

School of Electronics and Computer Science

**Web Intelligence
Professor Nigel Shadbolt**



Why distributed and digital is good...





The work of many people...

- Harith Alani
- Steve Harris
- Nick Gibbins
- Yannis Kalfoglou
- Kieron O'Hara
- David Dupplaw
- Bo Hu
- Paul Lewis
- Srinandan Dashamapatra
- Duncan Macrae-Spencer
- Hugh Glaser
- Les Carr
- David de Roure
- Wendy Hall
- Mike Brady
- David Hawkes
- Yorick Wilks
- Enrico Motta
- Carole Goble
- Simon Cox
- Andy Keane
- :





Drivers

- Moores Law and Powers of 10
- The WWW
- Making the Web Semantic
- Intelligence on the Web
- Research on the Semantic Web
- Futures Challenges



Alters our Perspectives

- In May 1997 Deep Blue beat world chess champion Kasparov
- It won the six-game rematch 3.5-2.5
- The system derives its playing strength mainly out of brute force computing power
- It is a massively parallel, 32-node, RS/6000, SP-based computer system enhanced with 256 special purpose VLSI chess processors
- Evaluating 100,000,000 positions per second.





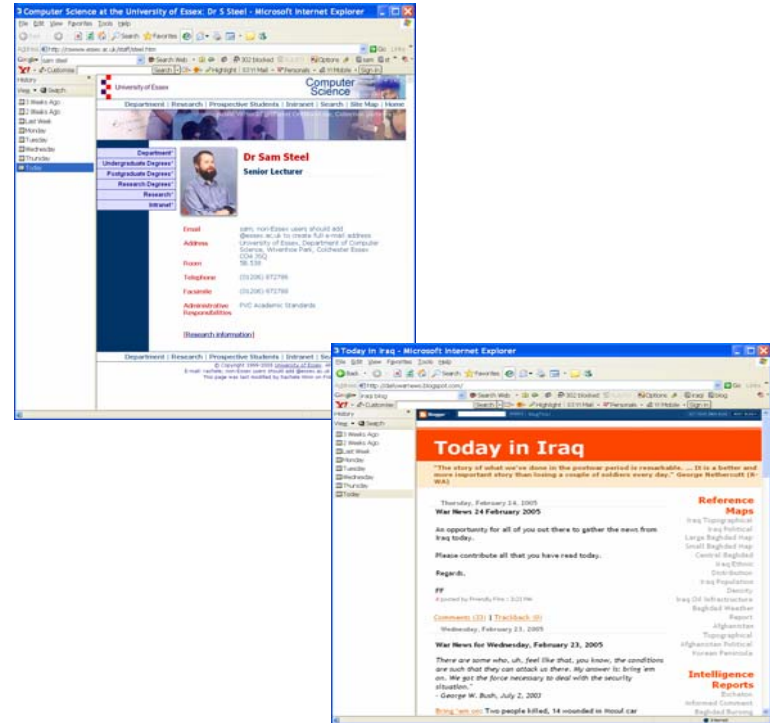
Drivers

- Moores Law and Powers of 10
- **The WWW**
- Making the Web Semantic
- Intelligence on the Web
- Research on the Semantic Web
- Futures Challenges



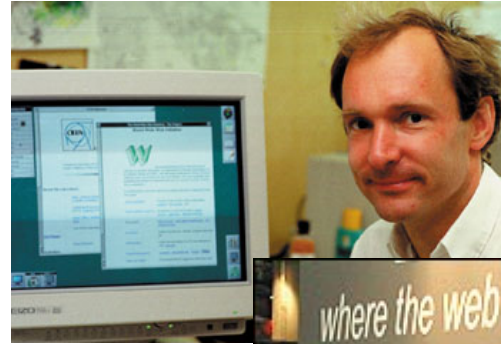
The Magic of the Network Effect

- On the order of 10^8 users
- Used in **every** country on Earth
- A tiny percentage is “trained” in any way
- On the order of 10^{10} indexed web resources (text) in Google etc
- Massively distributed and open
- A set of protocols and languages driven by a strong standards approach



Exponential Development

- A new era 1990
- nxoc01.cern.ch – the first web server
- <http://nxoc01.cern.ch/hypertext/WW/TheProject.html> - the first web page
- In 1992 – 26 web servers
- 1993 – 200 web servers
- 1994 – Mosaic browser – 1000 times more traffic than on the first web server
- 1998 -329 million pages
- The International Telecommunications Union estimated 665 million users 2002





Drivers

- Moores Law and Powers of 10
- The WWW: Blessing and Curse
- **Making the Web Semantic**
- Intelligence on the Web
- Research on the Semantic Web
- Futures Challenges



Making the Web Semantic...

<http://www2002.org>

WWW 2002



THE ELEVENTH INTERNATIONAL
WORLD WIDE WEB CONFERENCE

Sheraton Waikiki Hotel
Honolulu, Hawaii, USA
7-11 May 2002



HAWAII

1 LOCATION. 5 DAYS. LEARN. INTERACT.

Conference
Proceedings

Call for
Participation

Program

Registration
Information

Hotel
Accommodation

Conference
Committee

Sponsorship/
Exhibition
Opportunities

Volunteer
Information

Information
about Hawaii

Previous & Future
WWW Conferences

Registered participants coming from:

Australia · Canada · Chile · Denmark · France · Germany · Ghana · Hong Kong · India · Italy · Ireland · Japan · Malta · New Zealand · The Netherlands · Norway · Singapore · Switzerland · The United States · Vietnam · Zambia

[REGISTER NOW](#)

On 7-11 May 2002, Honolulu, Hawaii will provide the backdrop for The Eleventh International World Wide Web Conference. This prestigious series of the International World Wide Web Conference Committee (IW³C²) attracts participants from around the world, and it provides a public forum for the World Wide Web Consortium (W3C) through the annual W3C track.

The conference is being organized by the [International World Wide Web Conference Committee \(IW³C²\)](#), the [University of Hawaii](#) and the [Pacific Telecommunications Council \(PTC\)](#).

FEATURED SPEAKERS (CONFIRMED)



Tim Berners-Lee, inventor of the World Wide Web and Director of the W3C who now holds the 3Com Founders chair at the Laboratory for Computer Science (LCS) at the Massachusetts Institute of Technology (MIT).



Richard A. DeMillo, vice president and chief technology officer for Hewlett-Packard Company.



Ian Foster, guru of "Grid Computing", associate



McArthur Prize Winner,



That is machine readable....

<http://www2002.org>

WWW2002

THE ELEVENTH INTERNATIONAL WORLD WIDE WEB CONFERENCE

Sheraton Waikiki Hotel
Honolulu, Hawaii, USA

CONFERENCE ORGANIZERS
International World Wide Web Conference Committee

1 LOCATION. 5 DAYS. LEARN. INTERACT.

Participants coming from:
Chile · Denmark · France · Germany · Ghana · Hong Kong · India · Italy · Ireland · Japan · Malta · New Zealand · The Netherlands · Norway · Singapore · Switzerland · The United States · Vietnam · Zambia

REGISTER NOW

Honolulu, Hawaii will provide the backdrop for The Eleventh International World Wide Web Conference. This year, the International World Wide Web Conference Committee (IW³C²) attracts participants from around the world, and provides a public forum for the World Wide Web Consortium (W3C) through the annual W3C track.

The conference is being organized by the International World Wide Web Conference Committee (IW³C²), the University of Hawaii and the Pacific Telecommunications Council (PTC).

FEATURED SPEAKERS (CONFIRMED)

Tim Berners-Lee, inventor of the world wide web and Director of the W3C who now holds the 3Com Founders chair at the Laboratory for Computer Science (LCS) at the Massachusetts Institute of Technology (MIT).

Richard A. DeMillo, vice president and chief technology officer for Hewlett-Packard Company.

Ian Foster, guru of "Grid Computing", associate professor at the University of Cambridge.

McArthur Prize Winner, ...

This is a type of object event and this is its title

This is the URL of the web page for the event

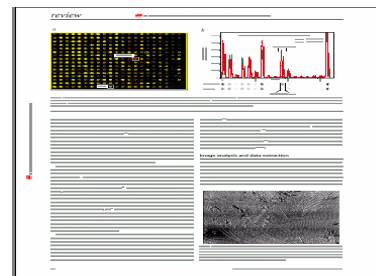
This is a type of object photograph and the photograph is of Tim Berners-Lee

Tim Berners-Lee is an invited speaker at the event

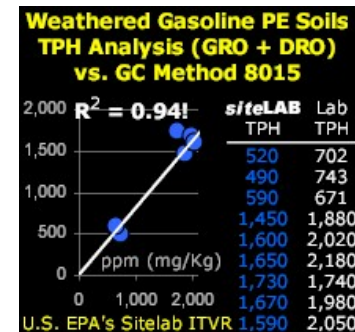


Can Annotate Anything

→ Publications...

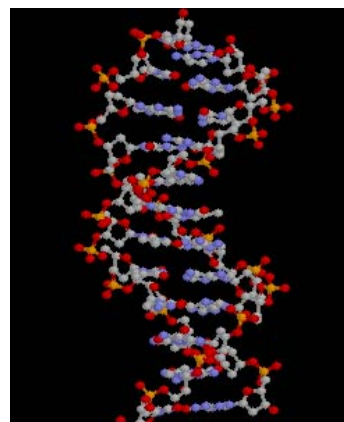


→ Databases...



Web data set (XHTML)

→ Metadata on
scientific
structures



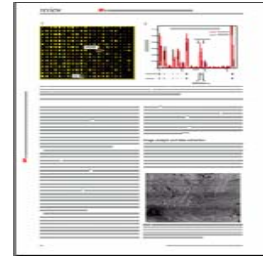
The SW Community: Structured Spaces

→ Linkage of heterogeneous information

- web content
- databases
- meta-data repository
- multimedia

→ Via ontologies as information mediation structures

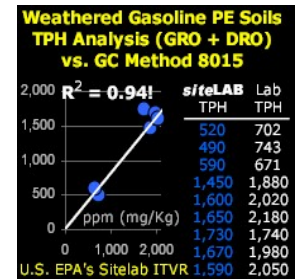
→ Using Semantic Web languages



Cancer Risk		
Cancer risk estimates do not reach zero no matter how low the level of exposure to a carcinogen. Terms used to describe this risk are defined below as the number of excess cancers expected in a lifetime:		
Term		# of Excess Cancers
moderate	is approximately equal to	1 in 1,000
low	is approximately equal to	1 in 10,000
very low	is approximately equal to	1 in 100,000
slight	is approximately equal to	1 in 1,000,000

Oncogene(MYC):
Found_In_Organism(Human).
Gene_Has_Function(Transcriptional_Regulation).
Gene_Has_Function(Gene_Transcription).
In_Chromosomal_Location(8q24).
Gene_Associated_With_Disease(Burkitts_Lymphoma).

```
<meta>  
<classifications>  
  <classification type="MYC" subtype="old_arx_id">bcr-2-1-059</classification>  
</classifications>  
</meta>
```



Web data set (XHTML)





Ontologies offer....

→ Communication

- Normative models
- Networks of relationships
- Consistent and unambiguous
- Integrate multiple perspectives

→ Inter-operability and Integration: Sharing & Reuse

- Inter-lingua
- Specifications
- Reliability

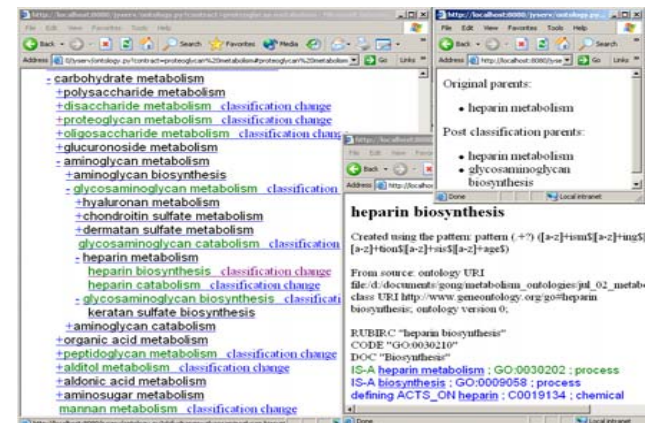
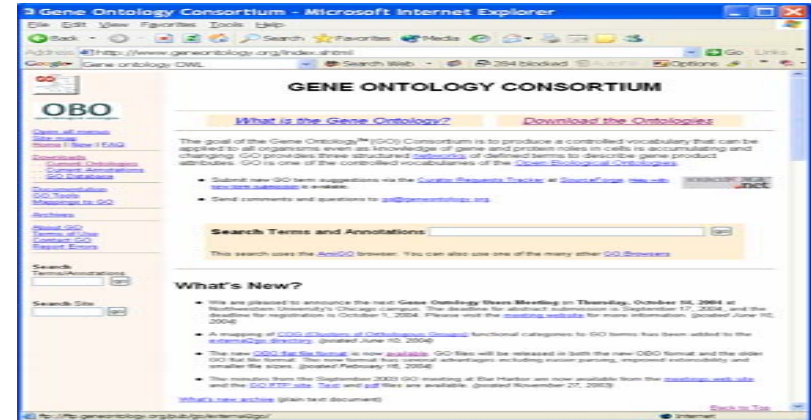
→ Control

- Controlled vocabularies
- Accurate data collection or retrieval
- Classification
- Finding, sharing, discovering, navigation, indexing



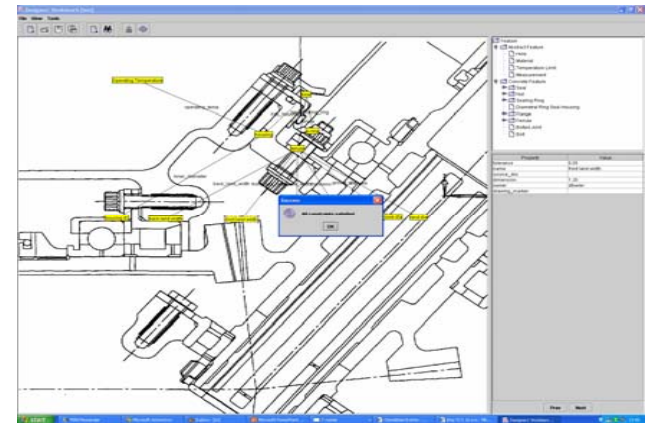
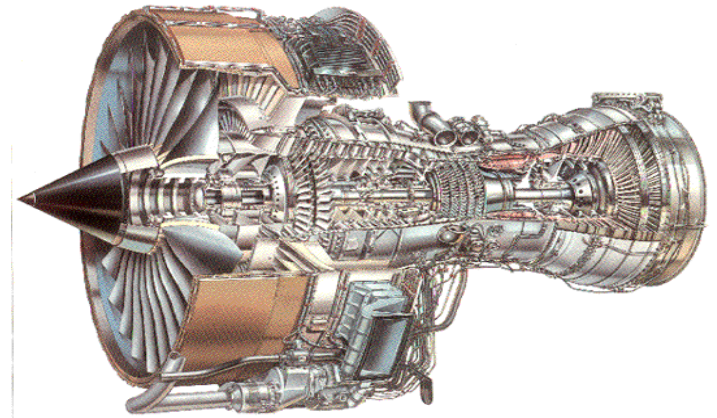
Genetics: Gene Ontology

- One of the earliest examples of the benefits of ontologies
- Integration and interoperability were big wins
- Specific tool support
- Considerable resources invested and continuing in maintenance
- Translation into DLs
- Spawned more generic biological ontology efforts



Manufacturing: Aerospace

- Considerable work on ontologies for products and components
- Used in all stages of the life cycle, from design to in service maintenance
- Need for multiple perspectives e.g.
 - Whole engine
 - Heat transfer
 - Cost model
 - Manufacturing
 - Assembling/Maintenance



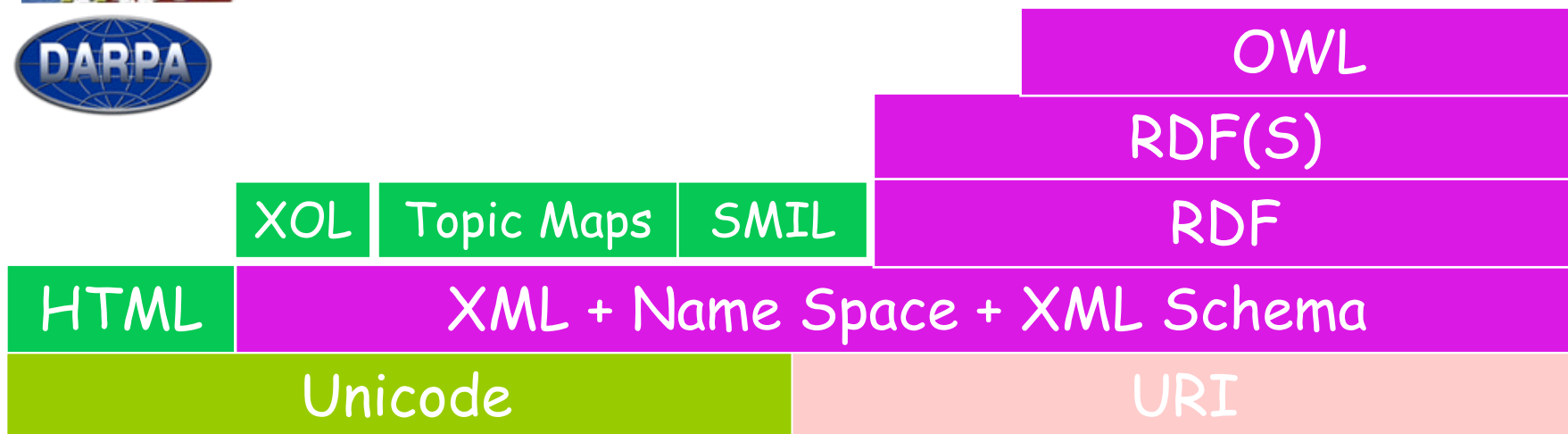


Ontologies: Observations

- In any domain
 - Usually highly implicit
 - Poorly documented
 - Likely to be ambiguous, vague, inconsistent
- When modelling
 - Interaction Problem: tasks influence ontologies
 - Integration Problem: integrating multiple ontologies
 - Modularity Problem: how to modularise and what grain size?
- Maintenance
 - Ongoing maintenance overhead
 - Ontologies evolve and change
 - Design rationale is important
- Upside
 - They do facilitate interoperability
 - They do enhance reuse
 - They are becoming part of the infrastructure



Standards are fundamental





Drivers

- Moores Law and Powers of 10
- The WWW: Blessing and Curse
- Making the Web Semantic
- **Intelligence on the Web**
- Research on the Semantic Web
- Futures Challenges





Advanced Knowledge Technologies IRC



AKT started Sept 00, 6 years, £8.8 Meg, EPSRC

www.aktors.org

Around 65 investigators and research staff





Infrastructures and Components

- Built core infrastructures
- Constructed component technologies that cover the knowledge life cycle in a number of applications

Six challenges define the Life Cycle:
Acquire • Model • Reuse • Retrieve • Publish • Maintain knowledge

The collage illustrates the six challenges of the knowledge life cycle through various visual elements:

- STRATEGY**: A flowchart showing the process from strategy to experience, F-Map, and knowledge.
- EXPERIENCE**: A screenshot of a user interface showing a profile and various data fields.
- F-Map**: A screenshot of a data table with multiple columns and rows.
- KNOWLEDGE**: A screenshot of a complex data visualization or dashboard.
- Acquire**: A screenshot of a search or discovery interface with a search bar and results.
- Model**: A screenshot of a network diagram or map.
- Reuse**: A screenshot of a '3store' interface with the text 'Scalable storage solutions for next generation knowledge services'.
- Retrieve**: A screenshot of a data table with multiple columns and rows.
- Publish**: A screenshot of a user interface showing a profile and various data fields.
- Maintain knowledge**: A screenshot of a complex data visualization or dashboard.



Exemplar Technology: ClassAKT

http://eprints.aktors.org/archive/00000268/01/iesl03.pdf - Microsoft Internet Explorer provided by ...

File Edit View Favorites Tools Help

Address http://eprints.aktors.org/archive/00000268/01/iesl03.pdf

Google Search Web Search Site PageRank Options

GATE: A Unicode-based Infrastructure Supporting Multilingual Information Extraction

Kalina Bontcheva and Diana Maynard and Valentin Tablan and Hamish Cunningham
Dept. of Computer Science, University of Sheffield
Royal Court, 211 Portobello St, Sheffield, S1 4DP, UK
[K.Bontcheva, D.Maynard, V.Tablan, H.Cunningham]@ics.shef.ac.uk

Abstract

NLP infrastructures with comprehensive multilingual support can substantially decrease the overhead of developing Information Extraction (IE) systems in new languages by offering support for different character encodings, language-independent components, and clean separation between linguistic data and the algorithms that use it. This paper will present GATE - a Unicode-aware infrastructure that offers extensive support for multilingual Information Extraction with a special emphasis on low-overhead portability between languages. GATE has been used in many research and commercial projects at Sheffield and elsewhere, including Information Extraction in Bulgarian, Romanian, Russian, and many other languages.

1 Introduction

GATE[Cunningham 02]¹ is an architecture, development environment and framework for building systems that process human language. It has been in development at the University of Sheffield since 1995, and has been used for many R&D projects, including Information Extraction in multiple languages and for multiple tasks and clients.

The GATE architecture defines almost everything in terms of components - reusable units of code that are specialised for a specific task. There are three main types of components:

- **Language Resources (LRs)** store some kind of linguistic data such as documents, corpora, ontologies and provide services for accessing it. At the moment all the predefined LRs are text based but the model doesn't restrict the data format so the framework could be extended to handle multimedia documents as well.
- **Processing Resources (PRs)** are resources whose character is principally programmatic or

GATE is implemented in Java and is freely available from <http://gate.ac.uk> as open-source free software under the GNU library licence.

algorithmic such as a POS tagger or a parser. In most cases PRs are used to process the data provided by one or more LRs but that is not a requirement.

- **Visual Resources (VRs)** are graphical components that are displayed by the user interface and allow the visualisation and editing of other types of resources or the control of the execution flow.

The GATE framework defines some basic language resources such as documents and corpora, provides resource discovery and loading facilities and supports various kinds of input/output operations such as format decoding, file or database persistence.

GATE uses a single unified model of *annotation* - a modified form of the TIPSTER format [Grisman 97] which has been made largely compatible with the Atlas format [Bird & Liberman 99], and uses the now standard mechanism of 'stand-off markup' [Thompson & McKelvie 97]. Annotations are characterised by a *type* and a set of *features* represented as attribute-value pairs. The annotations are stored in structures called *annotation sets* which constitute independent layers of annotation over the text content.

The advantage of converting all formatting information and corpus markup into a unified representation, i.e. the annotations, is that NLP applications do not need to be adapted for the different formats of each of the documents, which are catered for by the GATE format filters (e.g. some corpora such as BNC come as SGML/XML files, while others come as email folders, HTML pages, news wires, or Word documents).

The work for the second version of GATE started in 1999 and led to a complete redesign of the system and a 100% Java implementation. One of the additions brought by version 2 is full support for Unicode data allowing the users to open, visualise and process documents in languages dif-

Classifier - Microsoft Internet Explorer provided by AOL

File Edit View Favorites Tools Help

Address http://robin.ecs.soton.ac.uk:8000/classifier/index.jsp?url=http%3A%2F%2Feprints.aktors.org%2Farchive%2F00000268%2F01

Google Search Web Search Site PageRank Options

Submit Reset

http://eprints.aktors.org/archive/00000268/01/iesl03.pdf

Document Classified as:

- H. Information Systems
 - H.3 INFORMATION STORAGE AND RETRIEVAL

Distributions:

A. General Literature	0.000
B. Hardware	0.000
C. Computer Systems Organization	0.000
D. Software	0.000
E. Data	0.000
F. Theory of Computation	0.000
G. Mathematics of Computing	0.000
H. Information Systems	1.000
I. Computing Methodologies	0.000
J. Computer Applications	0.000
K. Computing Milieux	0.000

H.1 MODELS AND PRINCIPLES	0.000
H.2 DATABASE MANAGEMENT	0.000
H.3 INFORMATION STORAGE AND RETRIEVAL	1.000
H.4 INFORMATION SYSTEMS APPLICATIONS	0.000
H.5 INFORMATION INTERFACES AND PRESENTATION (I.7)	0.000

Last updated 23/08/2003: ClassAKT is currently configured to classify computer related pdf documents according to the ACM classification scheme. In particular, classification can currently be made with regards to all the top level ACM classes (listed below) together with their immediate sub-classes:

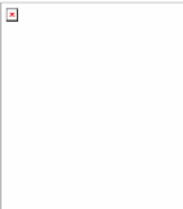


Component Technologies: Modelling



Annotation Key:
Person **Organization** **Address**

Professor Nigel Shadbolt



Address: **Department of Electronics and Computer Science**
University of Southampton
Highfield, Southampton
SO17 1BJ
United Kingdom

Email: nrs@ecs.soton.ac.uk

Tel: +44 23 8059 7682 (direct)
+44 23 8059 4505 (secretary)

Fax: +44 23 8059 3313

Directions: [to the University](#) ; [to my Office](#) (Room 3027, Building 59)

[[Projects](#) | [Publications](#) | [Presentations](#) | [CV](#)]

Nigel Shadbolt is Professor of Artificial Intelligence (AI) in the **Department of Electronics and Computer Science** at **Southampton University**. He is a member of the [Intelligence, Agents, Multimedia](#) Group.

His research concentrates on two ends of the spectrum of AI - namely, **Knowledge Technologies** and Biorobotics. For fifteen years as Director of the **AI Group** at the **School of Psychology** at the **University of Nottingham** he established an international reputation for work in **Knowledge Technologies**. In particular developing methods, tools and techniques to support the construction of knowledge-based systems. Much of this work was undertaken collaboratively and in many cases led to deployed applications. In 2000 he led a consortium of five Universities that secured an **EPSRC** Interdisciplinary Research Collaboration in **Advanced Knowledge Technologies**. **Professor Shadbolt** is the Director of this eight million pound, six-year research programme that is pursuing basic and applied research in the provision of technologies to support **Knowledge Management** and realise the promise of the Semantic Web. The **AKT** project has attracted additional funding including two grants for which **Professor Shadbolt** is also the Principal Investigator. The first of these (**CoAKTinG**) is investigating the use of knowledge technologies in a collaborative context using multicast videoconference technologies such as the Access Grid. The second (**MLAKT**) is attempting to support medical decision-making through the use of **AKT's** tools, methods and techniques. He is also an investigator on a **University of Southampton e-Science** project **GEODISE** where he is directing the effort to integrate knowledge engineering methods into design optimisation tasks that exploit Grid computing. He is also working with his colleague **Professor Nick Jennings** on a **Hewlett-Packard** sponsored project **ANNA** that aims to develop a framework within which software agents can be furnished with knowledge acquisition capabilities.

His work on biorobotics investigates how we can draw inspiration from real animal systems in the construction of robots and biologically inspired neural modelling. Research work has included investigations into simple sensory systems for mobile robots and hybrid architectures for autonomous robot systems. He works with **Dr Terry Elliott** on the development of computational models of neural plasticity that reflect the dynamic nature of anatomical change in the nervous system. This work is also now being applied to robots to understand how we might build developing artificial

<http://www.soton.ac.uk/menus/mimenu/hfield.html>



Component Technologies: Modelling

The screenshot displays the Protégé-2000 ontology editor interface. The main window is titled "AKT_Ontology_4 Protégé-2000 (C:\Program Files\Protege-1.8beta\examples\AKT_Ontology_3\AKT_Ontology_4.pprj)". The interface is divided into several panes:

- Relationships:** A tree view on the left showing the ontology's class hierarchy, including categories like .THING, base:Document, base:Person, and person:Employee.
- Instances:** A central list of instances, with "Alani" selected. The list includes names such as Addis, Adhianto, Al-Khoury, Alani, Alavi, Allen, Atta, De Roure, Shadbolt, Jennings, Carr, Moreau, Lewis, Crowder, Davis, Harnad, Braithwaite, Brown, Cecil-Wright, Chan, Charlton, Cho, Colbrook, Crouch, Cruickshank, Dattani, and Davidson.
- Get COP Results Table:** A table showing the results of a query. The table has columns for "Instance" and "Weight".
- Selected Slots:** A table showing the weights for selected slots.
- Auto Settings:** A section for configuring settings like "Freq. of Use" and "Class Selection".
- Temporal Intervals:** A section for setting time ranges, currently from 1980 to 2002.
- TGWikiBrowser:** A network diagram on the right showing relationships between various nodes, including "De Roure", "Carr", "Classer Alani", "Gibbins", "Hatchcock", "Crowder", and "Harnad".

Instance	Weight
Alani	38.366
AKT: Advanced Know... Intelligence, Agents a...	38.011
Hall	16.417
De Roure	12.066
Shadbolt	9.92
Jennings	9.56
Carr	9.475
Moreau	6.956
Lewis	6.895
Crowder	6.871
Davis	6.783
Harnad	6.118

Slots	Weight
base:hasAuthor	0.3
base:member_of_project	0.95
base:memberOf	0.2





Integrating Semantic Spaces

- Exploit core infrastructure
- Integrate component technologies from the knowledge life cycle for an application

Six challenges define the Life Cycle:
Acquire • Model • Reuse • Retrieve • Publish • Maintain knowledge

The collage illustrates the six challenges of the knowledge life cycle. It features a process flow diagram on the left, a user interface for a knowledge management system, a '3store' logo with the text 'Scalable storage solutions for next generation knowledge services', a network diagram, and a large clock face with a glowing arrow pointing upwards. The background is a dark blue grid with a glowing green circular path.





The CS AKTive Space: International Semantic Web Challenge Winner

- 24/7 update of content
- Content continually harvested and acquired against community agreed ontology
- Easy access to information gestalts - who, what, where
- Hot spots
 - Institutions
 - Individuals
 - Topics
- Impact of research
 - citation services etc
 - funding levels
 - Changes and deltas
- Dynamic Communities of Practice...



CS AKTiveSpace

AKT: CS Aktive Space - Mozilla

File Edit View Go Bookmarks Tools Window Help


Back Forward Reload Stop http://triplestore.aktors.org/demo/AKTiveSpace/ Search Print

AKT CS AKTive Space

Take a tour through CS AKTive Space

About this page research area/region region/research area

Radial: 100 miles Map: uk-political



Research area

- Computing Methodologies
 - pattern recognition
 - artificial intelligence
 - image processing and computer vision
 - general
- Data
 - coding and information theory
 - data encryption
- Theory of Computation
 - mathematical logic and formal languages
- Software
 - software engineering
 - general
- Computer Systems Organization
 - general
- Information Systems
 - information interfaces and presentation
- Hardware
 - performance and reliability
- undefined

Researcher

Top 5 10 20 unlimited

Order by Grant total RAE result

NR Shadbolt

- SJ Cox
- P Johnson
- D May
- LA Carr
- DC De Roure
- JF Baldwin
- NR Jennings
- L Moreau
- C Melhuish

CoP

- Kieron O'Hara
- Nicholas
- Gibbins
- Stephen
- Harris
- Nick Jennings
- David De
- Roure
- Hugh Glaser
- Paul Lewis
- Xudong Luo
- Wendy Hall
- Bo Hu
- Alexei Dingli
- Sam

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 interests Fluid Dynamics
Aerodynamics
Design and Testing Technology
Biological Sciences Domain

RDF W3C XHTML 1.0

Technology Overviews: CS AKTive Space | CoP | Armadillo | 3store



Extending the model

AKT: WUN Aktive Space - Mozilla
http://oteseer.ecs.soton.ac.uk/df/aktivespace/
WON AKT WUN AKTive Space Powered By CiteSeer

About this page research area/region region/research area

Radial: Map:

Research area:

Researcher: Top 5 10 20 unlimited
Order by Funding Citation Impact

CiteSeer.IST
Scientific Literature Digital Library

[Home](#) [Submit Documents](#) [Statistics](#) [About](#) [Feedback](#) [Help](#)

Documents indexed by CiteSeer.IST Citations made by indexed documents

Copyright [NEC](#) and [IST](#) | [Privacy Policy](#) | [OAI Compliance](#)

PubMedBrowser V1.0 - Microsoft Internet Explorer
http://www.pmbrowser.info/pubmed3.php?start=12020448

Network diagram illustrating relationships between research topics:

- Synaptic plasticity and the organization of behaviour after early and late brain
- Models describing nonlinear interactions in graded neuron synapses.
- Reverse Hebb plasticity leads to optimization and association in a simulated
- Dynamic model of long-term synaptic plasticity: the role of nonlinear plasticity and unreliable synapses.
- Competitive Hebbian learning
- Synaptic heterogeneity and stimulus-induced modulation of depression in central
- Equilibrium properties of temporally asymmetric Hebbian plasticity.
- Fluency of a neurobiological theory of the hippocampus: the role of
- Self-regulation mechanisms of temporally asymmetric Hebbian plasticity.
- Stable Hebbian learning from spike timing dependent plasticity.
- An analysis of synaptic normalization in a general class of Hebbian models.
- Learning input correlates through nonlinear temporal rate underlie a nonlinear asymmetric Hebbian plasticity.
- Homeostatic stabilization of output rates by spike-based Hebbian learning.
- Converging evidence for a simplified biophysical model of synaptic plasticity.
- Spike-driven synaptic plasticity for learning correlated patterns of mean
- Competitive anatomical and physiological plasticity: a neurotrophic bridge.
- Competition for neurotrophic factors: mathematical analysis.
- A mathematical model of activity-dependent, anatomical segregation.
- Role of neurotrophins in synapse development and plasticity.
- Have we been hobnobbing down the wrong path?
- Neurotrophins in the central nervous system: do synapses divide?
- Disassociating ocular dominance development and plasticity: a
- Competition for neurotrophic factors: mathematical analysis.
- Role of neurotrophins in synapse development and plasticity.
- Have we been hobnobbing down the wrong path?
- Neurotrophins in the central nervous system: do synapses divide?

GoogleBrowserApplet started

Announcements



Microsoft Research

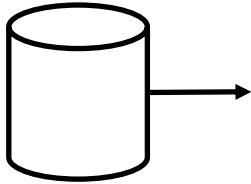
IST

NEC

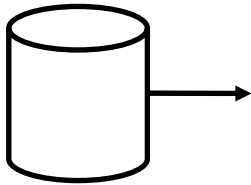


Mediate and Aggregate: UK Research Councils

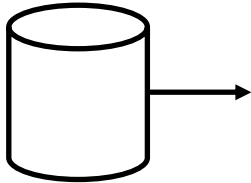
EPSRC



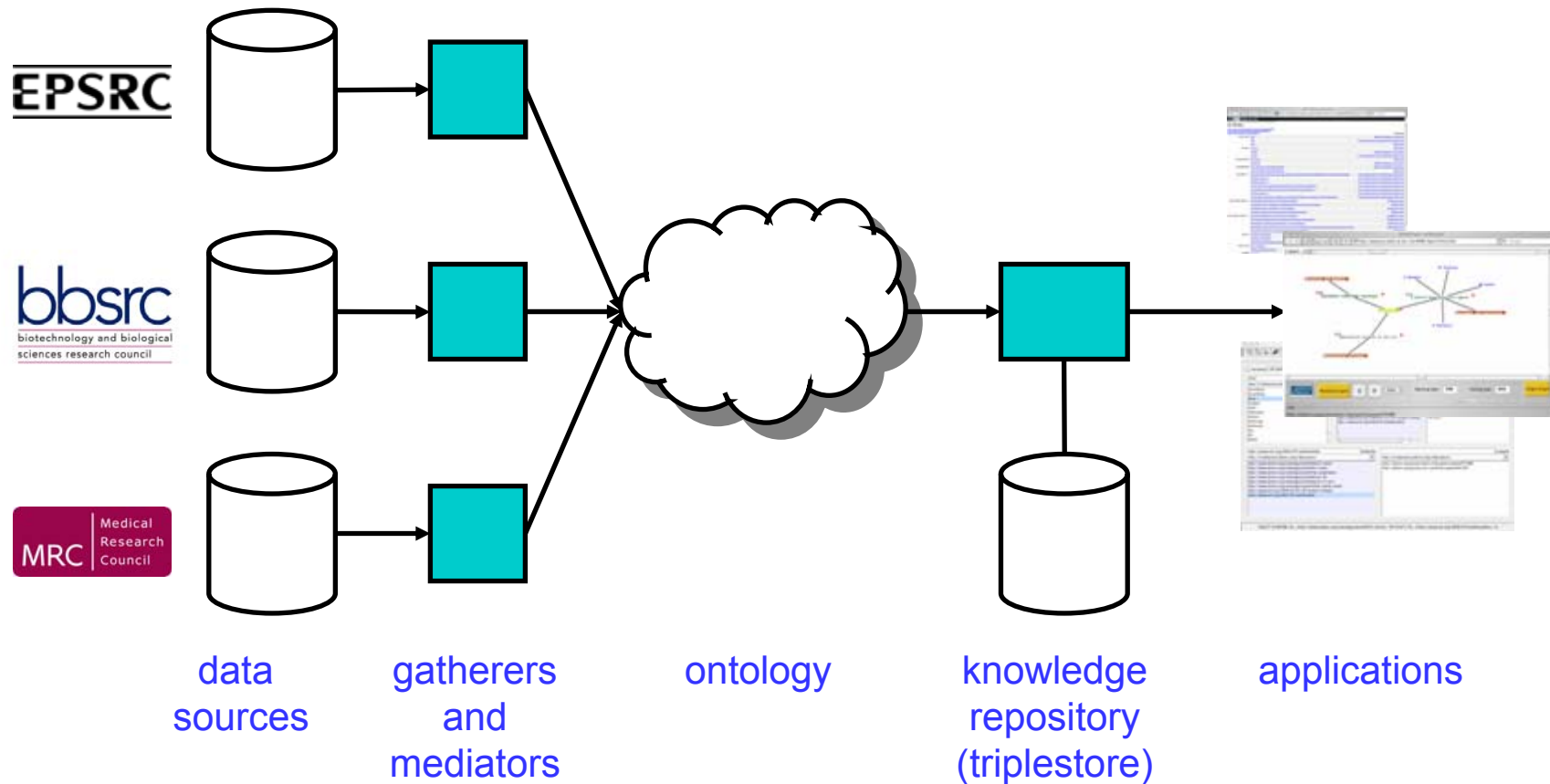
bbsrc
biotechnology and biological
sciences research council



MRC | Medical
Research
Council



EPSRC: Knowing what they know



Mediate and Aggregate: Ontologies

```
Index of /demo/EP5RC/data/raw - Mozilla [Build ID: 2004092716]
File Edit View Go Bookmarks Tools Window Help Debug QA
http://triplestore.aktors.org/demo/EP5RC/data/raw/AKT-basicgrant.txt

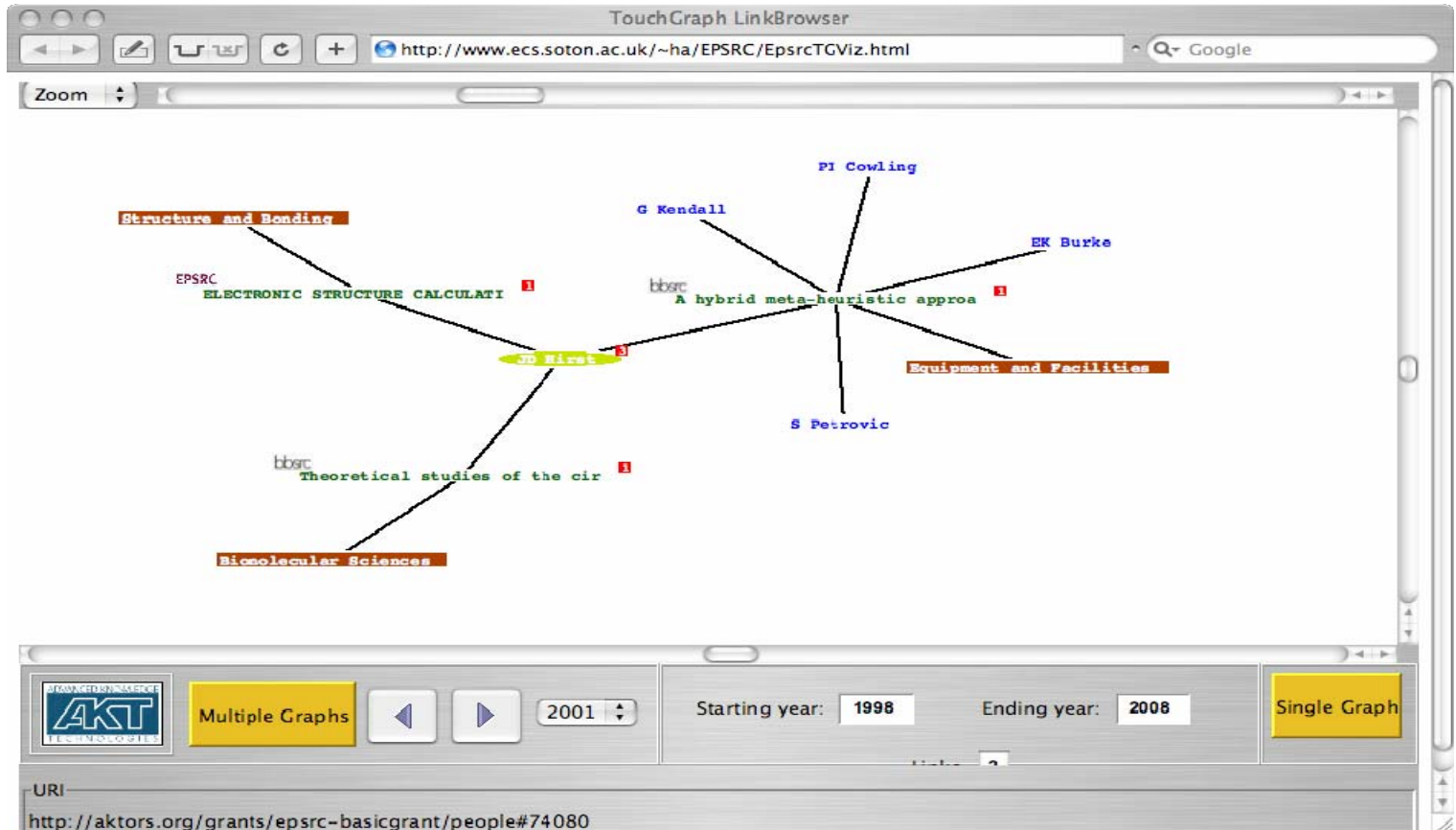
"GrantRefNumber", "HoldingOrganisationId", "HoldingOrganisationName", "Holding
"GR/E79682/01", 1, "Bolton Institute", 3460, "Civil & Environmental Eng Subject
"GR/F27789/01", 1, "Bolton Institute", 3460, "Civil & Environmental Eng Subject
"GR/G55372/01", 1, "Bolton Institute", 3460, "Civil & Environmental Eng Su
"GR/H48736/01", 1, "Bolton Institute", 3615, "Mechanical & Automobile Eng
"GR/M26633/01", 12, "University of Dundee", 3251, "Applied Computing", 1, "A
"GR/G02024/01", 16, "University of Warwick", 1197, "Physics", 2, "Standard",
"GR/J20449/01", 1, "Bolton Institute", 3615, "Mechanical & Automobile Eng
"GR/J42854/01", 1, "Bolton Institute", 3615, "Mechanical & Automobile Eng
"GR/J46951/01", 1, "Bolton Institute", 4088, "Unknown", 2, "Standard", 0, 0, 1
"GR/J91289/01", 1, "Bolton Institute", 3460, "Civil & Environmental Eng Su
"GR/K12007/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 2, "St
"GR/K47405/01", 1, "Bolton Institute", 3615, "Mechanical & Automobile Eng
"GR/K90364/01", 16, "University of Warwick", 5210, "Chemistry", 2, "Standard
"GR/K87319/02", 1, "Bolton Institute", 6298, "Faculty of Technology", 1, "Am
"GR/K87531/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 1, "Am
"GR/L56770/01", 1, "Bolton Institute", 5026, "Unknown", 2, "Standard", 1, 0, 1
"GR/L84711/01", 1, "Bolton Institute", 5026, "Unknown", 2, "Standard", 1, 0, 1
"GR/L98435/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 2, "St
"GR/N01088/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 2, "St
"GR/P00260/01", 1, "Bolton Institute", 6521, "DTA Department", 2, "Standard"
"GR/G42501/01", 16, "University of Warwick", 5210, "Chemistry", 1, "Amendabl
"GR/N31122/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 1, "Am
"GR/R19397/01", 1, "Bolton Institute", 6298, "Faculty of Technology", 2, "St
"GR/S24350/01", 1, "Bolton Institute", 15858, "Computing & Electronic Tech
```

```
Mozilla [Build ID: 2004092716]
File Edit View Go Bookmarks Tools Window Help Debug QA
http://triplestore.aktors.org/demo/EP5RC/data/rdf/bbsrc.rdf

<owl:imports rdf:resource="http://www.aktors.org/ontology/extension"/>
</owl:Ontology>
- <rdf:Description rdf:about="http://aktors.org/grants/bbsrc/#B15240">
- <support:has-pretty-name>
  Theoretical studies of the circular dichroism of peptides and proteins
</support:has-pretty-name>
- <support:has-time-interval>
- <support:Time-Interval
  rdf:about="http://www.aktors.org/ontology/date#2001-08-06/2004-08-06">
- <support:begins-at-time-point>
- <support:Calendar-Date
  rdf:about="http://www.aktors.org/ontology/date#2001-08-06">
  <support:has-pretty-name>2001-08-06</support:has-pretty-name:
  <support:day-of>06</support:day-of>
  <support:month-of>08</support:month-of>
  <support:year-of>2001</support:year-of>
</support:Calendar-Date>
</support:begins-at-time-point>
- <support:ends-at-time-point>
- <support:Calendar-Date
  rdf:about="http://www.aktors.org/ontology/date#2004-08-06">
```



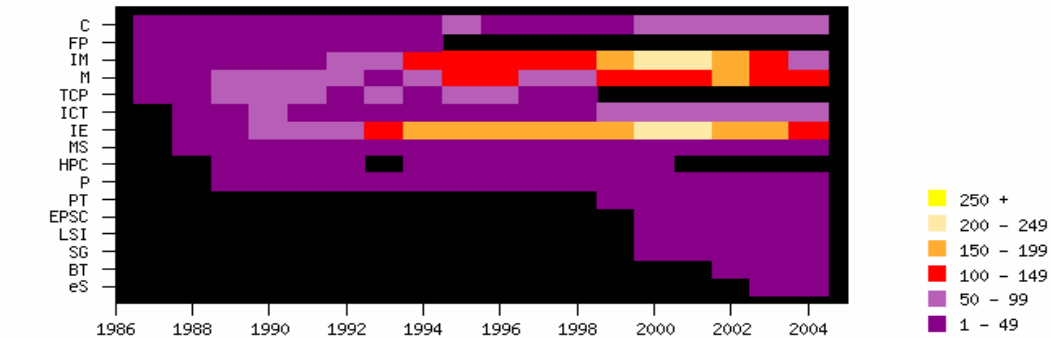
Visualising Interaction



Visualising Interaction: Programmes

Collaboration over time

View: by year by area



BT C E EPSC FP HPC ICT IE IM LSI M MS P PT SG TCP eS

- +

- | | | | |
|------|---|-----|--------------------------|
| BT | Basic Technology | LSI | Life Sciences Interface |
| C | Chemistry | M | Materials |
| E | Engineering | MS | Mathematical Sciences |
| EPSC | Engineering & Physical Sciences Council | P | Physics |
| FP | Facilities Programme | PT | Postgraduate Training |
| HPC | High Performance Computing | SG | Schemes Group |
| ICT | Information and Communications Technology | TCP | Teaching Company Project |
| IE | Infrastructure and Environment | eS | e-Science |
| IM | Innovative Manufacturing | | |





Integrating Semantic Spaces

- Exploit core infrastructure
- Integrate component technologies from the knowledge life cycle for an application

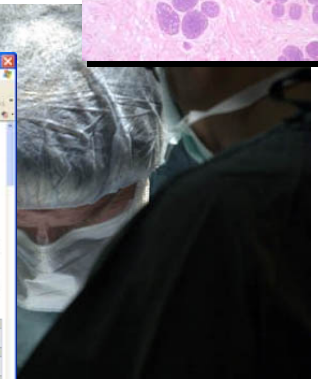
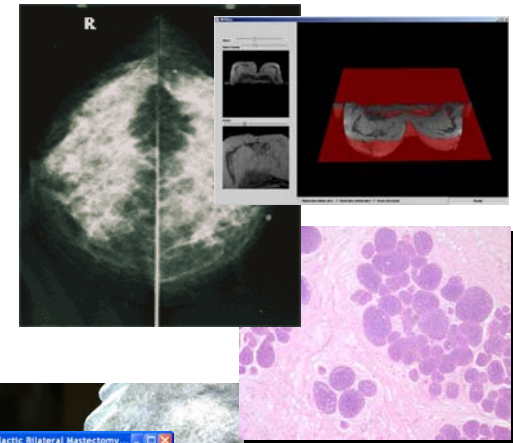
Six challenges define the Life Cycle:
Acquire • Model • Reuse • Retrieve • Publish • Maintain knowledge

The collage illustrates the six challenges of the knowledge life cycle. It features a central clock face with a glowing arrow pointing upwards, symbolizing the progression of time and knowledge. Surrounding the clock are several images: a process flow diagram with steps like 'STRATEGY', 'EXPERIENCE', 'F-Map', and 'KNOWLEDGE'; a user interface for a knowledge management system; a '3store' logo with the text 'Scalable storage solutions for next generation knowledge services'; a network diagram; and a large clock face with a glowing arrow pointing upwards.



Collaborative Medical Decision Making MIAKT: Integrating Semantic Spaces

- Diverse and heterogeneous content
- Clinical examination
 - Notes
- Imaging
 - X-ray,
 - Ultrasound
 - MRI
- Microscopy
 - Histopathology
- Treatment
 - Protocol Records
 - Re-assessment
- Medical Records
 - Case sets
 - Individual patient records
- Published background
 - Epidemiology
 - Medical Abstracts





Why This Domain?

→ The user and patient perspective

- $\approx 10\%$ of women develop breast cancer during their lives
- 3 million screening cases in UK per year: 8~25% cancers are missed by radiologists; 70~80% biopsies turn out to be benign; inter- and intra-radiologist variability is typically 30%.

→ AKT perspective

- Large amounts of data, information and knowledge
- Collaboration among different domain experts but little technical support for the collaborative elements of the problem and little technical support for the semantics of the domain

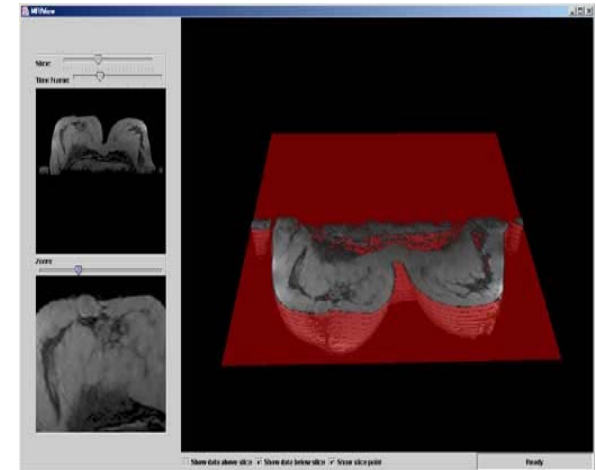
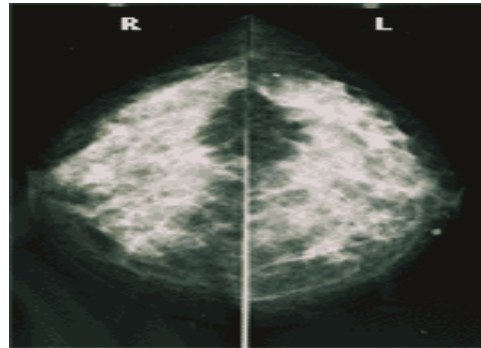
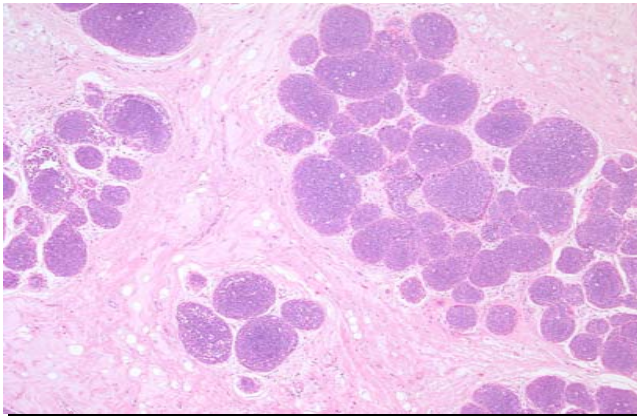
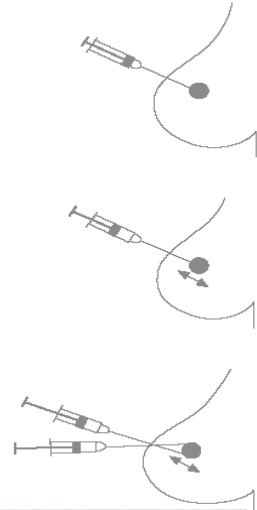
→ MIAS perspective

- Deployment of methods as web and grid services
- Need for information management



Multi-disciplinary Assessment: The Medics

- Different domains of expertise overlaid
- Breast imaging – X-ray, ultrasound, MRI
- Clinical examination
- Non-aggressive methods – Fine Needle Aspiration
- Histopathology Slides
- Microscopy – cells and tissues (also, hormone receptors)
- Prognosis, staging and statistical windows into the future



- **Image Analysis Services**
 - Oxford's XRay Mammogram Analyser
 - KCL MRI Mammogram Analyser/Classifier
- **Classification Services**
 - Abnormality Naïve Bayes Classifier (Soton)
 - MRI Lesion Classifier (KCL)
- **Patient Data Retrieval Services (OU)**
 - For example, "Find Patients With Same Age"
- **Image Registration (KCL)**
 - GRID service invoked via web-service
- **Natural Language Report Generation (Sheffield)**
 - Generate a patient report from RDF description
- **UMLS Lookup (Sheffield)**
 - Lookup term definitions in the UMLS
- **Patient Records also accessed through web-service (Soton)**
 - Web-service enabled AKT 3store





Patient Cases in RDF

```
<rdf:Description rdf:about='#g1p78_patient'>
  <rdf:type rdf:resource='#Patient'/>
  <NS2:has_date_of_birth>01.01.1923</NS2:has_date_of_birth>
  <NS2:involved_in_ta rdf:resource='#ta_soton_000130051992'/>
</rdf:Description>
```

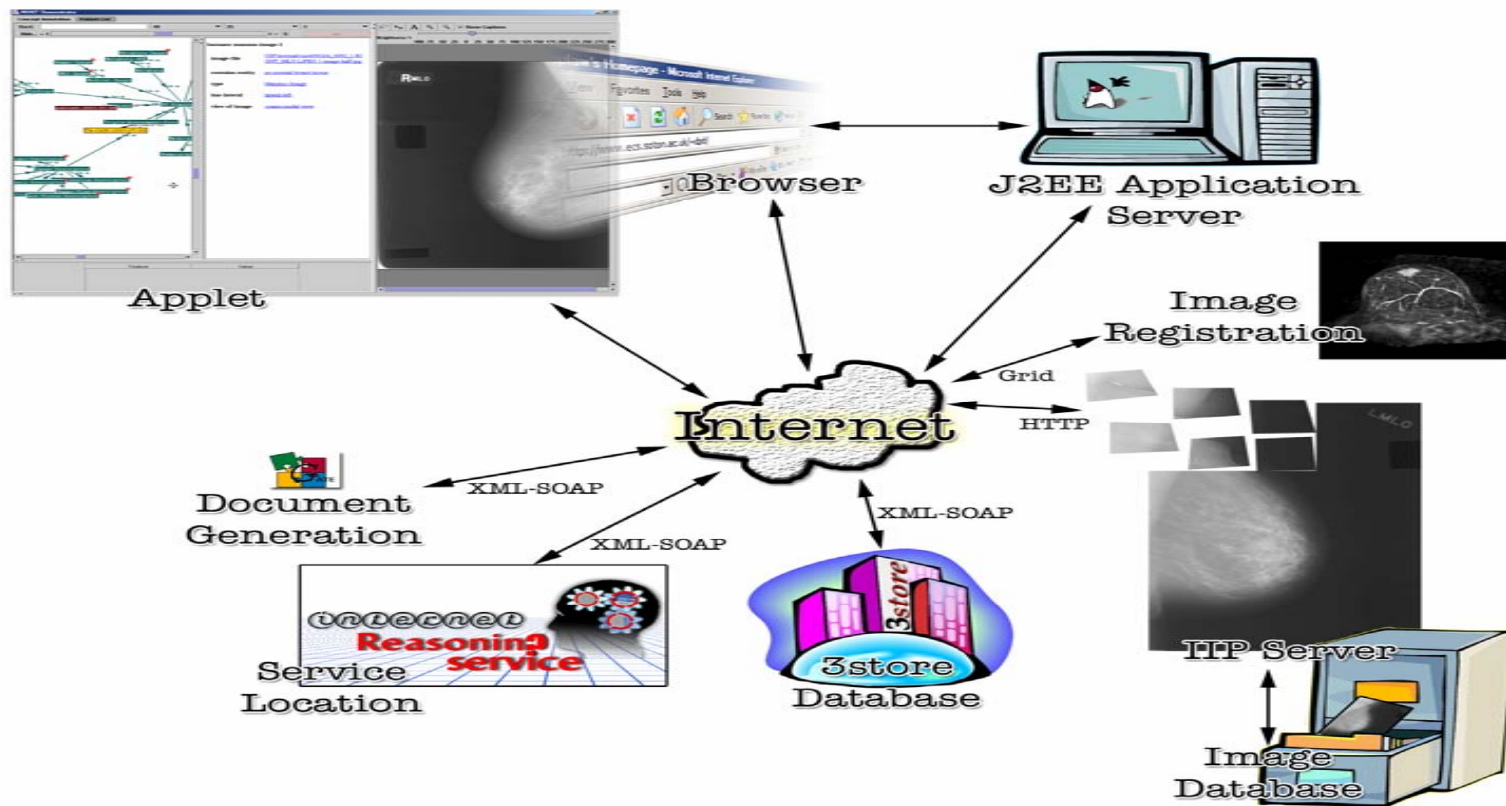
```
<rdf:Description rdf:about='#ta_soton_000130051992'>
  <rdf:type rdf:resource='#Multi_Disciplinary_Meeting_TA'/>
  <NS2:involve_patient rdf:resource='#g1p78_patient'/>
  <NS2:consist_of_subproc rdf:resource='#oe_00103051992'/>
  <NS2:consist_of_subproc rdf:resource='#hp_00117051992'/>
  <NS2:consist_of_subproc rdf:resource='#ma_00127051992'/>
  <NS2:has_overall_impression rdf:resource='#assessment_b5_malignant'/>
  <NS2:has_overall_diagnosis>invasive carcinoma</NS2:has_overall_diagnosis>
</rdf:Description>
```

```
<rdf:Description rdf:about='#oe_00103051992'>
  <rdf:type rdf:resource='#Physical_Exam'/>
  <NS2:has_date>03.05.1992</NS2:has_date>
  <NS2:produce_result rdf:resource='#oereport_glp78_1'/>
  <NS2:carried_out_on rdf:resource='#g1p78_patient'/>
</rdf:Description>
```

```
<rdf:Description rdf:about='#oereport_glp78_1'>
  <NS2:type rdf:resource='#Lateral_OE_Report'/>
  <NS2:contains_roi rdf:resource='#oe_roi_00103051992'/>
  <NS2:has_lateral rdf:resource='#lateral_left'/>
</rdf:Description>
```



The MIAKT Framework



Demonstration

The screenshot displays the MIAKT Demonstrator interface, which is used for exploring medical image analysis knowledge. It is divided into three main sections:

- Concept Browser (Left):** A hierarchical tree view of the ontology. The root is 'BC_Domain_Top', which branches into 'Medical_Exam', 'MRI_Contrast_Media', 'Medical_Descriptor', 'MRI_Process_Descriptor', 'Image_Descriptor', 'Patient', and 'Findings'. The 'Patient' node is currently selected.
- Central Ontology Diagram:** A network graph showing relationships between concepts. 'BC_Domain_Top' is the central node, with 'is-a' relationships to 'Medical_Exam', 'MRI_Contrast_Media', 'Medical_Descriptor', 'MRI_Process_Descriptor', 'Image_Descriptor', 'Patient', 'Findings', 'Lateral_Side', 'Breast_Disease', 'Triple_Assessment_Proc', 'Medical_Im', 'Region_Of_Interest', and 'Clinician'. Other nodes like 'Findings' and 'Medical_Exam' also have their own sub-relationships.
- Instance Details Panel (Right):** A panel showing details for a specific instance. It has tabs for 'Instances of Patient' and 'Instance 00071_patient'. The 'Instance 00071_patient' tab is active, showing:
 - Instance 00071_patient:** type [Patient](#), has_age 57, involved_in [ta-soton-1070478266177](#).
 - Instance ta-soton-1070478266177:** type [Triple_Assessment_Proc](#), consist_of_subproc [00071_mammography](#), involve_patient [00071_patient](#).
 - Instance 00071_mammography:** produce_result [image_00071_right_mlo](#), [image_00071_right_cc](#), [image_00071_left_mlo](#), [image_00071_left_cc](#); type [Mammography](#); carried_out_on [00071_patient](#); has_date 27 12 1992.





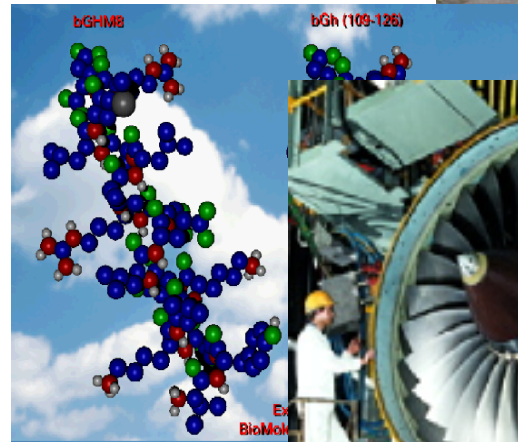
Drivers

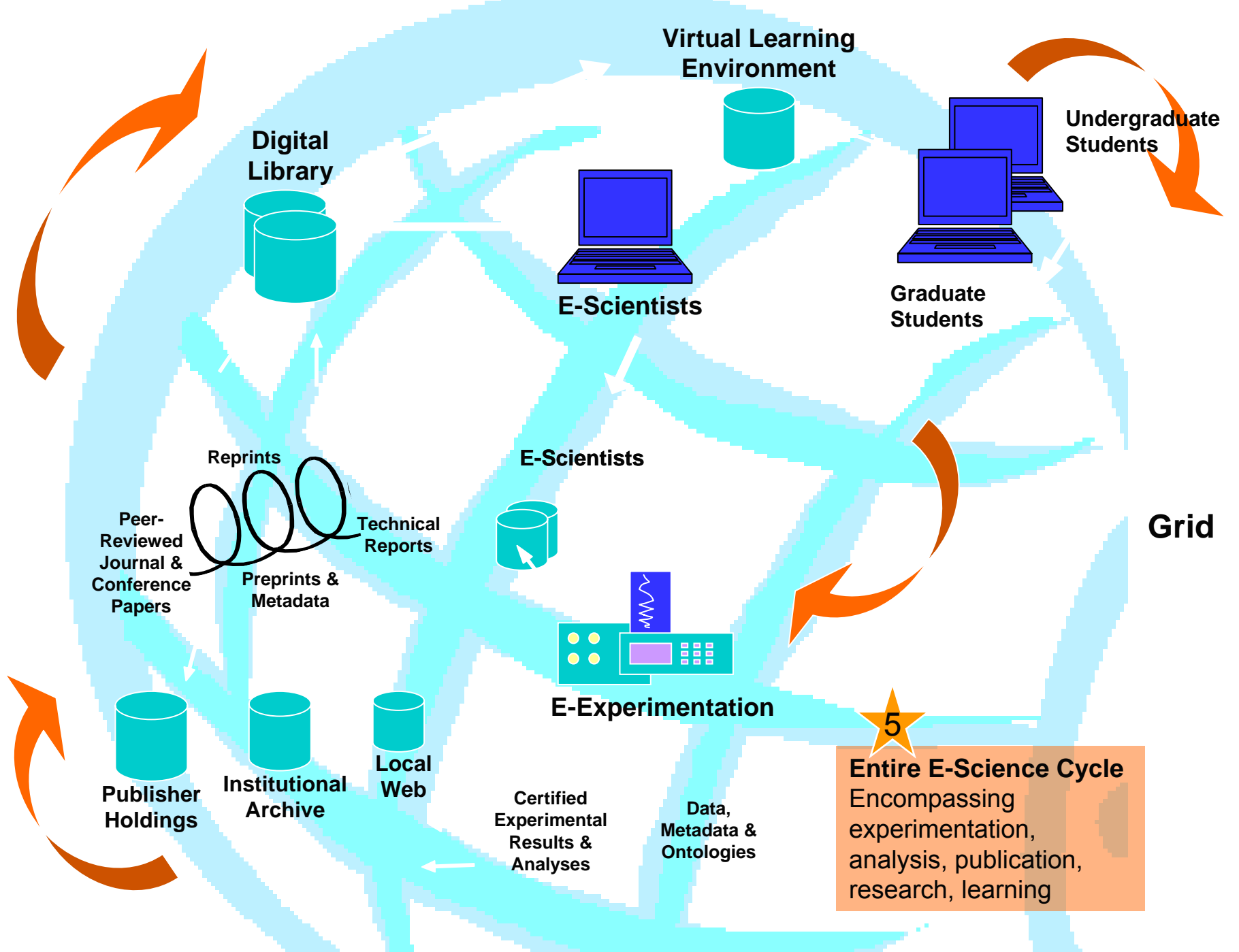
- Moores Law and Powers of 10
- The WWW: Blessing and Curse
- Making the Web Semantic
- Intelligence on the Web
- **Research on the Semantic Web**
- Futures Challenges



New ways of discovery: e-Science

- A large part of scientific discovery is now a joint human machine endeavour
- Without considerable compute power no hope of progress
- Examples from physics, astronomy, biology, chemistry and engineering





Virtual Learning Environment

Digital Library

Undergraduate Students



E-Scientists



Graduate Students

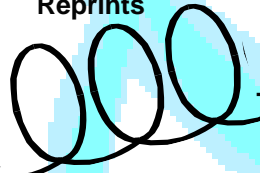
E-Scientists



E-Experimentation

Grid

Reprints



Technical Reports

Peer-Reviewed Journal & Conference Papers

Preprints & Metadata



Publisher Holdings



Institutional Archive



Local Web

Certified Experimental Results & Analyses

Data, Metadata & Ontologies

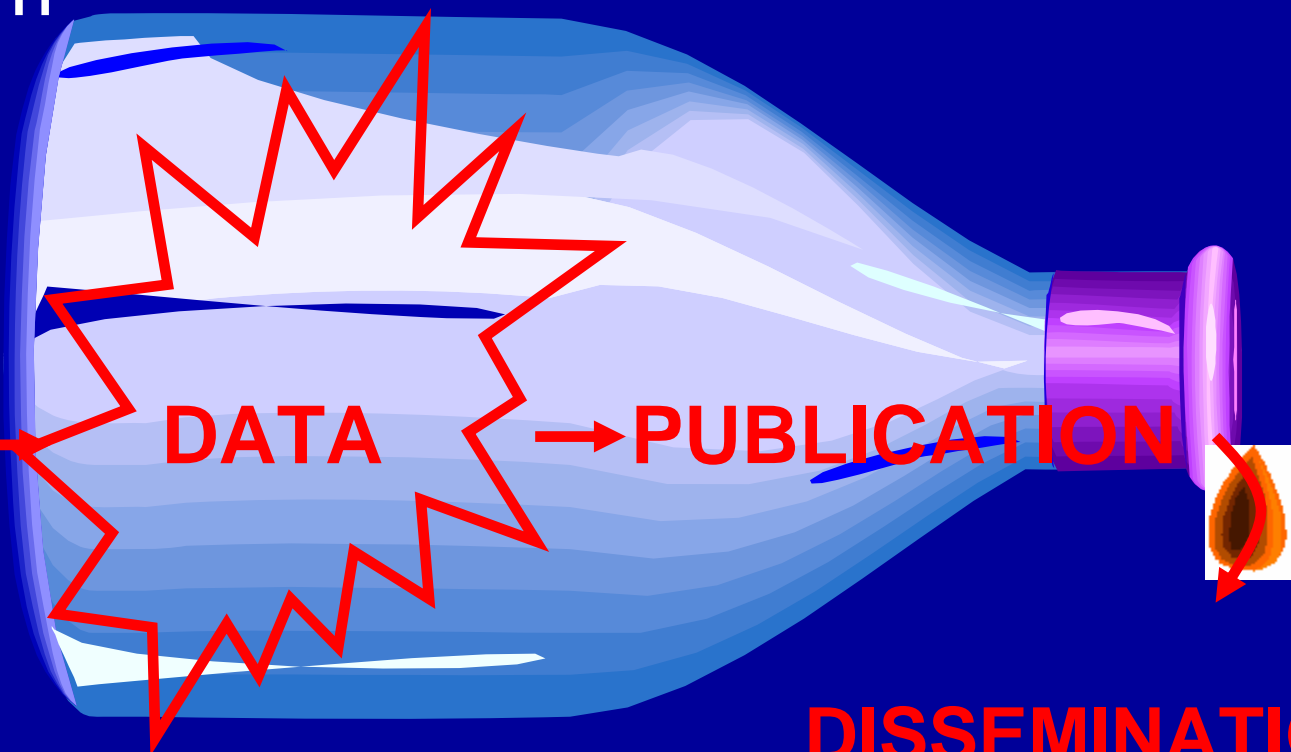
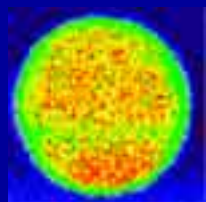


Entire E-Science Cycle
Encompassing
experimentation,
analysis, publication,
research, learning

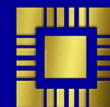


The need for *x/tl*-Prints

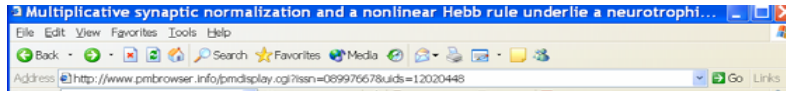
Combechem



Combichem

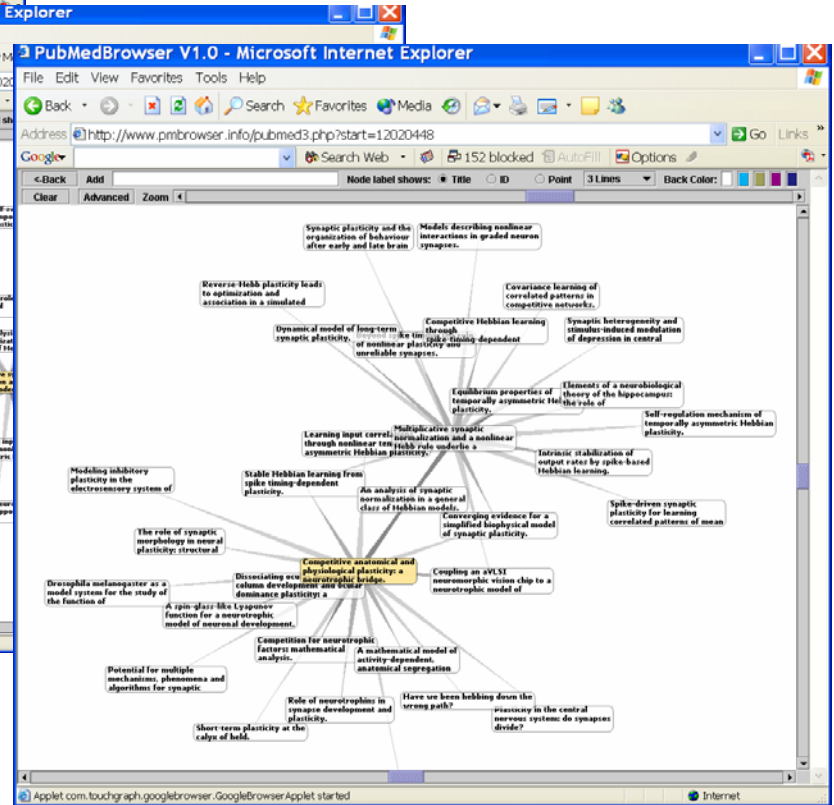


Increasing Use of Value Added Services



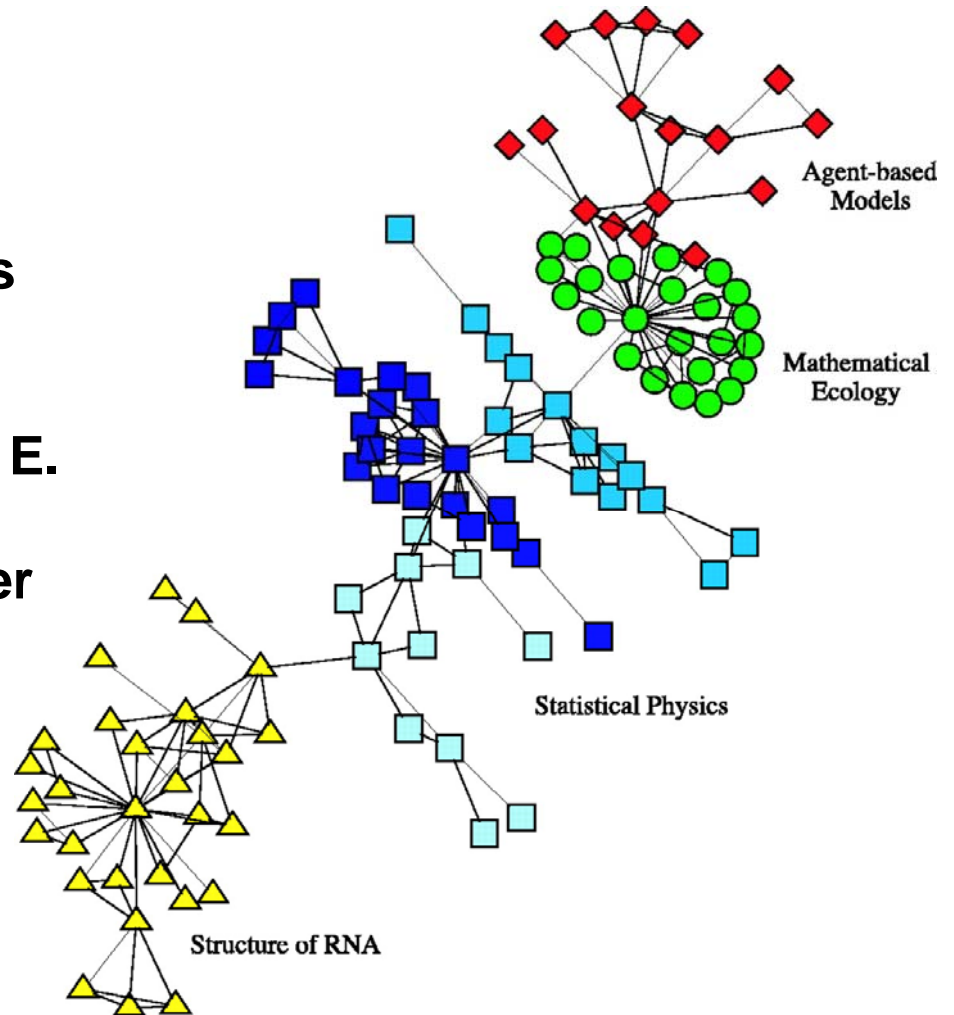
Multiplicative synaptic normalization and a nonlinear Hebb rule underlie a neurotrophin...
Elliott T. Shadloft NR
Neural Comput. 2002 Jun; 14(6): 1311-22

Synaptic normalization is used to enforce competitive dynamics in linear and semilinear Hebbian models, multiplicative activity patterns are positively correlated. To achieve this, normalization must be used instead. Our own model of competitive synaptic normalization, was developed in part in need for synaptic normalization altogether. However, we need two decoupled subspaces, with competitive dynamics being rule and multiplicative synaptic normalization. This normalization these observations permit biologically plausible forms of descriptions of the underlying biology in certain scaleless networks.



Communities of Authors

- An example of a small coauthorship network depicting collaborations among scientists at a private research institution. Newman, M. E. J. (2004)
- Web services to run over archives at varying grainsize





Hubs and Authorities

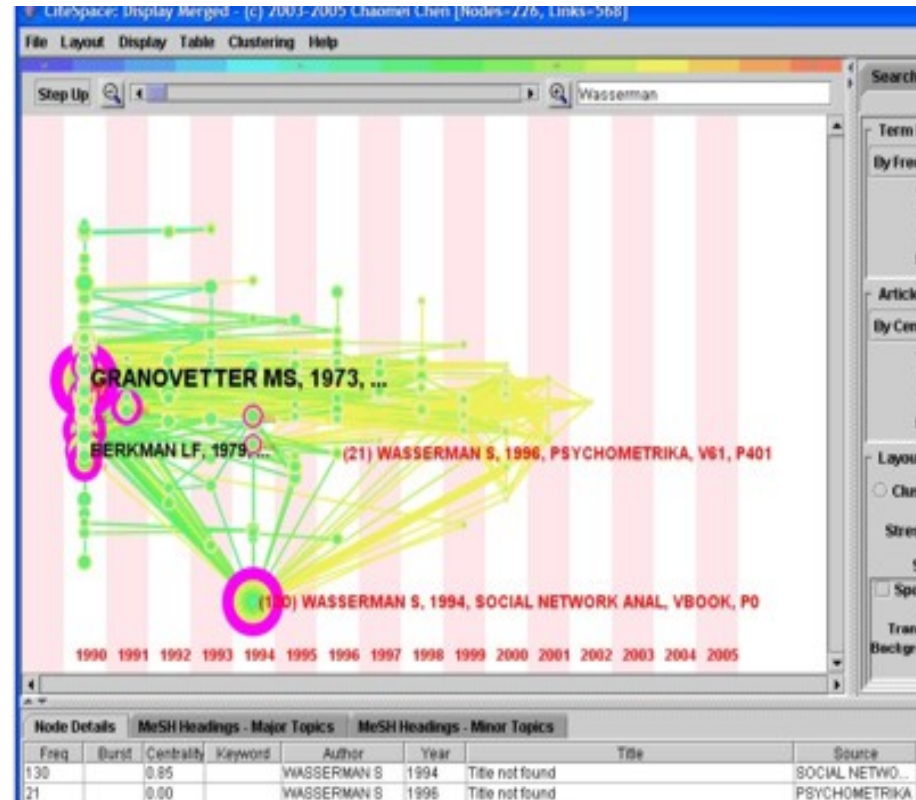
- Begin with existing measures: document count and citation count.
- Apply Kleinberg (1998) 'hubs/authorities' analysis to data.
- Note that higher citation count may not mean higher authority rating: quality citations are what count.

Author Name	Citeseer DocID	Citations	Authority Rating
Anant Agarwal	agarwal91mit	24	6.5161
Anant Agarwal	agarwal95mit	9	6.4675
Anant Agarwal	kranz93integrating	13	5.1927
Anant Agarwal	agarwal88evaluation	21	2.9159



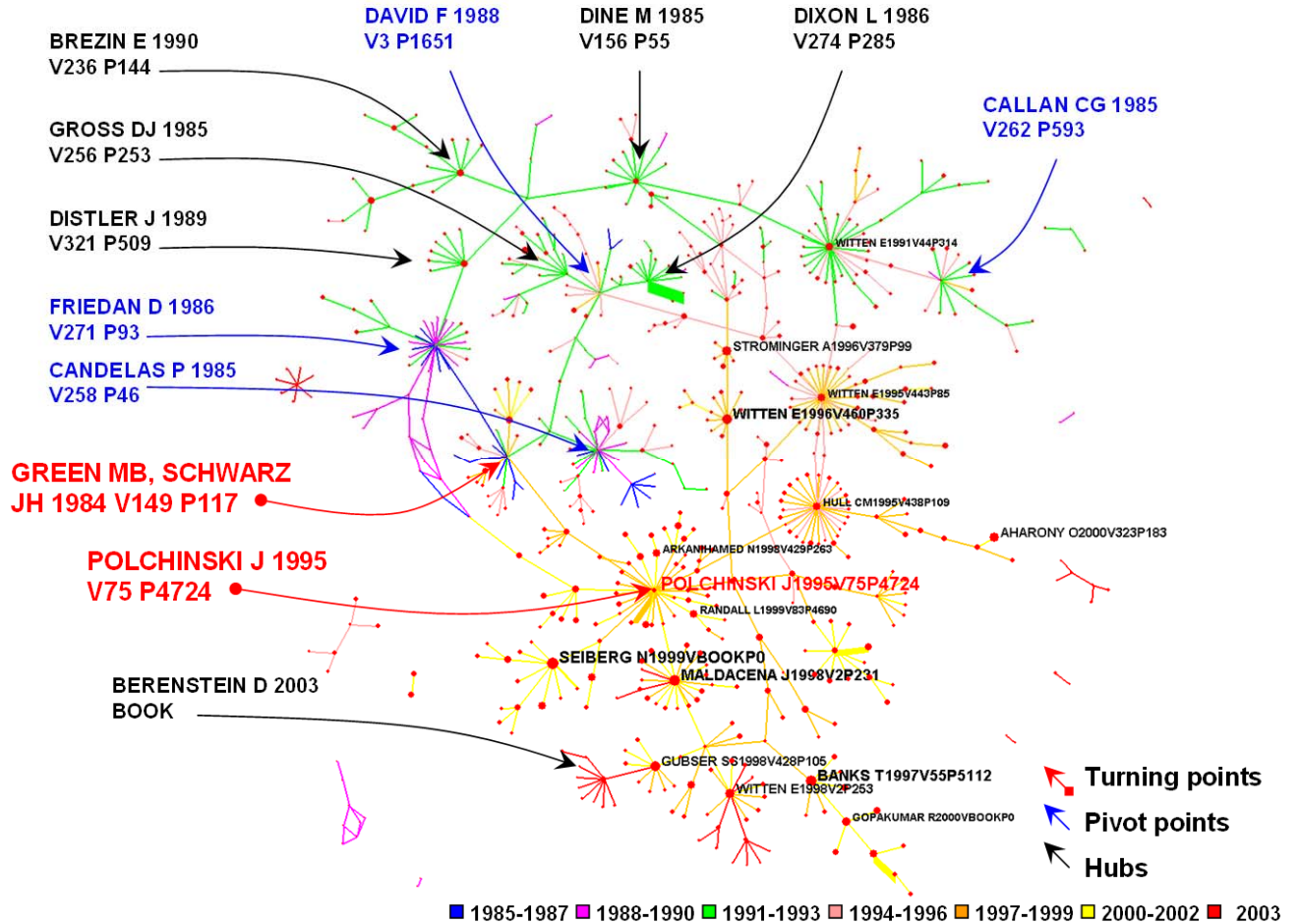
Turning Points and Centrality

- Allows us to find turning points in scientific development: Kuhn's "paradigm shift" moment.
- 'Centrality' measure to be applied to same Citeseer data.



Detecting Key Moments

→ Cause and effect chains in content sets





Future Challenges

- Developing, Deploying, Managing and Reusing Ontologies
- Co reference resolution and referential integrity on the SW
- “Inference” on the web
- The annotation bottleneck
- Annotating Multimedia Content
- Composing Knowledge Services
- Trust
 - Representing provenance
 - Mechanisms for trust



More Information

www.aktors.org

