



Intuitive and machine understandable representation of the bioinformatics domain and of related resources with Resourceomes

[Nicola Cannata](#), [Flavio Corradini](#), [Sergio Gabrielli](#), [Luana Leoni](#), [Emanuela Merelli](#), [Francesca Piersigilli](#), [Leonardo Vito](#)

- Mathematics and Computer Science Department, University of Camerino, Italy

Houston, we have a problem

Meaning

Originally a genuine report of a life-threatening fault. Now used humorously to report any kind of problem.



How many minutes (hours) do we spend every day in Google and bibliographic searches?

In life sciences and “-omics” disciplines
we are becoming used to deluges and overflows...

1: [Healthc Inform.](#) 2004 Mar;21(3):26-30.

Data overflow. Options are available to hold the tide of images and information.

[Hagland M.](#)

PMID: 15040049 [PubMed - indexed for MEDLINE]



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Biomedical Informatics 39 (2006) 314–320

Journal of
Biomedical
Informatics

www.elsevier.com/locate/yjbin

Beyond the data deluge: Data integration and bio-ontologies

Judith A. Blake *, Carol J. Bult

The Jackson Laboratory, Bar Harbor, ME, USA

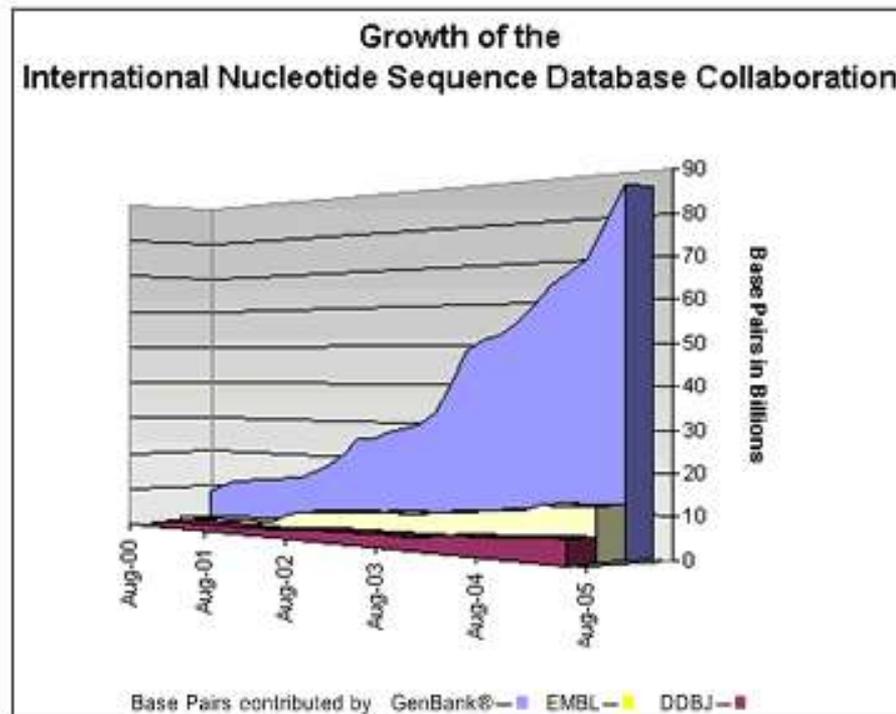
Received 31 August 2005

Available online 21 February 2006

The data overflow...

International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related [press release](#).



Genome Technology

Next-Gen Sequencing: The Waiting Game

June 2007

By Meredith Salisbury

Roche

- ◆ Platform cost: \$500,000
- ◆ Read length: 250 bp
- ◆ Cost per run: \$16,000
- ◆ Megabases per day: 200
- ◆ Cost per megabase: \$160

Illumina

- ◆ Platform cost: \$395,000
- ◆ Read length: 40-50 bp
- ◆ Cost per run: \$5,000
- ◆ Megabases per day: 333
- ◆ Cost per megabase: \$5

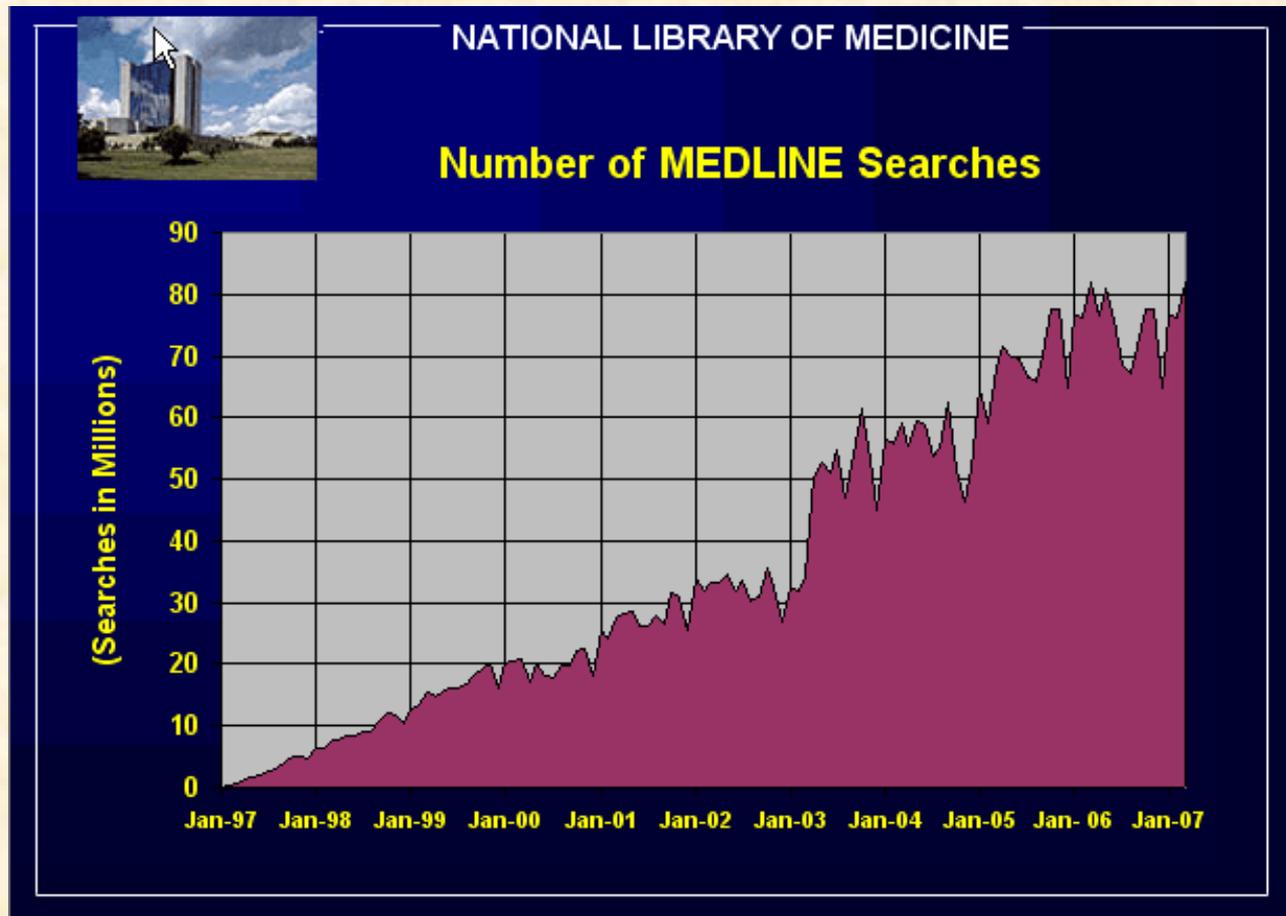
ABI 3730

- ◆ Platform cost: \$350,000
- ◆ Read length: 650+ bp
- ◆ Cost per run: \$55
- ◆ Megabases per day: 1.4
- ◆ Cost per megabase: \$880

Source: Elaine Mardis

Besides data overflow we are experimenting also **resources** overflow.

E.g. academic **Articles**,...



Now science is becoming e-science...

Expert Opinion

E-Science: The Grid and the Semantic Web

David De Roure, *University of Southampton*
James A. Hendler, *University of Maryland*

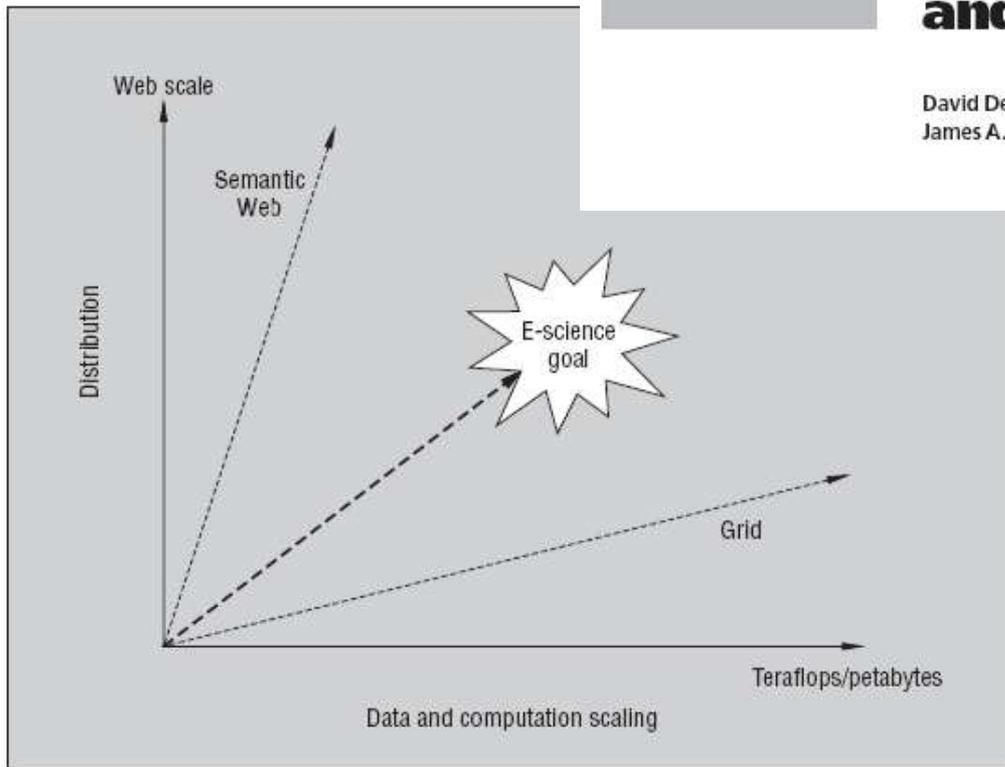


Figure 1. E-science needs aren't totally met by either the Grid or the Semantic Web. Rather, a Semantic Grid vision requires features of both.

The Future Interconnection Environment

Hai Zhuge
Chinese Academy of Sciences

Networks pervade nature, society, and virtual worlds, giving structure and function to a variety of resources and behaviors.

Discovering the rules that govern the future interconnection environment is a major challenge.

In 1960, Marvin Minsky predicted that computers would be as smart as humans within three to eight years. Nearly half a century later, however, computing systems still cannot pass the Turing test. Despite impressive achievements in robotics, mathematical theorem proving, scientific classification, and advanced user interfaces,¹ artificial intelligence remains elusive.

Scientists and engineers have nearly realized Vannevar Bush's dream of a universal multimedia data-processing machine with the Internet and the World Wide Web. Extending this vision into the future, Microsoft researcher Jim Gray foresees the development of highly secure, highly available, self-programming, self-managing, and self-replicating computer networks.² Gray imagines a system, akin to Bush's memex device, that can automatically organize, index, digest, evaluate, and abstract information. However, creating intelligent networks that can program, manage, and replicate themselves is a major challenge.

The China Knowledge Grid Research Group (<http://kg.ict.ac.cn>), established in 2001, is exploring the operating principles of this future interconnection environment.

Border between society and virtual environment

Limitation of change

Semantic grid

Knowledge grid

Present

Semantic Web

Web

Grid

Internet2

Internet

Expansion

Compression

Virtual environment

Local area network

Time

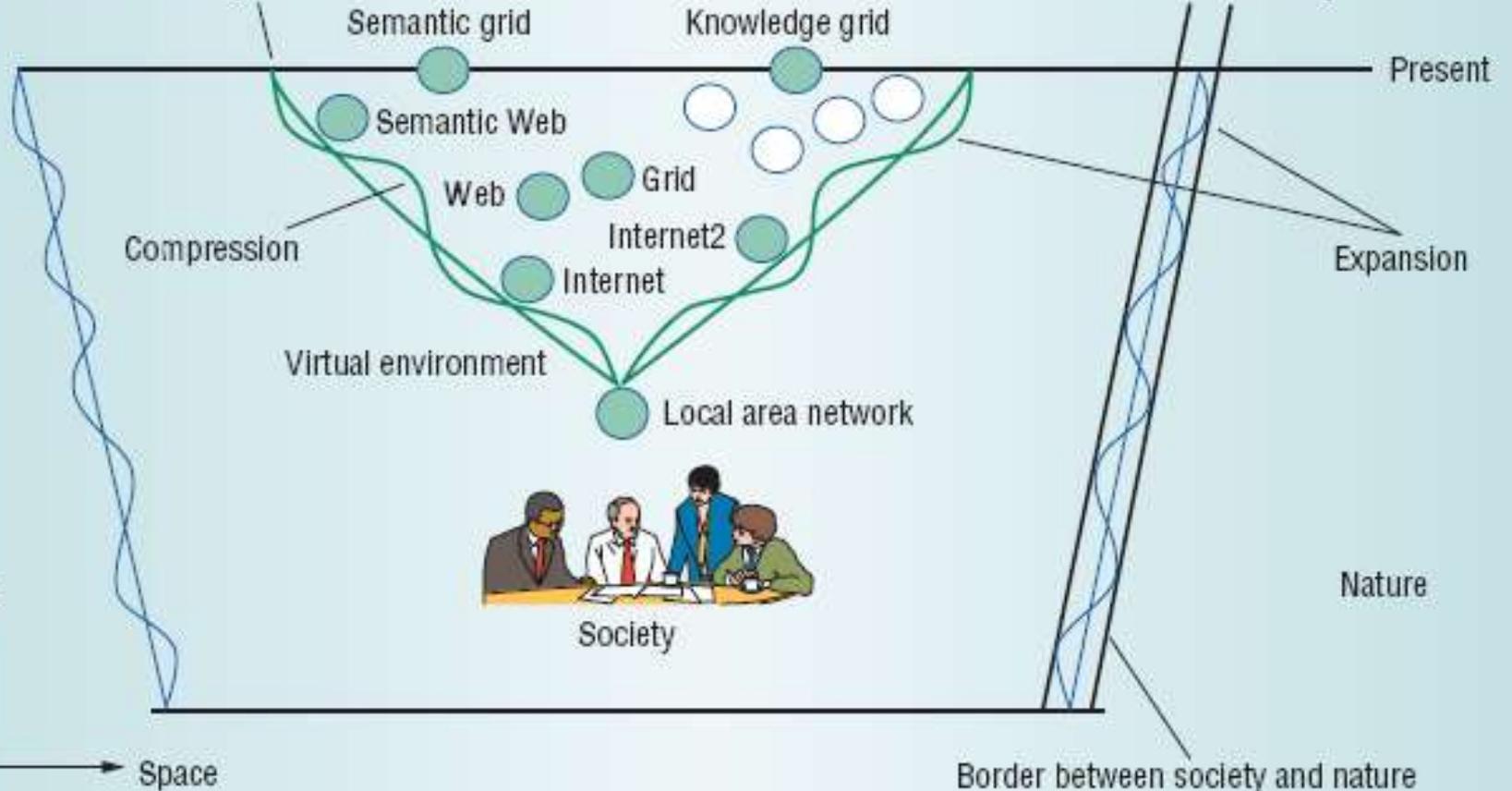
Nature

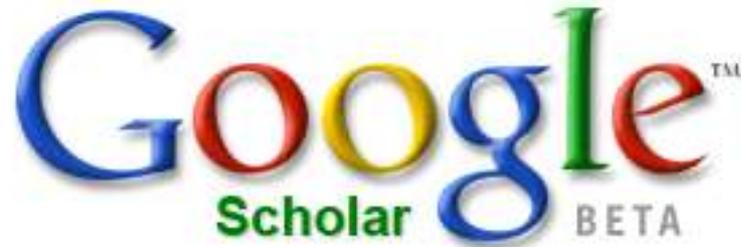


Society

Space

Border between society and nature





[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

Search

[Advanced Scholar Search](#)
[Scholar Preferences](#)
[Scholar Help](#)

Stand on the shoulders of giants

[Google Home](#) - [About Google](#) - [About Google Scholar](#)

©2007 Google

Will e-science become g-science...?

Databases are essential resources for bioinformatics

Published online 5 December 2006

Nucleic Acids Research, 2007, Vol. 35, Database issue **D3–D4**
doi:10.1093/nar/gkl1008

The Molecular Biology Database Collection: 2007 update

Michael Y. Galperin*

National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, Bethesda, MD 20894, USA

Received November 1, 2006; Revised and Accepted November 2, 2006

ABSTRACT

The NAR online Molecular Biology Database Collection is a public resource that contains links to the databases described in this issue of *Nucleic Acids Research*, previous NAR database issues, as well as a selection of other molecular biology databases that are freely available on the web and might be useful to the molecular biologist. The 2007 update includes 968 databases, 110 more than the previous one. Many databases that have been described in earlier issues of NAR come with updated summaries, which reflect recent progress and, in some instances, an expanded scope of these databases. The complete database list and summaries are available online on the *Nucleic Acids Research* web site <http://nar.oxfordjournals.org/>.

COMMENTARY

The current issue of the *Nucleic Acids Research* features 174 databases, of which 106 are new and 68 are updates of previ-

total database list, which again held very nicely and showed surprising resilience.

In the comment to the last year's release of the NAR database collection (1), I have discussed the citation rates for various papers in the 2004 NAR database issue and noted that the high-citation rate of certain databases reflects their worldwide acceptance as *de facto* standards of protein functional annotation [UniProt, <http://www.uniprot.org>, No. 318, Ref. 2], domain structure [<http://www.sanger.ac.uk/Software/Pfam/>, No. 210, Ref. 3] and biomedical terminology [Gene Ontology, <http://www.geneontology.org/>, No. 487, Ref. 4]. However, citation data can be biased; e.g. in many articles use of information from publicly available databases is acknowledged by providing their URLs, or not acknowledged at all. Besides, some databases could be cited on the web sites and in new or obscure journals, not covered by the ISI Citation Index. With this in mind, I have tried here to use additional metrics for assessing the popularity of the NAR database issue. First, I have checked the citations of the database papers listed on the Google Scholar web site, which reflects citations on the web sites. In addition, I have looked at the number of times that the full text of each paper (in PDF or HTML versions) was downloaded from the PubMed

Nucleic Acids Research

[ABOUT THIS JOURNAL](#) [CONTACT THIS JOURNAL](#) [SUBSCRIPTIONS](#)

[CURRENT ISSUE](#) [ARCHIVE](#) [SEARCH](#)

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > [Database Summary Paper Categories](#)

NAR Database Categories List

[Nucleotide Sequence Databases](#)
[RNA sequence databases](#)
[Protein sequence databases](#)
[Structure Databases](#)
[Genomics Databases \(non-vertebrate\)](#)
[Metabolic and Signaling Pathways](#)
[Human and other Vertebrate Genomes](#)
[Human Genes and Diseases](#)
[Microarray Data and other Gene Expression Databases](#)
[Proteomics Resources](#)
[Other Molecular Biology Databases](#)
[Organelle databases](#)
[Plant databases](#)
[Immunological databases](#)

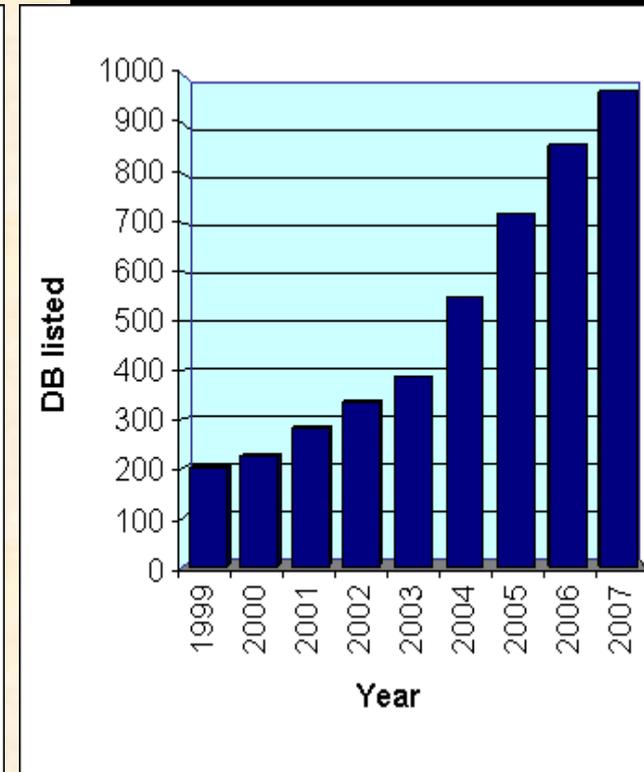
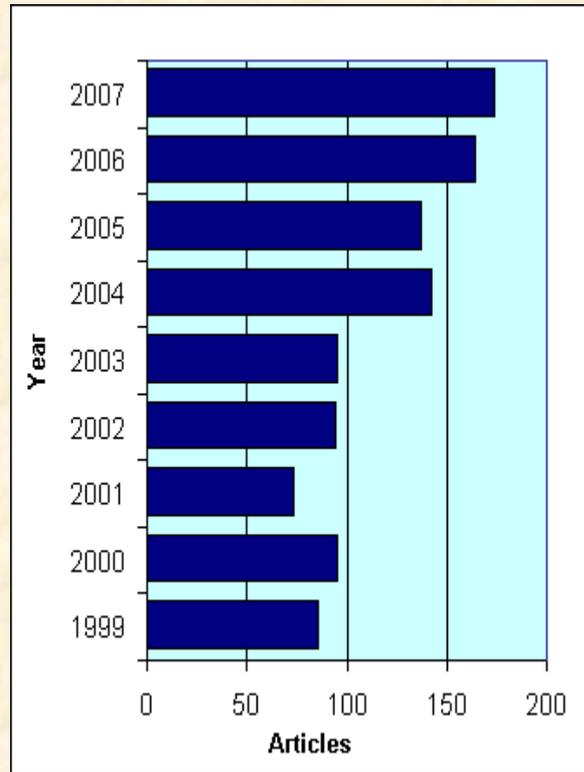
- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

Nucleic Acids Research (NAR) - Oxford Journals

Database special issues

Year	Articles	DB listed
2007	174	968
2006	164	858
2005	137	719
2004	142	548
2003	95	386
2002	94	335
2001	73	281
2000	95	226
1999	86	201



Web Servers are common tools for bioinformaticians

Nucleic Acids Research, 2006, Vol. 34, Web Server issue W3–W5
doi:10.1093/nar/gkl379

A compilation of molecular biology web servers: 2006 update on the Bioinformatics Links Directory

Joanne A. Fox, Scott McMillan and B. F. Francis Ouellette*

UBC Bioinformatics Centre, University of British Columbia, Michael Smith Laboratories, 2185 East Mall, Vancouver, BC, Canada V6T 1Z4

Received April 29, 2006; Accepted May 2, 2006

ABSTRACT

The Bioinformatics Links Directory is a public online resource that lists the servers published in this and all previously published *Nucleic Acids Research* Web Server issues together with other useful tools, databases and resources for bioinformatics and molecular biology research. This rich directory of tools and websites can be browsed and searched with all listed links freely accessible to the public. The 2006 update includes the 149 websites highlighted in the July 2006 issue of *Nucleic Acids Research* and brings the total number of servers listed in the Bioinformatics Links Directory to over 1000 links. To aid navigation through this growing resource, all link entries contain a brief synopsis, a citation list and are classified by function in descriptive biological categories. The most up-to-date version of this actively maintained listing of bioinformatics resources is available at the Bioinformatics Links Directory website, http://bioinformatics.ubc.ca/resources/links_directory/.

example, if you search for 'bioinformatics servers' using Google, you will face the daunting task of sorting through almost a million different websites. In an effort to provide clarity, *Nucleic Acids Research* has devoted several special issues to compiling molecular biology Web Servers (1) and Databases (2) helping researchers to quickly locate peer-reviewed tools and resources that directly apply to the changing bioinformatics landscape. Over the past 4 years, the NAR Web Server special issue has published a rich collection of over 500 different internet-based resources. This year, the 2006 Web Server issue highlights 149 bioinformatics and molecular biology servers that are all openly available to the world-wide research community. A complete listing of servers from the 2006 Web Server issue can be accessed online at http://bioinformatics.ubc.ca/resources/links_directory/narweb2006/ and in Supplementary Table 1. Together with the long standing Database issue (2), these special issues at *Nucleic Acids Research* represent a valuable directory of resources for the global life sciences research community.

The Bioinformatics Links Directory, http://bioinformatics.ubc.ca/resources/links_directory/, is a public resource that lists the servers published in this and previous issues of *Nucleic Acids Research* Web Server issues together with

NAR Web Server Issue (July 1, 2006)

 Jump to: 

[Links Directory Homepage](#)

Nucleic Acids Research

Source: http://nar.oxfordjournals.org/content/vol34/suppl_2/index.dtl

 [NAR 2006 RSS](#)

[list by category](#)

(PS)2 Protein Structure Prediction Server

(PS)2 Protein Structure Prediction Server performs automated homology modeling by combining PSI-BLAST, IMPALA, and T-Coffee for template selection and target-template alignment. The final three-dimensional (3D) structure is built using RAMP or MODELLER.

[Protein](#) > [3-D Structure Prediction](#)

[website](#) | 

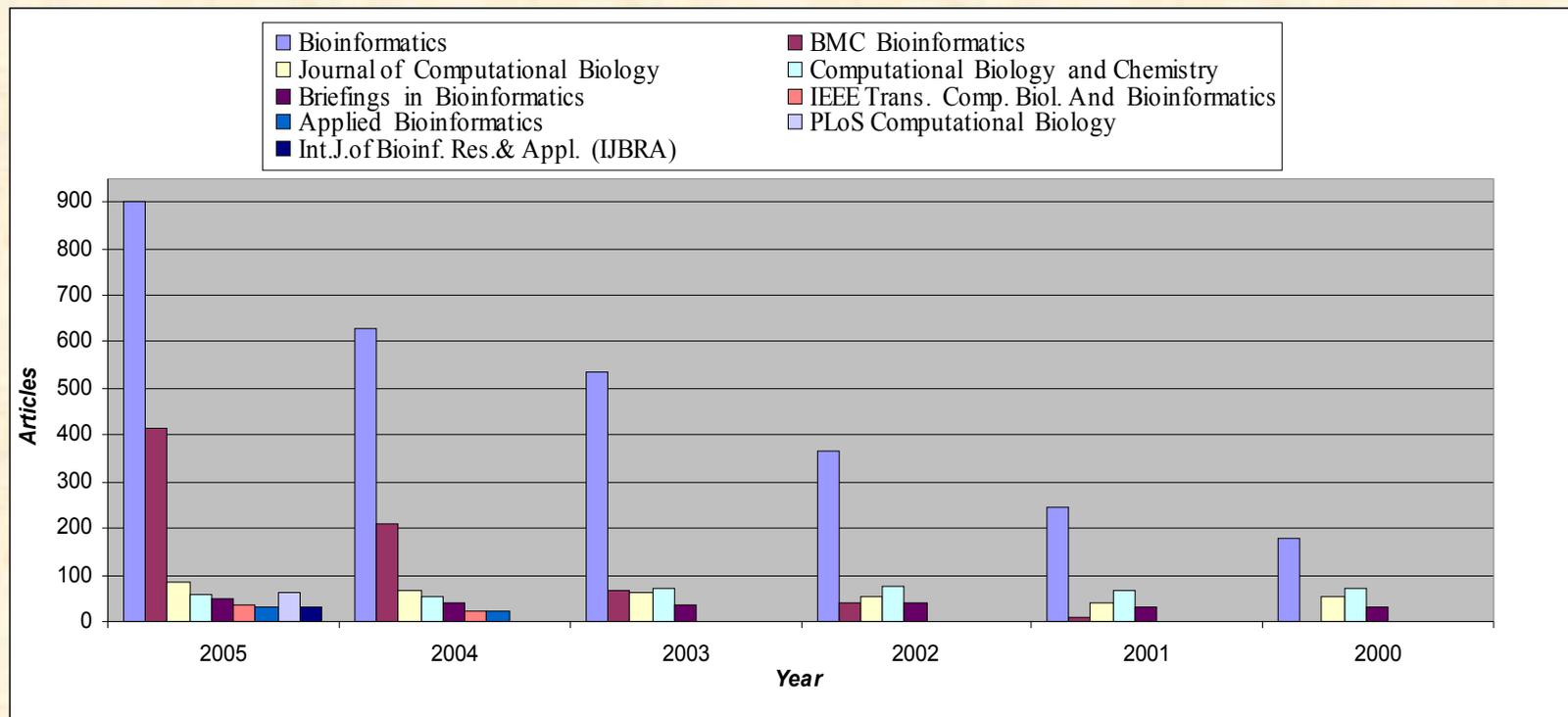
3DSS

3-Dimensional Structural Superposition (3DSS) is a tool for superposing two or more protein structures that uses RASMOL for visualization; some browser configuration is necessary.

[Protein](#) > [3-D Structure Comparison](#)
[Protein](#) > [3-D Structure Retrieval, Viewing](#)

[website](#) | 

Year	Bioinformatics	BMC Bioinformatics	Journal of Computational Biology	PLoS Computational Biology	Computational Biology and Chemistry	Briefings in Bioinformatics	IEEE Trans. Comp. Biol. And Bioinformatics	Applied Bioinformatics	Int.J.of Bioinf. Res.& Appl. (JBRA)
2005	900	414	83	61	58	47	34	32	32
2004	627	209	69		53	40	22	24	
2003	534	66	61		71	35			
2002	365	40	52		77	41			
2001	245	9	39		66	31			
2000	178	1	52		72	33			



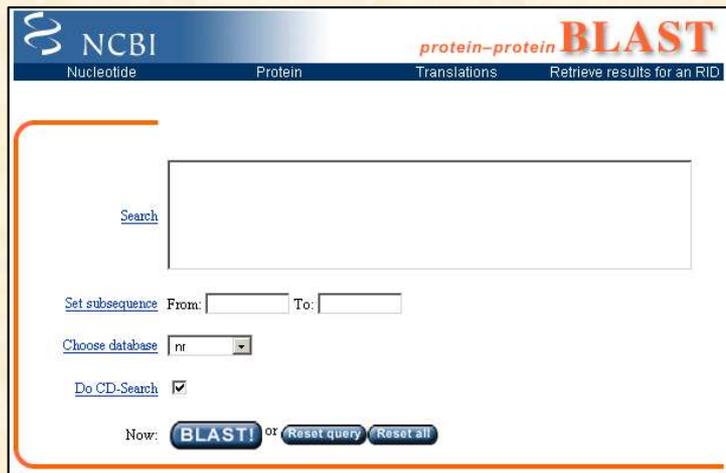
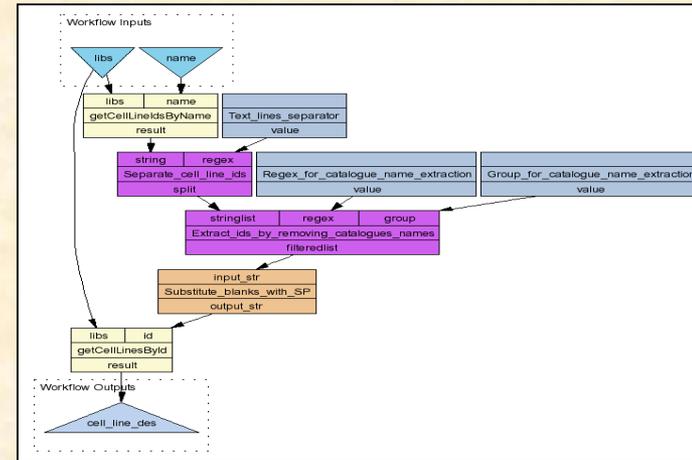
Articles published "ONLY" from main bioinformatics journals

Bioinformatics is evolving

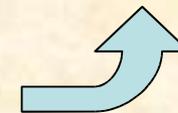
```
BLASTN 2.2.4 [Aug-26-2002]
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.
Query= gi|8777285|gb|AB044774.1|AB044774 AB044774 Panax ginseng hairy
root Panax ginseng cDNA clone HR36, mRNA sequence
(223 letters)
Database: /local/wwwstud/html/bioinfo3-64/EST-ginseng.fas
21 sequences; 7481 total letters
Searching done
Sequences producing significant alignments:
Score E
(bits) Value
gi|8777285|gb|AB044774.1|AB044774 AB044774 Panax ginseng hairy r... 442 e-128
gi|8777269|gb|AB044758.1|AB044758 AB044758 Panax ginseng hairy r... 24 0.077
gi|8777271|gb|AB044760.1|AB044760 AB044760 Panax ginseng hairy r... 22 0.30
>gi|8777285|gb|AB044774.1|AB044774 AB044774 Panax ginseng hairy root
Panax ginseng cDNA clone HR36, mRNA sequence
Length = 223
Score = 442 bits (223), Expect = e-128
Identities = 223/223 (100%)
Strand = Plus / Plus
Query: 1 ggctcccatgtttgttaaaaaatagtgtccctgctgtgctgaagattctggaatctatag 60
Sbjct: 1 ggctcccatgtttgttaaaaaatagtgtccctgctgtgctgaagattctggaacctatag 60
```

From command line... (till 90s)

“IN SILICO EXPERIMENTS”



To Web Services and workflows (now)



to web interfaces and (perl) scripts... (with the advent of the WWW)

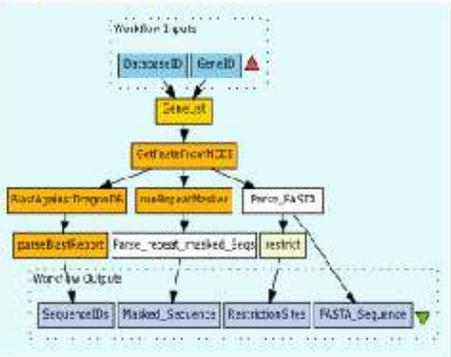
In this new bioinformatics **Web Services** play a key role. Programs interact with programs on the web...

Tired of This?

Protocol

Create a gene list in Excel
Go to NCBI
Retrieve FASTA for each gene
DragonDB Blast each sequence
Copy/paste IDs into a spreadsheet
Run Repeat Masker on each sequence
copy/paste masked sequence into Excel
Run MacVectors cut each seq with EcoRI

Try This!



Worldwide Distribution of MOBY Services



Archives

- » June 2007
- » April 2006

Categories

- » General (19)
- » News (1)

Downloads, Docs, and Stats

- » Code Releases
- » CVS Code Repository
- » General Docs
- » Currently 'dead' Services

RSS Feeds

feeds are updated hourly...

- » MOBY Services
- » MOBY Objects

Meta

- » Register
- » Login
- » WordPress

Why to search for resources?

- you have to develop a program, a database, possibly avoiding to re-invent the wheel...

- interdisciplinarity...

introduction to a new domain. Obtain a fast overview of a (new, for you) scientific domain (preferably in a visual fashion)

So, **where** to search for (bioinformatics) resources
-that of course you are not aware of - ?

In search engines?... Good luck!

 **Web** [Immagini](#) [Gruppi](#) [News](#) [Desktop](#) [altro »](#)

bioinformatics resources [Ricerca avanzata](#)
[Preferenze](#)

Cerca: il Web pagine in Italiano pagine provenienti da: Italia

Web Personalizzazione Risultati **1 - 10** su circa **2.850.000** per **bioinformatics resources**. (0,18 secondi)

[1.425 risultati memorizzati sul computer](#) - [Nascondi](#) - [Informazioni su](#)

 [Cannata2007-NETTAB.ppt](#) - the **bioinformatics** domain and of related **resources** with

Bioinformatics Resources - [[Traduci questa pagina](#)]
Resource for searching genetic databases, online tools, journals, and recent news.
[www.genet.sickkids.on.ca/bioinfo_resources/](#) - 2k -
[Copia cache](#) - [Pagine simili](#) - [Salva risultato](#)

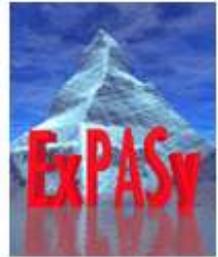
Bioinformatics Resources - [[Traduci questa pagina](#)]
Course materials, book recommendations and links related to **bioinformatics**.
[www.bioinformaticscourses.com/](#) - 9k - [Copia cache](#) - [Pagine simili](#) - [Salva risultato](#)

Biotech: **Bioinformatics** - [[Traduci questa pagina](#)]
Very good basic explanation of **Bioinformatics** - from the BioTech **Resources** Web Project at Indiana Univ and the Univ. of Texas.
[biotech.icmb.utexas.edu/pages/bioinfo.html](#) - 1k - [Copia cache](#) - [Pagine simili](#) - [Salva risultato](#)

2Can Support Portal: Home Page - [[Traduci questa pagina](#)]
The **bioinformatics** educational **resource**. This site provides short and concise introductions to basic concepts in molecular and cell biology and ...
[www.ebi.ac.uk/2can/](#) - 11k - [Copia cache](#) - [Pagine simili](#) - [Salva risultato](#)

Bioinformatics resources - [[Traduci questa pagina](#)]
Its aims are to make essential **bioinformatics** tools easily accessible for the study of the relationships between the sequence, structure and function of ...
[www.biochem.ucl.ac.uk/~nagl/bioinformatics/bioinform_res.htm](#) - 13k -
[Copia cache](#) - [Pagine simili](#) - [Salva risultato](#)

In specialized web sites? ... First, find them!



ExPASy Proteomics tools

The tools marked by  are local to the ExPASy server. The remaining tools are developed and hosted on other servers.

[Protein identification and characterization] [DNA -> Protein] [Similarity searches] [Pattern and profile searches] [Post-translational modification prediction]
[Topology prediction] [Primary structure analysis] [Secondary structure prediction] [Tertiary structure] [Sequence alignment] [Phylogenetic analysis]
[Biological text analysis]

Protein identification and characterization

Identification and characterization with peptide mass fingerprinting data

- **Aldente**  - Identify proteins with peptide mass fingerprinting data. A new, fast and powerful tool that takes advantage of Hough transformation for spectra recalibration and outlier exclusion
- **FindMod**  - Predict potential protein post-translational modifications and potential single amino acid substitutions in peptides. Experimentally measured peptide masses are compared with the theoretical peptides calculated from a specified Swiss-Prot entry or from a user-entered sequence, and mass differences are used to better characterize the protein of interest.
- **FindPept**  - Identify peptides that result from unspecific cleavage of proteins from their experimental masses, taking into account artefactual chemical modifications, post-translational modifications (PTM) and protease autolytic cleavage
- **GlycoMod**  - Predict possible oligosaccharide structures that occur on proteins from their experimentally determined masses (can be used for free or derivatized oligosaccharides and for glycopeptides)
- **Mascot** - Peptide mass fingerprint from Matrix Science Ltd., London
- **PepMAPPER** - Peptide mass fingerprinting tool from UMIST, UK
- **PFMUTS** - Shows the possible single and double mutations of a peptide fragment from MALDI peptide mass fingerprinting
- **ProFound** - Search known protein sequences with peptide mass information from Rockefeller and NY Universities [or from Genomic Solutions]
- **ProteinProspector** - UCSF tools for peptide masses data (MS-Fit, MS-Pattern, MS-Digest, etc.)

In SIG web sites? ... Are they updated?



Biomedical Literature (and text) Mining Publications

BLIMP - See it all from here!

A Forum for collection, compilation and exchange of publications on biomedical text mining

[Contents \(by categories\)](#) * [Search](#)

[Mission Statement](#) * [Submit an Article](#)

[Paper of the Month \(Editor's choice\)](#) * [Editorial Board](#)

[Useful Links: Other Bioinfo-Literature-Mining Resources](#) * [Contact BLIMP](#)

BLIMP covers all publications related to the fast-growing field of biomedical literature and text mining. It is a one-stop resource, letting researchers find out who-does-what in the area and where it is published, bridging across the many discipline-specific venues in which biomedical text-mining papers are published.

BLIMP is committed to both coverage and relevance. It started with a large collection of publications, and constantly relies on authors to notify BLIMP through the [submission form](#) about their newly accepted papers. Relevance of submission is determined by the [editorial board](#) and each relevant paper is promptly added to the website.

[NLP and Ontologies in Biomedicine](#)

[Information Extraction in Biomedicine](#)

[Information Retrieval in Biomedicine](#)

[Text Categorization in Biomedicine](#)

[Hybrid Methods and Combined Data Sources](#)

[Background on NLP](#)

[Background on Information Extraction](#)

[Background in Information Retrieval and Text Categorization](#)

[Background on Text Mining](#)

[Reviews on Text Mining in Biomedicine](#)

[Other](#)

**8238 Visitors
Since February 2005**

Copyright 2004: BLIMP, Queen's University

The BLIMP project was partially funded by NSERC Discovery Grant # 298292-04, and by Queen's University Principal's Development Award

Last update on: July 25, 2005

Updated By: Anirudh Garg

Created By: Limin Zheng

The BioLink Homepage

Biological Literature, Information and Knowledge

· [CONTACT](#) ·

About BioLINK:

With the increasing availability of textual information related to biology, including [MedLine](#) abstracts and full-text journal articles, research on information extraction is rapidly becoming an essential component of various bioinformatics applications. It is expected that text mining in general, and information extraction in particular, will provide tools to facilitate the annotation of vast amounts of molecular information, including gene sequences, transcription profiles and biological pathways.

The Special Interest Group on Text Mining (or BioLINK) was created to address the need of communication and interchange of ideas in the field of text mining and information extraction applied to biology and biomedicine. Information extraction (IE) is an outgrowth of work in automated natural language processing, which began in the 1950s with work on transformational grammar by Zellig Harris and later Noam Chomsky. Information extraction technology made rapid progress starting in the late 1980s, thanks to a series of conferences focused on evaluation of IE: the Message Understanding Conferences (MUCs). There is also a long history of research on applications in medicine. Applications to the medical field focus on two distinct sub-problems: improved access to the medical literature and extraction of information from patient records.

Despite these successes in other fields Natural Language Processing (NLP) techniques were not introduced in biology until the late 90's (first publication around 1997 at [ISMB97](#)). The field is dominated by two, not necessarily convergent, views.



last update: 2003/07/15

© by Martin Krallinger



[Home](#)

[Key Dates](#)

[Program](#)

[Accepted Papers](#)

[Accepted Posters](#)

[Agenda](#)

[Paper Submission](#)

[Call for participation](#)

[Poster guidelines](#)

[Registration](#)

[Organization](#)

[Program Committee](#)

[Mailing List](#)

[Contact](#)

[Main Conference](#)

[ISMB'07](#)

BioLINK SIG: Linking Literature, Information and Knowledge for Biology

The Annual Meeting of
The ISMB BioLINK Special Interest Group on Text Data Mining

In Association with [ISMB 2007](#), Vienna, Austria

Thursday, July 19, 2007

With the increasing availability of textual information related to biology including Medline abstracts and full-text journal articles, the field of biomedical text mining has been rapidly maturing. It is concerned with using techniques from natural language processing, information extraction and information retrieval to automate knowledge discovery from biomedical text.

The BioLINK SIG meeting has been regularly held in association with the ISMB conference since 2001. The SIG focuses on the development and application of resources and tools for biomedical text mining. It is interdisciplinary in nature, and brings researchers applying natural language processing, text mining, information extraction and retrieval in the biomedical domain, together with scientists from bioinformatics and biology.

In the literature? ... When do you need them?

PubMed Update

Chart last updated: June 4, 2007

Data	Total Records	Dates Covered	Indexed Records Added	Date Added
PubMed	17,088,520	1951-present		
MEDLINE	15,883,894	1966-present	11,364	2007/05/28-2007/06/01
In Process	253,988			
OLDMEDLINE	420,961	1950-1965		2006/11/15

MEDLINE is generally updated each day, Tuesday-Saturday, with indexed citations. PubMed includes in-process citations which are added Tuesday - Saturday. These records are tagged [PubMed - in process]. Citations received electronically from publishers are tagged [PubMed -as supplied by publisher].

And also when articles are detected, maybe the presented resources are not there anymore!



404 not found: the stability and persistence of URLs published in MEDLINE

Jonathan D. Wren

Advanced Center for Genome Technology, Department of Botany and Microbiology,
The University of Oklahoma, 620 Parrington Oval Rm. 106, Norman, OK 73019, USA

Received on June 18, 2003; accepted on June 25, 2003

Advance Access publication January 22, 2004

ABSTRACT

Motivation: The advent of the World Wide Web has enabled unprecedented supplementation of traditional journal publications, allowing access to resources, such as video, sound, software, databases, datasets too large to publish, and even supplementary information and discussion. However, unlike traditional publications, continued availability of these online resources is not guaranteed. An automated survey was conducted to quantify the growth in Uniform Resource Locators (URLs) published to date in MEDLINE abstracts, their current availability and distribution by journal.

Results: Of 1630 unique URLs identified, formatting and/or spelling errors were detected within 201 (12%) of them as published. After corrections were made, a survey revealed that ~63% of these URLs were consistently available, and another 19% were available intermittently. The rate of failure was far worse for anonymous login to FTP sites, with only 12 of 33 sites (36%) responding. This survey also shows that journals vary disproportionately in the number of web citations published, suggesting policy implementation among a few could have a profound impact overall. Out of the 306 journals with a URL published in an abstract, *Bioinformatics* published the most (12% of total).

Availability of URLs published in MEDLINE

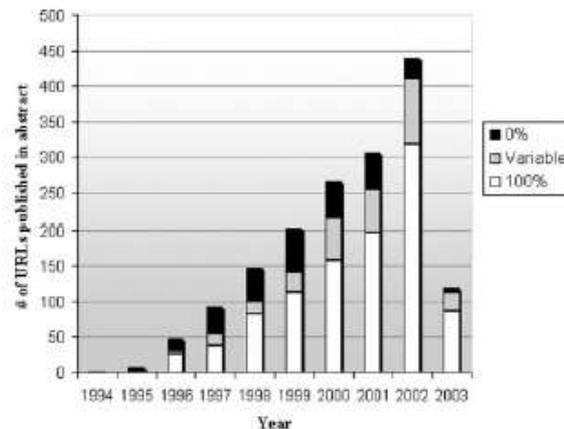


Fig. 1. The number of unique URLs published in MEDLINE, plotted as a function of time. The proportion of URLs currently accessible (as of this survey) is broken down by year and displayed as shading on each bar. Online resources are playing an increasingly important role in scientific research, as evidenced by their inclusion in the abstract. Not surprisingly, the more recent the publication, the more likely the

In the age of WWW resources appear... and disappear

POLICY FORUM

INFORMATION SCIENCE

Going, Going, Gone: Lost Internet References

Robert P. Dellavalle,^{1,2,3*} Eric J. Hester,² Lauren F. Heilig,²
Amanda L. Drake,² Jeff W. Kuntzman,⁴ Marla Graber,⁴ Lisa M. Schilling⁵

Internet references in medical and scientific periodicals may become more common as 7 million pages of new information, including data not available elsewhere, appear daily on the World Wide Web (WWW) (1). The Internet, of which the Web is part, consists of a worldwide system of computer networks. The Internet promotes easy access to and revision of data and allows information formats not suitable for print media including high-resolution images, motion video, animations, simulations, and program source code. However, unlike hard copy references, Internet references may change and become inaccessible (2–6).

Nearly 20% of Internet addresses in a Web-rich high school science curriculum became inactive between August 2002 and March 2003. Addresses with “.com” and “.edu” top-level domains most frequently became inactive (6, 7). In another study, 108 of 184 Internet addresses for an herbal remedy, *Opuntia*, became inactive within 4 years (8). Furthermore, no consensus on Internet reference format exists, especially with regard to providing accession (or citation) dates that report when authors viewed the information online (9–11).

the 6-week period immediately antecedent to the initiation of the research on 16 January 2003. For comparison, issues from the same calendar period, 1 December to 15 January, for the preceding 2 years were also examined. A reference was defined as a numbered citation appearing at the end of an article. Advertisements and articles without references were excluded. Internet reference categories were (i) active Internet reference—Internet address citation yielding information other than an error message when accessed via an Internet browser (e.g., Internet Explorer) and (ii) inactive Internet reference—Internet address citation yielding an error message when accessed via an Internet browser.

Each inactive Internet reference was further categorized, by using www.archive.org and www.google.com, as a (i) recoverable Internet reference—inactive Internet address citation yielding recoverable information with Internet archiving systems or (ii) unrecoverable Internet reference—inactive Internet address yielding no recoverable information. The type of referenced Internet material (e.g., PDF, unpublished document, or conference proceedings) was recorded. A comparison was made between these re-

10% at 15 months and to 13% at 27 months after publication (13) (fig. S1). For articles 27 months old, *JAMA* had the greatest Internet reference inactivity (21%) compared with *NEJM* (13%) and *Science* (11%).

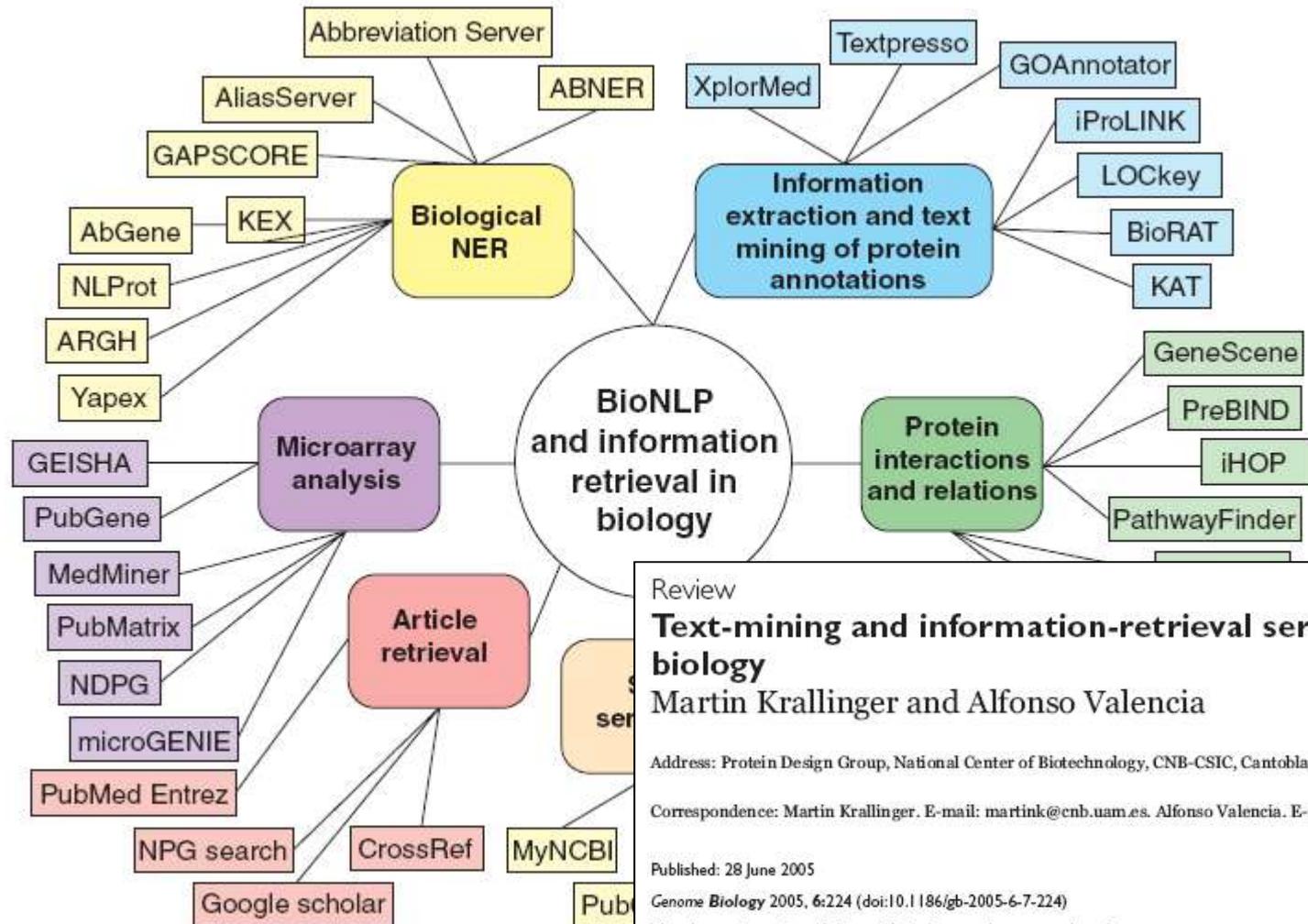
Inactive Internet references were most commonly “.com” addresses (46% lost after 27 months) followed by “.edu” (30%), other (20%), “.gov” (10%) and “.org” (5%) (see table). Book reviews had the greatest loss (17%) and opinion and news articles, the least (8%) (13) (table S1).

In contrast to *JAMA* and *NEJM*, some articles in *Science* contained Internet references directing readers to *Science*'s Web sites for supplemental material not published in the hard copy journal. These self-references accounted for approximately 1% of Internet references in *Science*. Excluding self-references raised the loss of Internet references in *Science*'s 27-month-old articles from 11.0 to 11.5%.

In September 2003, *NEJM*'s and *JAMA*'s, but not *Science*'s, instructions to authors requested accession dates for an Internet reference. Over the study period, accession dates became more common in *NEJM* (0 to 93% of Internet reference citations), less common in *JAMA* (100 to 73%), and changed minimally in *Science* (1 to 0%).

Although our study may not be representative of the entire field (13), Internet references occurred frequently and were often inaccessible within months after publication in the highest-impact U.S. medical and scientific journals. The problem of impermanent Internet references

If you are lucky you can find some nice and intuitive reviews



Review

Text-mining and information-retrieval services for molecular biology

Martin Krallinger and Alfonso Valencia

Address: Protein Design Group, National Center of Biotechnology, CNB-CSIC, Cantoblanco, E-28049 Madrid, Spain.

Correspondence: Martin Krallinger. E-mail: martink@cnb.uam.es. Alfonso Valencia. E-mail: valencia@cnb.uam.es

Published: 28 June 2005

Genome *Biology* 2005, 6:224 (doi:10.1186/gb-2005-6-7-224)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/7/224>

© 2005 BioMed Central Ltd

What's with the "staying updated"?

-TOC e-mail services

-Subject alert services

W16–W19 *Nucleic Acids Research*, 2004, Vol. 32, Web Server issue
DOI: 10.1093/nar/gkh453

PubCrawler: keeping up comfortably with PubMed and GenBank

Karsten Hokamp* and Kenneth H. Wolfe¹

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada and

¹Department of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

Received February 15, 2004; Revised and Accepted April 21, 2004

ABSTRACT

The free PubCrawler web service (<http://www.pubcrawler.ie>) has been operating for five years and so far has brought literature and sequence updates to over 22 000 users. It provides information on a personalized web page whenever new articles appear in PubMed or when new sequences are found in GenBank that are specific to customized queries. The server also acts as an automatic alerting system by sending out short notifications or emails with the latest updates as soon as they become available. A new output format and more flexibility for the email formatting help PubCrawler cope with increasing challenges arising from browser incompatibilities and mail filters, therefore making it suitable for a wide range of users.

Table 1. A list of free literature update alerting services

Service	URL
Amedeo	http://www.amedeo.com
BioMail	http://biomail.sourceforge.net/biomail
JADE	http://www.biodigital.org/jade
PubCrawler	http://www.pubcrawler.ie
PubMed Cubby	http://www.pubmed.gov
ScienceDirect	http://www.sciencedirect.com

that user, leaving only the new items to be compiled into a web page that closely resembles the look and feel of the familiar Entrez pages. Alerting occurs by email through short notifications or delivery of the complete results.

A number of other selective dissemination of information (SDI) services exist, both commercial and free to the public. Some examples include PubMed Cubby, BioMail, JADE, OVID and ScienceDirect (Table 1). Together with PubCraw-

Resource directories are good place to start...



Bioinformatics

[Science](#) > [Biology](#) > Bioinformatics

[Go to Directory Home](#)

Categories

- | | | |
|----------------------------------|--|--------------------------------------|
| Companies (146) | Hardware (1) | Programming (18) |
| Directories (14) | Human Genomics (224) | Publications (10) |
| Education (47) | Molecular Evolution (12) | Research Groups (31) |
| Employment (2) | Online Services (118) | Software (65) |

Related Categories:

- [Health](#) > [Medicine](#) > [Informatics](#) (338)
- [Science](#) > [Biology](#) > [Biochemistry and Molecular Biology](#) > [Biomolecules](#) > [Proteins and Enzymes](#) > [Proteomics](#) (159)
- [Science](#) > [Environment](#) > [Biodiversity](#) > [Informatics](#) (17)

Web Pages

Viewing in Google PageRank order

[View in alphabetical order](#)

- [The Swiss Institute of Bioinformatics Homepage \(SIB\)](#) - <http://www.isb-sib.ch/>
SIB operates the ExPASy proteomics server and the Swiss node of EMBnet. Teaching activities include a series of post-graduate courses given at the Universities of Geneva and Lausanne, as well as at the EPFL, and a Masters Degree in bioinformatics. Major research areas include the development of integrated databases and software resources in the field of proteomics.
- [The International Society for Computational Biology](#) - <http://www.iscb.org>
The International Society for Computational Biology is dedicated to advancing the scientific understanding of living systems through computation; the emphasis is on the role of computing and informatics in advancing molecular biology.
- [The Open Lab](#) - <http://bioinformatics.org/>
A community focused on the freedom of information as it pertains to the biosciences.
- [The Ensembl Project](#) - <http://www.ensembl.org/>

Bioinformatics Links Directory

The Bioinformatics Links Directory features curated links to molecular resources, tools and databases. All of the resources are free or available for a nominal fee. The links listed in this directory are selected on the basis of recommendations from bioinformatics experts in the field. We make every effort to list useful, well tested resources instead listing all tools. We also rely on input from our community of bioinformatics users for [suggestions](#). Starting in 2003, we have also started listing all links contained in the NAR Webserver issue. These NAR links are indicated with a NAR Webserver icon in the directory. The UBC Bioinformatics Centre does not specifically endorse or support any of the links in the directory, nor are we responsible for the content of any of these external sites.

 Jump to: 

[Computer Related](#) (64)

This category contains links to resources relating to programming languages often used in bioinformatics. Other tools of the trade, such as web development and database resources, are also included here.

[DNA](#) (392)

This category contains links to useful resources for DNA sequence analyses such as tools for comparative sequence analysis and sequence assembly. Links to programs for sequence manipulation, primer design, and sequence retrieval and submission are also listed here.

[Education](#) (76)

Links to information about the techniques, materials, people, places, and events of the greater bioinformatics community. Included are current news headlines, literature sources, educational material and links to bioinformatics courses and workshops.

[Expression](#) (241)

Links to tools for predicting the expression, alternative splicing, and regulation of a gene sequence are found here. This section also contains links to databases, methods, and analysis tools for protein expression, SAGE, EST, and microarray data.

[Human Genome](#) (114)

This section contains links to draft annotations of the human genome in addition to resources

[Model Organisms](#) (191)

Included in this category are links to resources for various model organisms ranging from mammals to microbes. These include databases and tools for genome scale analyses.

[Other Molecules](#) (12)

Bioinformatics tools related to molecules other than DNA, RNA, and protein. This category will include resources for the bioinformatics of small molecules as well as for other biopolymers including carbohydrates and metabolites.

[Protein](#) (687)

This category contains links to useful resources for protein sequence and structure analyses. Resources for phylogenetic analyses, prediction of protein features, and analyses of interactions are also found here.

[RNA](#) (92)

Resources include links to sequence retrieval programs, structure prediction and visualization tools, motif search programs, and information on various functional RNAs.

[Sequence Comparison](#) (210)

Tools and resources for the comparison of sequences including sequence similarity



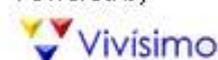
OBRC: Online Bioinformatics Resources Collection

OBRC

- ◆ [Email Suggestions](#)
- ◆ [Recommend a New Resource](#)

The Online Bioinformatics Resources Collection (OBRC) contains annotations and links for 1746 bioinformatics databases and software tools.

Powered by



Search examples: [transcription factors](#), [promoters](#), [RNAi](#)

Browse:

- ◆ [DNA Sequence Databases and Analysis Tools](#) (366)
- ◆ [Enzymes and Pathways](#) (147)
- ◆ [Gene Mutations, Genetic Variations and Diseases](#) (170)
- ◆ [Genomics Databases and Analysis Tools](#) (361)
- ◆ [Immunological Databases and Tools](#) (41)
- ◆ [Microarray, SAGE, and other Gene Expression](#) (121)
- ◆ [Organelle Databases](#) (27)
- ◆ [Other Databases and Tools \(Literature Mining, Lab Protocols, Medical Topics, and others\)](#) (105)
- ◆ [Plant Databases](#) (94)
- ◆ [Protein Sequence Databases and Analysis Tools](#) (316)
- ◆ [Proteomics Resources](#) (40)
- ◆ [RNA Databases and Analysis Tools](#) (147)
- ◆ [Structure Databases and Analysis Tools](#) (256)

But the amount of resources and their variety require that directories would be machine understandable

May 17, 2001 ¶

¶ The Semantic Web ¶

¶
A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities ¶

¶
By Tim Berners-Lee, James Hendler and Ora Lassila ¶

¶
The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other *local* devices that had a *volume control*. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring. ¶



¶
At the doctor's office, Lucy instructed her Semantic Web agent through her handheld Web browser. The agent promptly retrieved information about Mom's *prescribed treatment* from the doctor's agent, looked up several lists of *providers*, and checked for the ones *in-plan* for Mom's insurance within a *20-mile radius* of her *home* and with a *rating* of *excellent* or *very good* on trusted rating services. It then began trying to find a match between available *appointment times* (supplied by the agents of individual providers through their Web sites) and

Intelligent software agents will then be able to “reason” on the resources and to easily find them for you

Engineering in Genomics

the emerging in-silico scientist:
how text-based bioinformatics is bridging
biology and artificial intelligence

By Jonathan D. Wren

Perspective

Time to Organize the Bioinformatics Resourceome

Nicola Cannata, Emanuela Merelli, Russ B. Altman*

We will be witnessing the birth of the artificial, or in-silico, scientist. — J. D. Wren [1]

The field of bioinformatics has blossomed in the last ten years, and as a result, there is a large and increasing number of researchers generating computational tools for solving problems relevant to biology. Because the number of artifacts has increased greatly, it is impossible for many bioinformatics researchers to track tools, databases, and methods in the field—or even perhaps within their own specialty area. More critically, however, biologist users and scientists approaching the field do not have a comprehensive index of bioinformatics algorithms, databases, and literature annotated with information about their context and appropriate use. We suggest that the full set of bioinformatics resources—the “resourceome”—should be explicitly characterized and organized. A hierarchical and machine-understandable organization of the field, along with rich cross-links (an ontology!) would be a useful start. It is likely that a distributed development approach would be required so that those with focused expertise can classify resources in their area, while providing the metadata that would allow easier access to useful existing resources.

keyword searching [5]. However, the lack of standard terms makes sensitive and specific searches difficult. In addition, most search hits confound papers, Web sites, tools, departments, and people in a manner that makes extracting useful information very difficult.

Recognizing this limitation, there have been some grassroots attempts to organize the bioinformatics resourceome. Among the most famous are the “archaeological” Pedro’s List—a list of computer tools for molecular biologists (http://www.public.iastate.edu/~pedro/research_tools.html)—and the Expasy Life Sciences Directory, formerly known as the Amos’s WWW links page (<http://www.expasy.org/links.html>). The Bioinformatics Links Directory (http://www.bioinformatics.ubc.ca/resources/links_directory/) today contains more than 700 curated links to bioinformatics resources, organized into eleven main categories, including all the databases and Web servers yearly listed in the dedicated *Nucleic Acids Research* special issues [6]. The National Center for Biotechnology Institute has tried to make access to its suite of tools transparent, with moderate success. Many Web sites can be found listing “useful sites,” especially concerning special interest or limited topics (e.g., microarrays, text mining, and gene regulation). But all of these efforts are limited by the difficulty in maintaining

Two “orthogonal” classifications

- resources should be classified according their nature (a program, a database, a paper, a person...)
and according what they refer to, what they are for

Computer science subjects can be classified...

Top Two Levels of The ACM Computing Classification System (1998)

- A. General Literature
 - A.0 GENERAL
 - A.1 INTRODUCTORY AND SURVEY
 - A.2 REFERENCE (e.g., dictionaries, encyclopedias, glossaries)
 - A.m MISCELLANEOUS
- B. Hardware
 - B.0 GENERAL
 - B.1 CONTROL STRUCTURES AND MICROPROGRAMMING (D.3.2)
 - B.2 ARITHMETIC AND LOGIC STRUCTURES
 - B.3 MEMORY STRUCTURES
 - B.4 INPUT/OUTPUT AND DATA COMMUNICATIONS
 - B.5 REGISTER-TRANSFER-LEVEL IMPLEMENTATION
 - B.6 LOGIC DESIGN
 - B.7 INTEGRATED CIRCUITS
 - B.8 PERFORMANCE AND RELIABILITY **NEW!** (C.4)
 - B.m MISCELLANEOUS
- C. Computer Systems Organization
 - C.0 GENERAL
 - C.1 PROCESSOR ARCHITECTURES
 - C.2 COMPUTER-COMMUNICATION NETWORKS
 - C.3 SPECIAL-PURPOSE AND APPLICATION-BASED SYSTEMS (J.7)
 - C.4 PERFORMANCE OF SYSTEMS
 - C.5 COMPUTER SYSTEM IMPLEMENTATION
 - C.m MISCELLANEOUS
- D. Software
 - D.0 GENERAL
 - D.1 PROGRAMMING TECHNIQUES (E)
 - D.2 SOFTWARE ENGINEERING (K.6.3)
 - D.3 PROGRAMMING LANGUAGES
 - D.4 OPERATING SYSTEMS (C)
 - D.m MISCELLANEOUS

- E. Data
 - E.0 GENERAL
 - E.1 DATA STRUCTURES
 - E.2 DATA STORAGE REPRESENTATIONS
 - E.3 DATA ENCRYPTION
 - E.4 CODING AND INFORMATION THEORY (H.1.1)
 - E.5 FILES (D.4.3, F.2.2, H.2)
 - E.m MISCELLANEOUS
- F. Theory of Computation
 - F.0 GENERAL
 - F.1 COMPUTATION BY ABSTRACT DEVICES
 - F.2 ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY (B.6, B.7, F.1.3)
 - F.3 LOGICS AND MEANINGS OF PROGRAMS
 - F.4 MATHEMATICAL LOGIC AND FORMAL LANGUAGES
 - F.m MISCELLANEOUS
- G. Mathematics of Computing
 - G.0 GENERAL
 - G.1 NUMERICAL ANALYSIS
 - G.2 DISCRETE MATHEMATICS
 - G.3 PROBABILITY AND STATISTICS
 - G.4 MATHEMATICAL SOFTWARE
 - G.m MISCELLANEOUS
- H. Information Systems
 - H.0 GENERAL
 - H.1 MODELS AND PRINCIPLES
 - H.2 DATABASE MANAGEMENT (E.5)
 - H.3 INFORMATION STORAGE AND RETRIEVAL
 - H.4 INFORMATION SYSTEMS APPLICATIONS
 - H.5 INFORMATION INTERFACES AND PRESENTATION (e.g., HCI) (I.7)
 - H.m MISCELLANEOUS
- I. Computing Methodologies
 - I.0 GENERAL
 - I.1 SYMBOLIC AND ALGEBRAIC MANIPULATION Revised
 - I.2 ARTIFICIAL INTELLIGENCE
 - I.3 COMPUTER GRAPHICS
 - I.4 IMAGE PROCESSING AND COMPUTER VISION Revised

As well as Mathematics subjects



2000 Mathematics Subject Classification

[Search the MSC](#) || [Browse the MSC](#) || [Entire MSC2000 in PDF](#) || [How to Use the MSC](#)

The Mathematics Subject Classification (MSC) is used to categorize items covered by the two reviewing databases, Mathematical Reviews (MR) and Zentralblatt MATH (Zbl). The MSC is broken down into over 5,000 two-, three-, and five-digit classifications, each corresponding to a discipline of mathematics (e.g., 11 = Number theory, 11B = Sequences and sets, 11B05 = Density, gaps, topology).

The current classification system, 2000 Mathematics Subject Classification (MSC2000), is a revision of the 1991 Mathematics Subject Classification, which is the classification that has been used by MR and Zbl since the beginning of 1991. MSC2000 is the result of a collaborative effort by the editors of MR and Zbl to update the classification. The editors acknowledge the many helpful suggestions from the mathematical community during the revision process.

[Changes at the 2-digit level](#)

[Conversion Tables](#)

Browse the 2000 MSC

- [00-xx](#) General
- [01-xx](#) History and biography [See also the classification number -03 in the other sections]
- [03-xx](#) Mathematical logic and foundations
- [04-xx](#) This section has been deleted (For set theory see [03Exx](#))
- [05-xx](#) Combinatorics (For finite fields, see [11Txx](#))
- [06-xx](#) Order, lattices, ordered algebraic structures [See also [18B35](#)]
- [08-xx](#) General algebraic systems
- [11-xx](#) Number theory
- [12-xx](#) Field theory and polynomials
- [13-xx](#) Commutative rings and algebras
- [14-xx](#) Algebraic geometry
- [15-xx](#) Linear and multilinear algebra; matrix theory
- [16-xx](#) Associative rings and algebras (For the commutative case, see [13-xx](#))
- [17-xx](#) Nonassociative rings and algebras
- [18-xx](#) Category theory, homological algebra (For commutative rings see [13Dxx](#), for associative rings [16Exx](#), for groups [20Jxx](#), for topological groups and related structures [57Txx](#); see also [55Nxx](#) and [55Uxx](#) for algebraic topology)
- [19-xx](#) K -theory [See also [16E20](#), [18F25](#)]
- [20-xx](#) Group theory and generalizations
- [22-xx](#) Topological groups, Lie groups (For transformation groups, see [54H15](#), [57Sxx](#), [58-xx](#). For abstract harmonic analysis, see [43-xx](#))
- [26-xx](#) Real functions [See also [54C30](#)]
- [28-xx](#) Measure and integration (For analysis on manifolds, see [58-xx](#))
- [30-xx](#) Functions of a complex variable (For analysis on manifolds, see [58-xx](#))
- [31-xx](#) Potential theory (For probabilistic potential theory, see [60J45](#))
- [32-xx](#) Several complex variables and analytic spaces (For infinite-dimensional holomorphy, see [46G20](#), [58B12](#))
- [33-xx](#) Special functions (33-xx deals with the properties of functions as functions) (For orthogonal functions, see [42Cxx](#); for aspects of combinatorics, see [05Axx](#); for number-theoretic aspects, see [11-xx](#); for representation theory, see [22Exx](#))
- [34-xx](#) Ordinary differential equations
- [35-xx](#) Partial differential equations
- [37-xx](#) Dynamical systems and ergodic theory [See also [26A18](#), [28Dxx](#), [34Cxx](#), [34Dxx](#), [35Bxx](#), [46Lxx](#), [58Jxx](#), [70-xx](#)]
- [39-xx](#) Difference and functional equations
- [40-xx](#) Sequences, series, summability

But for Bioinformatics (and Life Sciences in general) is not existing any shared classification schema of the domain.

And life scientists like taxonomies...

*Biochemistry (Moscow), Vol. 69, No. 12, 2004, pp. 1403-1406.
Copyright © 2002 by CELL PRESS.*

DISCUSSIONS

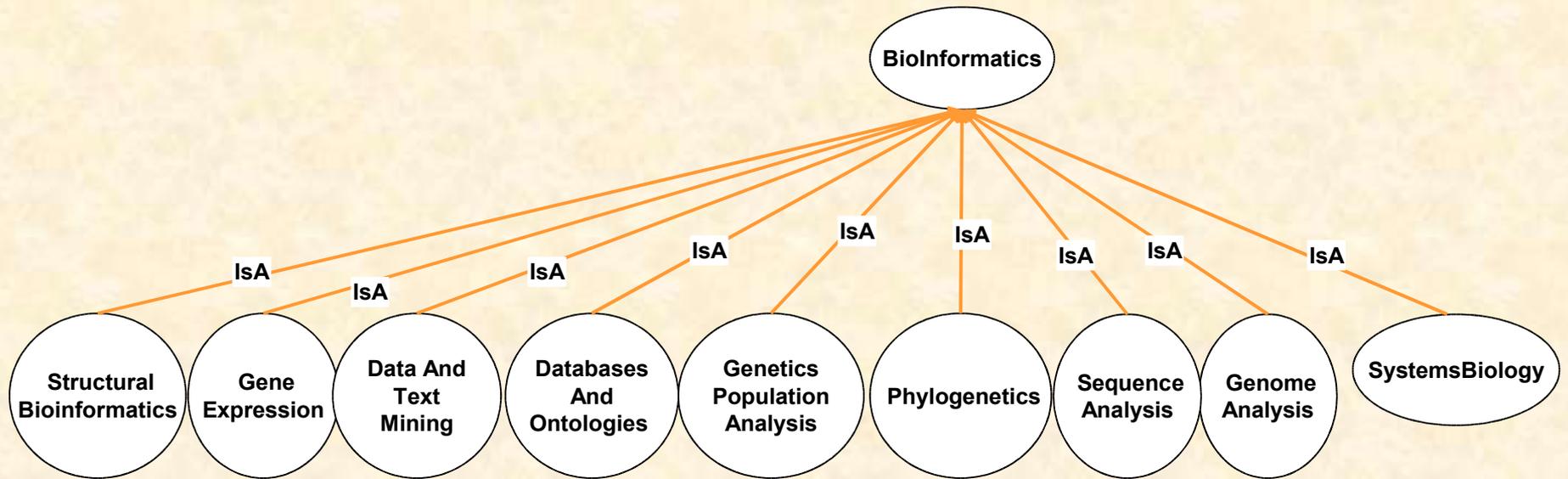
**Can a Biologist Fix a Radio? —
or, What I Learned while Studying Apoptosis**

Y. Lazebnik

*Cold Spring Harbor Laboratory, Cold Spring Harbor,
New York 11724, USA; E-mail: lazebnik@cshl.org*

This article by Yu. Lazebnik, “Can a Biologist Fix a Radio? — or, What I Learned while Studying Apoptosis” has already been published in English (*Cancer Cell*, 2002, **2**, 179-182) and in Russian (*Uspekhi Gerontologii*, 2003, No. 12, 166-171). Nevertheless, we have undertaken its secondary publication in our journal for two reasons: first, our journal has different readers, and, second, the great significance of this manifest of Yuri Lazebnik. The author in bright and clever form shows the emerging necessity to create formalized language designed to describe complicated systems of regulation of biochemical processes in living cells. The article is published with permission of *Cancer Cell* and *Uspekhi Gerontologii*.

Editor-in-Chief of Biokhimiya/Biochemistry (Moscow) V. P. Skulachev



The classification introduced for articles of “Oxford’s Bioinformatics” in 2005

Source vs. Resource Ontology

Adam Mathes
Knowledge Representation and Formal Ontology - LIS590KR
Graduate School of Library and Information Science
University of Illinois Urbana-Champaign
May 2005

What is a Resource?

The notion of a resource is fundamental in current networked information systems, and is used often, specifically in relation to the World Wide Web and the Web 3.0. Terms such as Resource Description Framework (RDF), Uniform Resource Identifier (URI), and Uniform Resource Locator (URL) are relatively simple terms that mask an exceptional amount of ambiguity.

What is a resource, exactly, in the context of electronic documents? The paper developed here attempts to explicate what a resource is and its relationship to other concepts.

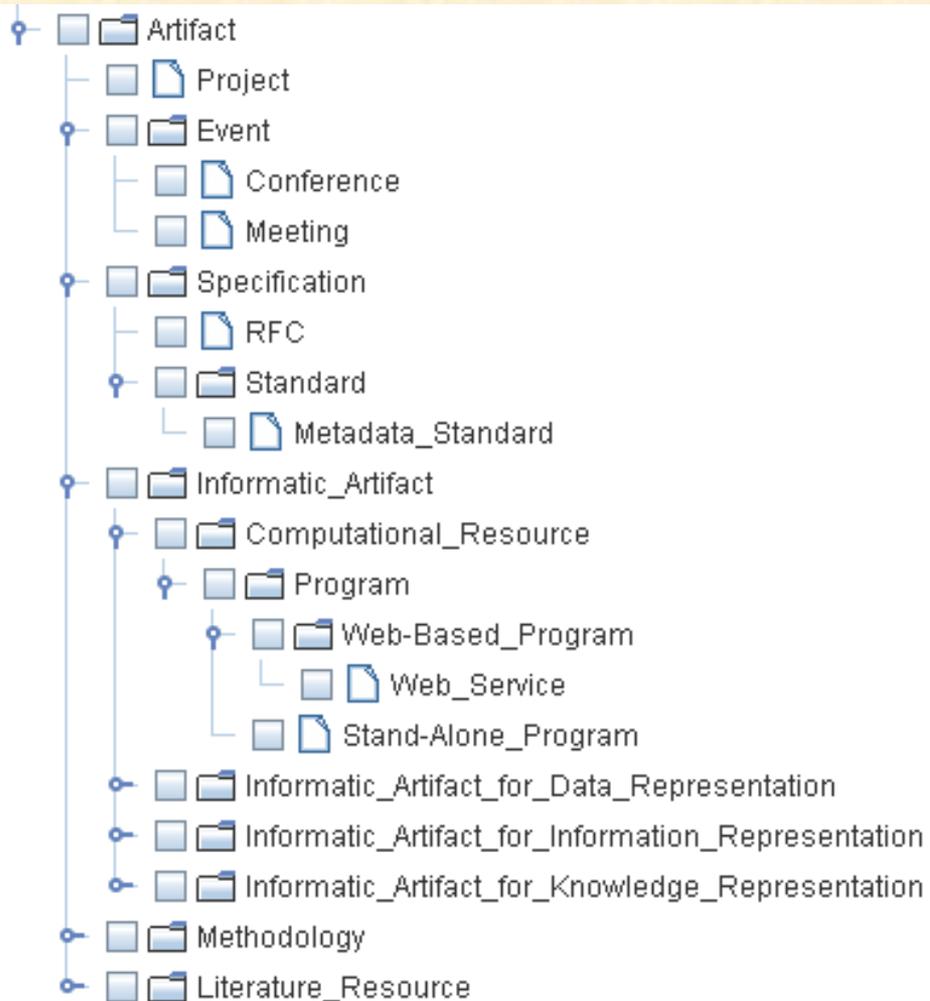
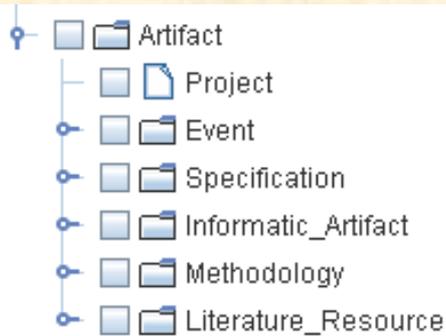
Although there is a stated definition of a resource in the URI RFC it is

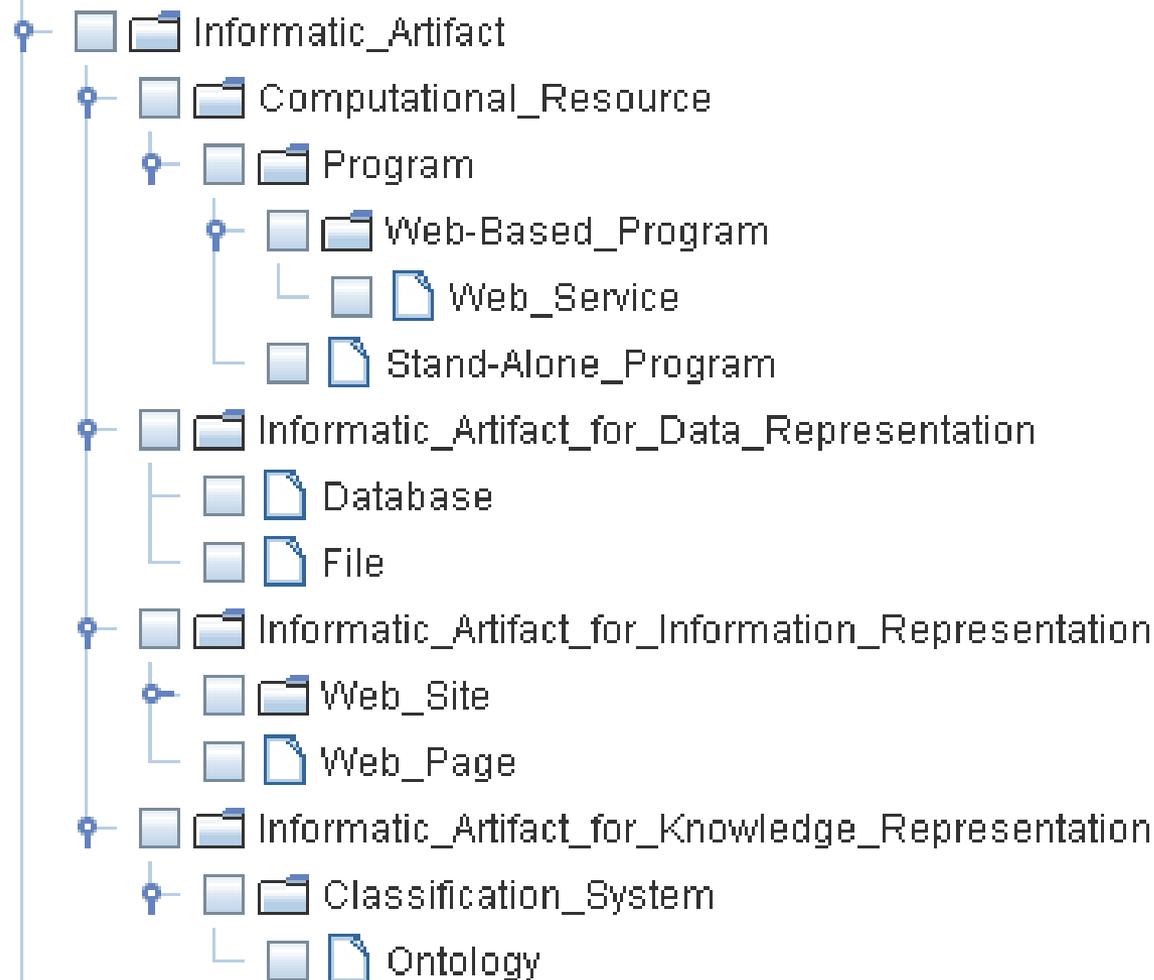
A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content-the entities to which it currently corresponds-changes over time, provided that the conceptual mapping is not changed in the process." [Berners-Lee]

RFC 2396 (RFC2396)
Internet RFC/STD/TF/VI/BCP Archives
[[RFC Index](#) | [RFC Search](#) | [Usenet FAQs](#) | [Web FAQs](#) | [Documents](#) | [Cities](#)]
Alternate Formats: [rfc2396.txt](#) | [rfc2396.html](#) | [rfc2396.txt.pdf](#)
RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax

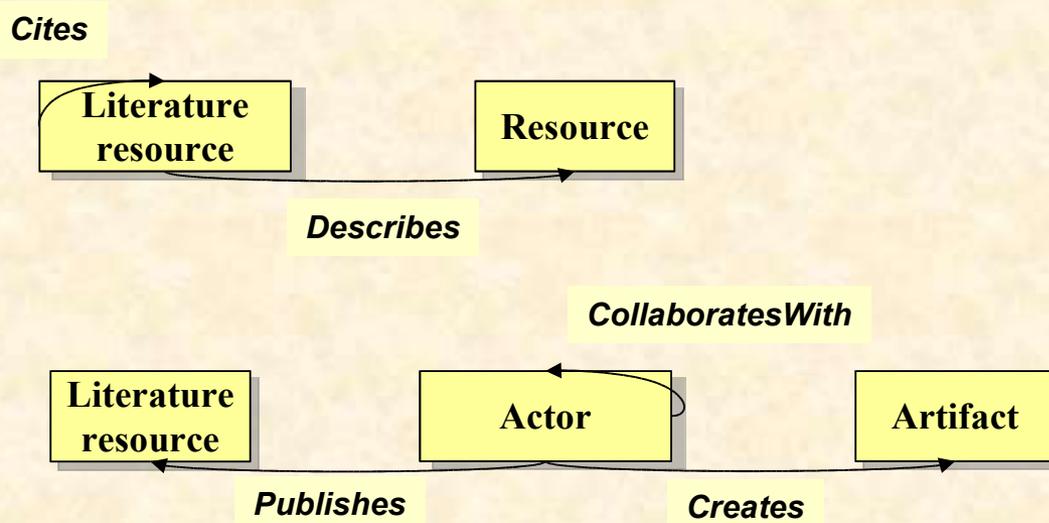


This is just our first prototype of resource ontology





Examples of semantic relationships between resources



Better this representation...

a)

= Full-text available
 = Non-subscribed
? What does this mean?

Articles in Press

Volume 40 (2007)
 Volume 39 (2006)

- Volume 39, Issue 6
pp. 573-738 (December 2006)
- Volume 39, Issue 5
pp. 465-572 (October 2006)
Dialog Systems for Health Communications
- Volume 39, Issue 4**
pp. 379-464 (August 2006)
- Volume 39, Issue 3
pp. 249-378 (June 2006)
Biomedical Ontologies
- Volume 39, Issue 2
pp. 103-248 (April 2006)
- Volume 39, Issue 1
pp. 1-102 (February 2006)
Phylogenetic Inference: Beyond Biology

Volume 38 (2005)
 Volume 37 (2004)
 Volume 36 (2003)
 Volume 35 (2002)
 Volume 34 (2001)

3. **WaveRead: Automatic measurement of relative gene expression levels from microarrays using wavelet analysis** • ARTICLE
Pages 379-388
Ghislain Bidaut, Frank J. Manion, Christophe Garcia and Michael F. Ochs
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(1237 K\)](#)

4. **Methods for reasoning from geometry about anatomic structures injured by penetrating trauma** • ARTICLE
Pages 389-400
Omolola Oguriyemi
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(975 K\)](#)

5. **Knowledge guided analysis of microarray data** • ARTICLE
Pages 401-411
Zhuo Fang, Jiong Yang, Youe Li, Qingming Luo and Lei Liu
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(529 K\)](#)

6. **Heuristic evaluation of paper-based Web pages: A simplified inspection usability methodology** • ARTICLE
Pages 412-423
Mureen Allen, Leanne M. Currie, Suzanne Bakken, Vimla L. Patel and James J. Cimino
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(973 K\)](#)

7. **A temporal constraint structure for extracting temporal information from clinical narrative** • ARTICLE
Pages 424-439
Li Zhou, Genevieve B. Melton, Simon Parsons and George Hripcsak
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(243 K\)](#)

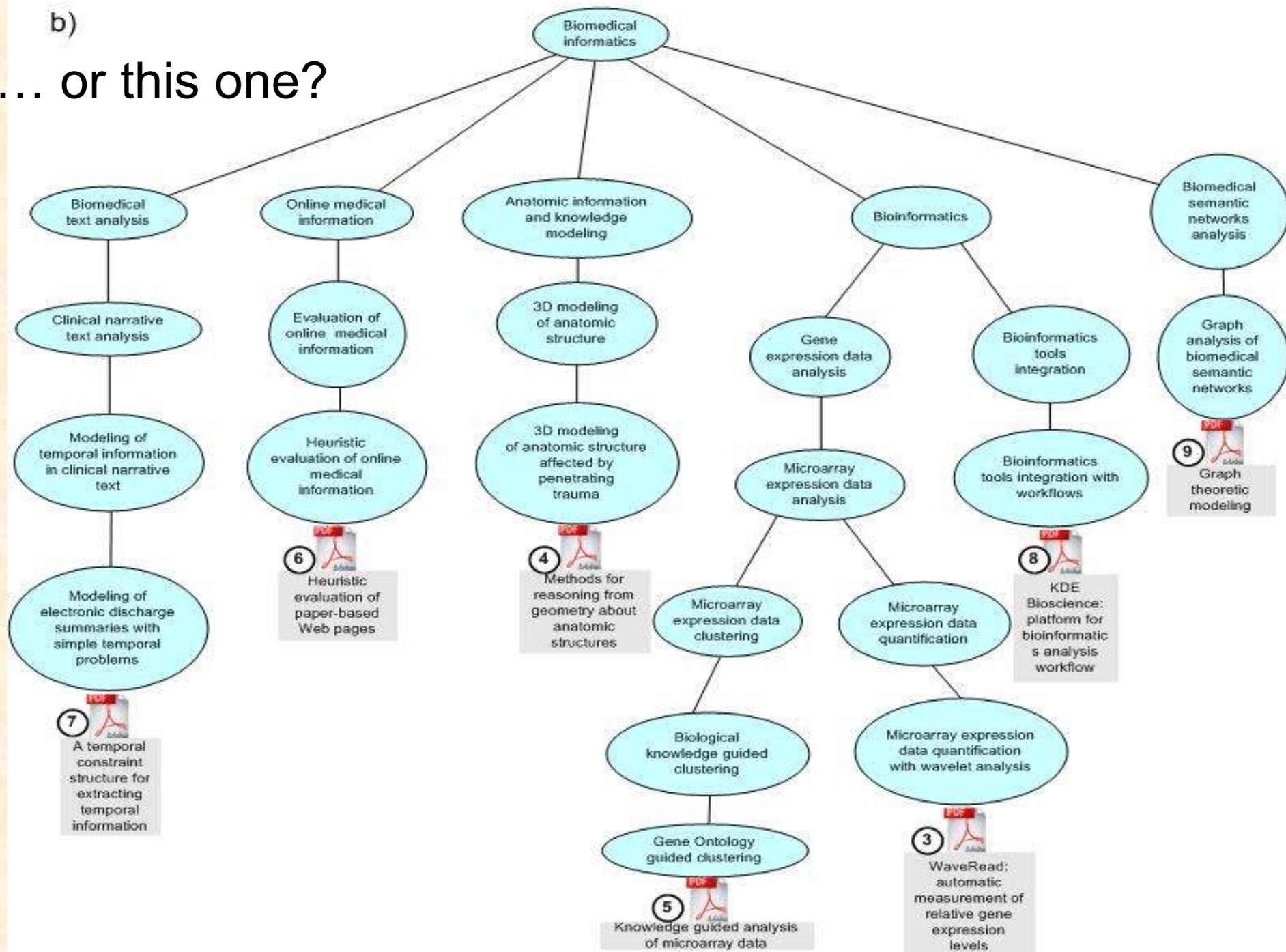
8. **KDE Bioscience: Platform for bioinformatics analysis workflows** • ARTICLE
Pages 440-450
Qiang Lu, Pei Hao, Vasa Curcin, Weizhong He, Yuan-Yuan Li, Qing-Ming Luo, Yi-Ke Guo and Yi-Xue Li
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(1015 K\)](#)

Methodological Review

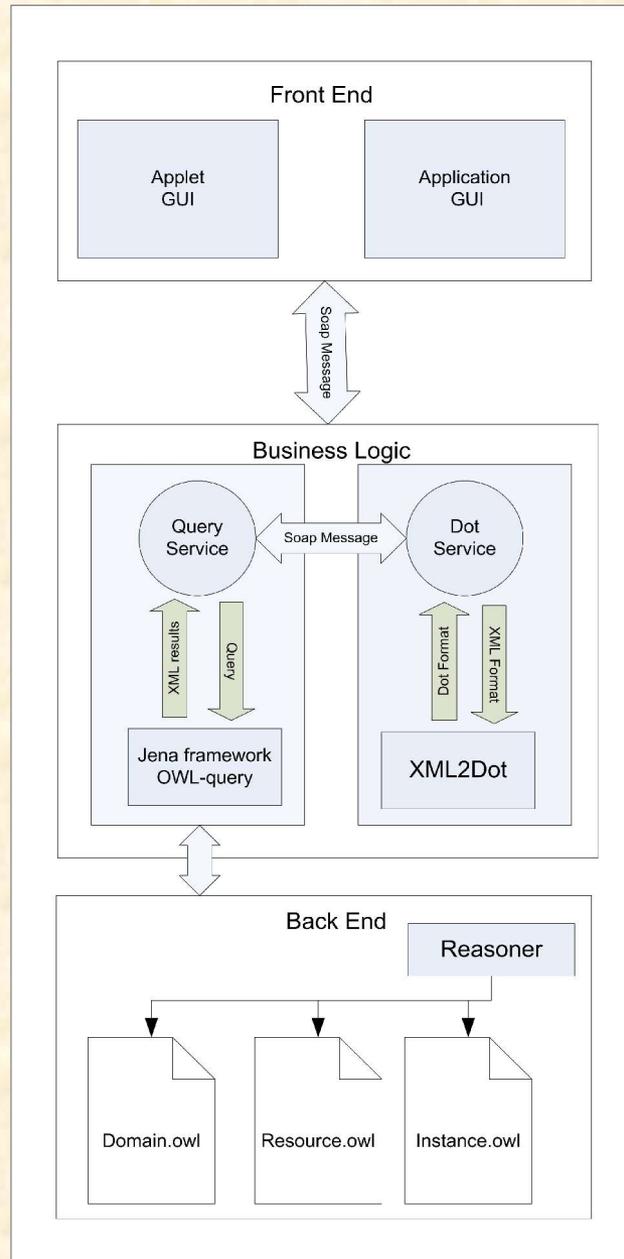
9. **Graph theoretic modeling of large-scale semantic networks** • REVIEW ARTICLE
Pages 451-464
Michael E. Bales and Stephen B. Johnson
[SummaryPlus](#) | [Full Text](#) • [Links](#) | [PDF \(213 K\)](#)

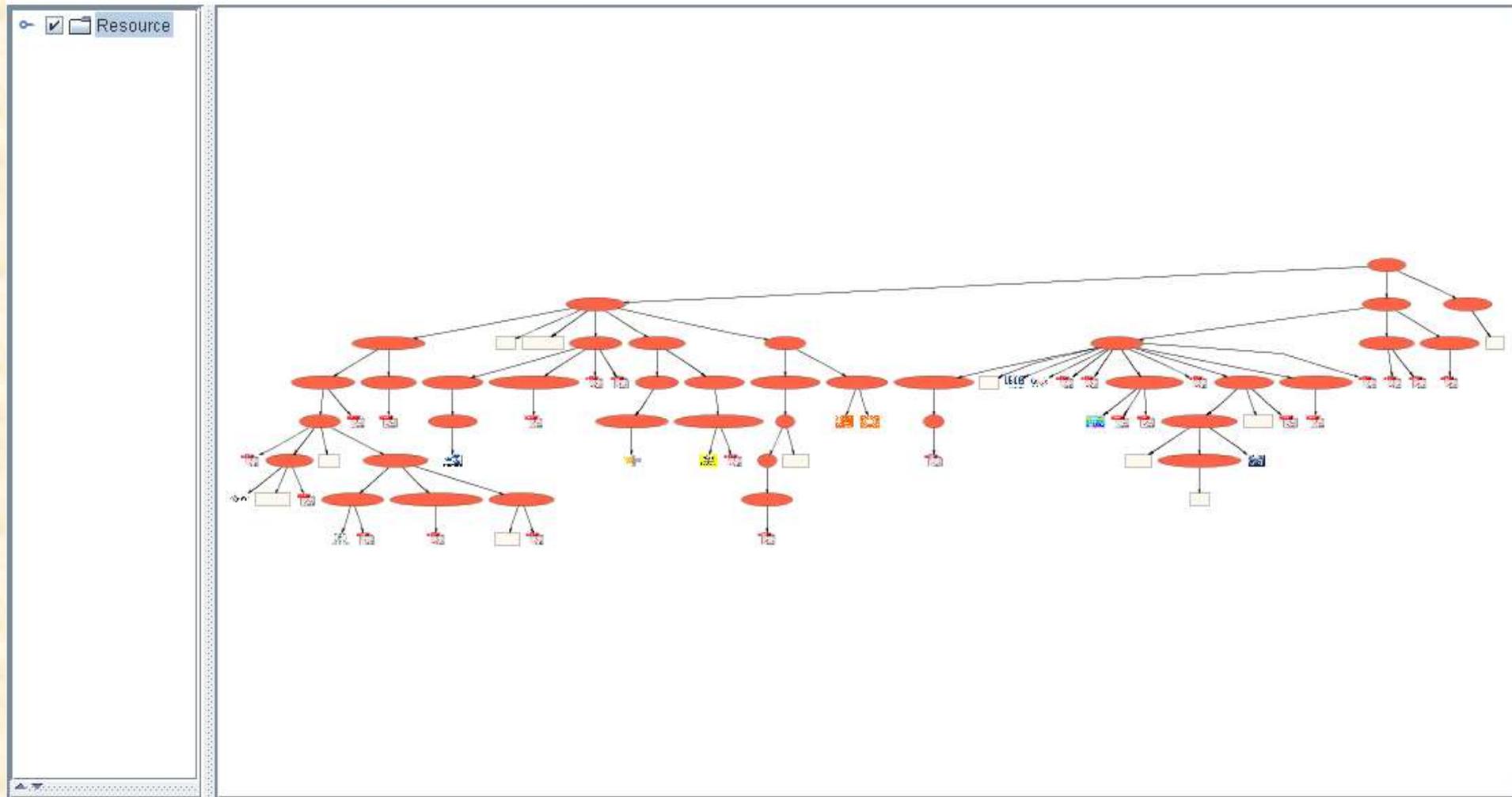
b)

... or this one?

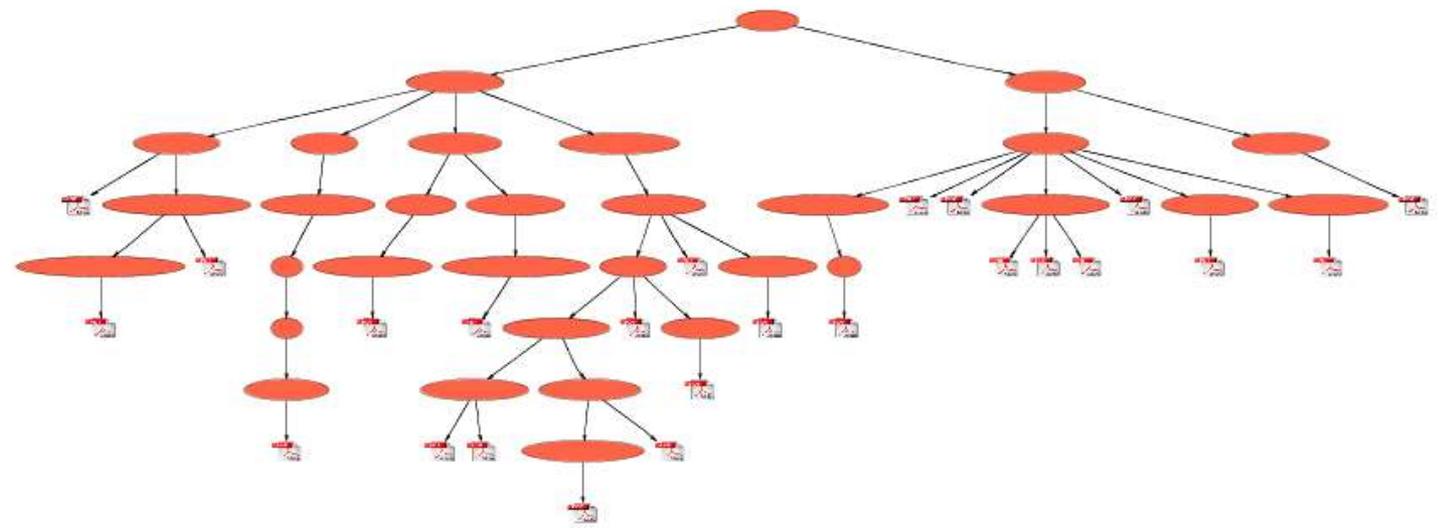
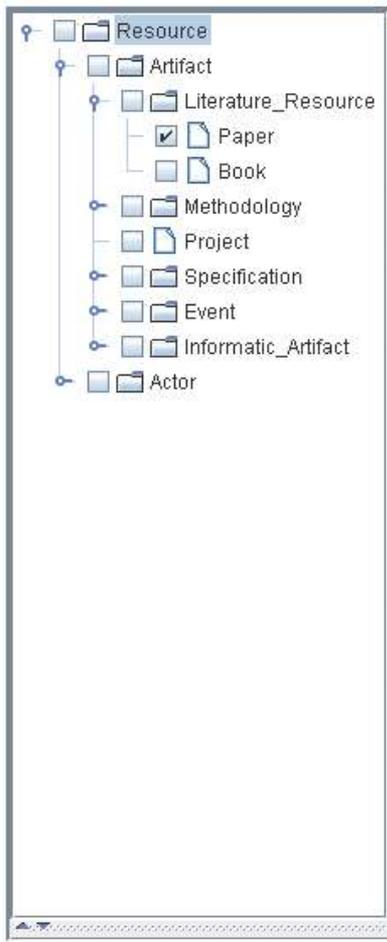


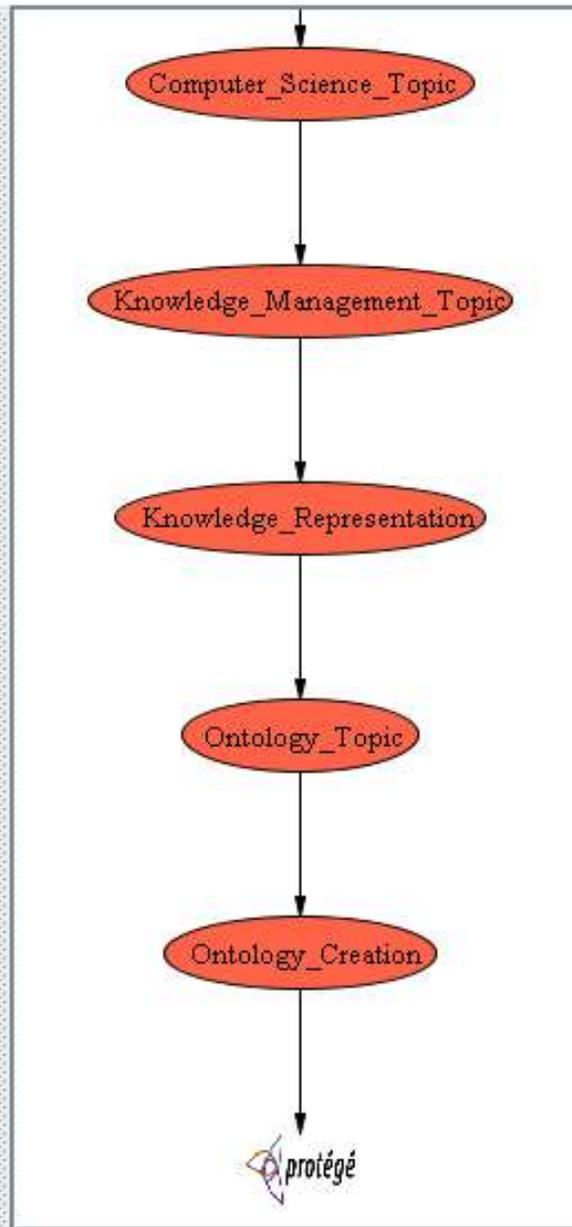
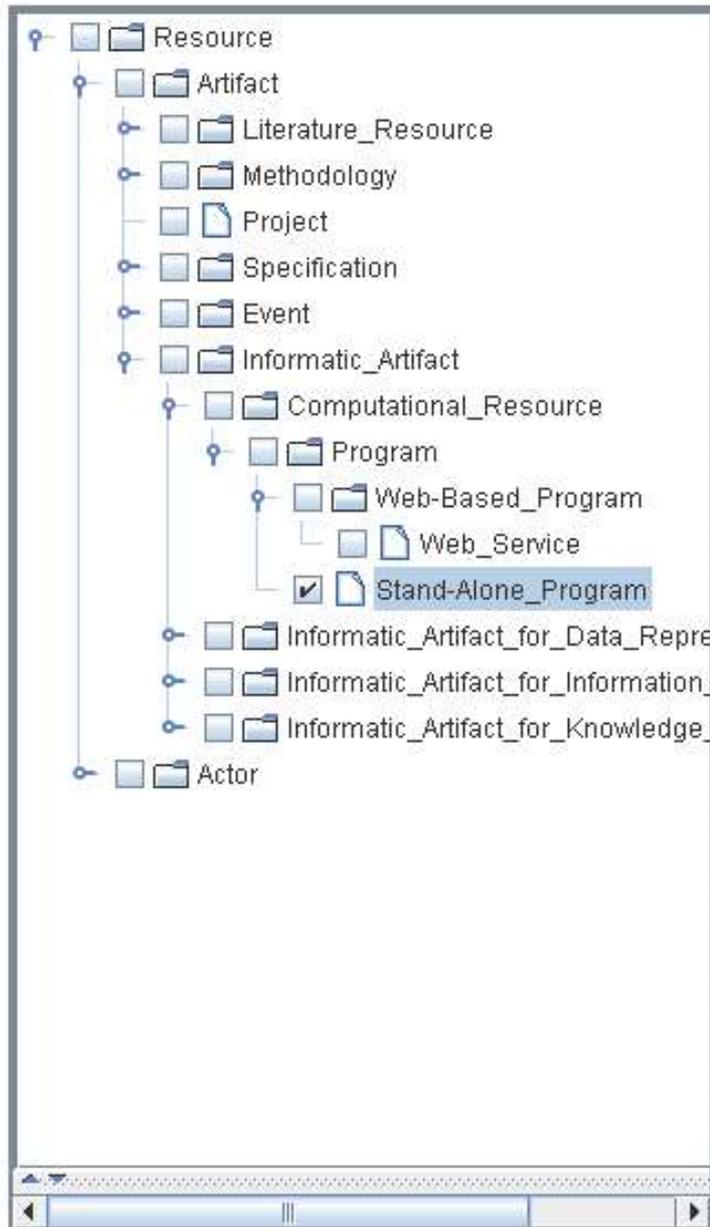
A web-based semantic browser for Resourceomes

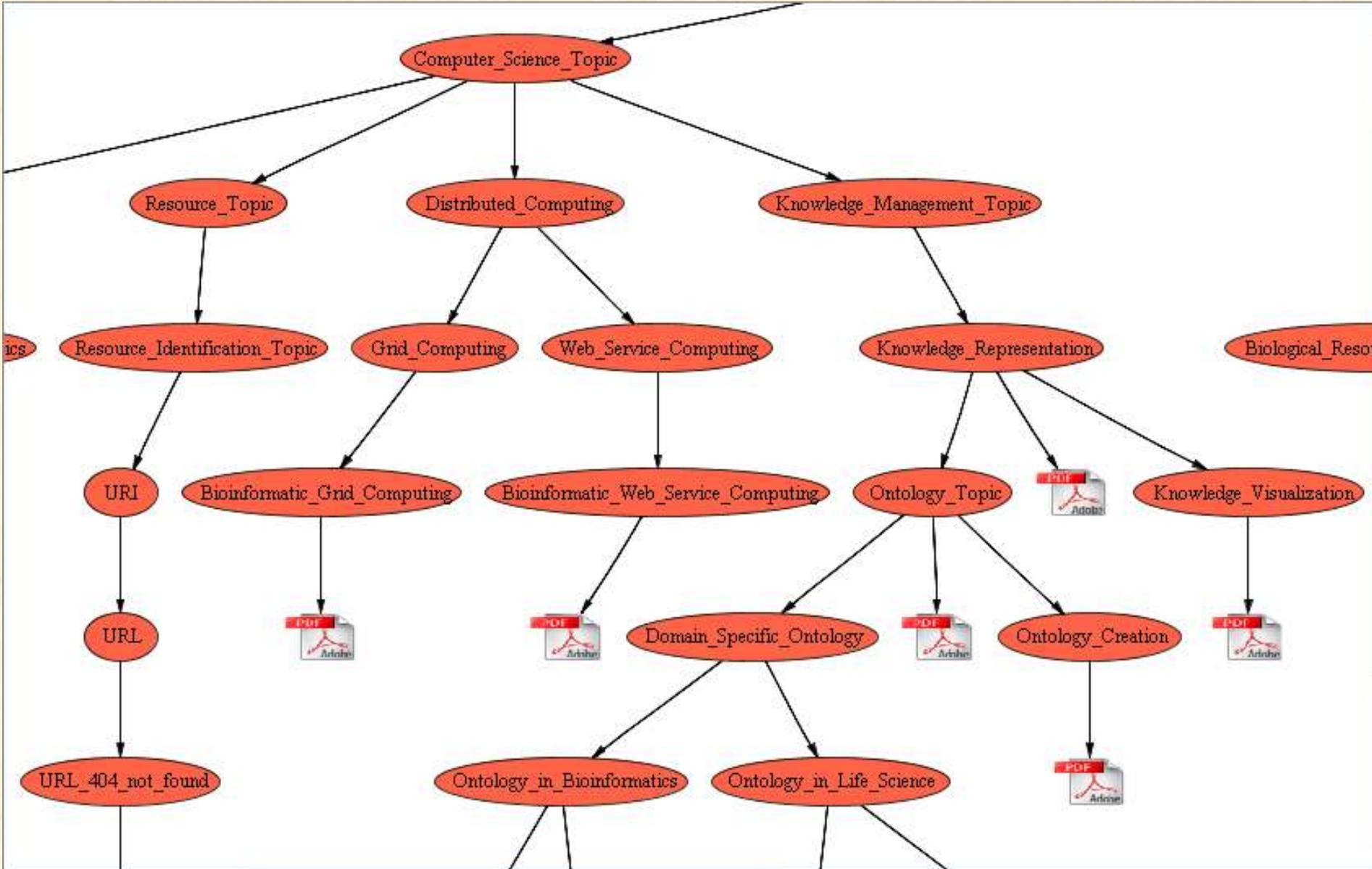


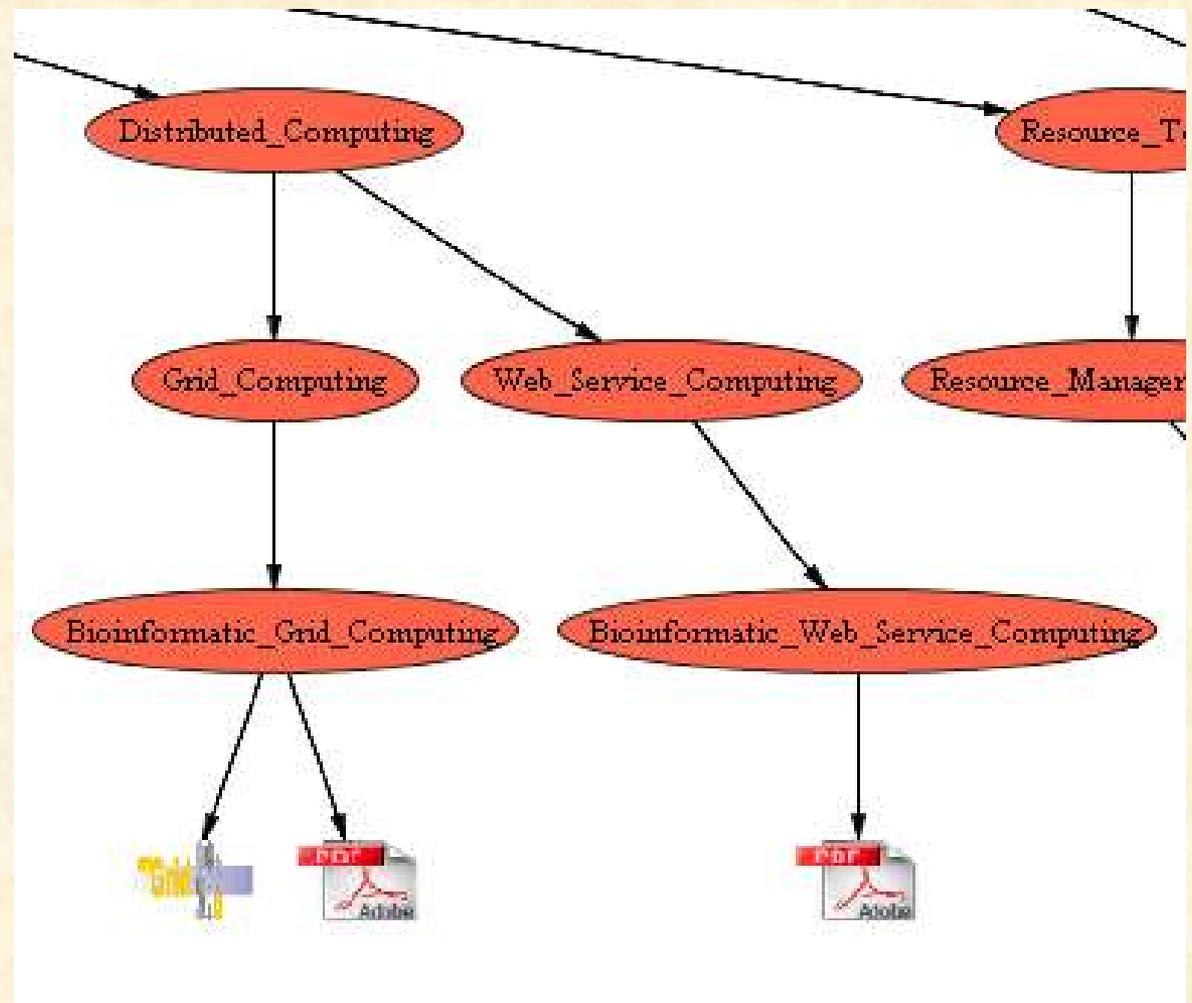
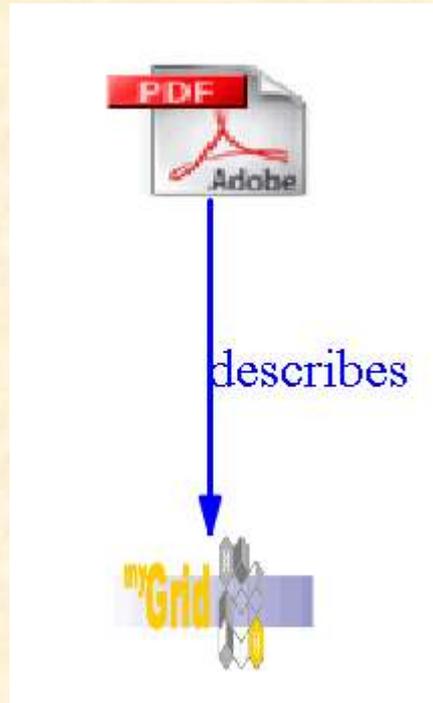


The first prototype of our browser

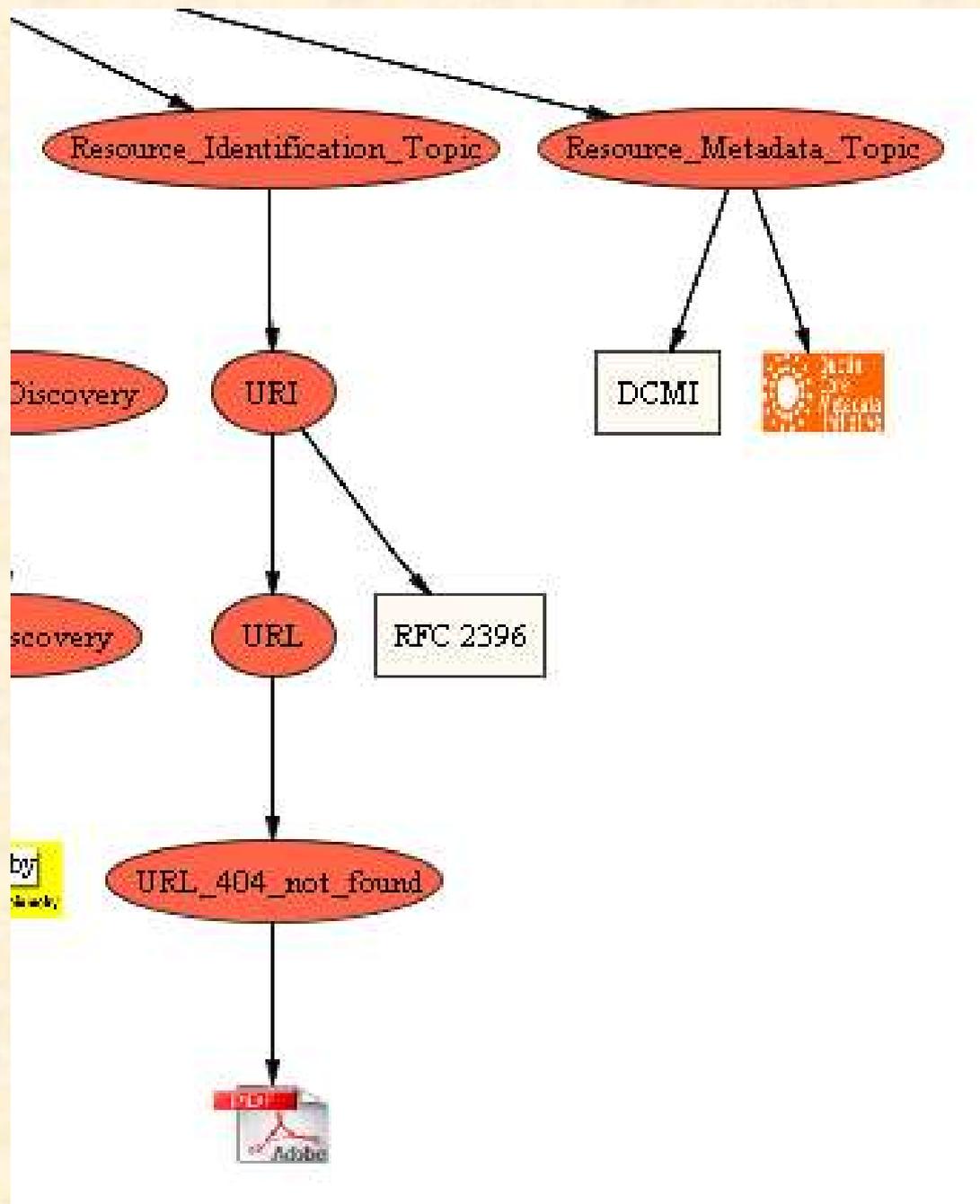








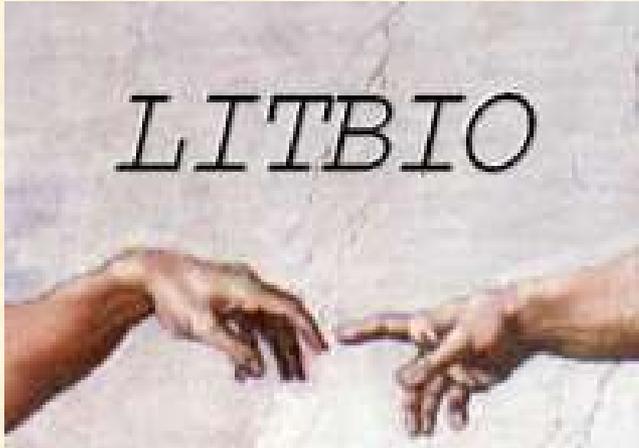
Passing with the mouse over the icons you can see a description and the URI of the resource



Many open issues:

- Annotation of resources (manual, semi-automatic, automatic)
- Representation of the domain (ontology, concept maps, topic maps, SKOS?)
- Ranking of resources (page rank, judgments)
- Status of resources (agents checking them)
- Graph visualization? (GRAPPA – Graphviz)
- Not only browser but also visual editor (GrOWL?)
- ...

Acknowledgment



www.litbio.org

This work is supported by the Italian Investment Funds for Basic Research (FIRB) project “Laboratory of Interdisciplinary Technologies in Bioinformatics” (LITBIO).

Hoping that the bioinformatics community could soon say

"Houston, We've Had a Problem"

Thank you for your attention!