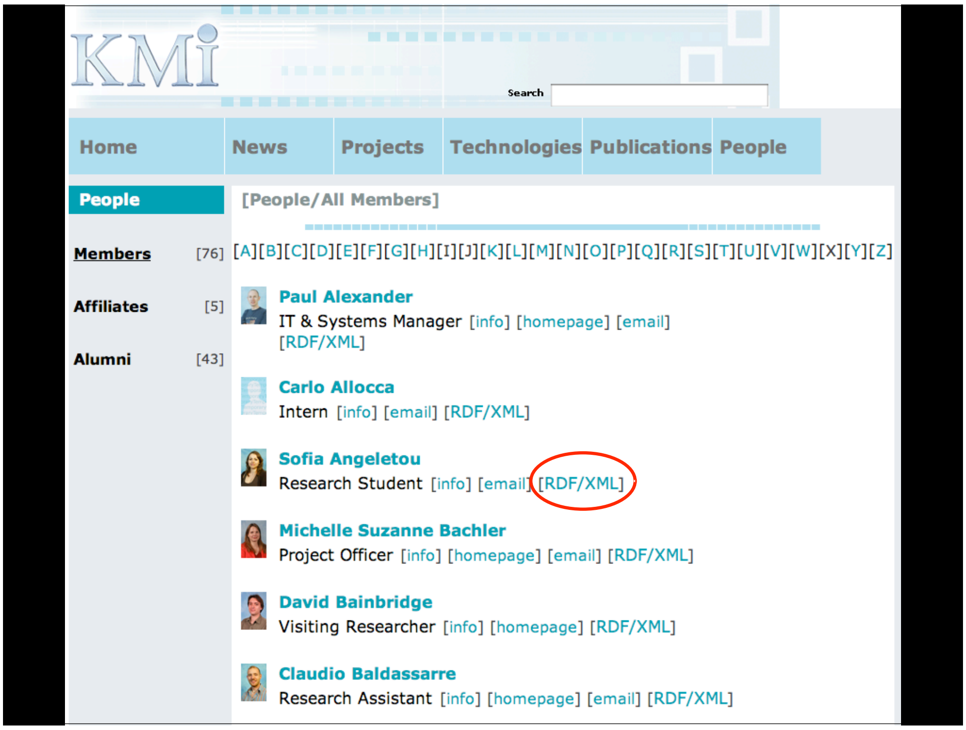
Semantic Web Applications

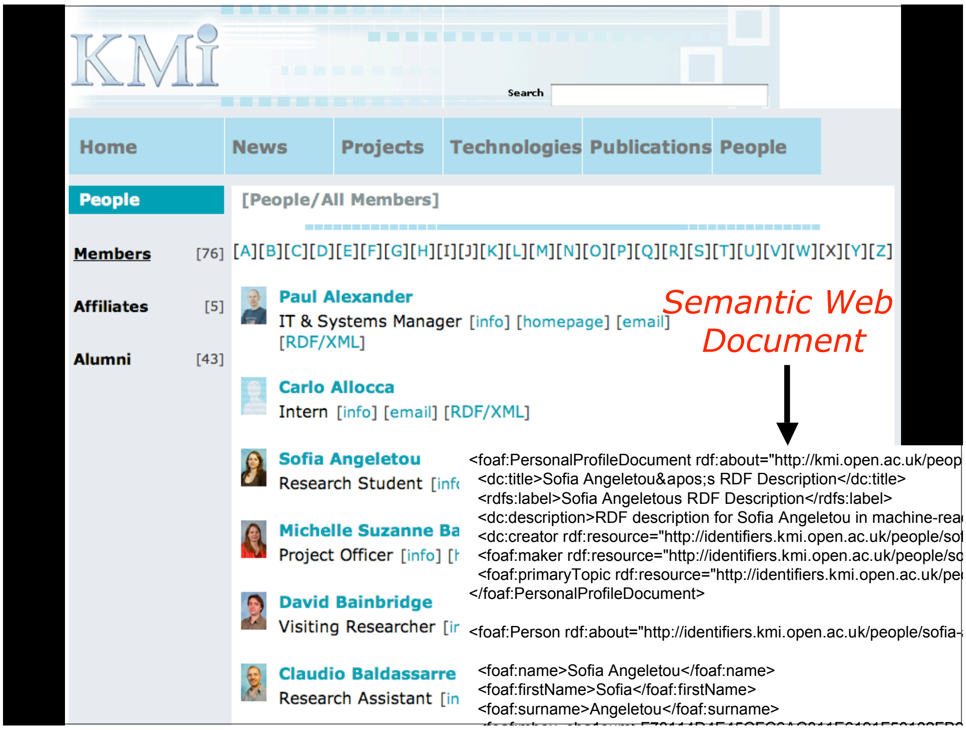
Prof. Enrico Motta, *PhD*

Knowledge Media Institute
The Open University
Milton Keynes, UK

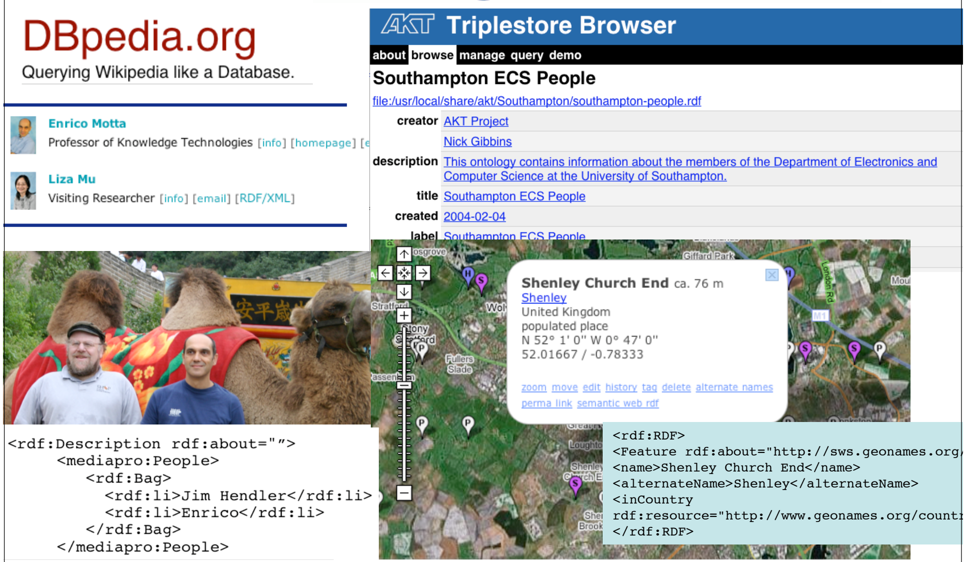
The Semantic Web

The collection of all formal, machine processable, web accessible, ontology-based statements (semantic metadata) about web resources and other entities in the world, expressed in a knowledge representation language based on an XML syntax (e.g., OWL, DAML, DAML+OIL, RDF, etc…)

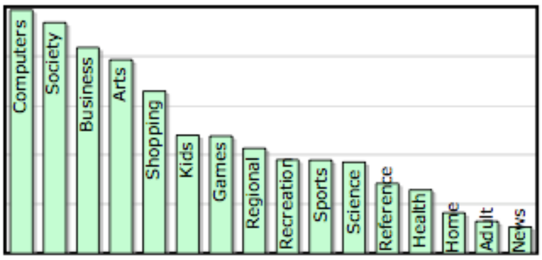

KMi
Search
Home
News
Projects
Technologies
Publications
People
People
[People/All Members]
Members [76]
Affiliates [5]
Alumni [43]
[A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][Q][R][S][T][U][V][W][X][Y][Z]
Paul Alexander
IT & Systems Manager [info] [homepage] [email] [RDF/XML]
Carlo Allocca
Intern [info] [email] [RDF/XML]
Sofia Angeletou
Research Student [info] [email] [RDF/XML]
Michelle Suzanne Bachler
Project Officer [info] [homepage] [email] [RDF/XML]
David Bainbridge
Visiting Researcher [info] [homepage] [RDF/XML]
Claudio Baldassarre
Research Assistant [info] [homepage] [email] [RDF/XML]


KMi
Search
Home
News
Projects
Technologies
Publications
People
People
[People/All Members]
Members [76]
Affiliates [5]
Alumni [43]
[A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][Q][R][S][T][U][V][W][X][Y][Z]
Paul Alexander
IT & Systems Manager [info] [homepage] [email] [RDF/XML]
Carlo Allocca
Intern [info] [email] [RDF/XML]
Semantic Web Document
<foaf:PersonalProfileDocument rdf:about="http://kmi.open.ac.uk/peop
<dc:title>Sofia Angeletou's RDF Description</dc:title>
<rdfs:label>Sofia Angeletous RDF Description</rdfs:label>
<dc:description>RDF description for Sofia Angeletou in machine-rea
<dc:creator rdf:resource="http://identifiers.kmi.open.ac.uk/people/so
<foaf:maker rdf:resource="http://identifiers.kmi.open.ac.uk/people/so
<foaf:primaryTopic rdf:resource="http://identifiers.kmi.open.ac.uk/pe
</foaf:PersonalProfileDocument>
<foaf:Person rdf:about="http://identifiers.kmi.open.ac.uk/people/sofia-
<foaf:name>Sofia Angeletou</foaf:name>
<foaf:firstName>Sofia</foaf:firstName>
<foaf:surname>Angeletou</foaf:surname>

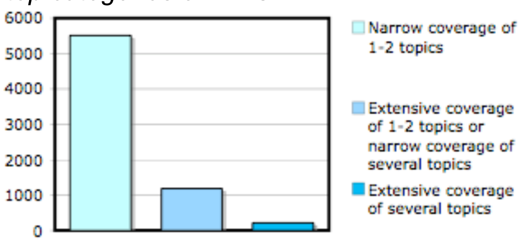
Semantic Content is Increasing….



Domain Coverage on the SW



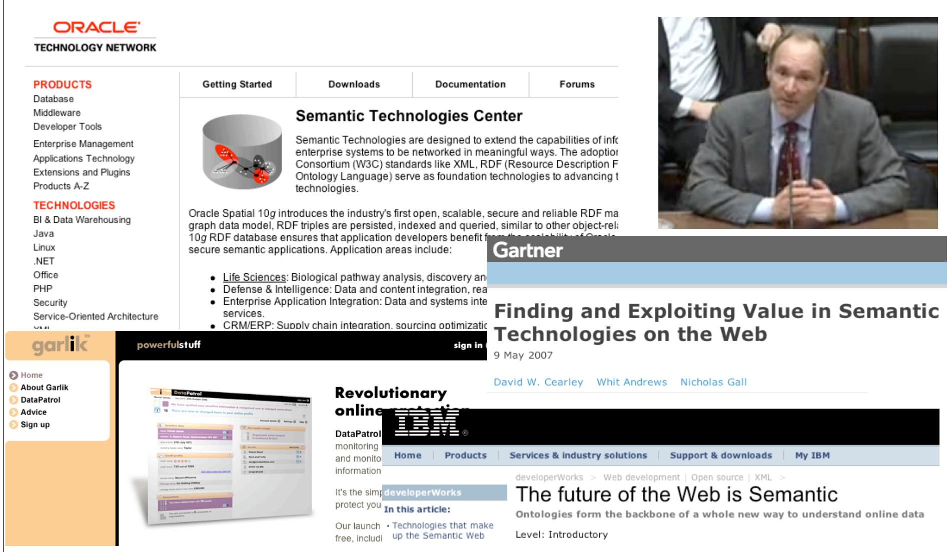
Distribution of documents in the 16 top categories of DMOZ



Distribution of the documents according to their coverage

- Great variety: Some topics are almost not covered (e.g. Health), while some are covered reasonably well (e.g. Society, Computers)
- As we can expect, a large number of narrow coverage documents and a small number of large coverage ones.

The Rise of Semantics



Key Points

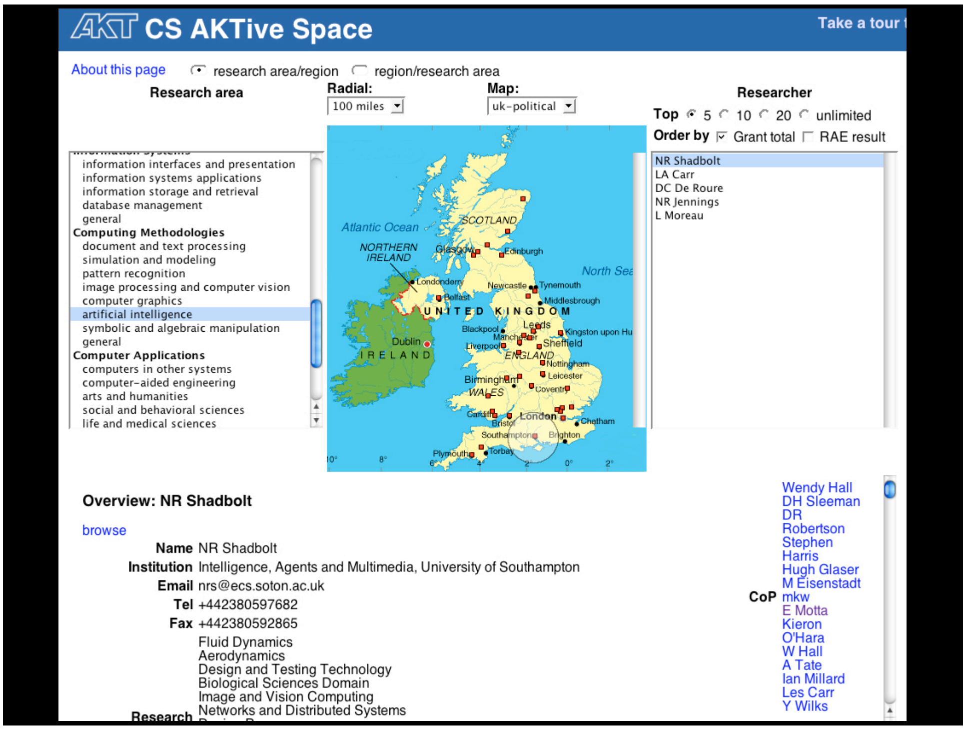
- The SW is less and less an aspiration and more and more a reality: there is already a large amount of semantic markup available on the web
- However the domain coverage of the semantic web is patchy. Some domains are well covered, others covered very little
- This emerging large scale semantics opens up new possibilities for building novel, knowledge-based applications
- In addition, it may also provide a solution to one of the holy grails of AI research: the availability of large-scale background knowledge to enable intelligent behaviour

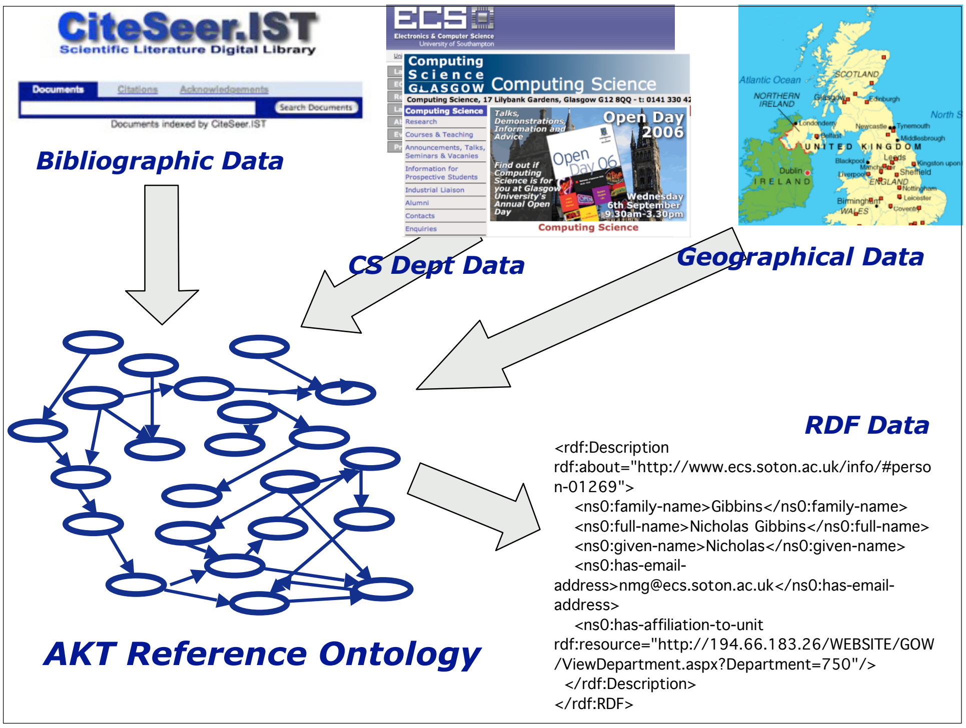
Semantic Web Applications

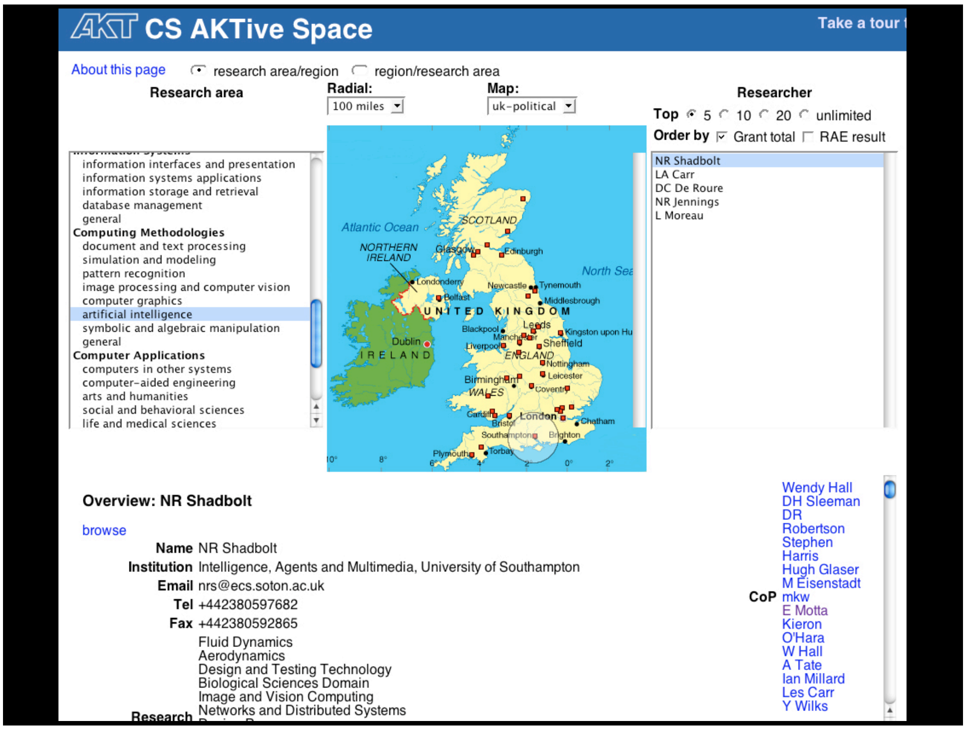
Application Types

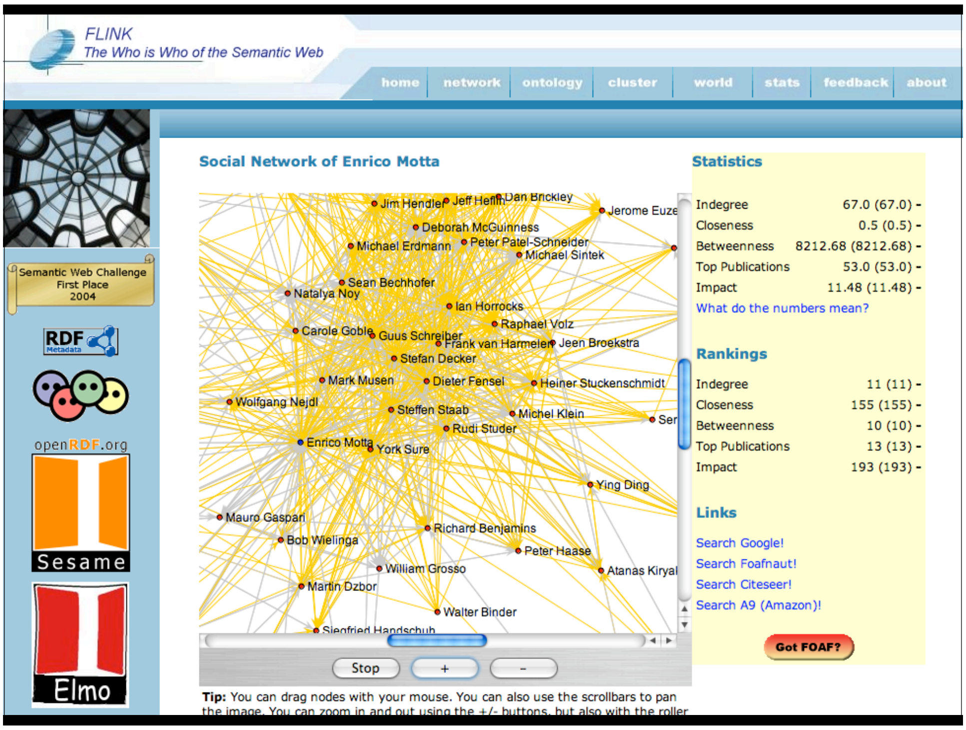
- Lots of different types…diversity is increasing…see “Intro to SW” talk…..
  - Enterprise Information Integration
  - Web 2.0 like
  - Devices
  - Agents
  - Etc…
- Here we focus on two broad classes of applications
  - (Semantically) Closed SW Applications *(1st generation)*
    - Typically in the domain of Enterprise Information Integration
  - (Semantically) Open SW Applications *(2nd generation)*

Closed SW Applications



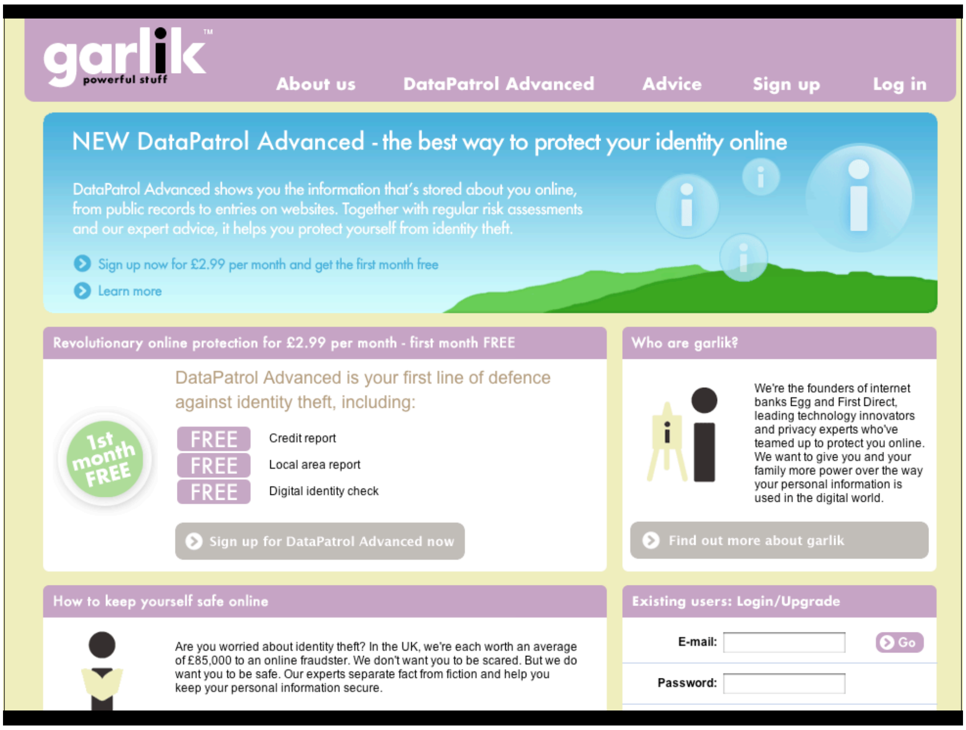

CiteSeer.IST
Scientific Literature Digital Library
Documents
Citations
Acknowledgements
Search Documents
Documents indexed by CiteSeer.IST
Bibliographic Data
ECS
Electronics & Computer Science
University of Southampton
Computing Science
Computing Science
Open Day 2006
Computing Science
CS Dept Data
Geographical Data
AKT Reference Ontology
RDF Data
<rdf:Description rdf:about="http://www.ecs.soton.ac.uk/info/#person-01269">
<ns0:family-name>Gibbins</ns0:family-name>
<ns0:full-name>Nicholas Gibbins</ns0:full-name>
<ns0:given-name>Nicholas</ns0:given-name>
<ns0:has-email-address>nmg@ecs.soton.ac.uk</ns0:has-email-address>
<ns0:has-affiliation-to-unit rdf:resource="http://194.66.183.26/WEBSITE/GOW/ViewDepartment.aspx?Department=750"/>
</rdf:Description>
</rdf:RDF>


AKT CS AKTive Space
Take a tour
About this page
research area/region
region/research area
Research area
Radial:
100 miles
Map:
uk-political
Researcher
Top 5 10 20 unlimited
Order by Grant total RAE result
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
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



Closed SW Applications

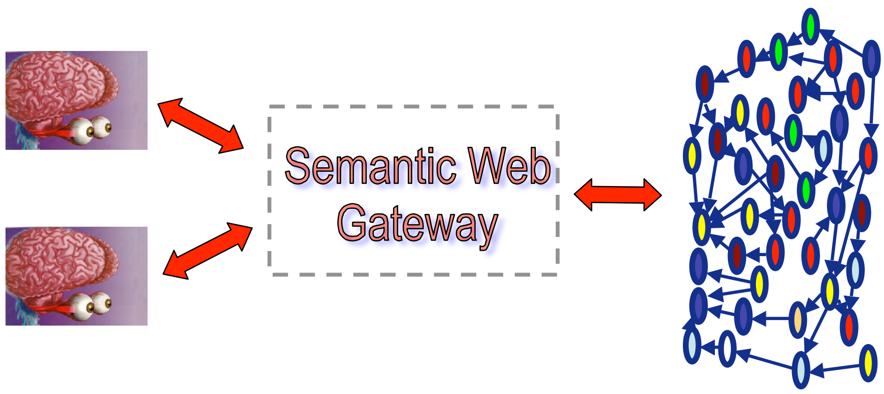
- A 'corporate ontology' is used to provide a homogeneous view over heterogeneous data sources.
- Limited use of existing SW data
  - Data are typically scraped from existing non-semantic sources
- Typically closed to additional (heterogenous) semantic resources
  - On-the-fly semantic data integration is usually not supported
- Often tackle Enterprise Information Integration scenarios
- Can be seen (to some extent) as the application of traditional knowledge engineering technologies in a web environment



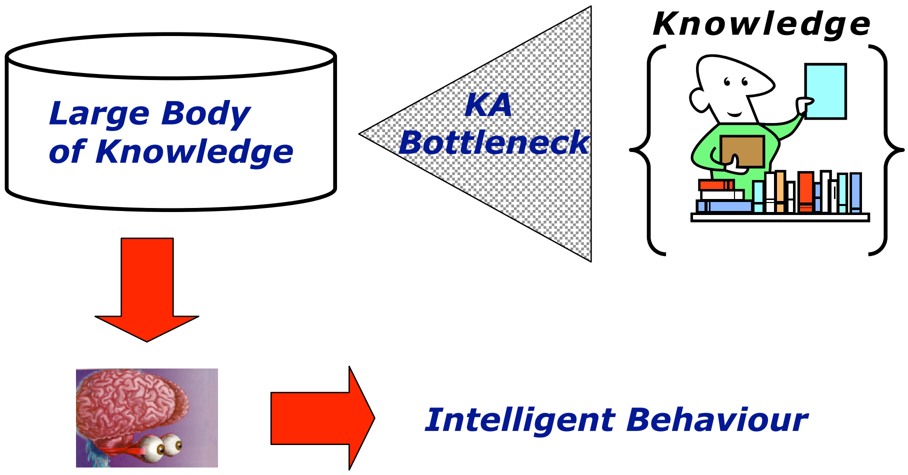
Open SW Applications

(2nd Generation SW Applications)

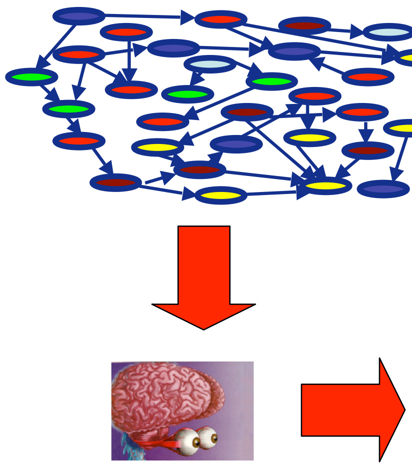
Architecture of OSW Apps



The Knowledge Acquisition Bottleneck

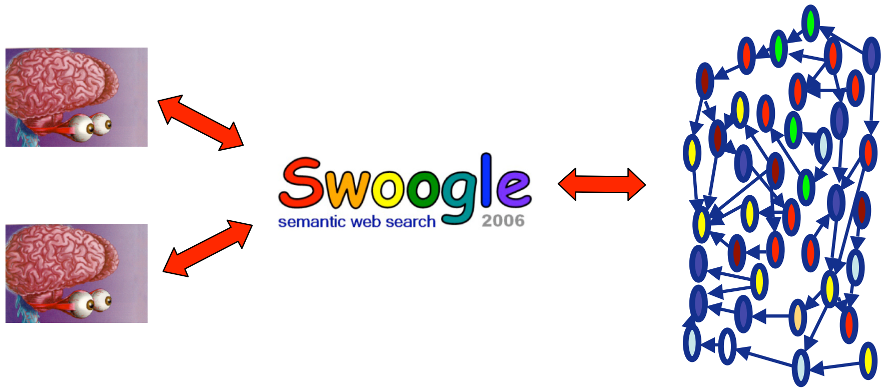


SW as Enabler of Intelligent Behaviour



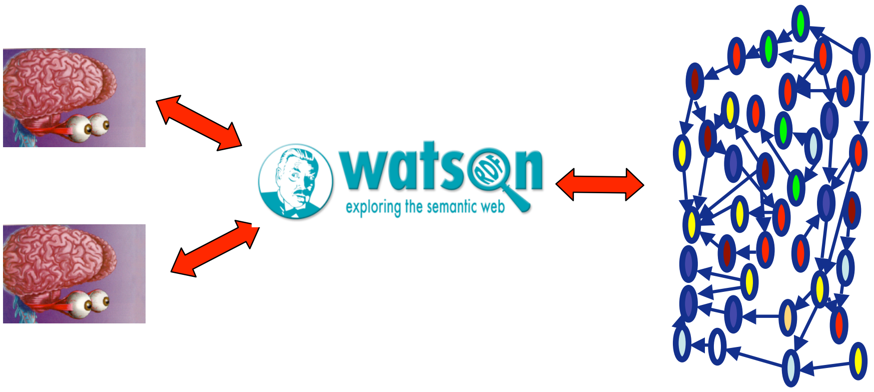
Intelligent Behaviour

Swoogle



http://swoogle.umbc.edu/

Watson

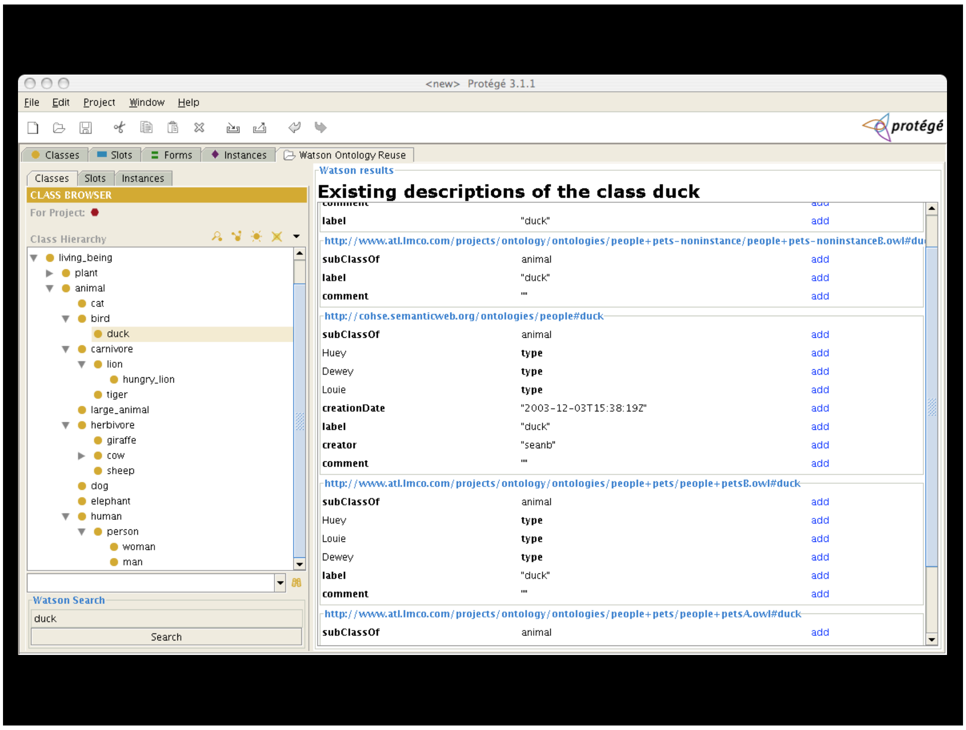


http://watson.kmi.open.ac.uk

watson

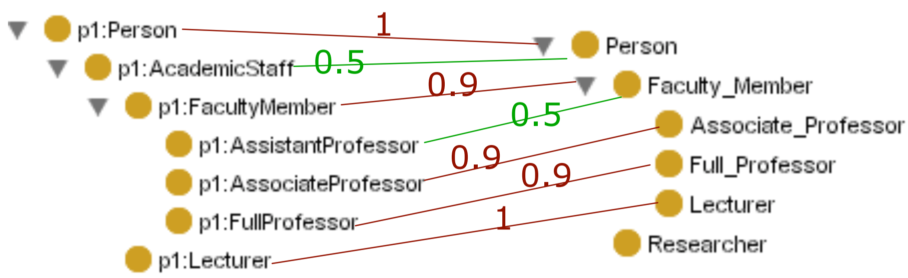
exploring the semantic web

- Sophisticated quality control mechanism
  - Detects duplications
  - Fixes obvious syntax problems
    - E.g., duplicated ontology IDs, namespaces, etc..
- Structures ontologies in a network
  - Using relations such as: *extends, inconsistentWith, duplicates*
- Provides efficient API
- Supports formal queries (SPARQL)
- Variety of ontology ranking mechanisms
- Modularization/Combination support
- Plug-ins for Protégé and NeOn Toolkit
- Very cool logo!



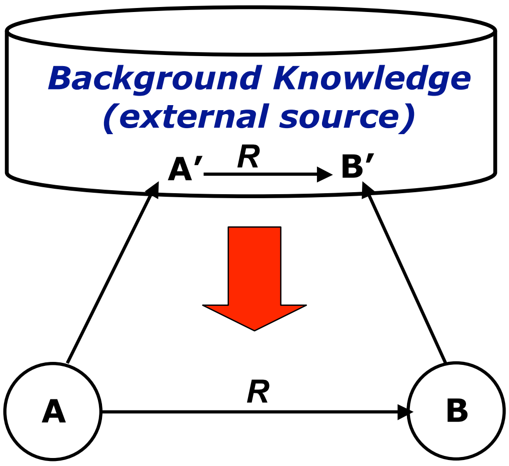
Example: Using the SW as Background Knowledge for Ontology Matching

Ontology Matching



- Label similarity methods
  - e.g., Full_Professor = FullProfessor
- Structure similarity methods
  - Using taxonomic/property related information

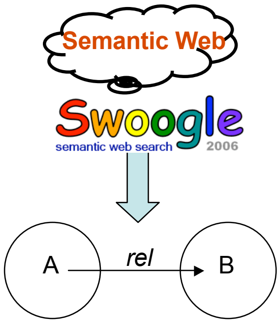
New paradigm: Use of Background Knowledge



External Source = SW

Proposal:

- rely on online ontologies (Semantic Web) to derive mappings
- ontologies are ***dynamically*** discovered and combined

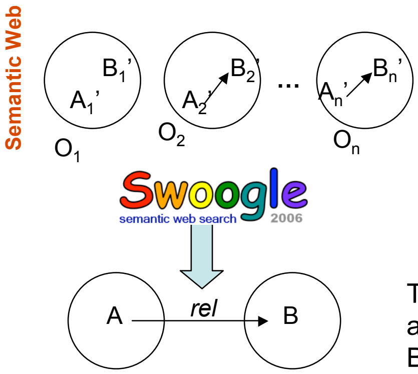


Does not rely on any pre-selected knowledge sources.

M. Sabou, M. d'Aquin, E. Motta, "*Using the Semantic Web as Background Knowledge in Ontology Mapping*", Ontology Mapping Workshop, ISWC'06. *Best Paper Award*

Strategy 1 - Definition

Find ontologies that contain equivalent classes for A and B and use their relationship in the ontologies to derive the mapping.



For each ontology use these rules:

$$A' \equiv B' \Rightarrow A \equiv B$$

$$A' \subseteq B' \Rightarrow A \subseteq B$$

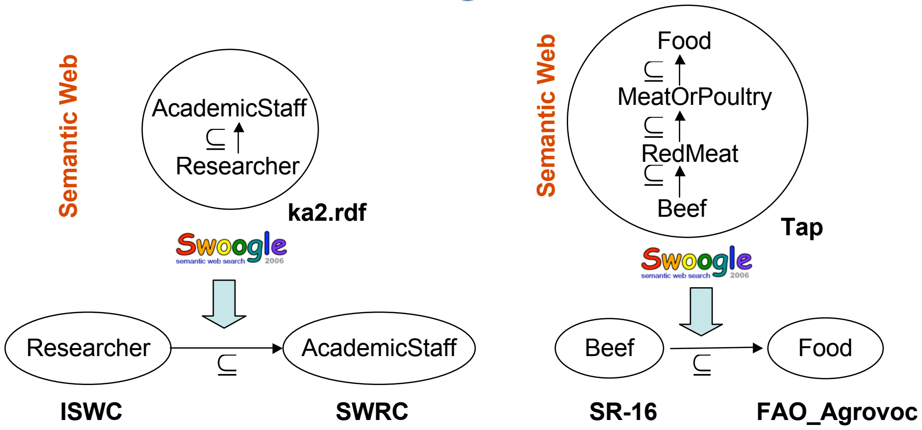
$$A' \supseteq B' \Rightarrow A \supseteq B$$

$$A' \perp B' \Rightarrow A \perp B$$

These rules can be extended to take into account indirect relations between A' and B', e.g., between parents of A' and B':

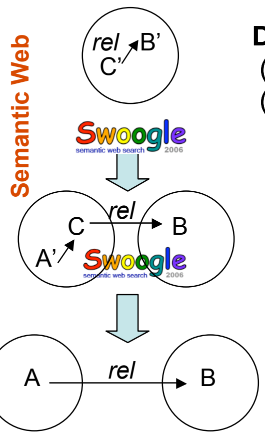
$$A' \subseteq C \wedge C \perp B' \Rightarrow A' \perp B'$$

Strategy 1- Examples



Strategy 2 - Definition

Principle: If no ontologies are found that contain the two terms then combine information from multiple ontologies to find a mapping.



Details:

(1) Select all ontologies containing A' equiv. with A

(2) For each ontology containing A':

(a) if $A' \subseteq C$ find relation between C and B.

(b) if $A' \supseteq C$ find relation between C and B.

$$(r1) A' \subseteq C \wedge C \subseteq B \Rightarrow A \subseteq B$$

$$(r2) A' \subseteq C \wedge C \equiv B \Rightarrow A \equiv B$$

$$(r3) A' \subseteq C \wedge C \perp B \Rightarrow A \perp B$$

$$(r4) A' \supseteq C \wedge C \supseteq B \Rightarrow A \supseteq B$$

$$(r5) A' \supseteq C \wedge C \equiv B \Rightarrow A \supseteq B$$

Strategy 2 - Examples

Ex1: $Chicken$ **Vs.** $Food$

$Chicken \subseteq Poultry$ **(midlevel-onto)**
$Poultry \subseteq Food$ **(Tap)**
(r1) ⇒ $Chicken \subseteq Food$

(Same results for Duck, Goose, Turkey)

Ex2: Ham **Vs.** $Food$

$Ham \subseteq Meat$ **(pizza-to-go)**
$Meat \subseteq Food$ **(SUMO)**
(r1) ⇒ $Ham \subseteq Food$

Ex3: Ham **Vs.** $Seafood$

$Ham \subseteq Meat$ **(pizza-to-go)**
$Meat \perp Seafood$ **(wine.owl)**
(r3) ⇒ $Ham \perp Seafood$

Large Scale Evaluation

Matching AGROVOC (16k terms) and NALT(41k terms)

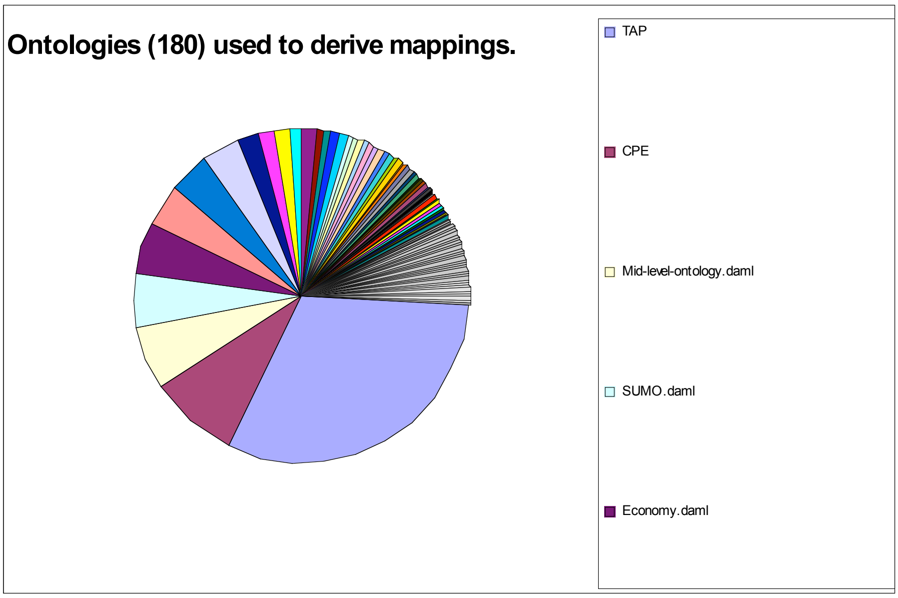
	Nr.	Examples
Subclass ($\xrightarrow{\sqsubseteq}$)	1477	$Lamb \xrightarrow{\sqsubseteq} Sheep$, $Soap \xrightarrow{\sqsubseteq} Detergent$, $Asbestos \xrightarrow{\sqsubseteq} Pollutant$ $Oasis \xrightarrow{\sqsubseteq} Ecosystem$, $RAM \xrightarrow{\sqsubseteq} ComputerEquipment$
SuperClass ($\xrightarrow{\sqsupseteq}$)	1857	$Shop \xrightarrow{\sqsupseteq} Supermarket$, $Spice \xrightarrow{\sqsupseteq} BlackPepper$, $Valley \xrightarrow{\sqsupseteq} Canyon$ $Infrastructure \xrightarrow{\sqsupseteq} Highway$, $Storm \xrightarrow{\sqsupseteq} Tornado$, $Rock \xrightarrow{\sqsupseteq} Crystal$
Disjoint ($\xrightarrow{\perp}$)	229	$Fluid \xrightarrow{\perp} Solid$, $Fluid \xrightarrow{\perp} Gas$, $Pond \xrightarrow{\perp} River$, $Plant \xrightarrow{\perp} Animal$ $Newspaper \xrightarrow{\perp} Journal$, $Fruit \xrightarrow{\perp} Vegetable$, $Female \xrightarrow{\perp} Male$
Total	3563	

(derived from 180 different ontologies)

Evaluation: 1600 mappings, two teams, 70% Precision

M. Sabou, M. d'Aquin, E. Motta. "Using the Semantic Web as Background Knowledge for Ontology Matching". To appear in the Journal of Data Semantics

Chart 2



Error Type	Nr./ %	Examples				
		AGROVOC Concept	Labels	Rel.	NALT Concept	Labels
Anchor	114, 53%	*c_6443*	Rams, Tups	$\sqsubseteq$	*memory*	memory
		O_1:ram $\sqsubseteq$ O_1:memory				
		O_1 = `http://www.arches.uga.edu/~gonen/qos_bilal.owl`				
Subsumption as generic relation	40, 18%	*c_3954*	Irrigation	$\sqsubseteq$	*agriculture*	agriculture
		O_1:Irrigation $\sqsubseteq$ O_1:SoilCultivation $\sqsubseteq$ O_1:Agriculture				
		O_1 = `http://sweet.jpl.nasa.gov/ontology/human_activities.owl`				
Subsumption as part-whole	16, 7%	*c_666*	Asia	$\sqsupseteq$	*Iran*	Iran
		O_1:Asia $\sqsupseteq$ O_1:WestAsia $\sqsupseteq$ O_1:Iran				
		O_1 = `http://islab.hanyang.ac.kr/damls/Country.daml`				
Subsumption as role	11, 5%	*c_11091*	Garlic	$\sqsubseteq$	*ingredients*	ingredients
		O_1:garlic $\sqsubseteq$ O_1:vegetable $\sqsubseteq$ O_1:ingredient				
		O_1 = `http://cvs.sourceforge.net/viewcvs.py/instancestore/instancestore/ontologies/Attic/pizza9.daml?rev=1.2`				
Inaccurate labeling	12, 5%	*c_1693*	Coal	$\sqsubseteq$	*industry*	industry
		O_1:coal $\sqsubseteq$ O_1:industry				
		O_1 = `http://www.aifb.uni-karlsruhe.de/WBS/meh/mapping/data/russia1a.rdf`				
Different View	12, 5%	*c_2943*	Fishes	$\sqsupseteq$	*lobsters*	lobsters
		O_1:Fish $\sqsupseteq$ O_1:MarineInvertebrate $\sqsupseteq$ O_1:Crustacean $\sqsupseteq$ O_1:Lobster				
		O_1 = `http://139.91.183.30:9090/RDF/VRP/Examples/tap.rdf`				

Conclusions

- Many ‘classic’ applications of SW technology use ontologies to support the integration of distributed data sources
  - These applications are typically ‘semantically closed’
- As more and more semantic information becomes available on the SW, researchers are also looking at ‘semantically open’ applications, able to exploit large scale semantics to support intelligent problem solving
- This approach is being used in a number of other scenarios, including:
  - Semantic Web Browsing
  - Question Answering
  - Integration of Folksonomies with the SW

Selected References

- Ontology Matching
  - Lopez, V., Sabou, M., Motta, E. (2006). ***"Mapping the real semantic web on the fly"***. ISWC 2006
  - Sabou, M., D'Aquin, M., Motta, E. (2006). ***"Using the semantic web as background knowledge for ontology mapping"***. ISWC 2006 Workshop on Ontology Mapping.
- Integration of Web2.0 and Semantic Web
  - L.Specia, E. Motta, ***"Integrating Folksonomies with the Semantic Web"***, ESWC 2007.
  - Angeletou, S., Sabou, M., Specia, L., and Motta, E., (2007). “***Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report***”. ESWC 2007 Workshop on Bridging the Gap between Semantic Web and Web 2.0.
- Watson
  - d’Aquin, M., Sabou, M., Dzbor, M., Baldassarre, C., Gridinoc, L., Angeletou, S. and Motta, E.: ***"WATSON: A Gateway for the Semantic Web"***. Poster Session at ESWC 2007

'Vision' Papers

- Motta, E., Sabou, M. (2006). ***"Next Generation Semantic Web Applications"***. 1st Asian Semantic Web Conference, Beijing.

- Motta, E., Sabou, M. (2006). ***"Language Technologies and the Evolution of the Semantic Web"***. LREC 2006, Genoa, Italy.

- Motta, E. (2006). ***"Knowledge Publishing and Access on the Semantic Web: A Socio-Technological Analysis"***. IEEE Intelligent Systems, Vol.21, 3, (88-90).

KNOWLEDGE MEDIA

KMi

INSTITUTE