiles of real-world things — things like movies, sports teams, celebrities, and restaurants. Including Open Graph tags on your Web page, makes your page equivalent to a Facebook Page. This means when a user clicks a Like button on your page, a connection is made between your page and the user. Your page will appear in the "Likes and Interests" section of the user's profile, and you have the ability to publish updates to the user. Your page will show up in the same places that Facebook pages show up around the site (e.g. search), and you can target ads to people who like your content. The structured data you provide via the Open Graph Protocol defines how your page will be represented on Facebook.





Contents

- **Big Data, Linked Data, Semantic Web**
- **An Example: Linked Data Querying**
- **The Big Data Added Value Pipeline**

Example problem

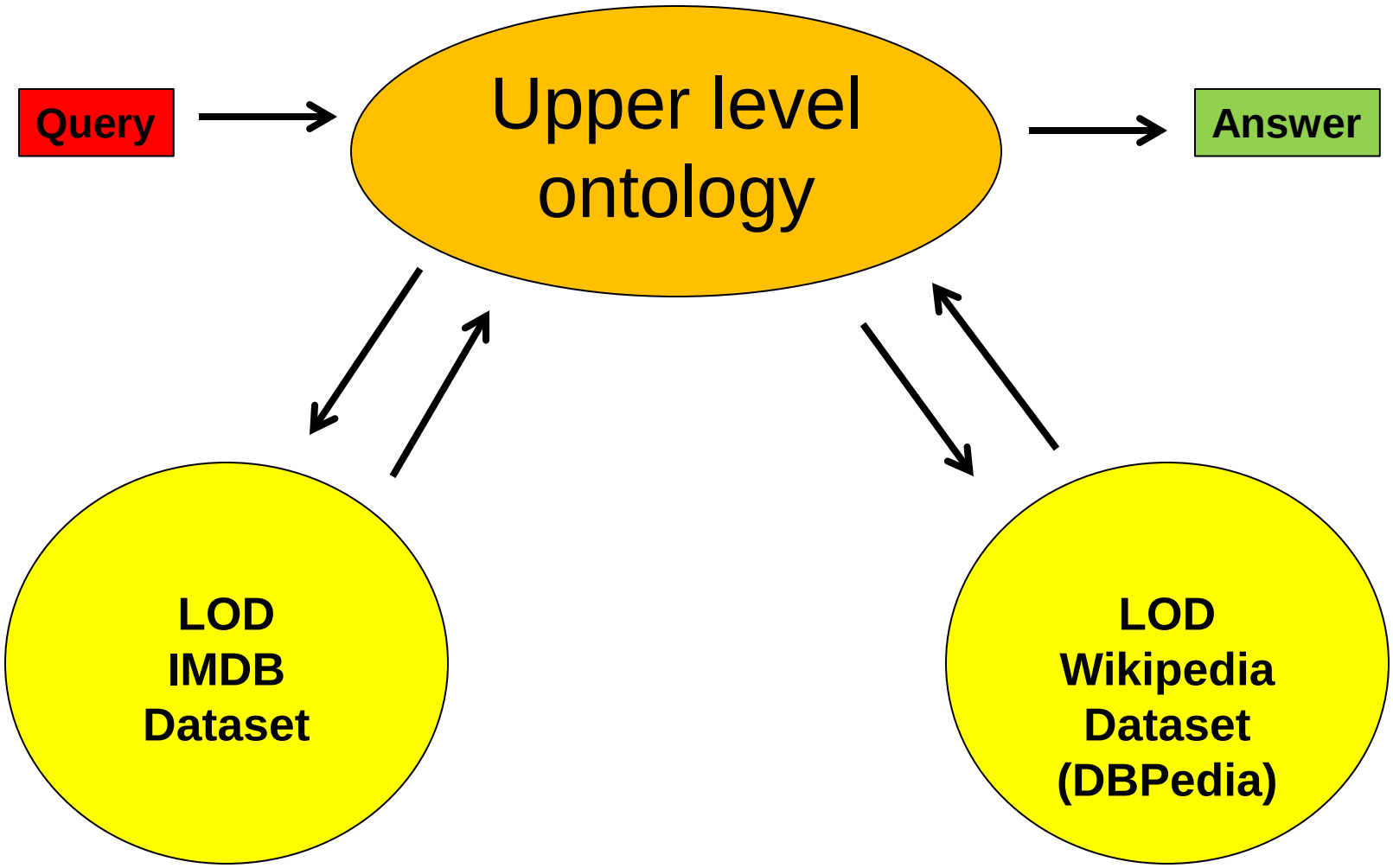
"Identify films, the nations where they were shot and the population of these countries"

Issues:

- **Where is the data? (what is in which linked dataset)**
- **How to query each dataset? (internal structure)**

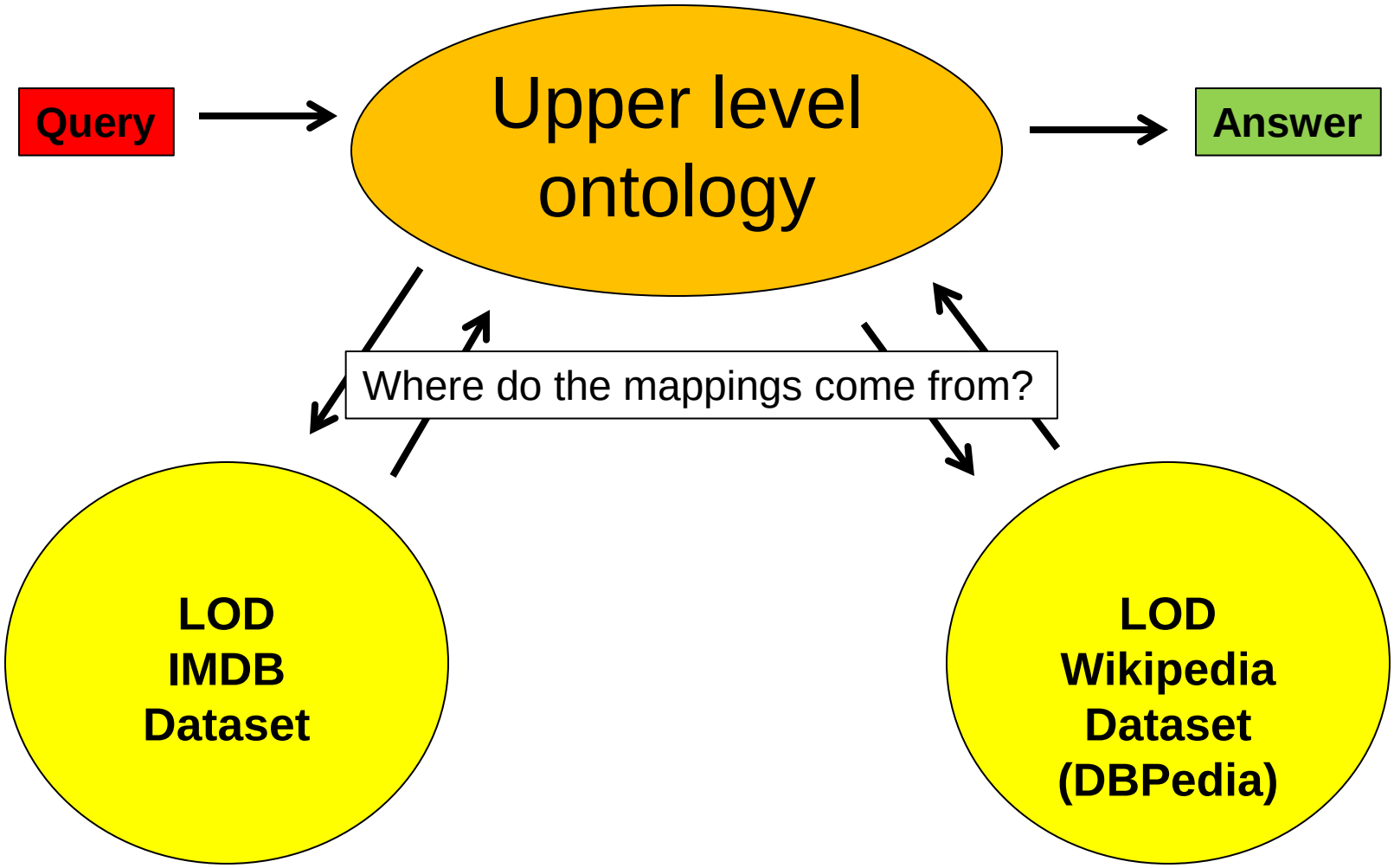
Here: Need to merge knowledge from IMDB and DBPedia datasets

Linked Data federated querying



Joshi, Jain, Hitzler et al. ODBASE 2012

Linked Data federated querying



Joshi, Jain, Hitzler et al. ODBASE 2012

Bootstrapping-based alignment

Table 4. Results of various systems for LOD Schema Alignment. Legends: Prec=Precision, Rec=Recall, M=Music Ontology, B=BBC Program Ontology, F=FOAF Ontology, D=DBpedia Ontology, G=Geonames Ontology, S=SIOC Ontology, W=Semantic Web Conference Ontology, A=AKT Portal Ontology, err=System Error, NA=Not Available

	Linked Open Data Schema Ontology Alignment									
	Alignment API		OMViaUO		RiMoM		S-Match		AROMA	
Test	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
M,B	0.4	0	1	0	err	err	0.04	0.28	0	0
M,D	0	0	0	0	err	err	0.08	0.30	0.45	0.01
F,D	0	0	0	0	err	err	0.11	0.40	0.33	0.04
G,D	0	0	0	0	err	err	0.23	1	0	0
S,F	0	0	0	0	0.3	0.2	0.52	0.11	0.30	0.20
W,A	0.12	0.05	0.16	0.03	err	err	0.06	0.4	0.38	0.03
W,D	0	0	0	0	err	err	0.15	0.50	0.27	0.01
Avg.	0.07	0.01	0.17	0	NA	NA	0.17	0.43	0.25	0.04

Jain, Hitzler et al, ISWC2010

BLOOMS

Table 1. Results on the oriented matching track. Results for RiMOM and AROMA have been taken from the OAEI 2009 website. Legends: Prec=Precision, A-API=Alignment API, OMV=OMViaUO, NaN=division by zero, likely due to empty alignment.

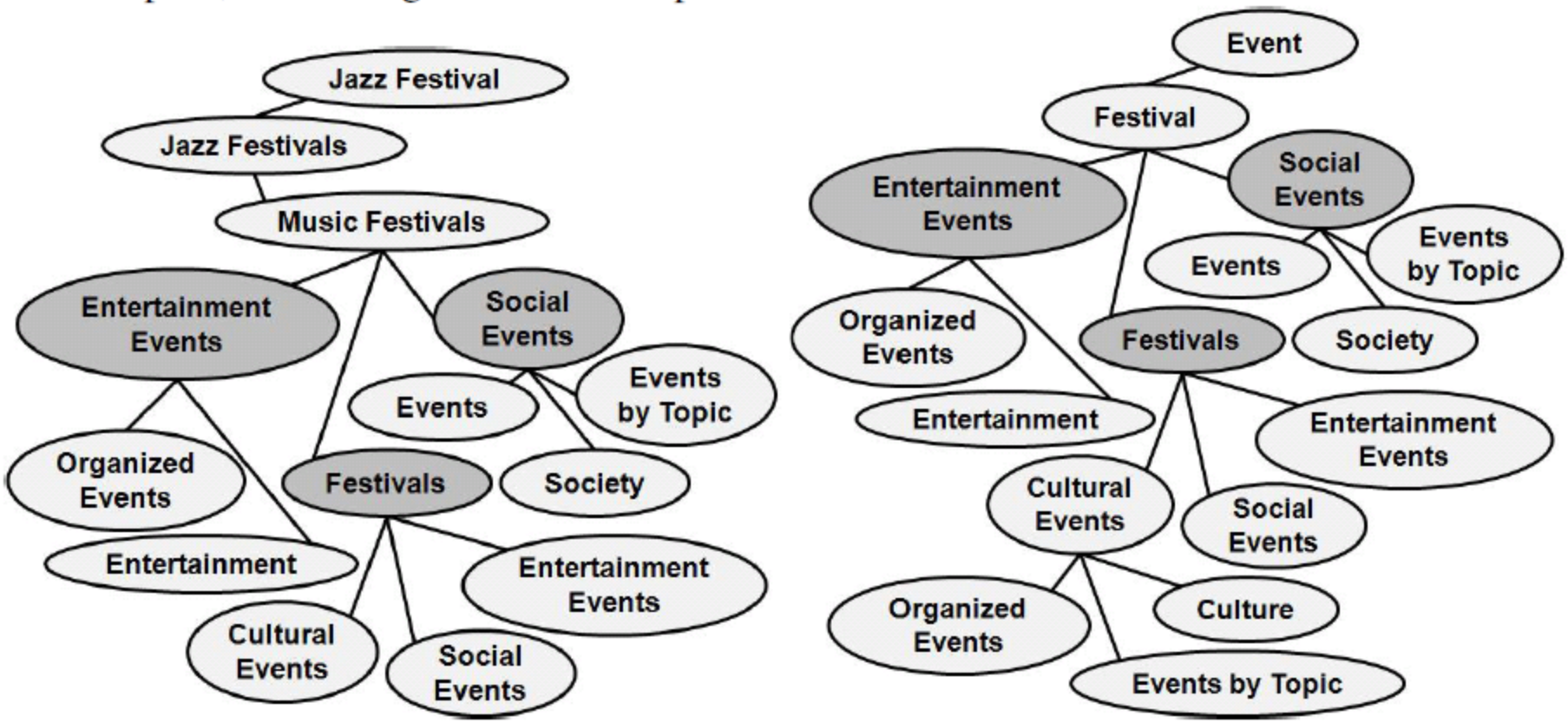
Ontology Alignment Initiative—Oriented Matching Track												
	A-API		OMV		S-Match		AROMA		RiMoM		BLOOMS	
Test	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec	Prec	Rec
1XX	0	0	0.02	0.06	0.01	0.71	NaN	0	1	1	1	1
2XX	0	0	0.01	0.03	0.05	0.30	0.84	0.08	0.67	0.85	0.52	0.51
3XX	0.01	0.03	0.02	0.047	0.01	0.14	0.72	0.11	0.59	0.81	1	0.84
Avg.	0.00	0.01	0.02	0.04	0.03	0.38	0.63	0.07	0.75	0.88	0.84	0.78

BLOOMS

1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.
2. **Construction of the BLOOMS forest** T_C for each class name C, using information from Wikipedia.
3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.
4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

BLOOMS trees

Fig. 1. BLOOMS trees for Jazz Festival with sense Jazz Festival and for Event with sense Event. To save space, some categories are not expanded to level 4.



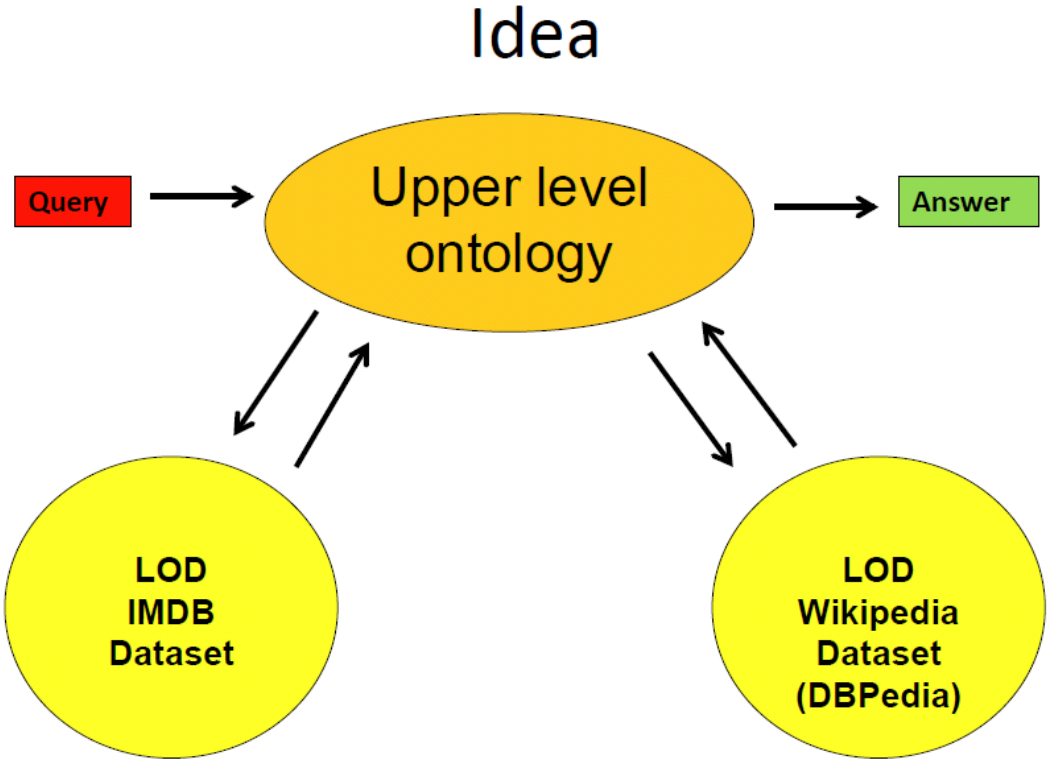
BLOOMS

1. **Pre-processing of the input ontologies** in order to (i) remove property restrictions, individuals, and properties, and to (ii) tokenize composite class names to obtain a list of all simple words contained within them, with stop words removed.
2. **Construction of the BLOOMS forest** T_C for each class name C, using information from Wikipedia.
3. **Comparison of constructed BLOOMS forests**, which yields decisions which class names are to be aligned.
4. **Post-processing** of the results with the help of the Alignment API and a reasoner.

Big Data Bootstrapping for Big Data

We

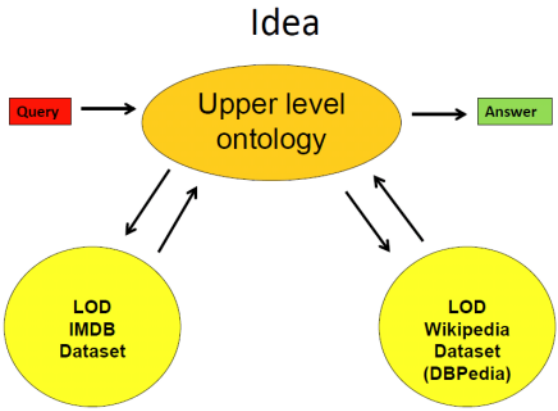
- **use big data**
- **for aligning big data**
- **in order to query big data**



Querying Illustration

"Identify films, the nations where they were shot and the population of these countries"

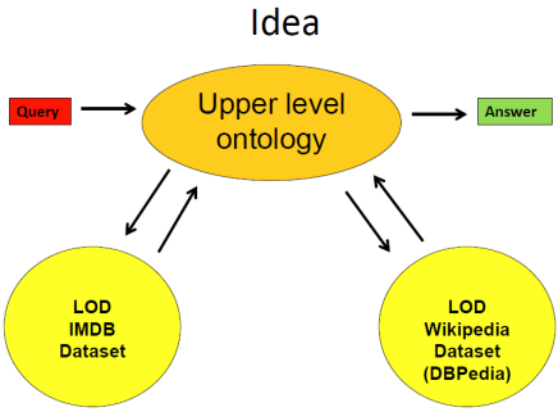
```
SELECT ?film ?nation ?pop

WHERE {
?film     protonu:ofCountry          ?nation.
?film     rdf:type                   protonu:Movie.
?film     rdfs:label                 ?film_name.
?nation protont:populationCount      ?pop.
}
```



Querying Illustration – Blooms Alignment

protonu:ofCountry	maps to	lmdb:country
protonu:Movie	maps to	lmdb:film
protont:populationCount	maps to	dbprop:populationCount

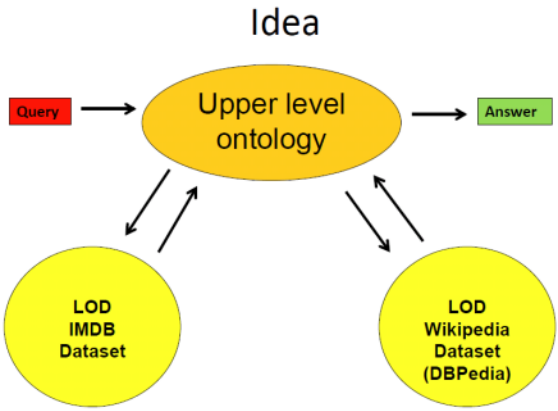
Alignment confidence > 0.9



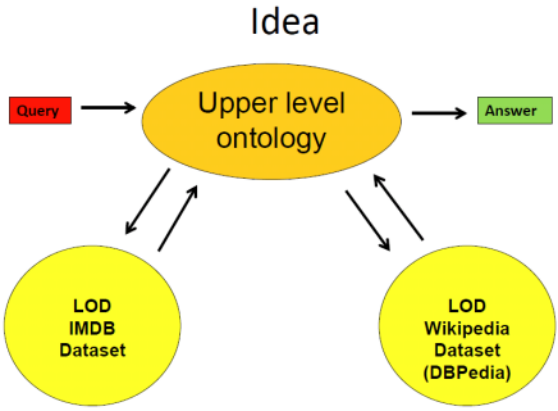
Querying Illustration – Sub Queries

(a) SELECT ?film ?nation ?pop
WHERE {
?film lmdb:country ?nation.
?film rdf:type lmdb:film.
?film rdfs:label ?film_name.
}

(b) SELECT ?nation ?pop
WHERE {
?nation dbprop:populationCensus ?pop.



Querying Illustration – Results



(a)

lmdb-film:11446	protonu:ofCountry	lmdb-country:IN.
lmdb-film:11446	rdf:type	protonu:Movie.
lmdb-film:11446	rdfs:label	"Run".
lmdb-film:17091	protonu:ofCountry	lmdb-country:LK.
lmdb-film:17091	rdf:type	protonu:Movie.
lmdb-film:17091	rdfs:label	"Getawarayo".
lmdb-film:16973	protonu:ofCountry	lmdb-country:IN.
lmdb-film:16973	rdf:type	protonu:Movie.
lmdb-film:16973	rdfs:label	"Kabeela".

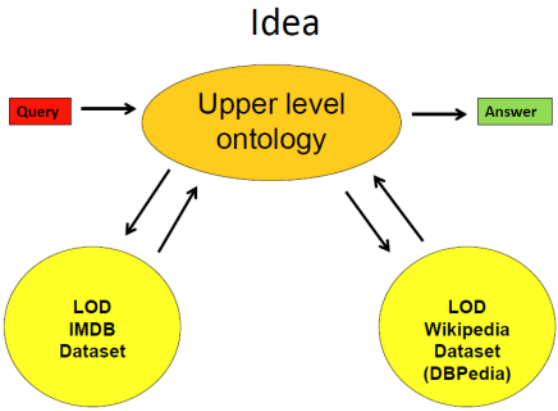
(b)

dbpedia:Sri_Lanka	protont:PopulationCount	21324791.
dbpedia:India	protont:PopulationCount	1210193422.

Querying Illustration – Result I

- **With proxy identifiers**

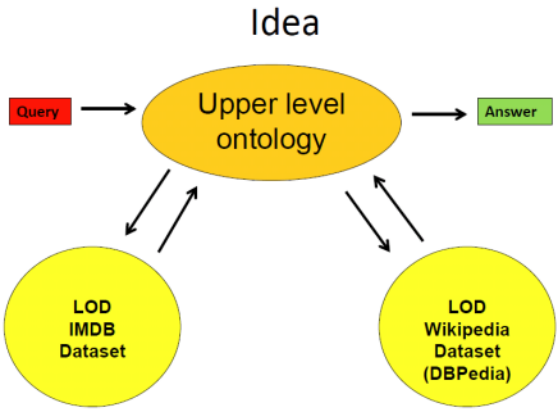
film	name	nation	population
lmdb-film:17091	"Getawarayo"	aloqus:2908ba82	21324791
lmdb-film:16973	"Kabeela"	aloqus:9bc35ca1	1210193422
lmdb-film:11446	"Run"	aloqus:9bc35ca1	1210193422



Querying Illustration – Result II

- **Without proxy identifiers**

film	name	nation	population
lmdb-film:17091	"Getawarayo"	lmdb-country:LK	21324791
lmdb-film:16973	"Kabeela"	lmdb-country:IN	1210193422
lmdb-film:11446	"Run"	lmdb-country:IN	1210193422
lmdb-film:11446	"Run"	nytimes:india_geo	1210193422

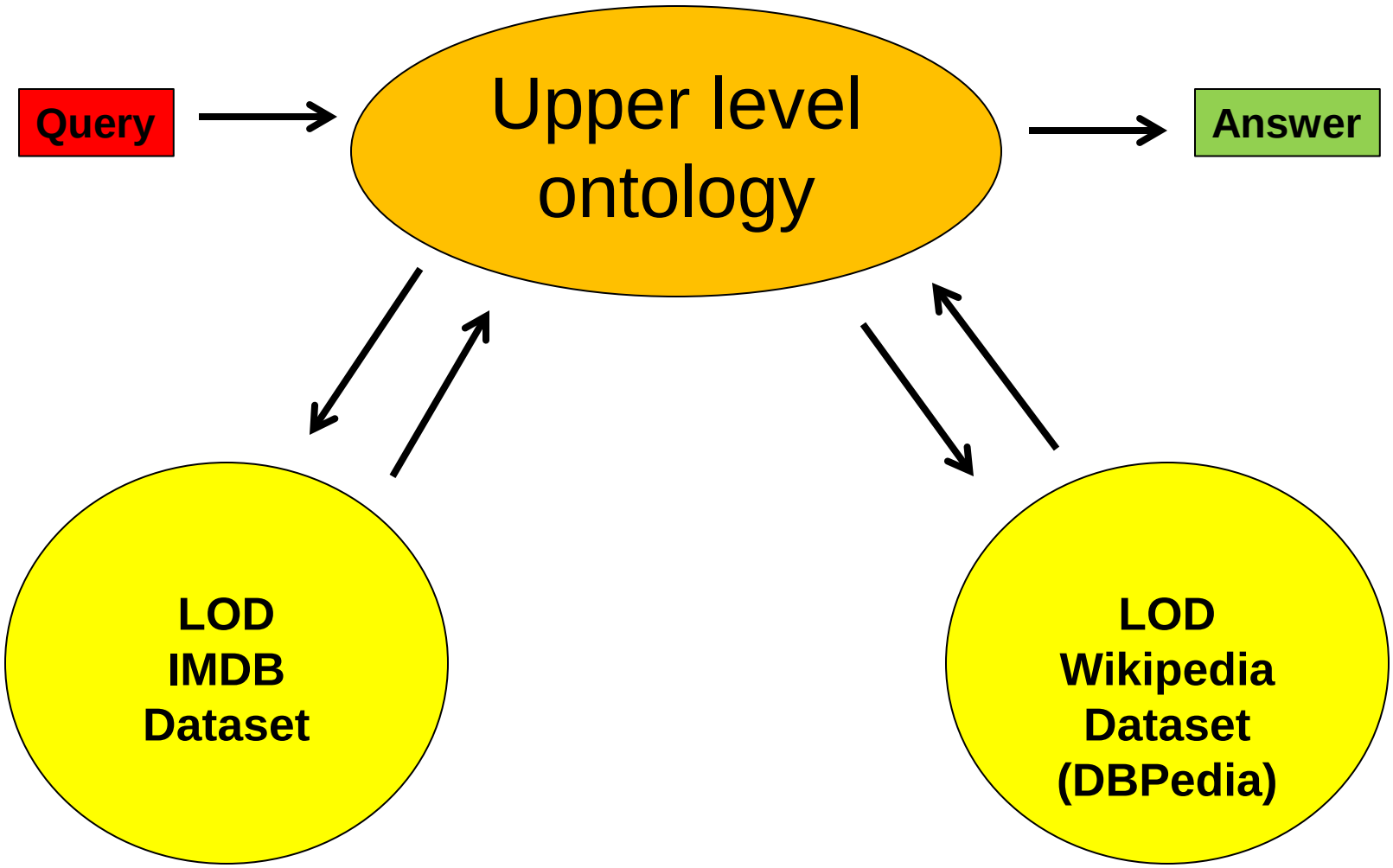


no.	Query	Datasets	Primary Ontology	Other LOD Ontologies
Q1	Identify movies, countries where they were shot and the latest population for these countries.	LinkedMDB, DBpedia	PROTON	N/A
Q2	List the semantic web people and their affiliation.	Semantic Web Dog Food	N/A	SWRC
Q3	Find all Jamendo artists along with their image, home page, and the population of city they are near.	Jamendo, Geonames	N/A	Music Ontology, Geonames
Q4	Software companies founded in the US	DBpedia	PROTON	DBPedia
Q5	Find list of movies, director and actors and the population of their birth cities.	DBpedia, LinkedMdb, Factbook	PROTON	LinkedMdb
Q6	List the countries, birth rates and sex ratios.	DBPedia, Factbook	PROTON	Factbook
Q7	Is Mayotte a country?	DBPedia	PROTON	N/A
Q8	Get the birthdates of folks who acted in Star Trek	DBPedia, LinkedMdb	PROTON	N/A
Q9	List Music artists and birth dates.	DBPedia, BBC Music, Jamendo	DBpedia	N/A
Q10	Find list of movies made in countries with population greater that 1 Billion.	DBpedia,LinkedMdb	DBPedia	N/A

Comparison with other systems

Features	ALOQUS	DARQ	SQUIN
Approach	Uses upper level ontology(PROTON) or any other ontology as primary ontology for query serialization and execution.	Requires formal description of datasets in the form of Service Description.	Requires an initial URI to execute queries.
Query Creation	Creates query corresponding to every mapping for a concept.	Creates queries only corresponding to the concepts mentioned in the query.	Creates queries only corresponding to the concepts mentioned in the query.
Failsafe	Executes all sub-queries for multiple mappings. Hence retrieves at least partial answers if a specific endpoint doesn't work.	X	X
Detect Entity co-references	Crawls and also consumes sameAs.org webservices.	X	X
Result Processing	Query answers, retrieved from different datasets are merged and presented to user.	Retrieves answers from multiple dataset based on service description.	Retrieves answers from multiple dataset through link traversal.
Write queries using ontology not present in LOD	Yes	X	X
Support for open-ended queries like ?s ?p ?o	Yes	X	X
Result Storage for later Retrieval	Yes	X	X
DESCRIBE Query Form	Yes	N/A	Yes

Linked Data federated querying

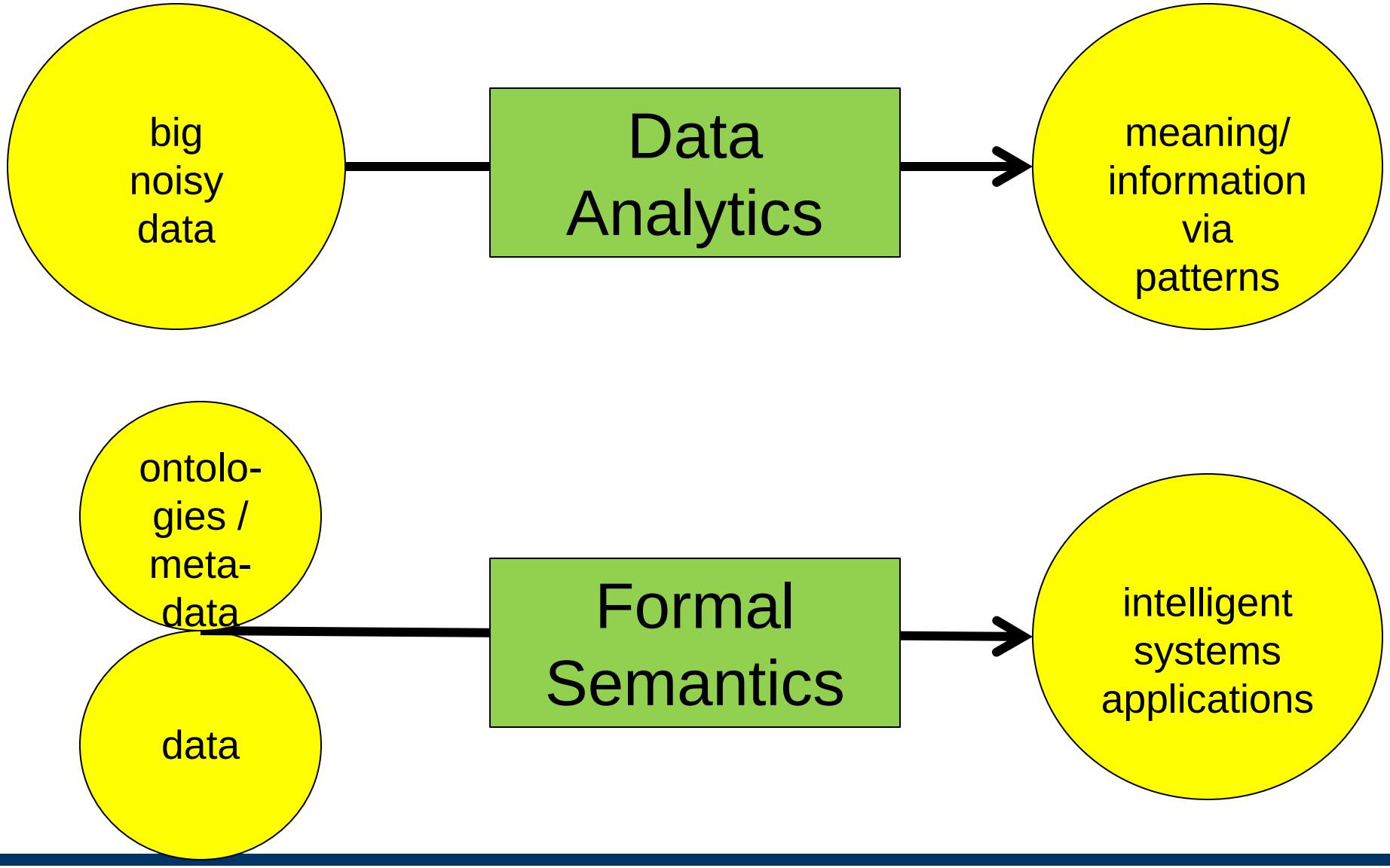


Joshi, Jain, Hitzler et al. ODBASE 2012

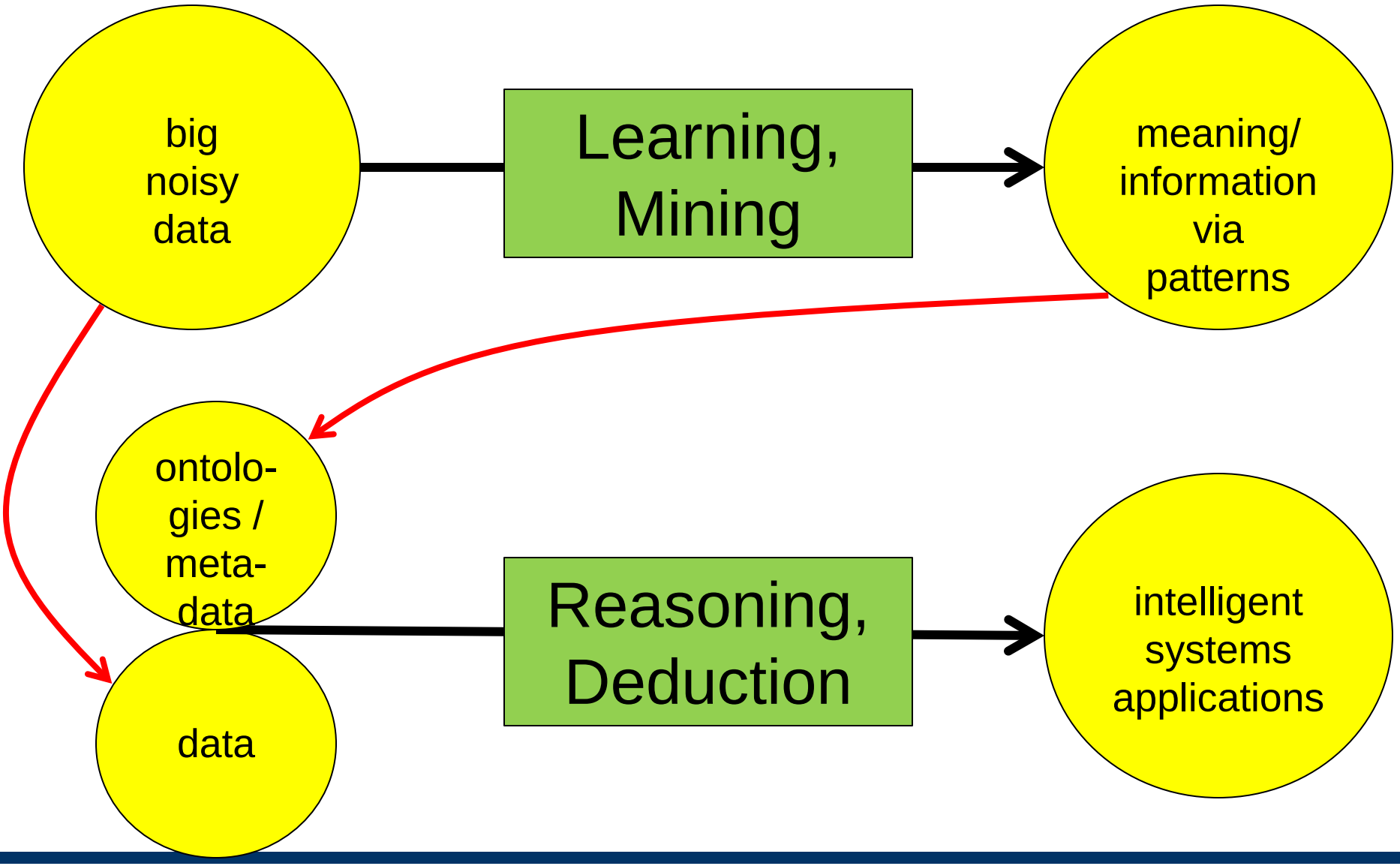
Contents

- **Big Data, Linked Data, Semantic Web**
- **An Example: Linked Data Querying**
- **The Big Data Added Value Pipeline**

The Big Data Added Value Pipeline



The Big Data Added Value Pipeline



Other stuff I'm doing

- **Pushing the limits of ontology modeling languages (OWL, RDF).**
- **Ontology reasoning algorithms.**
- **Ontology modeling.**
- **EarthCube.**

Thanks!

Happy to talk about collaboration opportunities.

References

- Pascal Hitzler, Knowledge Representation in the Big Data Age. In: NSF Workshop: Research Challenges and Opportunities in Knowledge Representation, Arlington, VA, February 7-8, 2013.
- Krzysztof Janowicz, Pascal Hitzler, *The Digital Earth as Knowledge Engine*. Semantic Web 3 (3), 213-221, 2012.
- Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, *Linked Data is Merely More Data*. In: Dan Brickley et al.: Linked Data Meets Artificial Intelligence. Technical Report SS-10-07, AAAI Press, Menlo Park, California, 2010, pp. 82-86. ISBN 978-1-57735-461-1. Proceedings of LinkedAI at the AAAI Spring Symposium, March 2010.
- Pascal Hitzler, Frank van Harmelen, *A reasonable Semantic Web*. Semantic Web 1(1-2), 39-44, 2010.
- Pascal Hitzler, Krzysztof Janowicz, *What's Wrong with Linked Data?* http://blog.semantic-web.at/2012/08/09/whats-wrong-with-linked-data/ , August 2012.

References

- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2009.
- Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, Sebastian Rudolph, OWL 2 Web Ontology Language: Primer (Second Edition). W3C Recommendation, 11 December 2012.
- Prateek Jain, Pascal Hitzler, Amit P. Sheth, Kunal Verma, Peter Z. Yeh, Ontology Alignment for Linked Open Data. In P. Patel-Schneider et al. (eds.), The Semantic Web - ISWC 2010. 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part I. Lecture Notes in Computer Science Vol. 6496. Springer, Berlin, 2010, pp. 402-417.
- Prateek Jain, Pascal Hitzler, Kunal Verma, Peter Yeh, Amit Sheth, Moving beyond sameAs with PLATO: Partonomy detection for Linked Data. In: Ethan V. Munson, Markus Strohmaier (eds.): 23rd ACM Conference on Hypertext and Social Media, HT '12, Milwaukee, WI, USA, June 25-28, 2012. ACM, 2012, pp. 33-42.

References

- Amit Krishna Joshi, Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, Amit P. Sheth, Mariana Damova, Alignment-based Querying of Linked Open Data. In: Meersman, R. et al. (eds.), On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012, Proceedings, Part II. Lecture Notes in Computer Science Vol. 7566, Springer, Heidelberg, 2012, pp. 807-824.
- Prateek Jain, Peter Z. Yeh, Kunal Verma, Reymonrod G. Vasquez, Mariana Damova, Pascal Hitzler, Amit P. Sheth, Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton. In: Grigoris Antoniou et al. (Eds.): The Semantic Web: Research and Applications – 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I. Lecture Notes in Computer Science 6643, Springer, 2011, pp. 80-92.