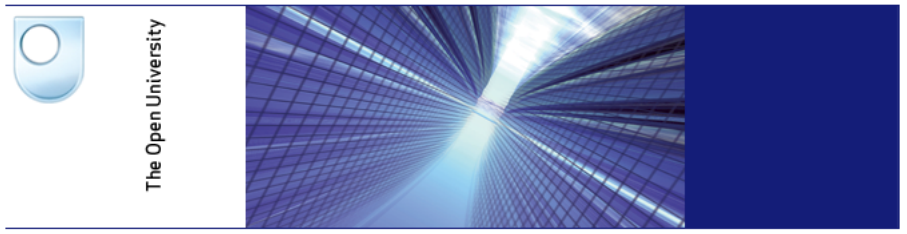


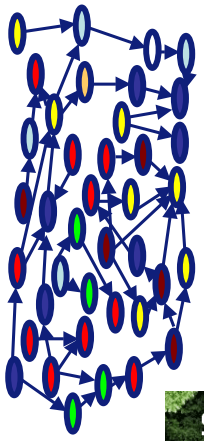
Understanding Ontologies

Enrico Motta
Professor of Knowledge Technologies
Knowledge Media Institute
The Open University

Preamble

Towards A New Generation of SW Applications

The Semantic Web



- 10-20 million semantic web documents
  - Expressed in RDF, OWL, DAML+OIL
- 7K-10K ontologies
  - These cover a variety of domains - multimedia, computing, management, bio-medical sciences, geography, entertainment, upper level concepts, etc…

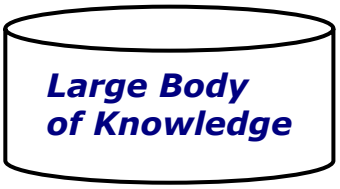
Ordnance Survey Ontologies

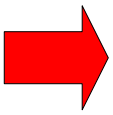
The above figures refer to resources which are publicly accessible on the web

The Semantic Web today

- The Semantic Web is already in place and is characterized by a widespread production of formalized knowledge models (ontologies and metadata), from a variety of different groups and individuals
  - ⇨ The SW provides an infrastructure for large scale, distributed knowledge publishing
    - "The Next Knowledge Medium - An information network with semi-automated services for the generation, distribution, and consumption of knowledge"
      - Stefik, 1986
    - "Knowledge modelling to become a new form of literacy?"
      - Stutt and Motta, 1997
- It can also be seen as a large scale, heterogeneous knowledge base, which can enable a new generation of intelligent applications
  - ⇨ Such availability of large-scale formalised knowledge is unprecedented in the history of AI
    - And indeed much of AI research for the past 30 years has tackled the so-called knowledge acquisition bottleneck: how to acquire and represent effectively large amounts of knowledge to enable intelligent behaviour

The Knowledge Representation Hypothesis

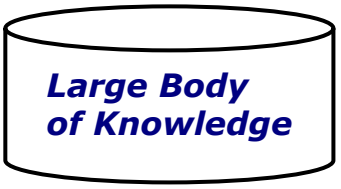


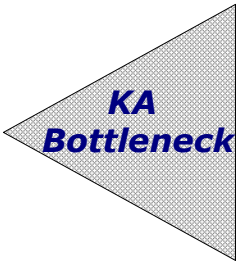


Intelligent Behaviour

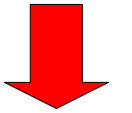
> ***"Any mechanically embodied intelligent process will be comprised of structural ingredients that we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and independent of such external semantic attribution, play a formal but causal and essential role in engendering the behaviour that manifests that knowledge"***
>
> ***Brian Smith, 1982***

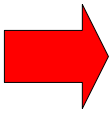
The Knowledge Acquisition Bottleneck





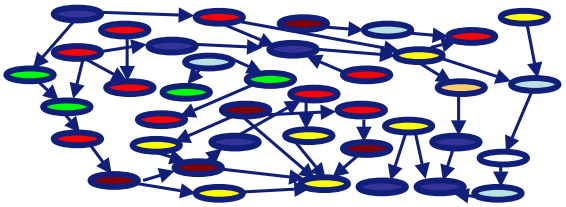
Knowledge

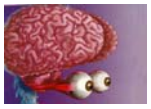




Intelligent Behaviour

SW as Enabler of Intelligent Behaviour





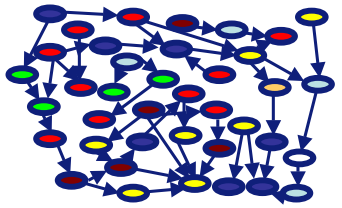
Intelligent Behaviour

Next Generation SW Apps

Two years ago we launched a research programme entitled "Next Generation Semantic Web Applications"

The goal was to investigate the development of applications which could exploit the SW as a large scale, heterogeneous source of knowledge

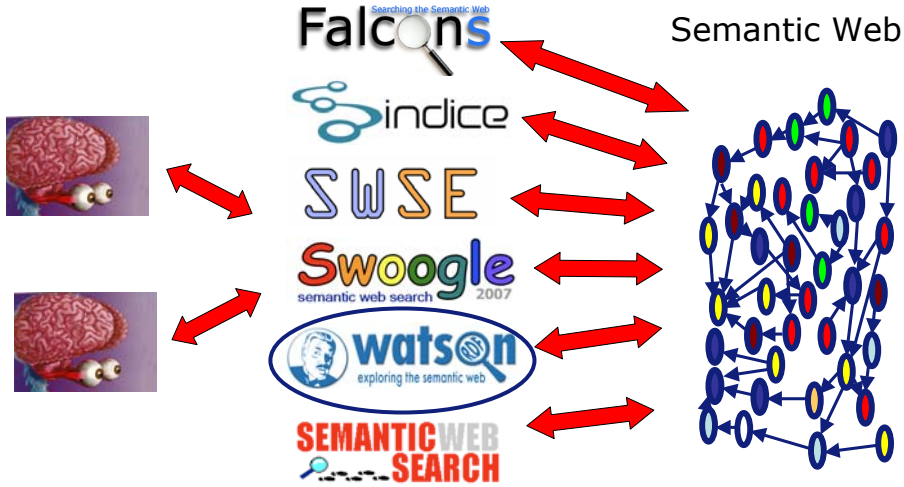
NG SW Application



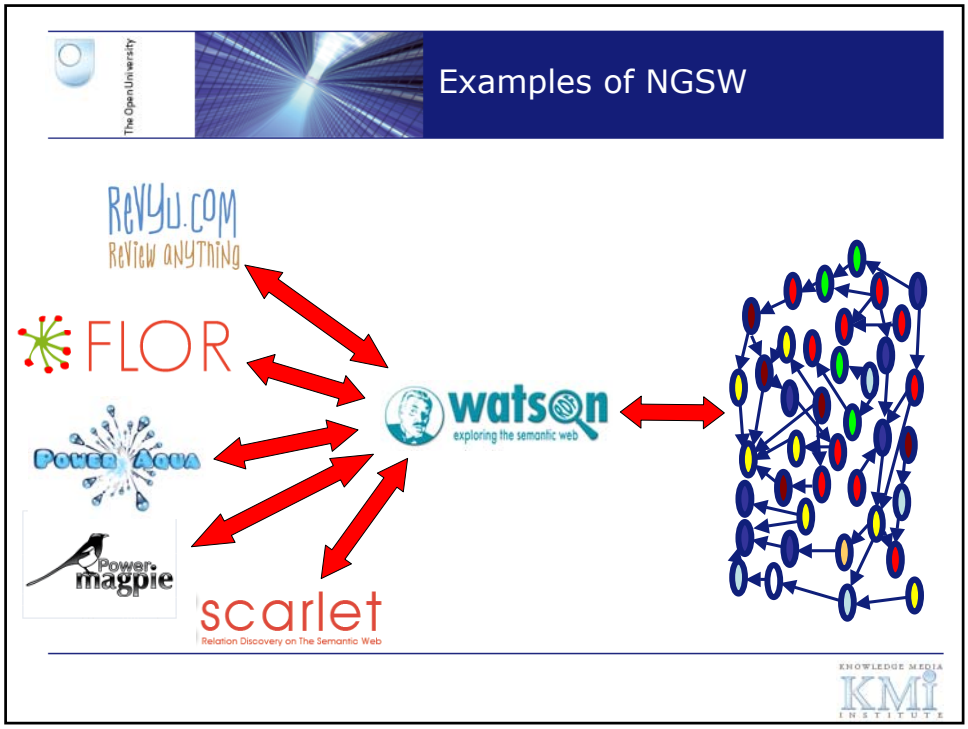
KBS vs NGSW Systems

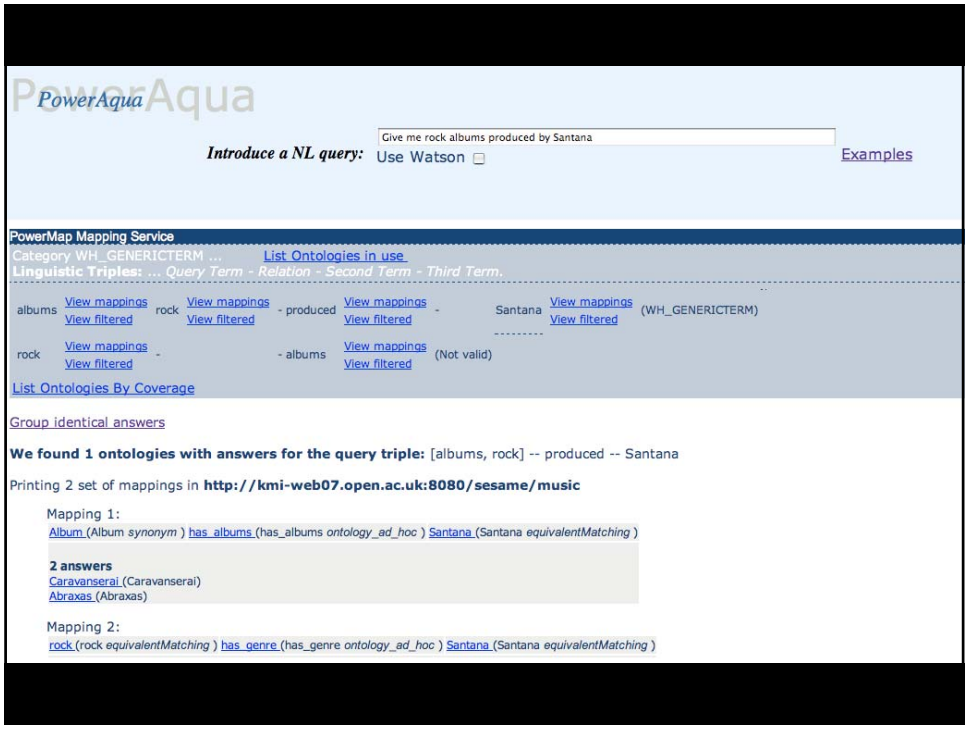
	Classic KBS	SW Systems
Ontology/Data Selection	Static/Design-time	Dynamic/Run-time
Provenance	Centralized	Distributed
Size	Small/Medium	Extra Huge
Repr. Schema	Homogeneous	Heterogeneous
Quality of data	High	Very Variable
Degree of trust	High	Very Variable

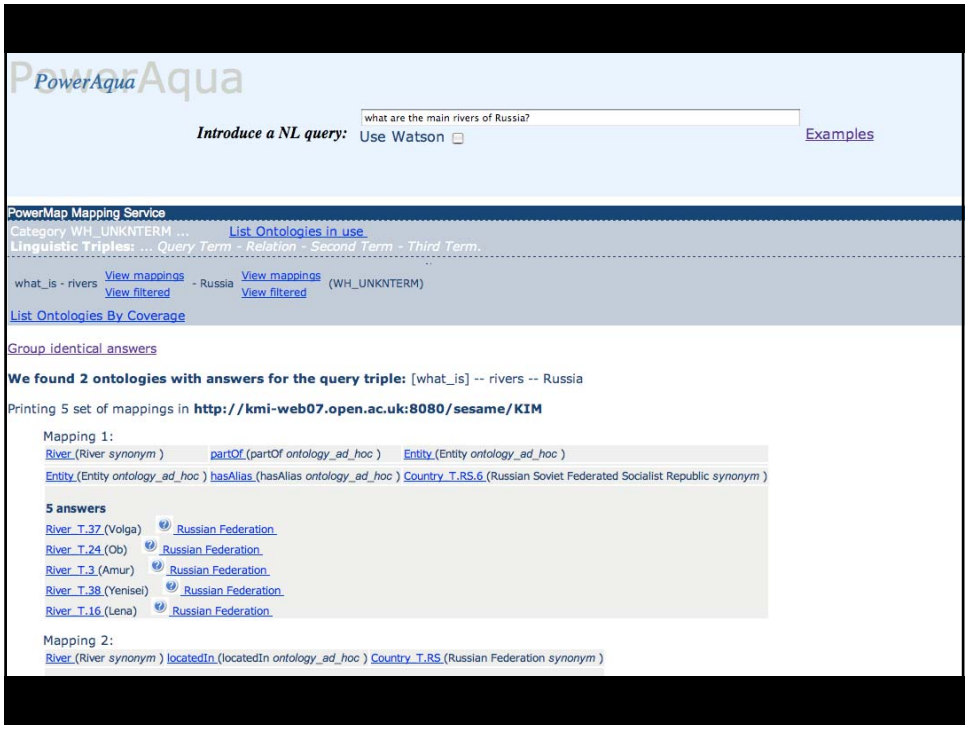
SW Gateways

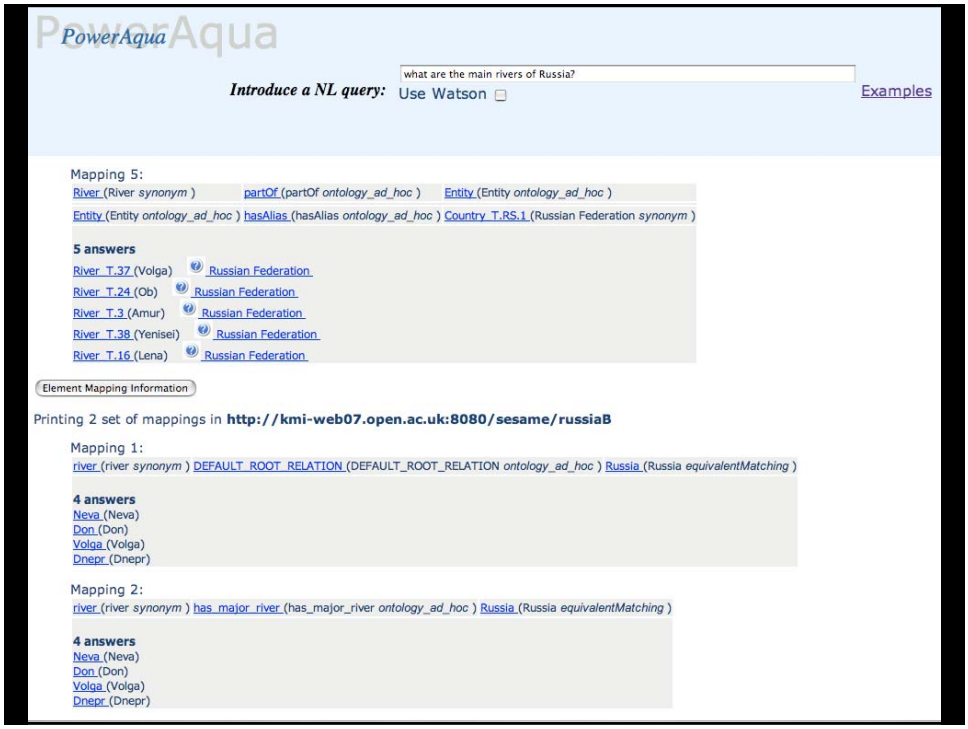


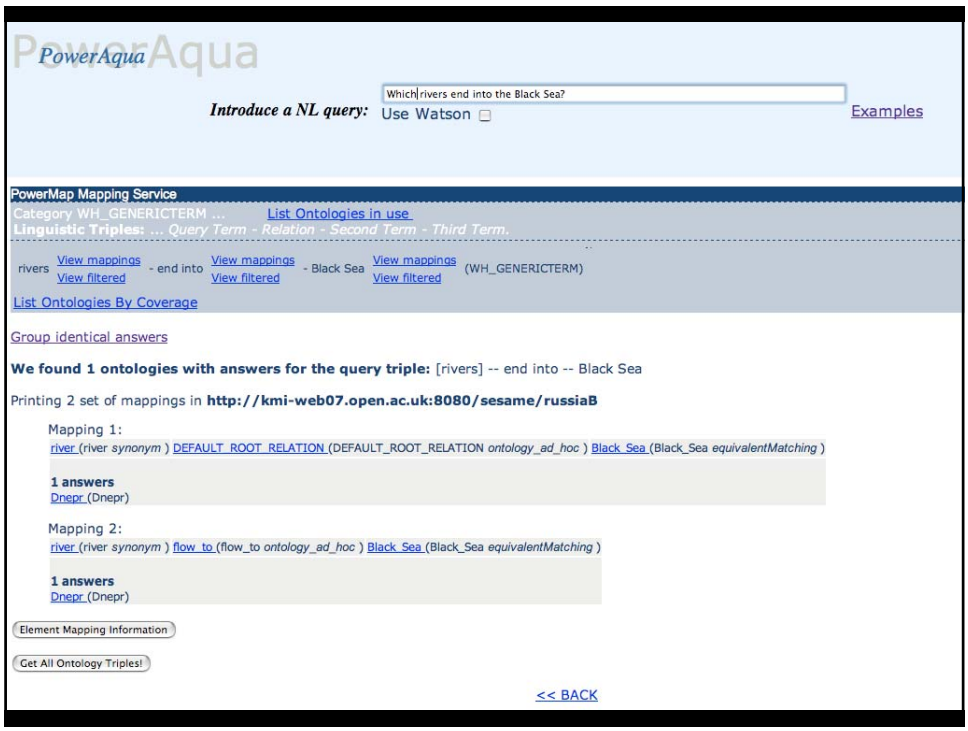
- Focus on SW documents viewed as *semantic content*
  - Not simply as documents
- Quality control
  - Detects duplications
  - Fixes obvious syntax problems
    - E.g., duplicated ontology IDs, namespaces, etc..
- Structures ontologies in a network
  - Using relations such as: imports, duplicates, *sameAs, extends, inconsistentWith,*
- Provides interfaces for both human users and software programs
- Provides efficient API
- Automatically generates OMV data
- Supports formal queries (SPARQL)
- Semantics-informed ontology ranking
- Integrated with review system (Revyu.com)
- Plug-ins for Protégé and NeOn Toolkit
  - These support ontology engineering through reuse of SW resources

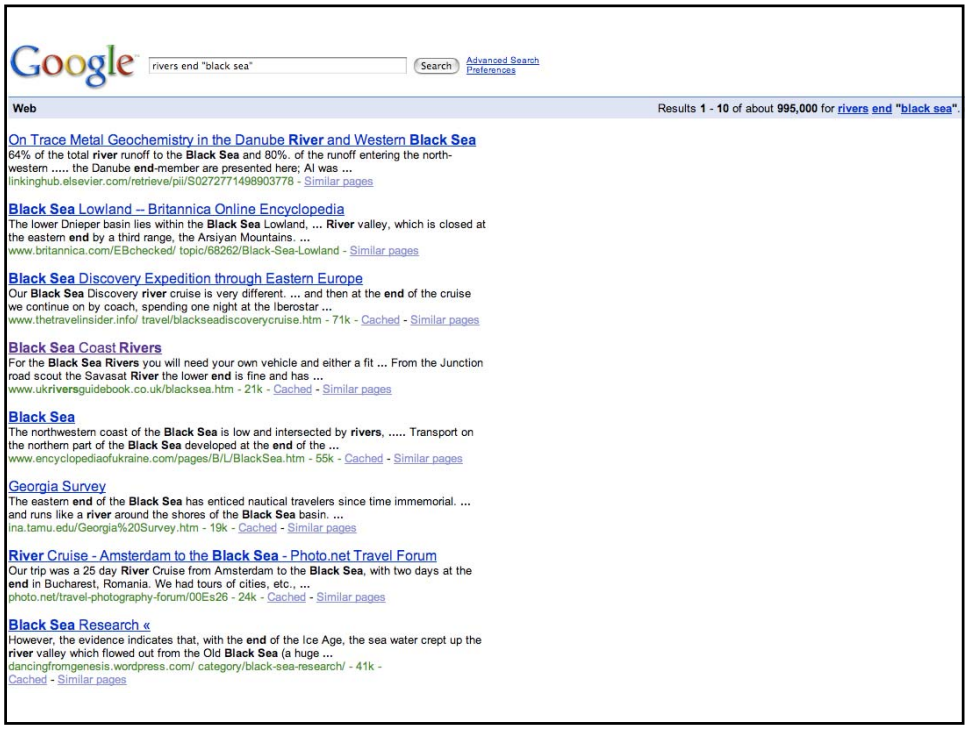



PowerAqua
Introduce a NL query:
Give me rock albums produced by Santana
Use Watson
Examples
PowerMap Mapping Service
Category WH_GENERICTERM ...
List Ontologies in use
Linguistic Triples: ... Query Term - Relation - Second Term - Third Term.
albums View mappings View filtered
rock View mappings View filtered
- produced View mappings View filtered
- Santana View mappings View filtered
(WH_GENERICTERM)
rock View mappings View filtered
- albums View mappings View filtered
(Not valid)
List Ontologies By Coverage
Group identical answers
We found 1 ontologies with answers for the query triple: [albums, rock] -- produced -- Santana
Printing 2 set of mappings in http://kmi-web07.open.ac.uk:8080/sesame/music
Mapping 1:
Album (Album synonym) has_albums (has_albums ontology_ad_hoc) Santana (Santana equivalentMatching)
2 answers
Caravanserai (Caravanserai)
Abraxas (Abraxas)
Mapping 2:
rock (rock equivalentMatching) has_genre (has_genre ontology_ad_hoc) Santana (Santana equivalentMatching)

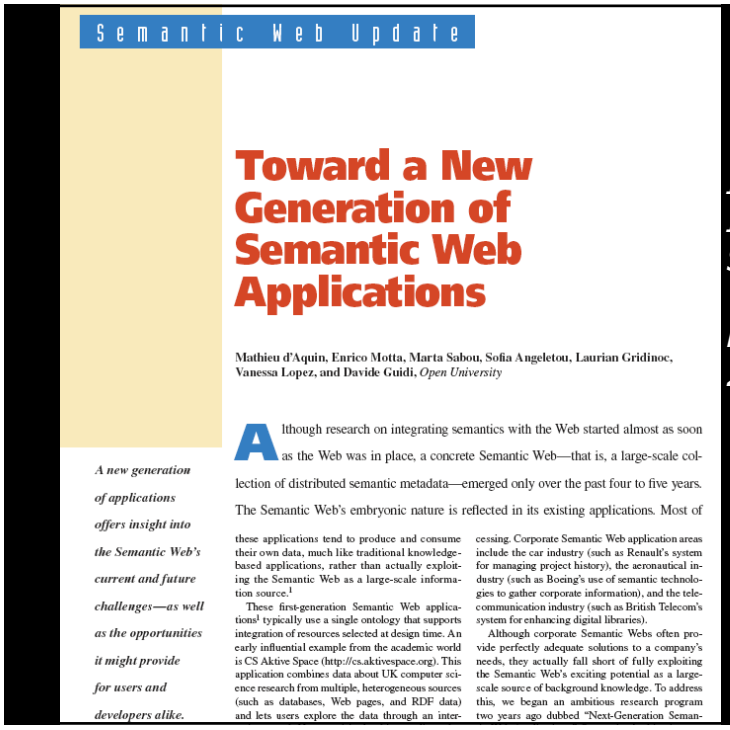

PowerAqua
Introduce a NL query:
what are the main rivers of Russia?
Use Watson
Examples
PowerMap Mapping Service
Category WH_UNKNTERM ...
List Ontologies in use
Linguistic Triples: ... Query Term - Relation - Second Term - Third Term.
what_is - rivers View mappings View filtered
- Russia View mappings View filtered
(WH_UNKNTERM)
List Ontologies By Coverage
Group identical answers
We found 2 ontologies with answers for the query triple: [what_is] -- rivers -- Russia
Printing 5 set of mappings in http://kmi-web07.open.ac.uk:8080/sesame/KIM
Mapping 1:
River (River synonym) partOf (partOf ontology_ad_hoc) Entity (Entity ontology_ad_hoc)
Entity (Entity ontology_ad_hoc) hasAlias (hasAlias ontology_ad_hoc) Country_T.RS.6 (Russian Soviet Federated Socialist Republic synonym)
5 answers
River_T.37 (Volga) Russian Federation
River_T.24 (Ob) Russian Federation
River_T.3 (Amur) Russian Federation
River_T.38 (Yenisei) Russian Federation
River_T.16 (Lena) Russian Federation
Mapping 2:
River (River synonym) locatedIn (locatedIn ontology_ad_hoc) Country_T.RS (Russian Federation synonym)


PowerAqua
Introduce a NL query:
what are the main rivers of Russia?
Use Watson
Examples
Mapping 5:
River (River synonym) partOf (partOf ontology_ad_hoc) Entity (Entity ontology_ad_hoc)
Entity (Entity ontology_ad_hoc) hasAlias (hasAlias ontology_ad_hoc) Country_T.RS.1 (Russian Federation synonym)
5 answers
River_T.37 (Volga) Russian Federation
River_T.24 (Ob) Russian Federation
River_T.3 (Amur) Russian Federation
River_T.38 (Yenisei) Russian Federation
River_T.16 (Lena) Russian Federation
Element Mapping Information
Printing 2 set of mappings in http://kmi-web07.open.ac.uk:8080/sesame/russiaB
Mapping 1:
river (river synonym) DEFAULT_ROOT_RELATION (DEFAULT_ROOT_RELATION ontology_ad_hoc) Russia (Russia equivalentMatching)
4 answers
Neva (Neva)
Don (Don)
Volga (Volga)
Dnepr (Dnepr)
Mapping 2:
river (river synonym) has_major_river (has_major_river ontology_ad_hoc) Russia (Russia equivalentMatching)
4 answers
Neva (Neva)
Don (Don)
Volga (Volga)
Dnepr (Dnepr)


PowerAqua
Introduce a NL query:
Which rivers end into the Black Sea?
Use Watson
Examples
PowerMap Mapping Service
Category WH_GENERICTERM ... List Ontologies in use
Linguistic Triples: ... Query Term - Relation - Second Term - Third Term.
rivers View mappings View filtered - end into View mappings View filtered - Black Sea View mappings View filtered (WH_GENERICTERM)
List Ontologies By Coverage
Group identical answers
We found 1 ontologies with answers for the query triple: [rivers] -- end into -- Black Sea
Printing 2 set of mappings in http://kmi-web07.open.ac.uk:8080/sesame/russiaB
Mapping 1:
river (river synonym) DEFAULT_ROOT_RELATION (DEFAULT_ROOT_RELATION ontology_ad_hoc) Black_Sea (Black_Sea equivalentMatching)
1 answers
Dnepr (Dnepr)
Mapping 2:
river (river synonym) flow_to (flow_to ontology_ad_hoc) Black_Sea (Black_Sea equivalentMatching)
1 answers
Dnepr (Dnepr)
Element Mapping Information
Get All Ontology Triples!
<< BACK


Google
rivers end "black sea"
Search
Advanced Search
Preferences
Web
Results 1 - 10 of about 995,000 for rivers end "black sea".
On Trace Metal Geochemistry in the Danube River and Western Black Sea
64% of the total river runoff to the Black Sea and 80%. of the runoff entering the north-western ..... the Danube end-member are presented here; Al was ...
linkinghub.elsevier.com/retrieve/pii/S0272771498903778 - Similar pages
Black Sea Lowland -- Britannica Online Encyclopedia
The lower Dnieper basin lies within the Black Sea Lowland, ... River valley, which is closed at the eastern end by a third range, the Arsiyan Mountains. ...
www.britannica.com/EBchecked/ topic/68262/Black-Sea-Lowland - Similar pages
Black Sea Discovery Expedition through Eastern Europe
Our Black Sea Discovery river cruise is very different. ... and then at the end of the cruise we continue on by coach, spending one night at the Iberostar ...
www.thetravelinsider.info/ travel/blackseadiscoverycruise.htm - 71k - Cached - Similar pages
Black Sea Coast Rivers
For the Black Sea Rivers you will need your own vehicle and either a fit ... From the Junction road scout the Savasat River the lower end is fine and has ...
www.ukriversguidebook.co.uk/blacksea.htm - 21k - Cached - Similar pages
Black Sea
The northwestern coast of the Black Sea is low and intersected by rivers, ... Transport on the northern part of the Black Sea developed at the end of the ...
www.encyclopediaofukraine.com/pages/B/L/BlackSea.htm - 55k - Cached - Similar pages
Georgia Survey
The eastern end of the Black Sea has enticed nautical travelers since time immemorial. ... and runs like a river around the shores of the Black Sea basin. ...
ina.tamu.edu/Georgia%20Survey.htm - 19k - Cached - Similar pages
River Cruise - Amsterdam to the Black Sea - Photo.net Travel Forum
Our trip was a 25 day River Cruise from Amsterdam to the Black Sea, with two days at the end in Bucharest, Romania. We had tours of cities, etc., ...
photo.net/travel-photography-forum/00Es26 - 24k - Cached - Similar pages
Black Sea Research «
However, the evidence indicates that, with the end of the Ice Age, the sea water crept up the river valley which flowed out from the Old Black Sea (a huge ...
dancingfromgenesis.wordpress.com/ category/black-sea-research/ - 41k - Cached - Similar pages


PowerAqua
Introduce a NL query:
Which rivers end into the Black Sea?
Use Watson
Examples
PowerMap Mapping Service
Category WH_GENERICTERM ...
List Ontologies in use
Linguistic Triples: ... Query Term - Relation - Second Term - Third Term
rivers View mappings View filtered - end into View mappings View filtered - Black Sea View mappings View filtered (WH_C
List Ontologies By Coverage
Group identical answers
We found 1 ontologies with answers for the query triple: [rivers] --
Printing 2 set of mappings in http://kmi-web07.open.ac.uk:8080/ses
Mapping 1:
river (river synonym) DEFAULT_ROOT_RELATION (DEFAULT_ROOT_RELATION ontology_ad_hoc)
1 answers
Dnepr (Dnepr)
Mapping 2:
river (river synonym) flow_to (flow_to ontology_ad_hoc) Black_Sea (Black_Sea eq
1 answers
Dnepr (Dnepr)
Element Mapping Information
Get All Ontology Triples!
<< BACK
Google
Dnepr rivers end "black sea"
Search
Advanced Search
Preferences
Web
Dnieper - MSN Encarta
Dnieper or Dnepr, river and important traffic artery of Ukraine, Belarus, ... and flows in a general southerly direction to empty into the Black Sea, ...
encarta.msn.com/encyclopedia_761558003/dnieper.html - 30k - Cached - Similar pages
Dnepr -- FREE Dnepr Information | Encyclopedia.com: Facts, Pictures ...
World Encyclopedia Dnieper (Dnepr) River in e Europe. Rising in the Valdai Hills, w of Moscow, it flows s through Belarus and Ukraine to the Black Sea. ...
www.encyclopedia.com/doc/1E1-X-Dnepr.html - 84k - Cached - Similar pages
Journal of Marine Systems : Reconnaissance of the main Black Sea's ...
Since the Dnepr mouth and the Dnestr mouth are very close to one another, the river discharge is allocated in the location of the river mouth of the biggest ...
linkinghub.elsevier.com/retrieve/pii/S0924796397001012 - Similar pages
Dnieper River -- Britannica Online Encyclopedia
Dnieper Riverriver, Europe Ukrainian Dnipro, Russian Dnepr, Belorussian Dnepro, ... The lower Dnieper basin lies within the Black Sea Lowland, ...
www.britannica.com/EBchecked/topic/167192/Dnieper-River - Similar pages
hidden europe magazine - Articles - get a flavour of hidden europe ...
The northern Black Sea region plays host to some remarkably resilient cultural .... This small railway station is at the south end of the Dnepr river dam, ...
www.hiddeneurope.co.uk/ carticle_info.php?articles_id=209 - 42k - Cached - Similar pages
AQUASTAT - FAO's Information System on Water and Agriculture
The Dnepr basin, covering about 65% of the country. The Dnepr River rises ... the small rivers which flow directly into the Sea of Azov and the Black Sea, ...
www.fao.org/nr/water/aquastat/ countries/ukraine/index.stm - 39k - Cached - Similar pages
Dnepr, River (Eng Dnieperancient Borysthenes) Online Encyclopedia ...
River in W Russia; rises in the S Valdayskaya Vozvyshennost range, flows S and W past ... End of Article: Dnepr, River (Eng Dnieperancient Borysthenes) ...
encyclopedia.jrank.org/Cambridge/ entries/030/Dnepr-River.html - 5k - Cached - Similar pages



Semantic Web Update

Toward a New Generation of Semantic Web Applications

Mathieu d'Aquin, Enrico Motta, Marta Sabou, Sofia Angeletou, Laurian Gridinoc, Vanessa Lopez, and Davide Guidi, *Open University*

A new generation of applications offers insight into the Semantic Web's current and future challenges—as well as the opportunities it might provide for users and developers alike.

Although research on integrating semantics with the Web started almost as soon as the Web was in place, a concrete Semantic Web—that is, a large-scale collection of distributed semantic metadata—emerged only over the past four to five years. The Semantic Web's embryonic nature is reflected in its existing applications. Most of these applications tend to produce and consume their own data, much like traditional knowledge-based applications, rather than actually exploiting the Semantic Web as a large-scale information source.[1]

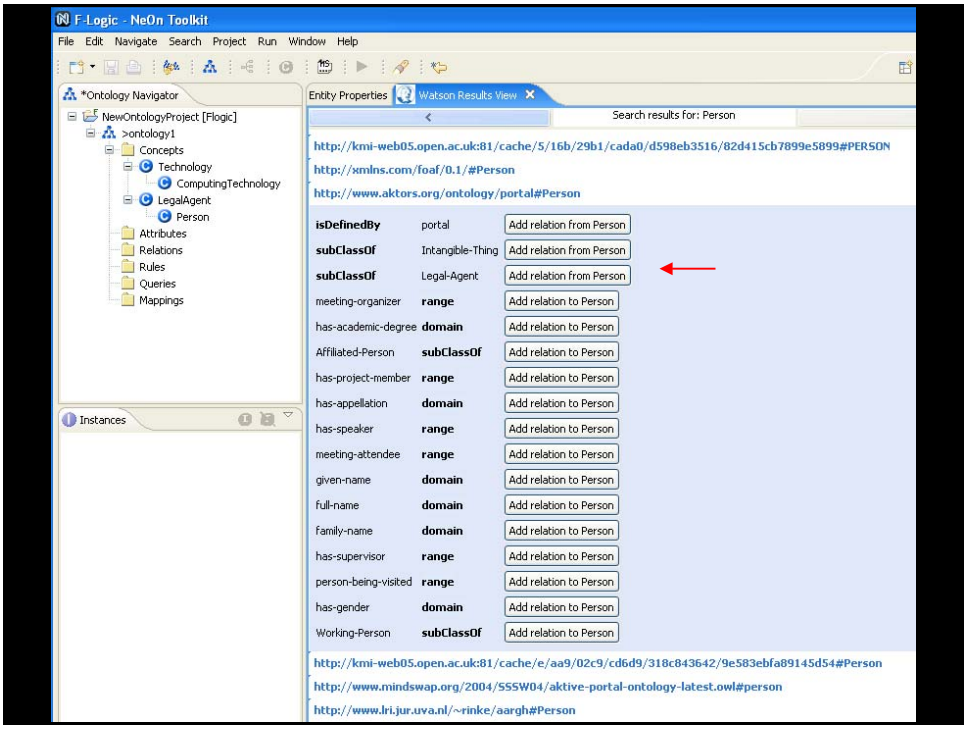
These first-generation Semantic Web applications[1] typically use a single ontology that supports integration of resources selected at design time. An early influential example from the academic world is CS Aktive Space (http://cs.aktivespace.org). This application combines data about UK computer science research from multiple, heterogeneous sources (such as databases, Web pages, and RDF data) and lets users explore the data through an interprocessing. Corporate Semantic Web application areas include the car industry (such as Renault's system for managing project history), the aeronautical industry (such as Boeing's use of semantic technologies to gather corporate information), and the telecommunication industry (such as British Telecom's system for enhancing digital libraries).

Although corporate Semantic Webs often provide perfectly adequate solutions to a company's needs, they actually fall short of fully exploiting the Semantic Web's exciting potential as a large-scale source of background knowledge. To address this, we began an ambitious research program two years ago dubbed "Next-Generation Seman-

IEEE Intelligent Systems

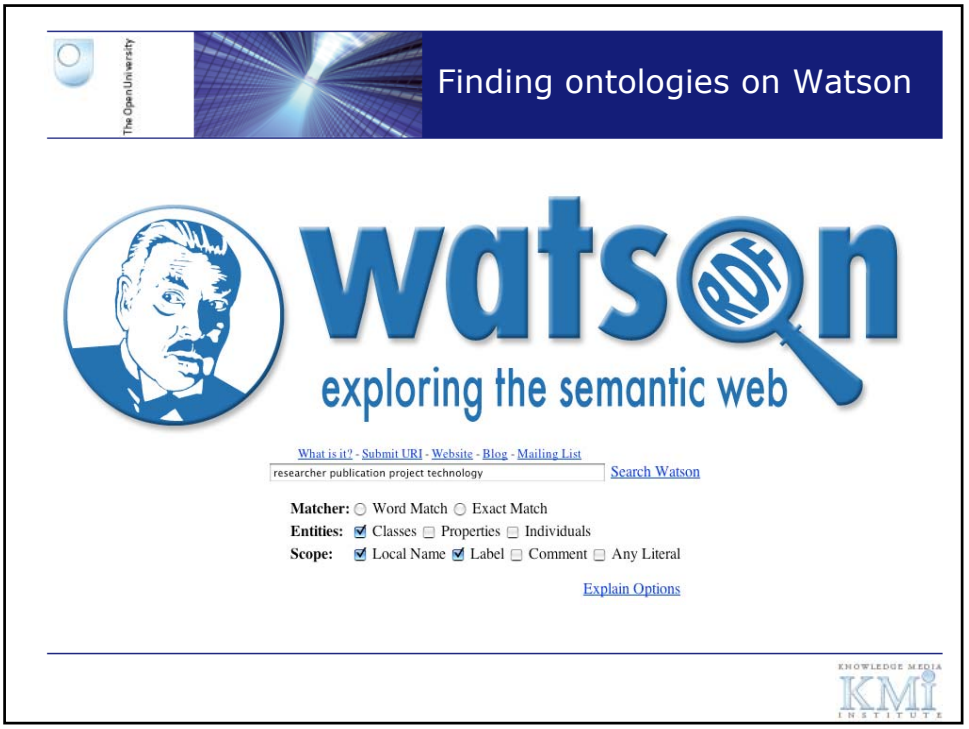
May/June 2008

Human-Ontology Interaction



Issue: How can we facilitate the process of finding the right ontology?


The Open University
Search
Home
News
Projects
Technologies
Publications
People
KNOWLEDGE MEDIA
KMi
INSTITUTE
About Us
What we Do
Join Us
Seminars [1]
Job Vacancies [2]
Studentships [3]
Visitors/Internships
Working With KMi
Find Us
CRC
MMIS
Spotlight
Buddyspace
Instant messenger / Presence
Magpie
The semantic filter
Hexagon
Community / Presence
Job / Studentship Vacancies
View the Job Vacancies
View the Studentship Vacancies
KMi employs 93 people, a mix of researchers, technologists, designers and administrative staff. We are in a phase of rapid expansion, and as a result job opportunities arise frequently...view details
KMi News
Global Sensemaking network launches
I'm delighted to say that the Global Sensemaking network has just gone public, after several months of community
[More]
Knowledge Media Institute (KMi®) Registered Trademark


The Open University
Finding ontologies on Watson
watson
RDF
exploring the semantic web
What is it? - Submit URI - Website - Blog - Mailing List
researcher publication project technology
Search Watson
Matcher: Word Match Exact Match
Entities: Classes Properties Individuals
Scope: Local Name Label Comment Any Literal
Explain Options
KNOWLEDGE MEDIA
KMi
INSTITUTE

What is it? - Submit URI - Website - Blog - Mailing List

researcher publication project technology Search Watson

Found 3 semantic documents - Search Options

1- http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#researcher*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#visiting-researcher*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#publication-reference*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#periodical-publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#serial-publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#composite-publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#electronic-publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#article-in-a-composite-publication*
- *http://www.mindswap.org/2004/SSSW04/aktive-portal-ontology-latest.owl#publication-type-event*

More...

2- http://www.aktors.org/ontology/portal
- *http://www.aktors.org/ontology/portal#Researcher*
- *http://www.aktors.org/ontology/portal#Researcher-In-Academia*
- *http://www.aktors.org/ontology/portal#Visiting-Researcher*
- *http://www.aktors.org/ontology/portal#Publication*
- *http://www.aktors.org/ontology/portal#Composite-Publication*
- *http://www.aktors.org/ontology/portal#Publication-Reference*
- *http://www.aktors.org/ontology/portal#Periodical-Publication*
- *http://www.aktors.org/ontology/portal#Serial-Publication*
- *http://www.aktors.org/ontology/portal#Publication-Type-Event*
- *http://www.aktors.org/ontology/portal#Article-In-A-Composite-Publication*

More...

3- http://www.aktors.org/ontology/portal.daml
- *http://www.aktors.org/ontology/portal#Researcher*
- *http://www.aktors.org/ontology/portal#Researcher-In-Academia*
- *http://www.aktors.org/ontology/portal#Visiting-Researcher*
- *http://www.aktors.org/ontology/portal#Publication*
- *http://www.aktors.org/ontology/portal#Publication-Reference*
- *http://www.aktors.org/ontology/portal#Serial-Publication*

Details for http://www.aktors.org/ontology/portal#Researcher (view as graph)
Back

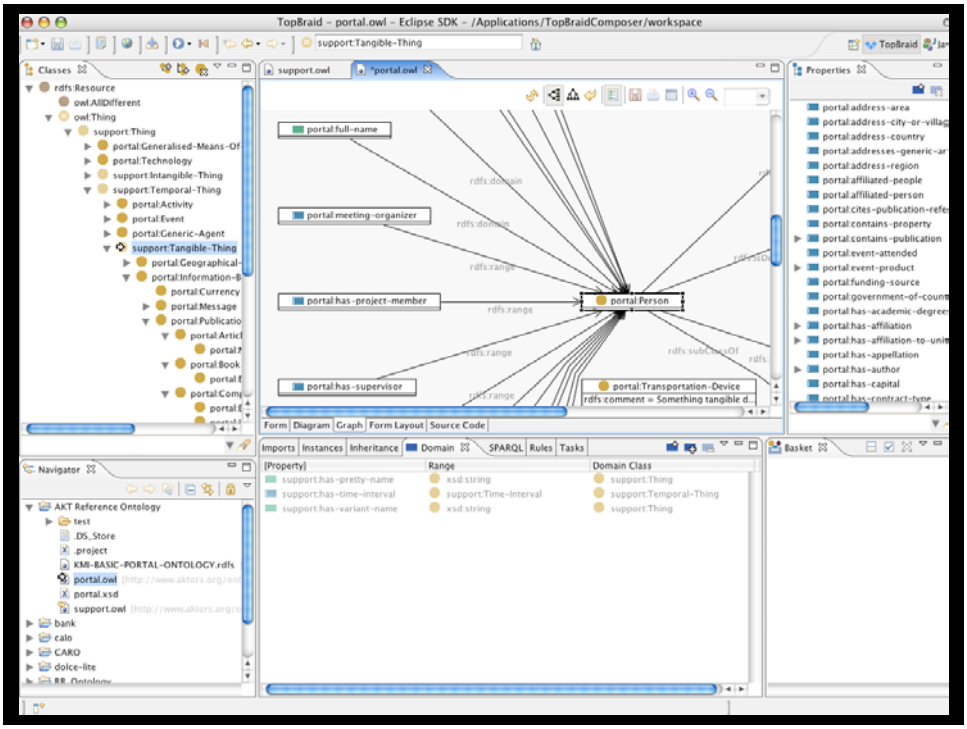
- In http://www.aktors.org/ontology/portal.daml
  - *Class*
  - **isDefinedBy**: http://www.aktors.org/ontology/portal
  - **subClassOf**: http://www.aktors.org/ontology/portal#Working-Person
  - http://www.aktors.org/ontology/portal#Visiting-Researcher: **subClassOf**
  - http://www.aktors.org/ontology/portal#Researcher-In-Academia: **subClassOf**
- In http://www.aktors.org/ontology/portal
  - *Class*
  - **isDefinedBy**: http://www.aktors.org/ontology/portal
  - **subClassOf**: http://www.aktors.org/ontology/portal#Working-Person
  - http://www.aktors.org/ontology/portal#Visiting-Researcher: **subClassOf**
  - http://www.aktors.org/ontology/portal#Researcher-In-Academia: **subClassOf**
- In http://triplestore.aktors.org/resolve/?resource=http%3A%2F%2Fwww.aktors.org%2Fontology%2Fportal
  - *Individual*
  - **isDefinedBy**: http://www.aktors.org/ontology/portal

Details for http://www.aktors.org/ontology/portal
Back

Get cached file - Query with SPARQL - Get OMV

Size of the file	89 KB
Number of statements	1480
Representation languages	RDF,OWL
Label	AKT Reference Ontology (Portal Ontology)
Comment	
Employed DL	ALCHOIN
Number of classes	152
Number of properties	122
Number of individuals	63
User Reviews	Loading... Review with Revyu.com
Locations	http://www.aktors.org/ontology/portal
Imports	http://www.aktors.org/ontology/support
Imported By	http://www.aktors.org/ontology/extension http://www.aktors.org/ontology/extension.daml http://www.aktors.org/ontology/rdfcompat http://www.hyphen.info/RDF/Southampton/southampton-people.rdf http://www.hyphen.info/RDF/RAE/rae-uoa02-people.rdf http://www.hyphen.info/RDF/RAE/rae-uoa02-groups.rdf http://www.hyphen.info/RDF/RAE/rae-uoa03-groups.rdf http://users.ecs.soton.ac.uk/~swh/result-size/source.rdf http://triplestore.aktors.org/resolve/?resource=http%3A%2F%2Fwww.aktors.org%2Fontology%2Fportal http://www.hyphen.info/RDF/RAE/rae-uoa01-people.rdf http://www.hyphen.info/RDF/Southampton/southampton-projects.rdf http://www.hyphen.info/RDF/RAE/rae-institutions.rdf http://www.hyphen.info/RDF/RAE/rae-uoa01-groups.rdf

INSTITUTE



Summarizing ontologies

- What you need is a way to quickly get a 'gestalt impression' of what an ontology is about
- People are able to summarize an ontology
  - "The AKT portal ontology can be used to describe academic organizations. It covers concepts such as event, person, technology, project, etc.."
- That is, they can extract the **key concepts** which can effectively summarise an ontology

Identifying Key Concepts

Project
Researcher
Phd Student
Technology
Publication
Educational Organization
.
.
.
.
.

Ontology Metadata

- Provides answers to pertinent questions about the content and provenance of the ontology in the library:
  - What is the domain covered by an ontology?
  - **What are the key classes and concepts?**
  - Who developed the ontology?
  - What is the policy for maintenance and distribution?
  - What is the format of the ontology (syntax, language, tools used to build it, etc.)

From Natasha's Talk

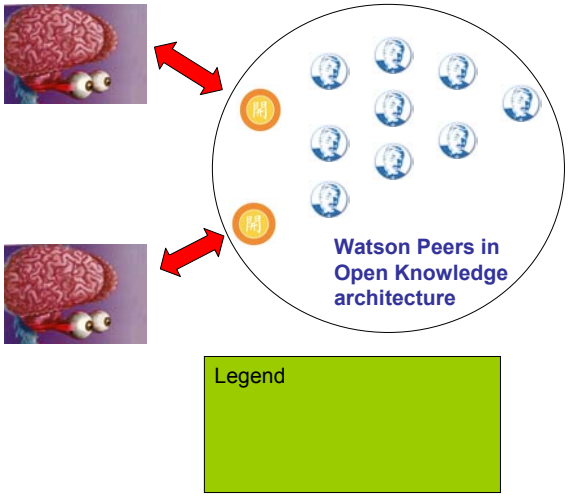
Research Issues

- What are the right concepts that could be used to describe an ontology concisely?
- Are there any principles/regularities in the way human beings extract 'key concepts' from an ontology?
- Can these principles be automated, to define algorithms that are able to characterize an ontology the way people do?

Possible Applications

- To support the ability to create meaningful 'snapshots' of ontologies
  - E.g., to support ontology understanding when browsing the SW with Watson
  - E.g. when using an ontology engineering toolkit, such as NeOn, Protégé, TopBraid, etc..
- To support new navigation/visualization mechanisms, which can improve over the rigid taxomonic displays provided by current ontology engineering tools
- To identify the 'best concepts' to classify, in automatic data classification scenarios
  - Tx to FVH for suggesting this application
- To provide mechanisms for knowledge providers to advertise knowledge contents, without publishing the whole ontology
  - 'cautious knowledge sharing scenarios'

Integration with peer-to-peer architecture

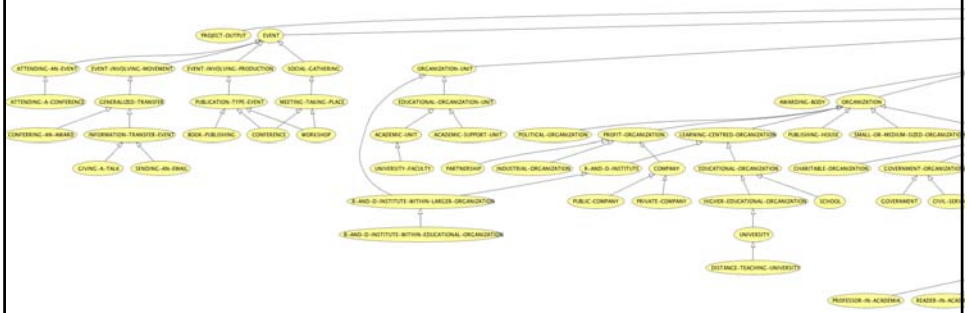


- The integration ensures that the OK architecture remains transparent to applications
  - The application peers mediate between the Mini Watsons and the external apps, forwarding queries and collecting results
- Each Watson peer defines a separate ontology provider
  - Architectural basis for "Cautious Knowledge Sharing" scenario

Choosing N concepts

- Challenge problem: Is there a way to automatically extract the 'best' N concepts which can be used to advertise an ontology?
- What are the criteria?…

Practical: Identifying Key Concepts



Generating key concepts from an ontology (V1)

- We developed a first version of the algorithm to identify the key concepts of an ontology, in order to test it with the selections made by human beings
- Specifically, our algorithm integrates the following three criteria:
  - **Density of concepts**
  - **Natural Categories**
  - **Coverage of isa hierarchy**

Density

- Density is a value related to any category
- It is worked out taking into consideration the number of sub-categories, the number of properties and the number of instances of each category
- We consider 2 different types of density: the global density and the local density

*Density(c) = (globalDensity(c) * weightG) + (localDensity(c) * weightL)*

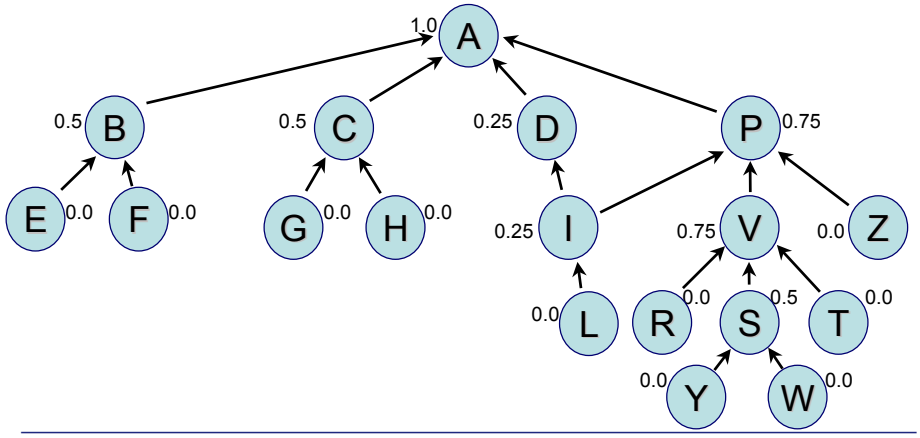
Global Density

- The global density is a [0,1] value that describes how much a category is dense taking into consideration the whole ontology

globalDensity(c) =
*(n° sub categories(c) * weightS) +*
*(n° properties(c) * weightP) +*
*(n° instances(c) * weightI)*

Global Density

- The global density is a [0,1] value that describes how much a category is dense taking into consideration the whole ontology



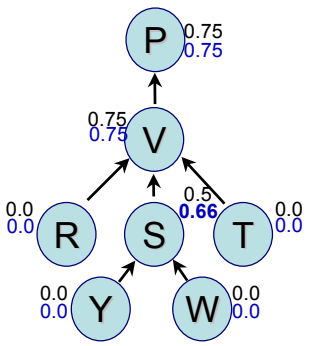
Local Density

- The local density is a [0,1] value that describes how much a category is dense taking into consideration a limited part of the ontology
- We use the local density to understand whether a category C (that does not have a good global density) is significantly dense compared with its "relatives" (i.e. all the categories that I can reach starting from C using at most two is-a relations)

Local Density

- Taking into consideration the previous example, only the categories "S" has a local density value that differs from their global density

localDensity(c) =
globalDensity(c) /
max{
globalDensity(c),
globalDensity(relatives (c))
}



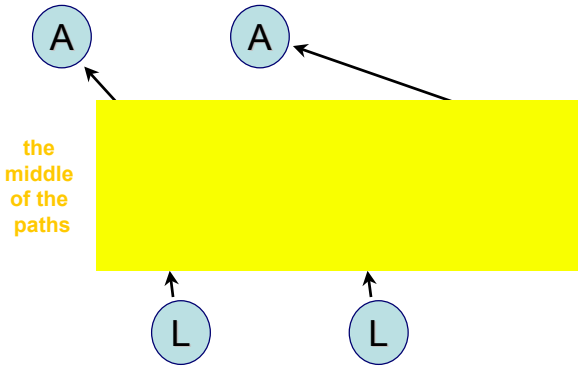
Natural Category (v1)

- Basic-level identification:
  "The level of specificity/generality of a concept within a conceptual hierarchy at which most people tend naturally to categorize it" [A Dictionary of Psychology, Oxford University Press, 2001]
- Label Simplicity: normally a natural category has a simple name. For example, "Chair" is 'more 'natural' than "KitchenChair"
- Other constraints:
  - root and leaf categories cannot be a natural category
  - there can be at most one natural category for each branch

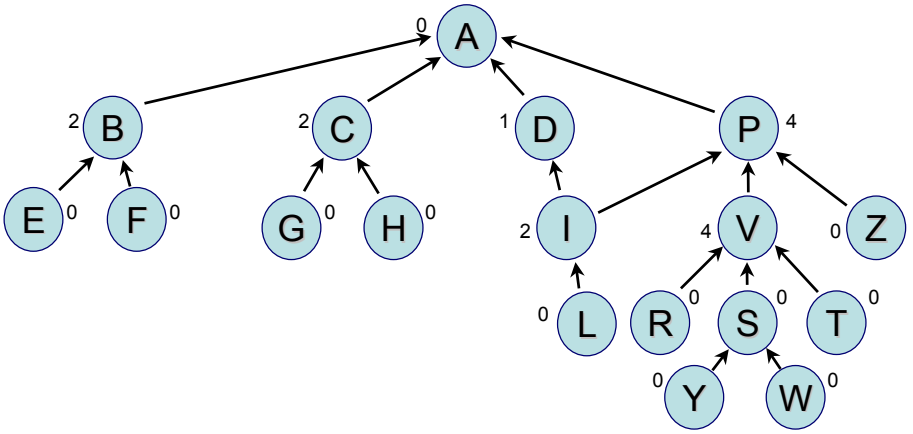
Basic Level

- The basic level of a category X is an integer value that denotes how many times X is in the middle of a path from the root to a leaf

Example of the 2 root-to-leaf paths that have the category "I" in the middle (for this reason the basic level of "I" specified in the previous slide is 2)



Basic Level



Name Naturalness

- The name 'naturalness' is a [0,1] value that describes whether a category name is a good name or not
- Normally, a natural category has a simple name
- All the categories with a compound name are penalized

Natural Category (v1)

*NaturalCategoryValue(c) = (basicLevel01(c) * weightL) + (nameGoodness(c) * weightN)*

basicLevel01(c) =
basicLevel(c) /
Max {
for each x in the ontology then
basicLevel(x)
}

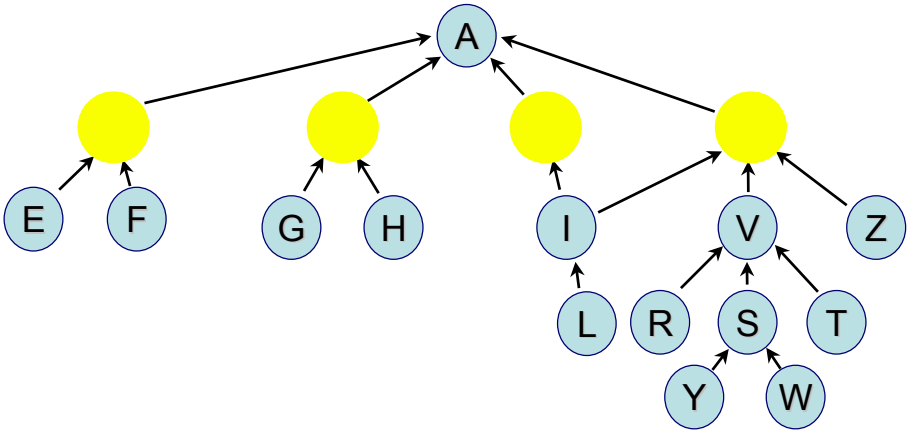
nameGoodness(c) =
1.0 -
*(0.3 * (count(words(name(c))) - 1))*

Coverage (v1)

- Taking into consideration a number n, the goal is to find one or more ontology coverages composed by exactly n categories
- There are 2 different kinds of coverage:
  - total coverage
  - partial coverage

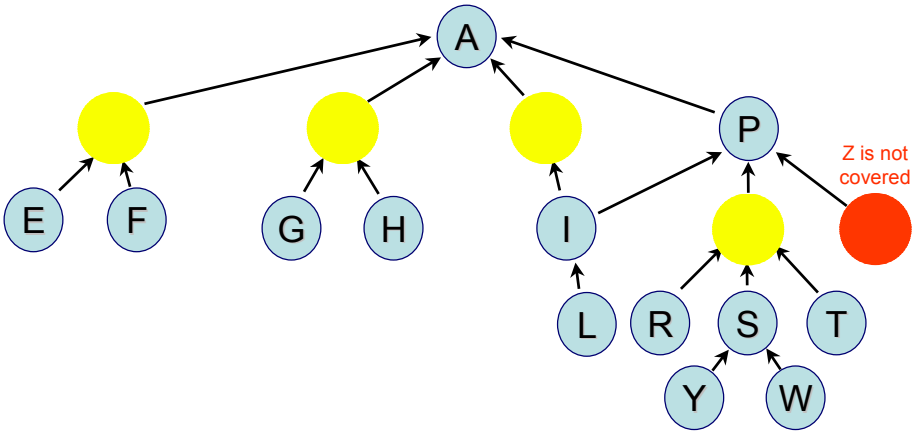
Total Coverage

- Example of a total coverage with 4 categories



Partial Coverage

- Example of a partial coverage with 4 categories



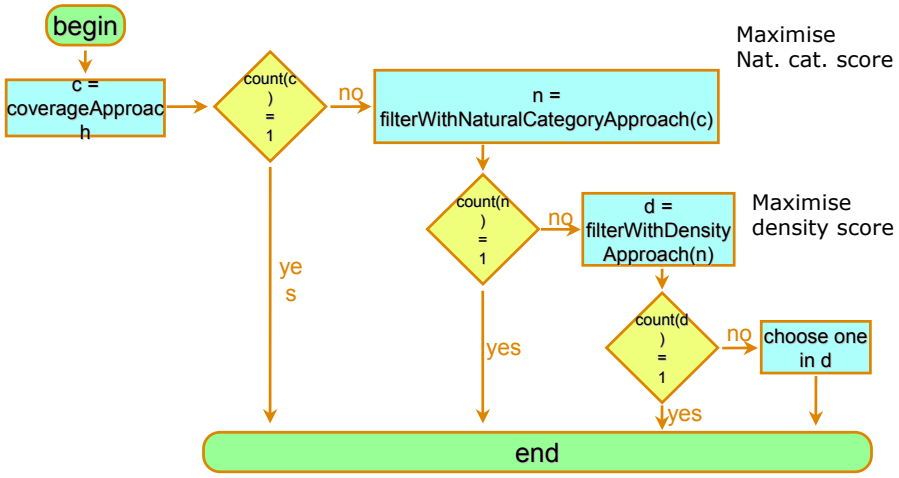
Coverage & Natural Category (v1)

- The coverage algorithm developed is based on a preventive application of the natural category approach
  - The natural categories identified represent a coverage for the ontology
- There can be 3 different cases (n is the number of the category for the coverage):
  - # natural category = n
  - # natural category > n
  - # natural category < n

Minimum Standard Deviation

- From a set of n-coverages with the same coverage affordance, we select the one with the minimum standard deviation
- The standard deviation for a coverage C is worked out taking into consideration the number of categories that are covered by each category in C

Combined Approach (first version): Work Flow



Empirical Analysis

- 4 ontologies have been used for our tests
- We asked 8 ontology engineers to choose 20 concepts they considered the most representative ones in the ontologies
- We also asked them to try and take into account the coverage criterion (but we did not mention any other criterion)

Results

Ontology	Number of categories	Common categories choosen by testers
biosphere	87	MarineAnimal, Plant, Animal, Vegetation, Bird, Fungi, Insect, Mammal, Microbiota, Reptile
music	91	Genre, Instrument, MusicArtist, MusicalExpression, Record, Medium, MusicGroup, Sound, Event
financial	188	bond, bank, contract, broker, dealer, stock, capital, financial_market, order
aktors portal	247	PUBLICATION-REFERENCE, PUBLICATION, EVENT, SOFTWARE-TECHNOLOGY, COMPUTING-TECHNOLOGY, ORGANIZATION, GEOPOLITICAL-ENTITY, PERSON

Testers' agreement

Ontology	% min. agreement	% max. agreement	% agreement among testers
biosphere	50	100	73,7499999
music	33,3333333	88,8888888	76,3888888
financial	44,4444444	100	75
aktors portal	37,5	87,5	73,6111111

mean: 74.68

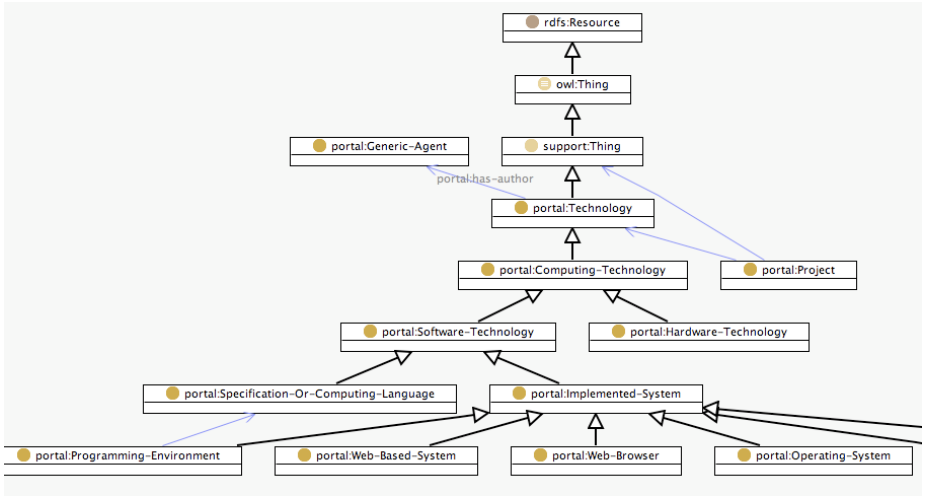
Correlation between algorithm and human experts

Ontology	Common choices between the testers and the algorithm	%
biosphere	Animal, MarineAnimal, Plant	30
music	Event, Genre, Instrument, MusicalExpression	44,4444
financial	broker, dealer, order	33,3333
aktors portal	COMPUTING-TECHNOLOGY, EVENT, ORGANIZATION, PERSON, PUBLICATION-REFERENCE	62,5

mean: 42.5

How people select key concepts

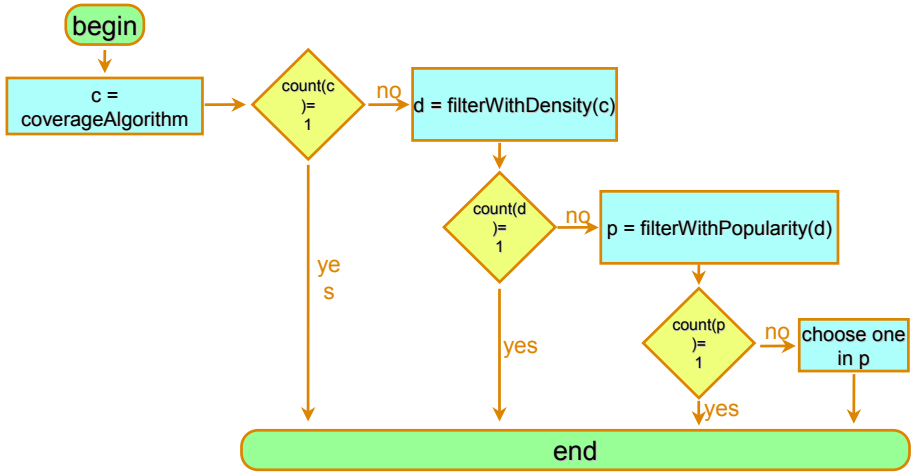
- Experts appear to use:
  - Density, particularly considering its sub categories
  - Concepts with simple labels
    - This tends to correlate with natural categories
  - *the popularity of the concept name*
- Coverage does not appear to be so important
  - Even though we explicitly asked people to use it
- In contrast with our algorithm, which tries to maximise coverage, people are prepared to select concepts that subsume each other, if they satisfy popularity and/or density and/or natural categories



Algorithm (v2)

- We have revised the algorithm, which now integrates the following four criteria:
  - Density
  - Popularity
    - The popularity is an integer value equal or greater than 0 retrieved observing how many results are returned by Google for a query composed by the name of a category
  - Natural Categories
    - We removed the constraint of at most 1 nat. cat. for isa path
  - Coverage

Combined Method (final version): Work Flow



Algorithm result

Ontology	Algorithm choices	% matches with testers' choices
biosphere	*Animal*, *Bird*, Coral, Crop, Crown, Dairy, Human, *Insect*, Leaf, Livestock, *MarineAnimal*, MicrobiotaTaxonomy, Moss, Mushroom, *Plant*, Poultry, Sponge, *Vegetation*	60
music	*Event*, TimeLine, TimeLineMap, Cd, CorporateBody, *Genre*, *Instrument*, *Medium*, *MusicalExpression*, MusicalManifestation, MusicalWork, Performance, *Record*, Show, Signal, Item, Work, Agent, Group, Person	66,666666666
financial	*bond*, *capital*, *contract*, contract_agent, cost, *dealer*, demander, *financial_market*, floor_broker, option_contract, *order*, organization, payment, price, security, specialist, *stock*, supplier, transaction, value	77,777777777
aktors portal	ACADEMIC-STAFF-MEMBER, COMPOSITE-PUBLICATION, COUNTRY, *EVENT*, HIGHER-EDUCATIONAL-ORGANIZATION-EMPLOYEE, IMPLEMENTED-SYSTEM, INFORMATION-BEARING-OBJECT, INFORMATION-TRANSFER-EVENT, INTANGIBLE-THING, INTEGER, LOCAL-DISTRICT, MONTH, MUNICIPAL-UNIT, *ORGANIZATION*, ORGANIZATION-UNIT, *PUBLICATION*, *SOFTWARE-TECHNOLOGY*, STUDENT, UNIVERSITY, WORKING-PERSON	50

mean: 63.61

AKTors portal without "INTANGIBLE-THING"

Ontology	Algorithm choices	% matches with testers' choices
aktors portal	ACADEMIC-STAFF-MEMBER, COMPOSITE-PUBLICATION, *EVENT*, GENERALISED-MEANS-OF-TRANSPORT, HIGHER-EDUCATIONAL-ORGANIZATION-EMPLOYEE, IMPLEMENTED-SYSTEM, INFORMATION-BEARING-OBJECT, INFORMATION-TRANSFER-EVENT, MUNICIPAL-UNIT, *ORGANIZATION*, *PUBLICATION*, *PUBLICATION-REFERENCE*, PUBLICATION-TYPE-EVENT, REFERENCE-TO-ITEM-IN-A-COMPOSITE-PUBLICATION, *SOFTWARE-TECHNOLOGY*, TEMPORAL-THING, UNIVERSITY, WORKING-PERSON	62,5

mean: 66.73

Conclusions (1)

- People appear to use reproducible mechanisms when selecting key concepts from an ontology, using both KR and psychological criteria
- The current version of the algorithm shows a good correlation with human experts
  - However it can be improved
- Additional empirical studies to evaluate the validity of the approach are needed

Conclusions (2)

- To a significant extent the Semantic Web is already in place
  - It is not a dream…
  - 10-20M semantic documents and 7-10K ontologies online
- NGSW apps are very different in nature from classical KBS.
  - Emphasis on handling scale, heterogeneity and quality issues
  - Deductive reasoning needs to be integrated with other technologies
- Our early prototypes give initial indications that relatively poor quality does not necessarily impede the development of useful apps
  - See IEEE IS paper for more details
- As the SW infrastructure becomes more robust, large scale access and exploitation of online knowledge may become the predominant paradigm for knowledge-intensive apps

KNOWLEDGE MEDIA INSTITUTE