Characterizing Semantic Web Applications

Prof. Enrico Motta
Director, Knowledge Media Institute
The Open University
Milton Keynes, UK

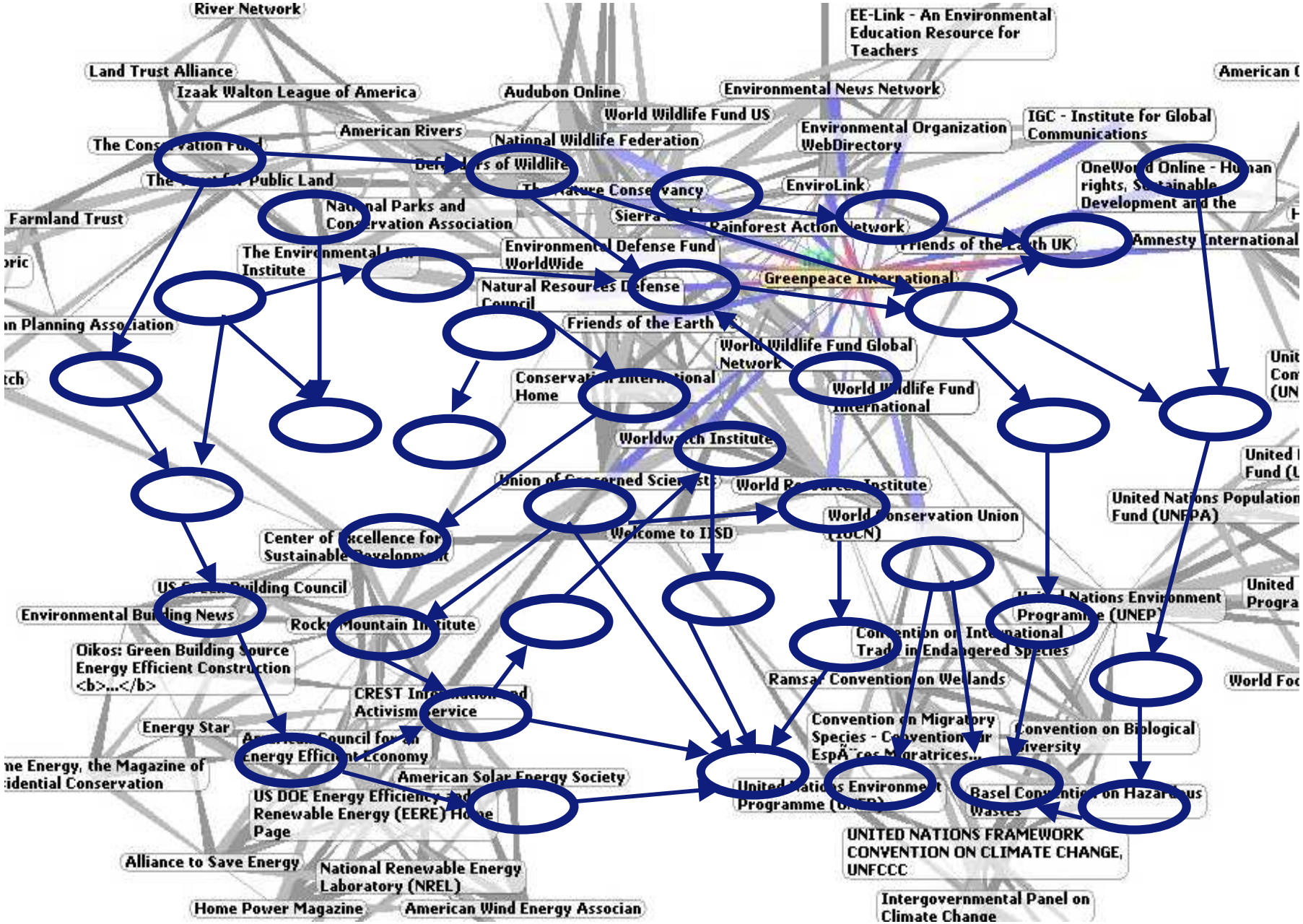
The Open University

Understanding the SW

- Issues
  - What is new/different about the semantic web?
  - What are the key aspects that characterize semantic web applications?
  - What are the key differences between semantic web applications and 'traditional' knowledge based systems?
- Results
  - A framework providing a characterization of semantic web applications
  - A classification of a representative sample of SW applications according to our framework
  - A blueprint (set of reqs) for designing SW applications

Semantics on the web
(The Semantic Web)

The Open University


River Network
EE-Link - An Environmental Education Resource for Teachers
Land Trust Alliance
Izaak Walton League of America
Audubon Online
Environmental News Network
World Wildlife Fund US
IGC - Institute for Global Communications
American Rivers
National Wildlife Federation
Environmental Organization WebDirectory
The Conservation Fund
Defenders of Wildlife
The Trust for Public Land
The Nature Conservancy
EnviroLink
OneWorld Online - Human rights, Sustainable Development and the
National Parks and Conservation Association
Sierra Club
Rainforest Action Network
Farmland Trust
Friends of the Earth UK
Amnesty International
The Environmental Law Institute
Environmental Defense Fund WorldWide
Natural Resources Defense Council
Greenpeace International
Friends of the Earth US
World Wildlife Fund Global Network
Conservation International Home
World Wildlife Fund International
Worldwatch Institute
Union of Concerned Scientists
World Resources Institute
United Nations Population Fund (UNFPA)
World Conservation Union (IUCN)
Welcome to IISD
Center of Excellence for Sustainable Development
US Green Building Council
Environmental Building News
Rocky Mountain Institute
United Nations Environment Programme (UNEP)
Oikos: Green Building Source Energy Efficient Construction ...
Convention on International Trade in Endangered Species
Ramsar Convention on Wetlands
CREST Information and Activism Service
Energy Star
Convention on Migratory Species - Convention sur Espèces Migratrices...
Convention on Biological Diversity
American Council for an Energy Efficient Economy
American Solar Energy Society
US DOE Energy Efficiency and Renewable Energy (EERE) Home Page
United Nations Environment Programme (UNEP)
Basel Convention on Hazardous Wastes
UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE, UNFCCC
Alliance to Save Energy
National Renewable Energy Laboratory (NREL)
Home Power Magazine
American Wind Energy Association
Intergovernmental Panel on Climate Change

Home

Enrico Motta

Professor of Knowledge Technologies
Director, Knowledge Media Institute, The Open University

How to Contact:

Knowledge Media Institute [**KMi**]
The Open University,
Milton Keynes,
MK7 6AA,
United Kingdom.

Email: e.motta@open.ac.uk
Phone: +44 (0) 1908 653506
Fax: +44 (0) 1908 653169

ADVANCED KNOWLEDGE AKT TECHNOLOGIES

NeOn

開

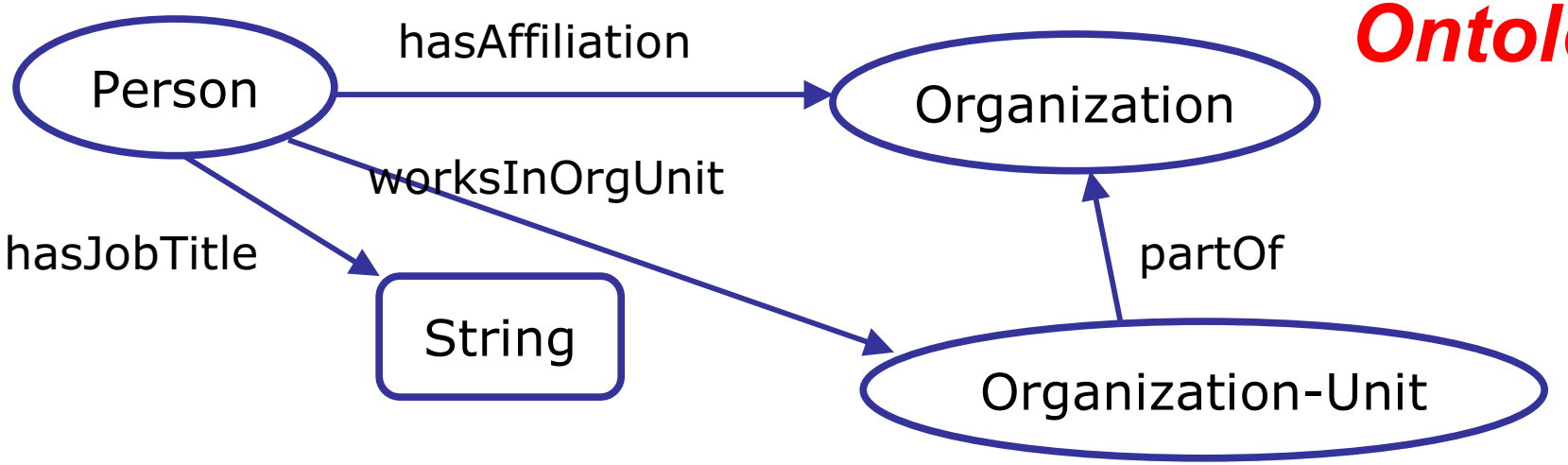
X-MEDIA

knowledgeweb

KNOWLEDGE MEDIA KMi INSTITUTE

The Open University


```
<akt:Person rdf:about="akt:EnricoMotta">
 <rdfs:label>Enrico Motta</rdfs:label>
 <akt:hasAffiliation rdf:resource="akt:TheOpenUniversity"/>
 <akt:hasJobTitle>kmi director</akt:hasJobTitle>
 <akt:worksInOrgUnit rdf:resource="akt:KnowledgeMediaInstitute"/
 <akt:hasGivenName>enrico</akt:hasGivenName>
 <akt:hasFamilyName>motta</akt:hasFamilyName>
 <akt:worksInProject rdf:resource="akt:Neon"/>
 <akt:worksInProject rdf:resource="akt:X-Media"/>
 <akt:hasPrettyName>Enrico Motta</akt:hasPrettyName>
 <akt:hasPostalAddress rdf:resource="akt:KmiPostalAddress"/>
 <akt:hasEmailAddress>e.motta@open.ac.uk</akt:hasEmailAddress>
 <akt:hasHomePage
   rdf:resource="http://kmi.open.ac.uk/people/motta/"/>
</akt:Person>
```



Ontology

Professor of Knowledge Technologies
Director, Knowledge Media Institute, The Open University

How to Contact:

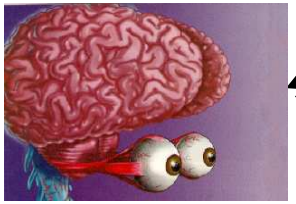
Knowledge Media Institute [**KMi**]
The Open University,
Milton Keynes,
MK7 6AA,
United Kingdom.

Email: e.motta@open.ac.uk
Phone: +44 (0) 1908 653506
Fax: +44 (0) 1908 653169

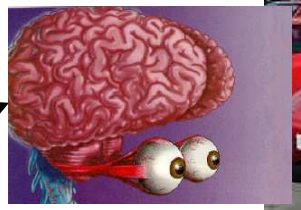
```
<akt:Person rdf:about="akt:EnricoMotta">
 <rdfs:label>Enrico Motta</rdfs:label>
 <akt:hasAffiliation rdf:resource="akt:TheOpenUniversity"/>
 <akt:hasJobTitle>kmi director</akt:hasJobTitle>
 <akt:worksInOrgUnit rdf:resource="akt:KnowledgeMediaInstit
 <akt:hasGivenName>enrico</akt:hasGivenName>
 <akt:hasFamilyName>motta</akt:hasFamilyName>
 <akt:worksInProject rdf:resource="akt:Neon"/>
 <akt:worksInProject rdf:resource="akt:X-Media"/>
 <akt:hasPrettyName>Enrico Motta</akt:hasPrettyName>
 <akt:hasPostalAddress rdf:resource="akt:KmiPostalAddress"/
 <akt:hasEmailAddress>e.motta@open.ac.uk</akt:hasEmailAddre
 <akt:hasHomePage
   rdf:resource="http://kmi.open.ac.uk/people/motta/"/>
</akt:Person>
```

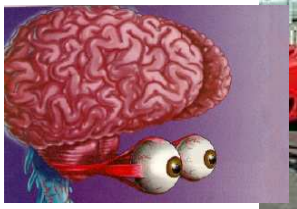
Agents on the SW

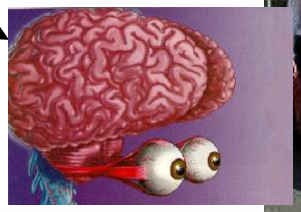
Please get me an appointment with a dealer within 50 miles of my home to arrange a test drive of a Ferrari F430 Spider for Saturday morning.



Enrico's Semantic Agent



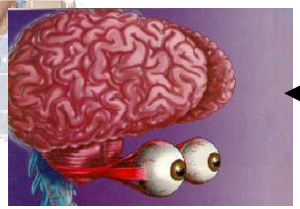


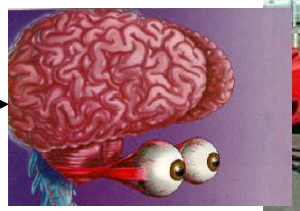


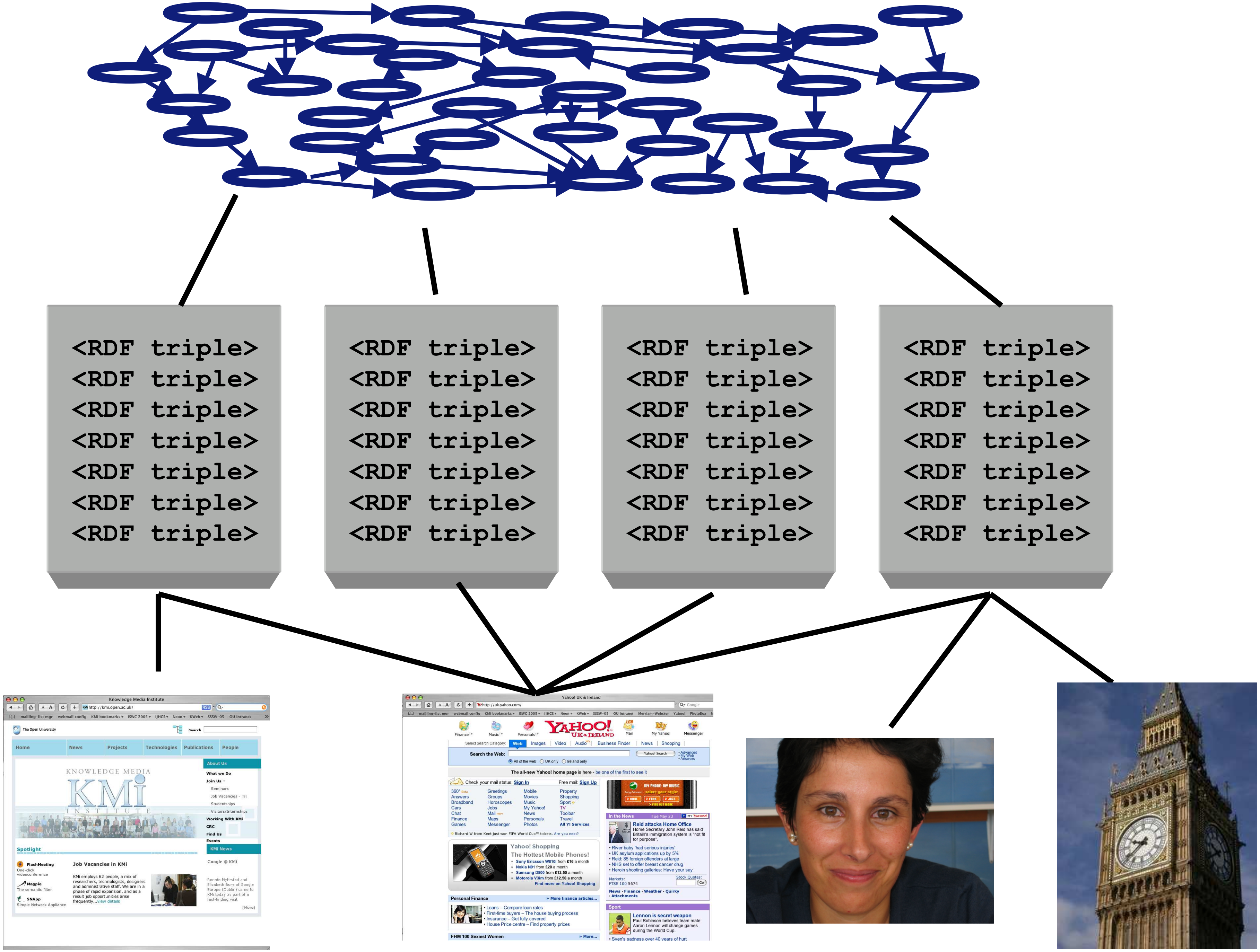
Conceptual Interoperability

Car-Dealership
hasAddress
hasWebAddress...
….
Schedule…..

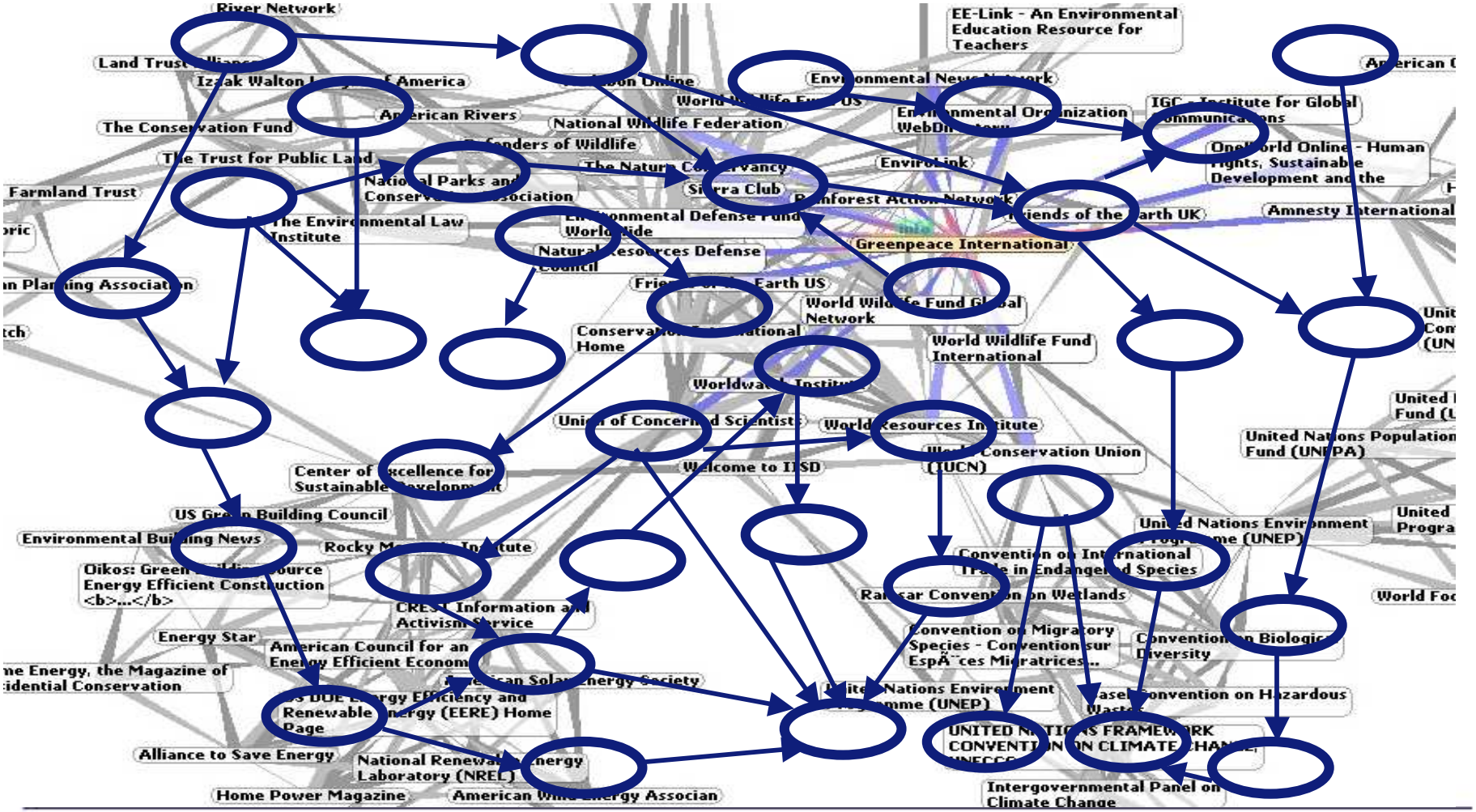
Car-Dealership
hasAddress
hasWebAddress...
….
Schedule…..





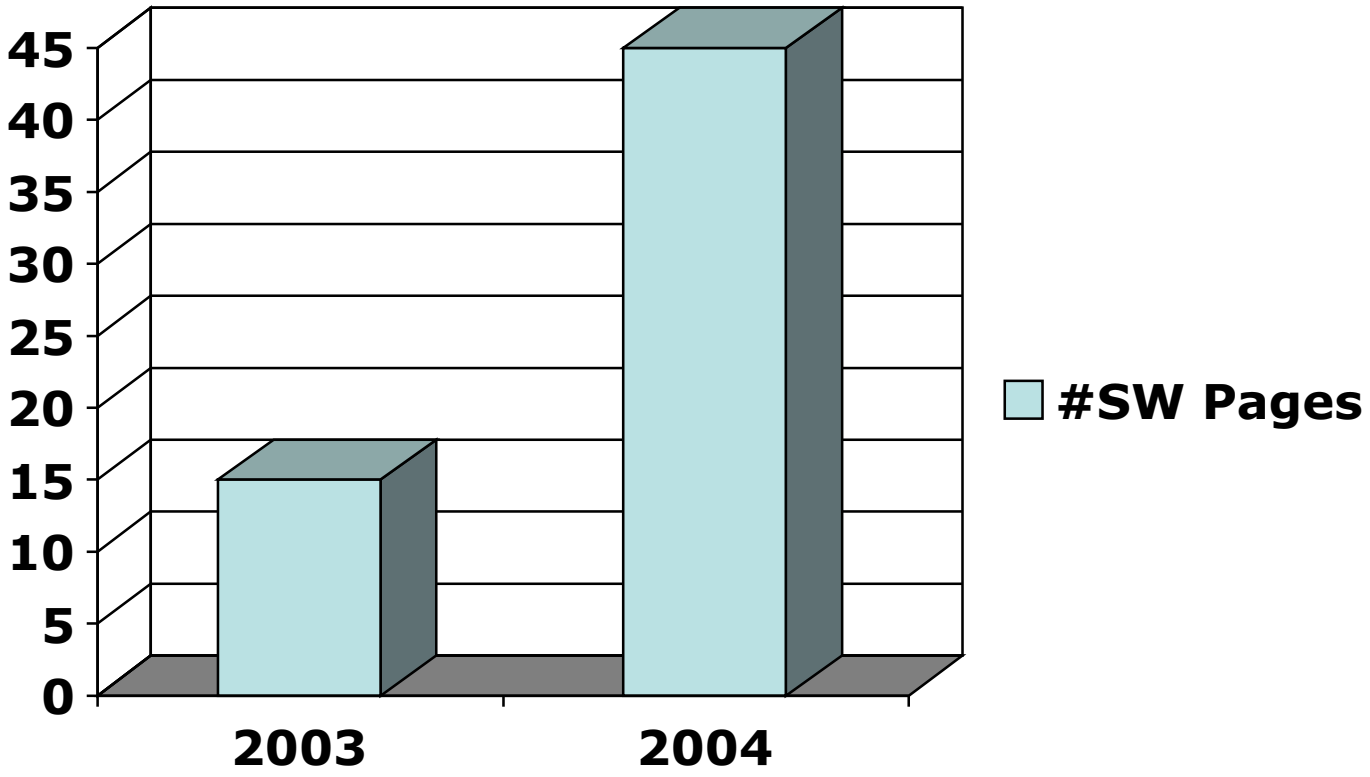

<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
KNOWLEDGE MEDIA
KMi
INSTITUTE
YAHOO!
UK & IRELAND

Key Aspect of SW #1: Hugeness



The Open University

Growth of the SW



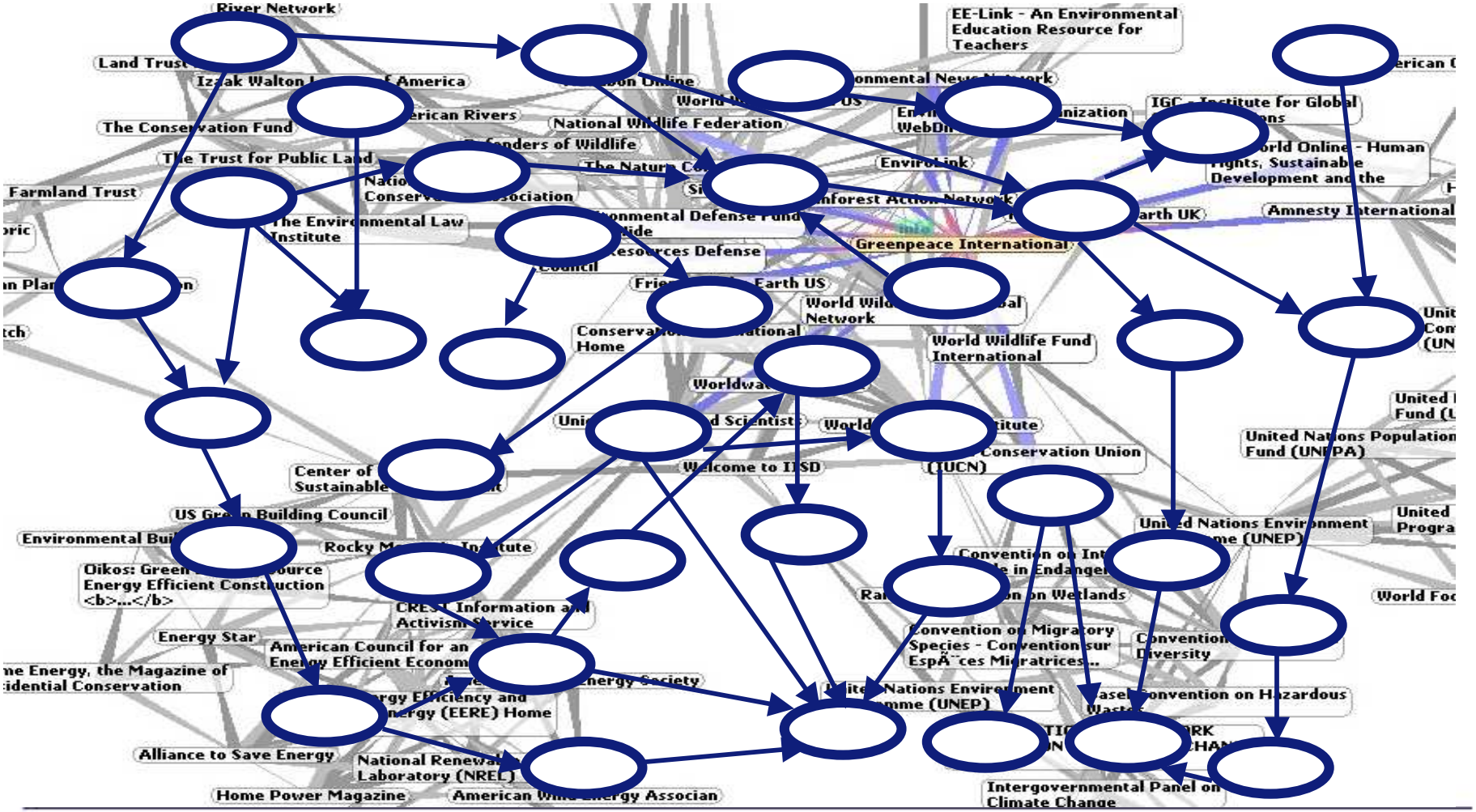
Swoogle

semantic web search 2006

Swoogle Today

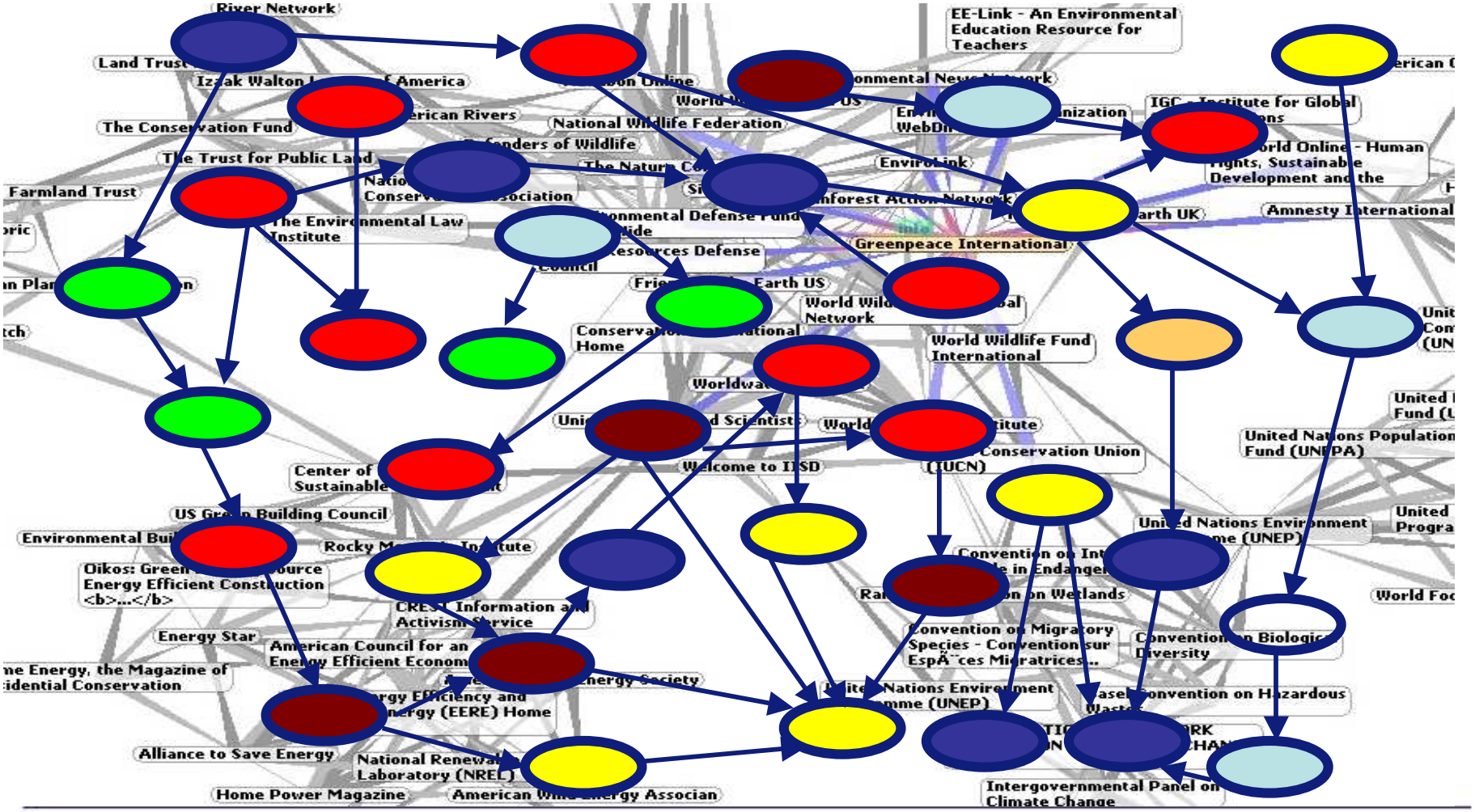
admin_dt	2006-06-27 23:58:30	Datetime Watched
url_total	3,946,561	Number of URLs being discovered
url_pinged	2,448,132	Number of URLs being pinged
total_swd	1,570,779	Number of Semantic Web Documents (regardless of embedded or containing some errors) be comfirmed.
total_swd_strict	1,003,095	Number of error-free pure Sematic Web Documents
total_swd_embed	407,408	Number of documents (except SWDs, PDF, and JPEG) embedding Semantic Web Data
triple_total	299,824,429	Number of triples could be parsed from all Semantic Web Documents.

Key Aspect of SW #2: Heterogeneity



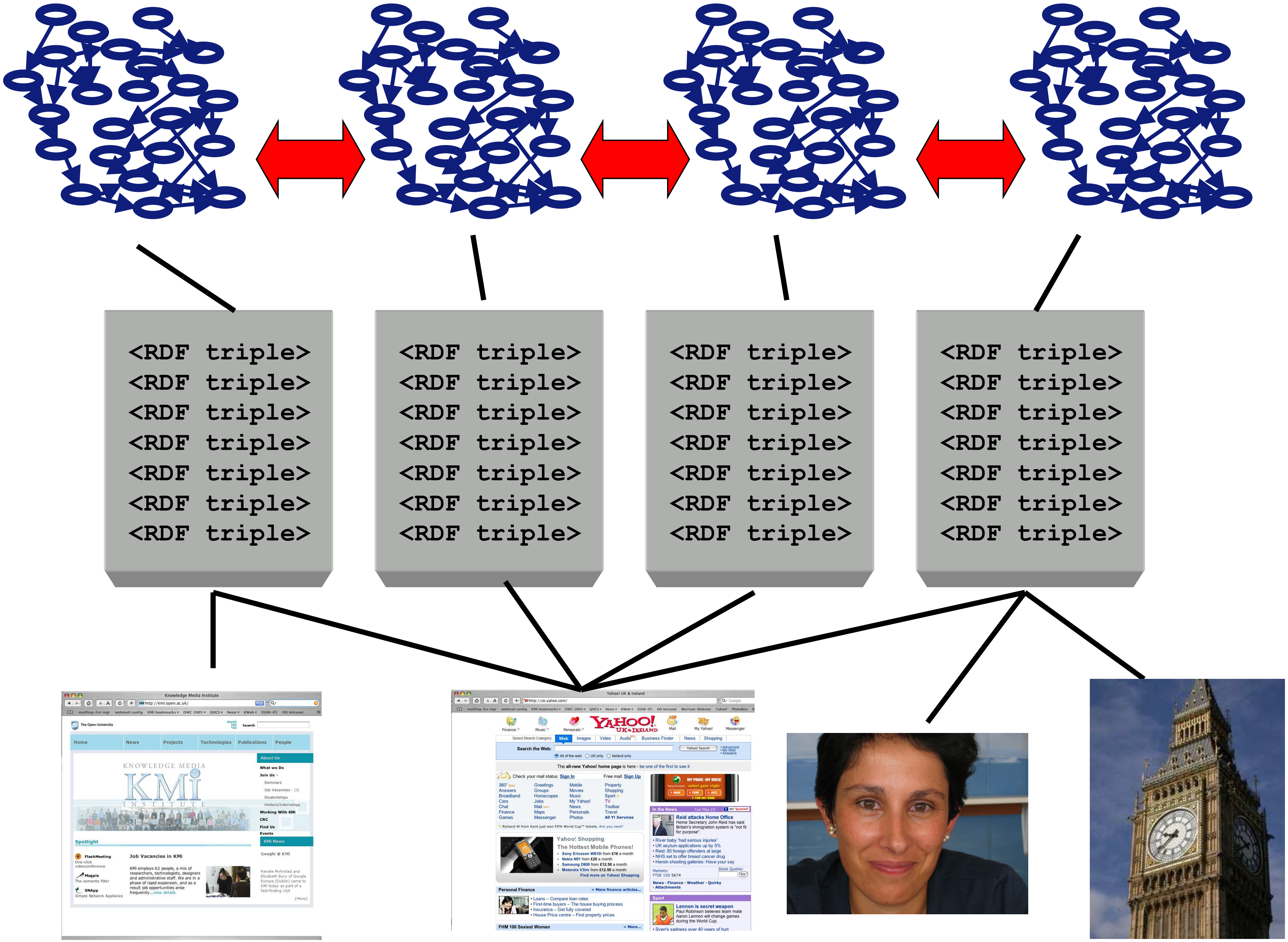
The Open University

Key Aspect of SW #2: Heterogeneity



The Open University


Swoogle
semantic web search
2006
ontology
document
term
more >>
Swoogle Search
Searching over 10,000 ontologies
news o faq o feedback o web-service o submit-url o sw-archive o swoogle-2005
Swoogle © 2004-2006, ebiquity group at UMBC
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License.


<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
<RDF triple>
Knowledge Media Institute
KMi
YAHOO! UK & IRELAND

Other key aspects of the SW

- Hugeness
  - Sem. markup of the same order of magnitude as the web
- Conceptual Heterogeneity
  - Sem. markup based on many different ontologies
- Very high rate of change
  - Semantic data generated all the time from web resources
- Heterogeneous Provenance
  - Markup generated from a huge variety of different sources, by human and artificial agents
- Various and subjective degrees of trust
  - Al-Jazeera vs CNN….
- Various degrees of data quality
  - No guarantee of correctness
- Intelligence a by-product of size and heterogeneity
  - rather than a by-product of sophisticated problem solving

Compare with traditional KBS

- Hugeness
  - KBS normally small to medium size
- Conceptual Heterogeneity
  - KBS normally based on a single conceptual model
- Very high rate of change
  - Change rate under developers' control (hence, low)
- Heterogeneous Provenance
  - KBS are normally created ad hoc for an application by a centralised team of developers
- Various and subjective degrees of trust
  - Centralisation of process implies no significant trust issues
- Various degrees of data quality
  - Centralisation guarantees data quality across the board
- Intelligence a by-product of size and heterogeneity
  - In KBS a by-product of complex, task-centric reasoning

Analysis of SW Applications

The Open University

Requirements for SW Applications

- Hugeness
  - SW applications should operate at scale
- Heterogeneity
  - SW applications should be able to handle multiple ontologies
- Very high rate of change
  - SW applications need to be open with respect to semantic resources
- Heterogeneous provenance
  - SW applications need to be open with respect to web resources

Additional Requirements

- SW is an extension of the web, so it makes sense to require that SW applications be compliant with key current web trends
  - Web 2.0 - i.e., providing interactive feature for *harnessing collective intelligence* (O'Reilly)
  - Web Services
    - Obviously it is also desirable that SW applications are also open with respect to web functionalities

Framework for characterizing SW applications

- Does app operate at scale?
- Can it handle multiple ontologies?
- Is it open to semantic resources?
- Is it open to web resources?
- Is it open to web services?
- Does it include Web 2.0 like features?

Applying the framework to six SW applications

CS AKTive Space, FLINK, Magpie, PiggyBank, AquaLog, PowerAqua

The Open University

AKT CS AKTive Space

Take a tour

About this page | research area/region | region/research area

Research area

- information interfaces and presentation
- information systems applications
- information storage and retrieval
- database management
- general

Computing Methodologies

- document and text processing
- simulation and modeling
- pattern recognition
- image processing and computer vision
- computer graphics
- artificial intelligence
- symbolic and algebraic manipulation
- general

Computer Applications

- computers in other systems
- computer-aided engineering
- arts and humanities
- social and behavioral sciences
- life and medical sciences

Radial: 100 miles

Map: uk-political

Researcher

Top 5 | 10 | 20 | unlimited

Order by Grant total | RAE result

- NR Shadbolt
- LA Carr
- DC De Roure
- NR Jennings
- L Moreau

Overview: NR Shadbolt

browse

Name NR Shadbolt

Institution Intelligence, Agents and Multimedia, University of Southampton

Email nrs@ecs.soton.ac.uk

Tel +442380597682

Fax +442380592865

Research

- Fluid Dynamics
- Aerodynamics
- Design and Testing Technology
- Biological Sciences Domain
- Image and Vision Computing
- Networks and Distributed Systems

CoP

- Wendy Hall
- DH Sleeman
- DR Robertson
- Stephen Harris
- Hugh Glaser
- M Eisenstadt
- mkw
- E Motta
- Kieron O'Hara
- W Hall
- A Tate
- Ian Millard
- Les Carr
- Y Wilks

AKT CS AKTive Space

Take a tour

About this page | research area/region | region/research area

Research area

Radial: 100 miles

- information interfaces and presentation
- information systems applications
- information storage and retrieval
- database management
- general

Computing Methodologies

- document and text processing
- simulation and modeling
- pattern recognition
- image processing and computer vision
- computer graphics
- artificial intelligence
- symbolic and algebraic manipulation
- general

Computer Applications

- computers in other systems
- computer-aided engineering
- arts and humanities
- social and behavioral sciences
- life and medical sciences

Overview: NR Shadbolt

browse

Name NR Shadbolt
Institution Intelligence, Agents and Multimedia,
Email nrs@ecs.soton.ac.uk
Tel +442380597682
Fax +442380592865
Research Fluid Dynamics
Aerodynamics
Design and Testing Technology
Biological Sciences Domain
Image and Vision Computing
Networks and Distributed Systems

A Tate
Ian Millard
Les Carr
Y Wilks

AKT Triplestore Browser

about | browse | manage | query | demo

E Motta

http://194.66.183.26/WEBSITE/GOW/ViewPerson.aspx?Person=34772
http://www.hero.ac.uk/rae/#Id-181275

family name	Motta
full name	E Motta
has appellation	Dr
has email address	e.motta@open.ac.uk Send email
has fax number	+441908653169 Make fax call
has telephone number	+441908653506 Make telephone call
involves person <—	KMi Planet : New Professorship for KMi
has author <—	MyPlanet
	Aqua
	WebOnto
	Magpie
	Semantic Annotation with MnM
	Internet Reasoning Service
	Ontology-Driven Document Enrichment Principles, Tools and Applications
	Solving VT in VITAL: A Study in Model Construction and Knowledge Reuse
	UPML: A Framework for knowledge system reuse
	A library of problem solving components based on the integration of the search
has project member <—	AN ONTOLOGY-BASED ENVIRONMENT FOR MANAGING DISTRIBUTED R
	INTERDISCIPLINARY RESEARCH COLLABORATION IN ADVANCED KNO
has research interest	Artificial Intelligence Technologies
	Information and Knowledge Management
	Education
works for	The Open University
works in unit	Knowledge Media Institute
	Educational Technology
	Dev. & Use of KM in Learning & Teaching
sameAs	E Motta

CS Aktive Space (2003)

Type	Aggregation and visualization of data from multiple sources
Operates at scale?	Yes, large numbers of data crawled from hundreds of different UK CS sites
Multi-ontology?	All data extracted and integrated into the AKT reference ontology
Open to semantic resources?	No, RDF data are generated by the system, rather than reused from existing repositories
Open to web resources?	No (it is not possible to indicate more sites to the system and expect it to add more data)
Open to web services?	No (there is no open architecture to add crawlers)
Web 2.0 like?	No (no tagging or interactive features)


W3C Technical Reports and Publications - Microsoft Internet Explorer provided by The Open University V 6.0 sp1
File Edit View Favorites Tools Help
Back Search Favorites
Address http://www.w3.org/TR/
Magpie Technologies Research areas Generic CS topics Community
magpie
The Semantic Filter
OWL Web Ontology Language Semantics and Abstract Syntax
10 February 2004, Patrick Hayes, Ian Horrocks, Peter F. Patel-Schneider - (Errata, Translations)
OWL Web Ontology Language Test Cases
10 February 2004, Jeremy J. Carroll, Jos De Roo - (Errata, Translations)
Resource Description Framework (RDF): Concepts and Abstract Syntax
10 February 2004, Graham Klyne, Jeremy J. Carroll - (Errata, Translations)
RDF Semantics
10 February 2004, Patrick Hayes - (Errata, Translations)
RDF Primer
10 February 2004, Frank Manola, Eric Miller - (Errata, Translations)
This specification supersedes Resource Description Framework (RDF) Model and Syntax Specification published on 22 February 1999
RDF Vocabulary Description Language 1.0: RDF Schema
10 February 2004, Dan Brickley, Ramanathan V. Guha - (Errata, Translations)
RDF/XML Syntax Specification (Revised)
10 February 2004, Dave Beckett - (Errata, Translations)
RDF Test Cases
10 February 2004, Jan Grant, Dave Beckett - (Errata, Translations)
OWL Web Ontology Language Use Cases and Requirements
10 February 2004, Jeff Hefl
Explain concept
Find learning resources in REASE
Find in ACM digital Library
People active in...
Namespaces in XML 1.1
4 February 2004, Andrew L
ard Tobin, Tim Bray - (Errata, Translations)
Extensible Markup Language
4 February 2004, Eve Maler, John Cowan, Jean Paoli, C. M. Sperberg-McQueen, François Yergeau, Tim Bray - (Errata, Translations)
Document Object Model (DOM) Level 3 Validation Specification
27 January 2004, Ben Chang, Rezaur Rahman, Joe Kesselman - (Errata, Translations)
Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0
http://www.w3.org/TR/2004/REC-webont-req-20040210/
Internet

Magpie (2003)

Type	Semantic Web Browser
Operates at scale?	Yes, large numbers of data crawled from publication archives, google, FOAF, etc..
Multi-ontology?	Partially. Can switch from one ontology to another, but only one ontology can be used at the time.
Open to semantic resources?	Yes
Open to web resources?	Yes (but quality can degrade as you move away from resources relevant to the current ontology)
Open to web services?	Yes
Web 2.0 like?	No (no tagging or interactive features)

FLINK
The Who is Who of the Semantic Web

home | network | ontology | cluster | world | stats | feedback | about

Semantic Web Challenge
First Place
2004

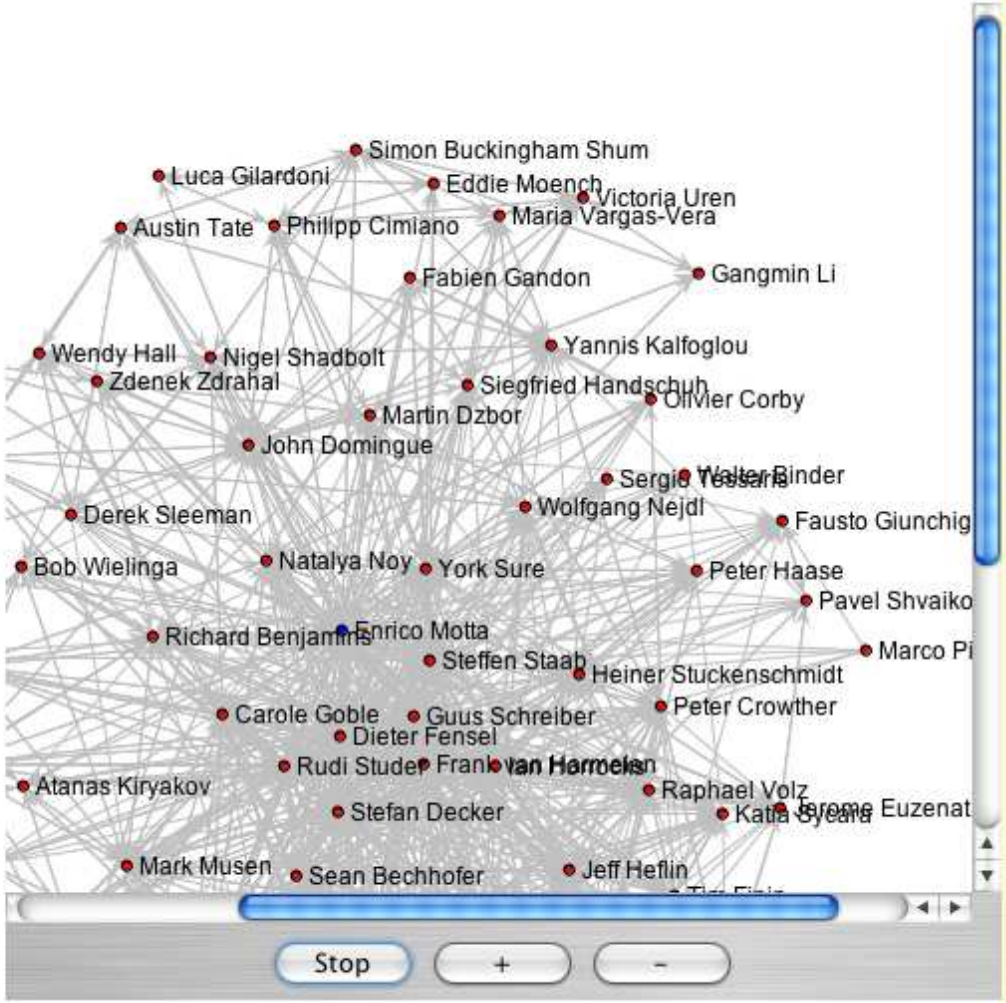
RDF Metadata

openRDF.org

Sesame

Elmo

Social Network of Enrico Motta



Tip: You can drag nodes with your mouse. You can also use the scrollbars to pan the image. You can zoom in and out using the + / - buttons, but also with the roller

Statistics

Indegree	67.0 (41.0) ↑
Closeness	0.5 (0.43) ↑
Betweenness	8212.68 (4982.34) ↑
Top Publications	53.0 (40.0) ↑
Impact	11.48 (9.21) ↑

What do the numbers mean?

Rankings

Indegree	11 (10) ↓
Closeness	155 (7) ↓
Betweenness	10 (10) -
Top Publications	13 (15) ↑
Impact	193 (177) ↓

Links

Search Google!
Search Foafnaut!
Search Citeseer!
Search A9 (Amazon)!

Got FOAF?

FLINK (2004)

Type	Aggregation and visualization of data from multiple sources
Operates at scale?	Yes, large numbers of data crawled from publication archives, google, FOAF, etc..
Multi-ontology?	No. All data extracted and integrated into a single ontology
Open to semantic resources?	No, RDF data are generated by the system, rather than reused from existing repositories
Open to web resources?	No (it is not possible to indicate more sites to the system and expect it to add more data)
Open to web services?	No
Web 2.0 like?	No (no tagging or interactive features)


Combined Information: restaurants + movies - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://127.0.0.1:1978/piggy-bank/model1116627657351?command=
Go
[My Piggy Bank] | Combined Information
PiggyBank
restaurants + movies
Combined Information
1 filter criterion
type: Show (remove) Restaurant (remove)
Order
Commands
View items as list
Map - Satellite New!
1 item(s)
Abacus Cuisine Of China
Markham
Toronto
Mississauga
Map data ©2005 NAVTEQ™, Tele Atlas
Show
Restaurant
Type here to search
cuisine
Type here to filter
"Asian" (28)
"Californian" (1)
"Canadian" (2)
"Chinese" (5)
"Desserts" (1)
"Dim Sum" (4)
date
By hour
Type here to filter
10:00 PM (138)
11:00 PM (20)
type
Type here to filter
Restaurant (28)
Show (158)
address
Type here to filter
Transferring data from mt.google.com...

PiggyBank (2005)

Type	Semantic Web Browser
Operates at scale?	Yes, data can be collected from of semantic and non-semantic sources
Multi-ontology?	Data can be brought in from different ontologies, unclear whether intg. support is provided
Open to semantic resources?	Yes
Open to web resources?	Yes (open to screen scraping mechanisms)
Open to web services?	Yes (open to screen scraping mechanisms)
Web 2.0 like?	Yes, supports tagging and sharing of bookmarks

The answer to the question:

peter-scott has-research-interest semantic-web-area

The value of has-web-address for peter-scott is/are:

"news.kmi.open.ac.uk/peterblog"

<< BACK

Powered by

AquaLog (2004)

Type	Question Answering System
Operates at scale?	Yes
Multi-ontology?	Partially. Can switch from one ontology to another with zero configuration effort, but only one ontology can be used at the time.
Open to semantic resources?	Yes
Open to web resources?	No
Open to web services?	No
Web 2.0 like?	Yes. No tagging, but learning mechanism supports mapping user terminologies to ontologies

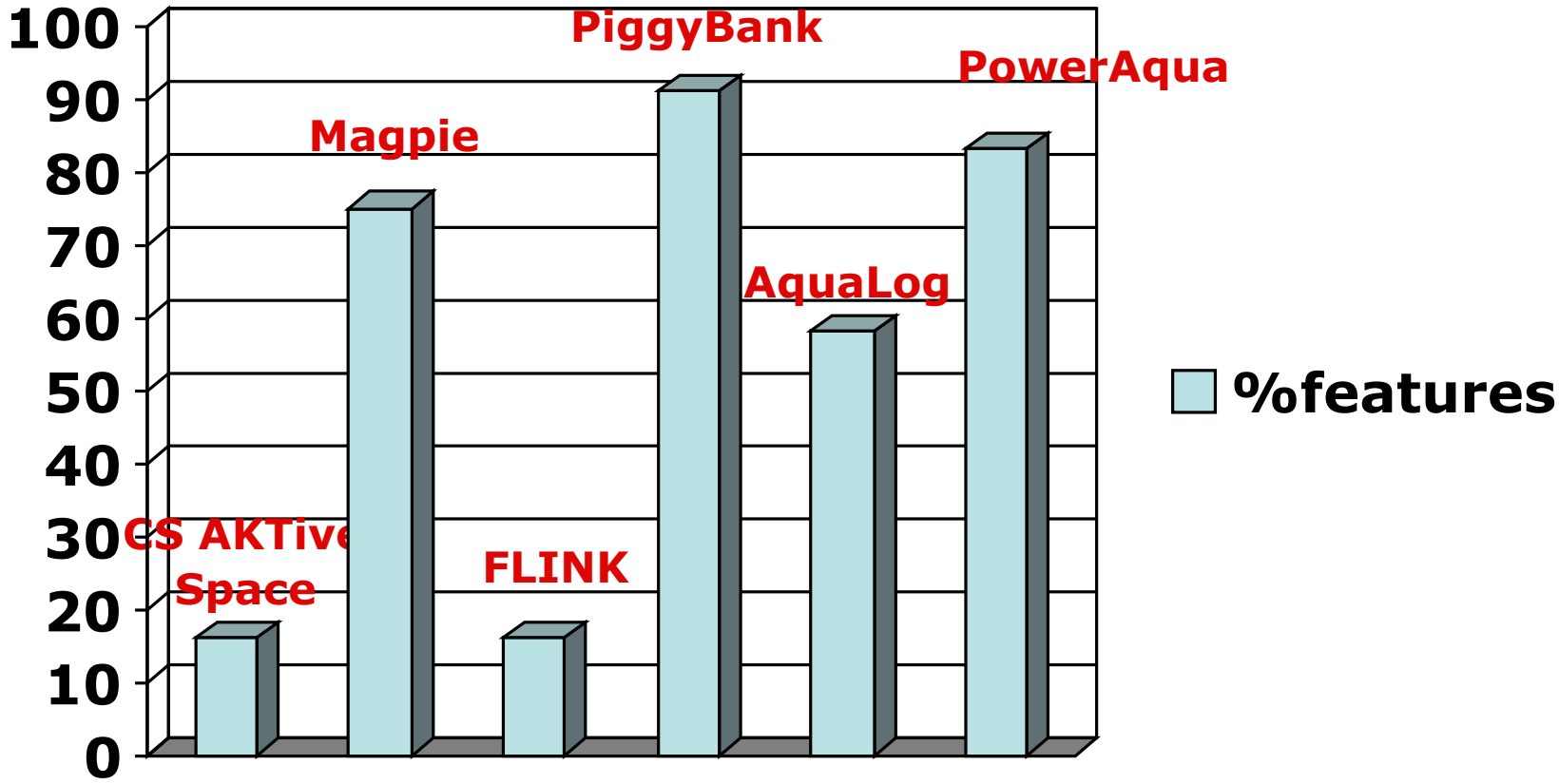
PowerAqua (2006)

Type	Question Answering System
Operates at scale?	Yes
Multi-ontology?	Yes
Open to semantic resources?	Yes
Open to web resources?	No
Open to web services?	Yes
Web 2.0 like?	Yes. No tagging, but learning mechanism supports mapping user terminologies to ontologies

Summary

Operates at scale?	All	100%
Multi-ontology?	PowerAqua, Magpie and AquaLog (partially), PiggyBank (unclear)	40%
Open to semantic resources?	PowerAqua, Magpie, AquaLog, PiggyBank	66%
Open to web resources?	PiggyBank, Magpie	33%
Open to web services?	PiggyBank, Magpie, PowerAqua	50%
Web 2.0 like?	PiggyBank, AquaLog, PowerAqua	50%

Graphical View



Conclusions

- Even the earliest SW applications recognised scale as a key requirement to address
- Semantic portals more similar to large scale KBs, than to our blueprint for SW applications
- The heterogeneous nature of the SW more and more taken into account by SW applications
- Overall trend is positive
  - Latest tools more closely address our requirements
- Automatic data acquisition remains the feature most often missing from SW applications
  - However, it may matter less and less…..


KNOWLEDGE MEDIA
KMi
INSTITUTE