

# **The Semantic Web: Some Issues in Supporting Integrated Web Access**

---

**Ling Liu**

***Georgia Institute of Technology***

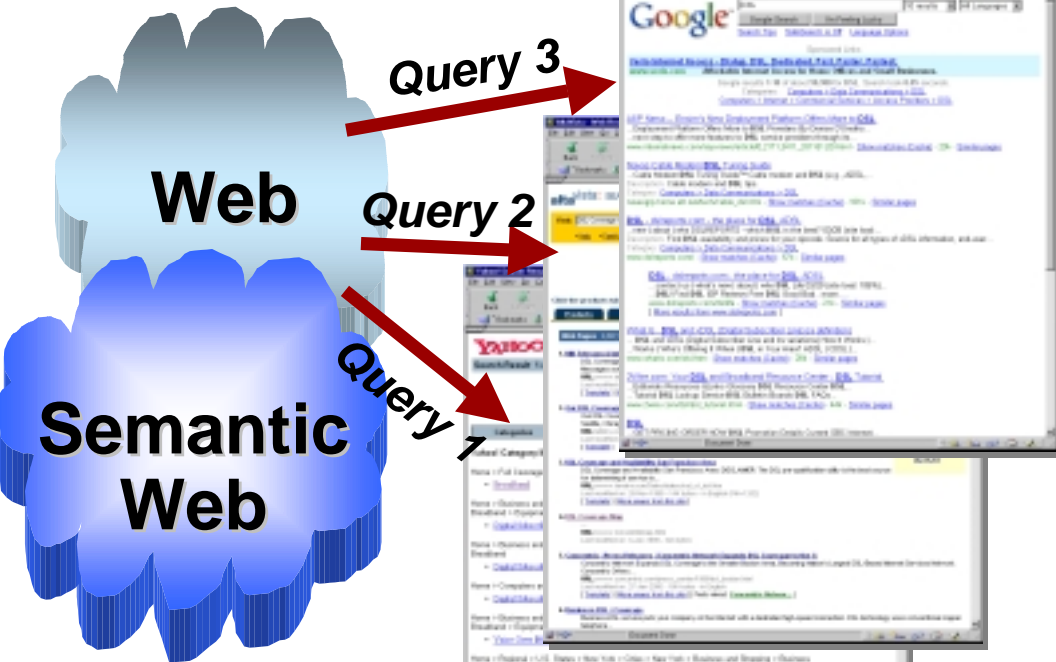
***NSF/OntoWeb Workshop on the Semantic Web, April 3-5, 2002***

# Three Big Questions

- How much will the Semantic Web facilitate Integrated Web Access
  - Web Information Analysis
  - Bioinformatics Application - A Scenario
- How do the Semantic Web and Ontologies relate to one another?
  - light-weight v.s. heavy-weight
- Is Peer-to-Peer Technology critical to building the Distributed Semantic Web?

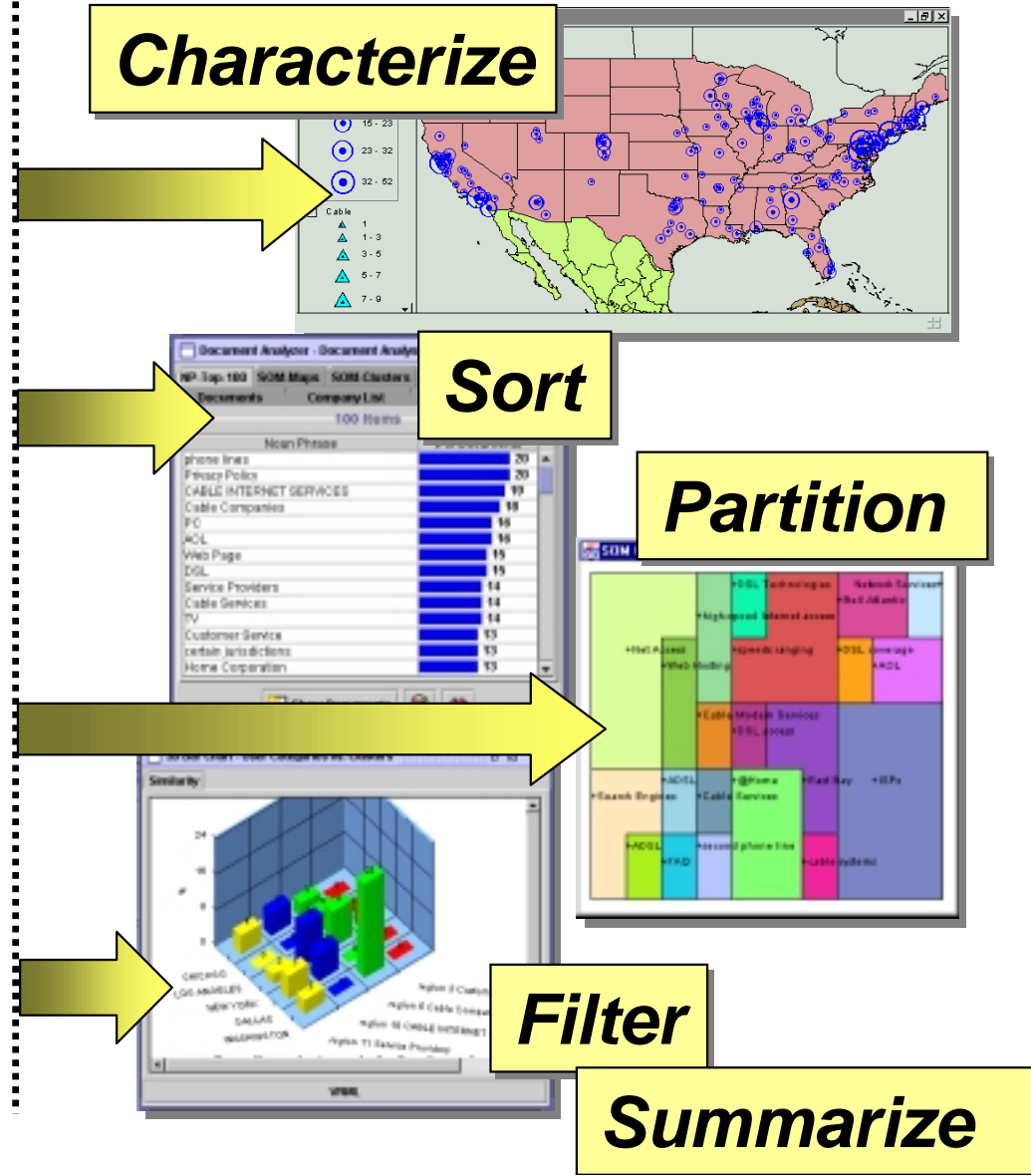
# Why Do We Need Integrated Web Access ?

## Simple Query-Based Searching



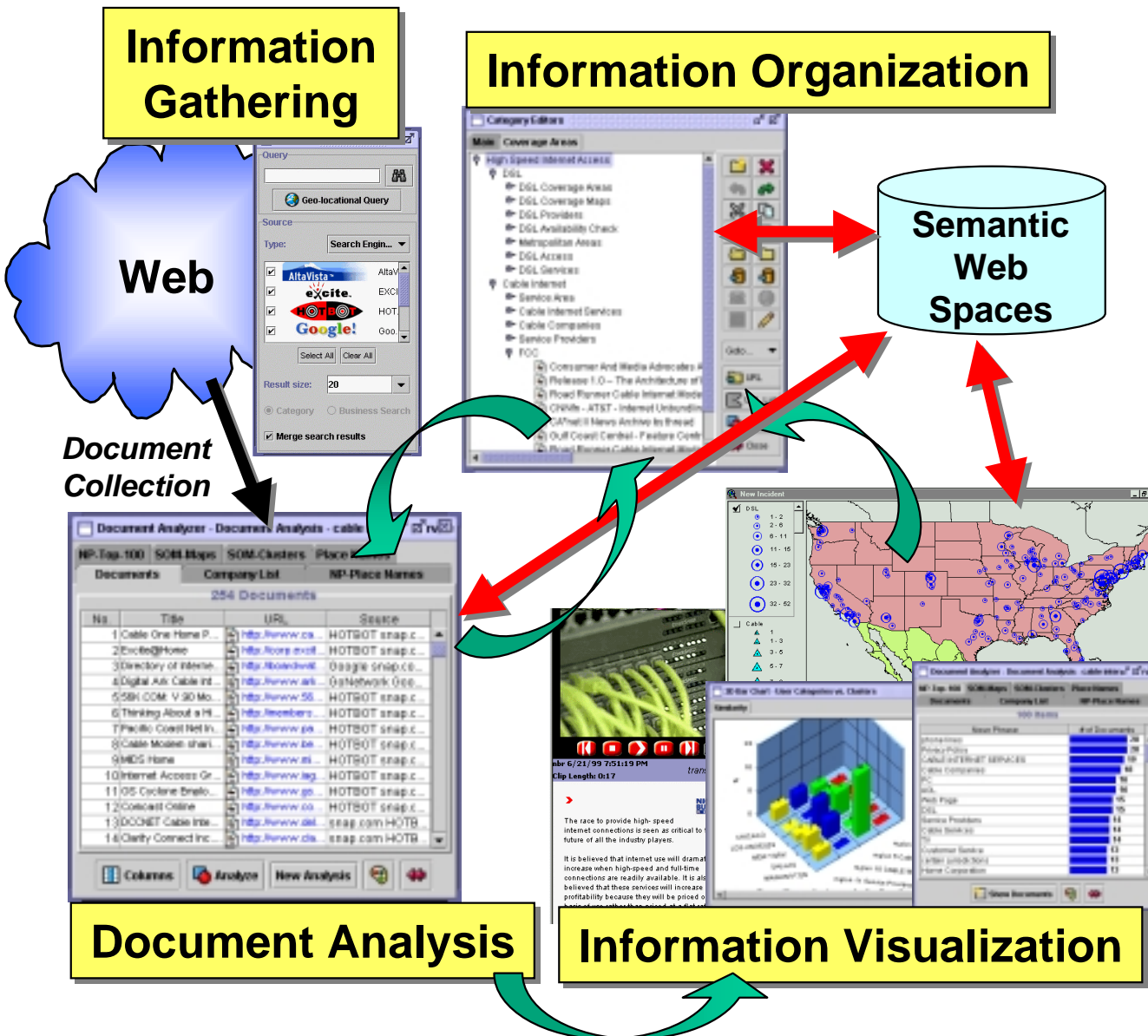
**Large & Unorganized Document Collections**

## Post Processing on Document Collections



[Adapted from Ko et.al 2000]

# Issues in Integrated Web Access



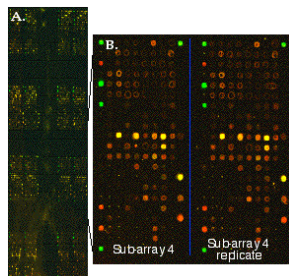
- **Context-sensitive service selections**
- **Context-Sensitive Service Composition**
- **Context-Sensitive data filtering, annotation, and fusion**
- **Context-Sensitive workflows**
- **Mechanisms for complex information access and integration**

# Experience with Biologists

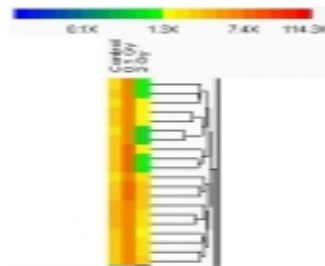
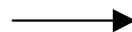
# Genome-scale Modeling of Low-Dose Irradiation Responses Using Microarray Based Gene Networks

*Hypotheses: Genes that show similar expression patterns in response to low-dose irradiation are components of coordinated expression groups (called synexpression groups) and that understanding the differential regulation of these synexpression groups will*

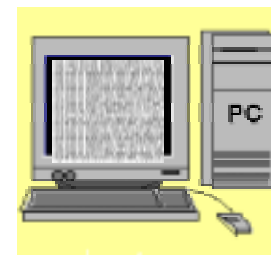
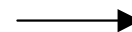
- *provide DNA-sequenced-based understanding of the complex biological processes associated with low-dose radiation and*
- *identify determinants of radiation dose and genetic susceptibility to radiation damage.*



Microarray analysis

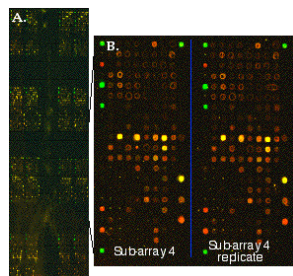


Statistical Clustering of genes

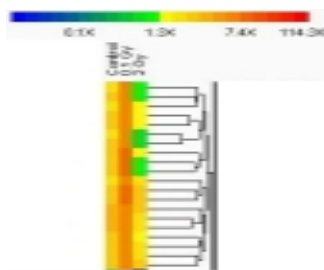


Database search for common promoter elements to link new candidate genes

# A Bioinformatics Application: Low-dose radiation on human genes



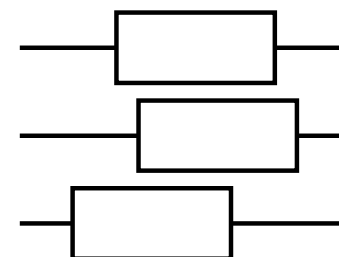
Microarray analysis



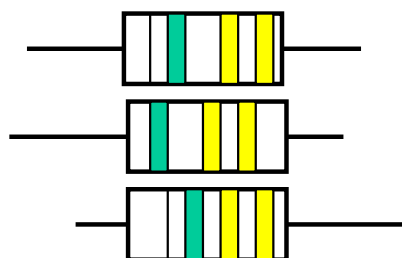
cDNA Cluster



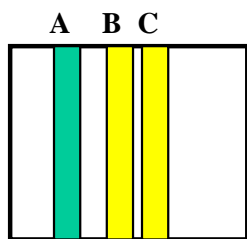
Database search for promoter identification



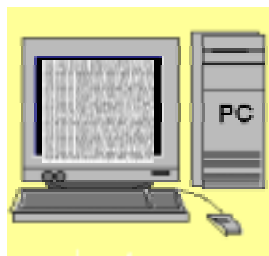
Promoter sequences



Common promoter alignment



Promoter model



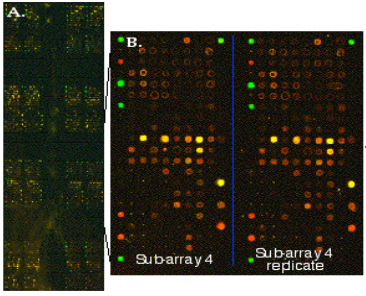
Database search



```

11005-D30037.FEATURE=.DEFINITION=HUMP17PB Human mRNA
11006-U77949.FEATURE=.DEFINITION=HSU77949 Human Cdc6-
11007-Cluster Incl. AB002450.Homo sapiens mRNA from chromo
* 11008-Cluster Incl. AB009262.Homo sapiens mRNA for cytochrome
* 11009-Cluster Incl. U35459.Human bombain mRNA, complete cds.
* 11010-Cluster Incl. U18271.Human thymopoietin (TMPO) gene 3'UT
11011-Cluster Incl. AF023203.Homo sapiens horieobox protein Oj
11012-Cluster Incl. AB3926.H.sapiens mRNA for transcription fact
11013-Cluster Incl. D63257.Homo sapiens mRNA for CDC2 delta T
11014-Cluster Incl. X79291.H.sapiens mRNA for SVT 1(ov=3,117
11015.X77794.FEATURE=cds.DEFINITION=HSC YCO1 H.sapiens
11016-Cluster Incl. AJ011915.Homo sapiens mRNA for synapso
11017-MI1507.Human transferrin receptor mRNA, complete cds (
* 11018-Cluster Incl. W25932.15b1.Homo sapiens cDNA, bp=49259.
* 11019-Cluster Incl. X16983.Human mRNA for integrin alpha-4 sub
11020-Cluster Incl. M74091.Human cyclin mRNA, (cyc=)ANK33AN
11021-Cluster Incl. AB007896.Homo sapiens KIAA0436 mRNA, pt
11022-Cluster Incl. X78925.H.sapiens HLF 2 mRNA for zinc finger
* 11023-Cluster Incl. AF001175.Homo sapiens ribonuclease P prote
11024-MI1507.Human transferrin receptor mRNA, complete cds (
    
```

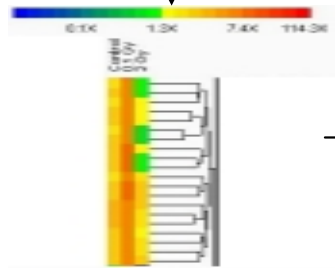
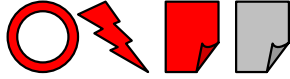
\* - New candidate target genes



Microarray analysis

Genes that changed significantly

XWrap Composer  
CLUSFAVOR

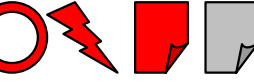


Statistical Clustering of genes

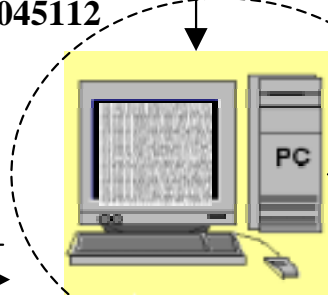
Gene ids



AA045112

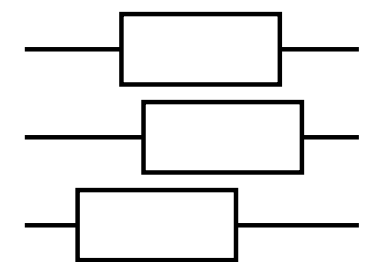
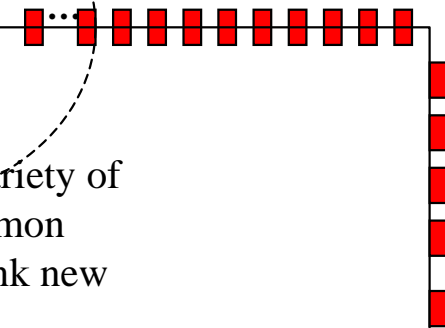


XWrap Composer  
NCBI BLAST

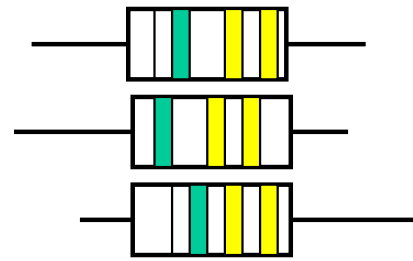


BLAST search over a variety of data sources for common promoter elements to link new candidate genes

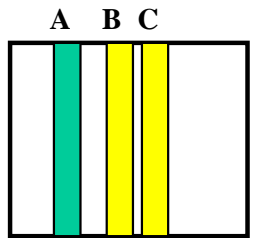
A list of related sequences



Promoter sequences



Common promoter alignment



Promoter model

Consensus sequences

- \* 11005-D30037 FEATURE= DEFINITION=HUMPTFB Human mRNA
- \* 11006-U77949 FEATURE= DEFINITION=HSU77949 Human Cdc6
- \* 11007-Cluster incl. AB002450 Homo sapiens mRNA from chromo
- \* 11008-Cluster incl. AB009202 Homo sapiens mRNA for cytochrome
- \* 11009-Cluster incl. U05459 Human bombain mRNA, complete cds.
- \* 11010-Cluster incl. U18271 Human thymopoietin (TMPO) gene Xclt
- \* 11011-Cluster incl. AF023203 Homo sapiens homeobox protein Oj
- \* 11012-Cluster incl. X83928 H.sapiens mRNA for transcription fact
- \* 11013-Cluster incl. D83257 Homo sapiens mRNA for CDC2 delta T
- \* 11014-Cluster incl. X75201 H.sapiens mRNA for SYT Jcds(3,117
- \* 11015-X77794 FEATURE=cds DEFINITION=HSC Y001 H.sapiens
- \* 11016-Cluster incl. AJ011915 Homo sapiens mRNA for synapto
- \* 11017-M11507 Human transferrin receptor mRNA, complete cds (
- \* 11018-Cluster incl. Y25932 15b1 Homo sapiens cDNA, bp=42592
- \* 11019-Cluster incl. X16903 Human mRNA for integrin alpha-4 sub
- \* 11020-Cluster incl. M74091 Human cyclin mRNA, cds=UNKNOWN
- \* 11021-Cluster incl. AB007896 Homo sapiens KIAA0436 mRNA, pt
- \* 11022-Cluster incl. X76925 H.sapiens HLF2 mRNA for zinc finger
- \* 11023-Cluster incl. AF001175 Homo sapiens ribonuclease P prote
- \* 11024-M11507 Human transferrin receptor mRNA, complete cds (

\*- New candidate target genes

<http://genomatix.ge>

# Technical Challenges

- ***Context-sensitive workflows/interactions management***
- ***Context-sensitive Service Selection/Routing and Service Composition***
- ***Context-Sensitive information extraction and information filtering***
- ***Mechanisms for complex information access/analysis/integration***
- ***...***

# The Semantic Web and Ontologies



*Towards more meaningful  
Internet Systems*

# Semantic Web Need Ontologies and Tools

## ■ Light-weight Ontology

- concepts, atomic types
- Is-A hierarchy among concepts
- associations between concepts

## ■ Tools

- There are tools to support light-weight ontologies
  - ➔ Examples:
    - Protege, Stanford
    - OntoEdit, Karlsruhe
    - UML-Tools, several

# Complex Ontologies

## ■ Heavy-weight Ontology

- cardinality constraints
- taxonomy of **ontological relationships**
- ontologies for events, interactions, processes

## ■ Heavy-weight

- Layering/ontological compositions are necessary
  - ➔ self-configuring v.s. pre-defined
- Need tools to effectively encode domain experts' knowledge
  - ➔ domain-specific semantic axioms/ontological relationships
  - ➔ domain-specific processes/interactions/workflows

# Some Critical Research Challenges

- ‘‘No Method/Language fits all’’ different applications‘ need
  - ➔ different representation languages often provide their own underlying reasoning services
  - ➔ are we creating more problems than we are trying to solve?
- Ontology development should be aware of
  - Mechanisms to access/integrate several ontologies
    - ➔ distributed on the web
    - ➔ identified by (XML-) namespaces
  - Mechanisms for self-configuring of available ontologies
  - Pramatic approaches to engineering efficient, scalable, and yet easy-to-use systems for managing ontologies

**Peer-to-Peer Technology**  
for  
**Building Distributed Semantic Web**

# Large-scale Web Info. Sys.

## ■ The Web and its enabling technology

- Search Engine Technology, HTTP, Web Servers/Proxy Servers
  - ➔ Performance
  - ➔ Scalability (#access/sec., #users/sec., 24\*7, etc.)
- First and last stops for any surfers

## ■ The Semantic Web

- What are the enabling Technology (???)
- How to perform a Semantic Search
  - ➔ Performance
  - ➔ Scalability
- Who are the users?

# What can Peer-to-Peer offer?

## ■ Benefits

- Decentralized Infrastructure
- Inherent Scaling Capability
- Low Lost in Start and Maintenance
- ...

## ■ Challenges

- P2P technology is still in its infancy
  - ➔ P2P File Sharing Systems – A success story
- Efficiency Issues
  - ➔ Service partitioning and lookup
  - ➔ data placement and location
- Trust Issue
  - ➔ context, usage, ...

# References

1. Tim Berners-Lee. Semantic Web Roadmap. World Wide Web Consortium (W3C), 1998.
2. Stefan Decker, Michael Erdmann, Dieter Fensel, and Rudi Studer. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In R. Meersman et al. (eds.), *Semantic Issues in Multimedia Systems*, Proceedings of DS-8, pp. 351-369, Kluwer Academic Publisher, Boston, 1999.
3. Harry S. Delugach. An Exploration Into Semantic Distance. Lecture notes in artificial intelligence, No.754, pp. 119-124, Springer-Verlag, Berlin, 1993.
4. Jeff Heflin, James Hendler, and Sean Luke. SHOE: A Knowledge Representation Language for Internet Applications. Technical Report CS-TR-4078 (UMIACS TR-99-71). 1999.
5. Vipul Kashyap and Amit Sheth. Semantics-Based Information Brokering. In Proceedings of the Third International Conference on Information and Knowledge Management (CIKM), Gaithersburg, MD, November, 1994.
6. Li Xiong and Ling Liu. PeerTrust: Supporting Trust in a Peer-to-Peer Information Sharing System. Technical Report, Georgia Institute of Technology, Jan. 2002.
7. Sergey Melnik *et al.* Generic Interoperability Framework, Working Paper, Department of Computer Science, Stanford University.
8. Marian Nodine, William Bohrer, Anne Hee Hiong Ngu. Semantic Brokering over Dynamic Heterogeneous Data Sources in InfoSleuth<sup>TM</sup>. 15th International Conference on Data Engineering, March, 1999, Sydney, Australia.
9. Amit P. Sheth. Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics, *Interoperating Geographic Information Systems*. M. F. Goodchild, M. J. Egenhofer, R. Fegeas, and C. A. Kottman (eds.), Kluwer, 1998.
10. John F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA, ©2000.
11. Resource Description Framework (RDF) Model and Syntax, World Wide Web Consortium (W3C) Recommendation, February 22, 1999.
12. Ke-Thia Yao, In-Young Ko, Ragy Eleish, and Robert Neches. Asynchronous Information Space Analysis Architecture Using Content and Structure Based Service Brokering. In Proceedings of Fifth ACM Conference on Digital Libraries (DL 2000), San Antonio, Texas, June 2000.
13. A. Ankolekar, M. Burstein, J. Hobbs, O. Lassila, D. Martin, S. McIlraith, S. Narayanan, M. Paolucci, T. Payne, K. Sycara, H. Zeng, DAML-S: Semantic Markup for Web Services, 2001, <http://www.daml.org>
14. Tim Berners-Lee (with Mark Fischetti), *Weaving the Web, The original design and ultimate destiny of the World Wide Web*, Harper, 1999.
15. D. Fensel and M. Musen, Eds. "The Semantic Web: A Brain for Humankind," IEEE Intelligent Systems, March/April 2001.
16. R. Meersman, *Ontologies and Databases: More than a Fleeting Resemblance*, Proceedings of the OES-SEO 2001 Rome Workshop", A. d'Atri and M. Missikoff (eds), LUISS Publications, 2001.