Realizing Semantic Information
Brokering and Semantic Web

by
Amit Sheth
Director, Large-Scale Distributed Information Systems Lab.
University of Georgia, Athens, GA USA
http://lsdis.cs.uga.edu
Founder/Chairman, Taalee, Inc.
http://www.taalee.com

Special thanks, Tarcisio Lima

ITC-IRST|University of Trento Seminar Series on Perspectives on Agents: Theories and Technologies

April, 27, 2000 - Trento, Italy
perspective I: State-of-the-art

FROM:

Traditional environments

- documents
- document links
- structured data

TO:

Computer-assisted environments
Current learning on the Web

ESRI Online University Database

Do you want to know whether a college or university offers a GIS program near you?

Here at ESRI, we've created an online database that we hope will eventually contain a comprehensive list of universities and colleges that offer GIS-related courses, certificates, or degrees.

You can search this database to locate GIS programs in your local area or you can locate geography departments or engineering schools that teach GIS. There are a number of ways that the database can be searched. Please keep in mind that our database is always expanding, so be sure to check back for new entries.

The database currently includes:

- University of Georgia
- University of California, Santa Barbara
- University of Connecticut
- University of Delaware
- University of Georgia

For more information about the ESRI Online University Database, please visit our website at http://www.esri.com/education/databases.

Department of Geography

University of Georgia
Geography/GEOlgy Building
Room 306
Athens, GA 30602
Phone: 706-542-2966
Fax: 706-542-2398

News & Events
- Upcoming Events
- Departmental News/Announcements

People
- Faculty
- Staff
- Graduate Students

Research, Computing, and Instructional Facilities

UGA Information
- University of Georgia
- Foreign Language Center
- College of Arts and Sciences
- Graduate School
- UGA Libraries

Helpful Geography Links

GEOG 1111
Introduction to Physical Geography
Spring Semester 1999

Instructor: Robert Maxwell Beavers

Course Syllabus
Lecture and Reading Schedule

Wednesday, May 5th, 12 to 3 pm
(Bring two Number 2 pencils and Student ID Card)
Current learning on the Web

BUT we should explore richer and more complex new information requests.
perspective II: Current search issues

FROM:

Computer-assisted environments

TO:

The Web-based environment

- semantic models/ontologies
- services/agents
Taking advantage of the Web now

BLEND BROWSING & QUERYING INTERFACE

ATTRIBUTE & KEYWORD QUERYING

SEMANTIC BROWSING

uniform view of worldwide distributed assets of similar type

Targeted e-shopping/e-commerce

assets access
Creating a Web of related information

(a commercial perspective)

What can a context do?
Rocker looks for mercy from foreign arbitrator

http://foxsports.com

John Rocker arrived in New York Monday, seeking to overturn his suspension for making racist remarks about 'foreigners' in New York. Rocker was heckled by New York fans in the lobby of the MLB building and reportedly yelled back at them. Commissioner Bud Selig was the first man to appear at the hearing, and afterwards discussed the hearings.
Big Cat clears fence

Andres Galarraga hits his second dinger of the season out of Wrigley Field.

Big Cat goes yard

Andres Galarraga homers against the Braves.

Big Cat goes yard

Andres Galarraga hits his second dinger of the season out of Wrigley Field.
Looking ahead

FROM:
- Browsing
- Lexical search
- Data exchange
- Data retrieval

TO:
- Information requests
- Content search
- Semantic retrieval
- Interpretation
- Knowledge creation
- Knowledge sharing

perspective III: Semantic Information Brokering, Semantic Web
Evolving targets and approaches in integrating data and information *(a personal perspective)*

- **Generation I (multidatabases)**
  - 1980s
  - Mermaid
  - DDTS
  - Multibase, MRDSM, ADDS, IISS, Omnibase, ...

- **Generation II (mediators)**
  - 1990s
  - VisualHarness
  - InfoHarness
  - InfoSleuth, KMed, DL-I projects
  - Infoscopes, HERMES, SIMS, Garlic, TSIMMIS, Harvest, RUFUS, ...

- **Generation III (information brokering)**
  - 1997...
  - ADEPT
  - VideoAnywhere
  - InfoQuilt
  - DL-II projects, MediaAnywhere

The diagram illustrates the evolution of approaches and projects in the context of integrating data and information, highlighting different generations of technologies and methodologies.
Comprehensive knowledge-based, semantic information modeling, with multiple domain ontologies as a starting point, and

Distributed agents, to analyze Web-based content and establish semantic relationships.
Semantic Information Brokering
Semantic Web

enablers of the emerging concepts

- Terminology (and language) transparency
- Comprehensive metadata management
- Context-sensitive information processing
- Semantic correlation
Semantic Web

• “A Web in which machine reasoning will be ubiquitous and devastatingly powerful.”

• “A place where the whim of a human being and the reasoning of a machine coexist in an ideal, powerful mixture.”

• “A semantic Web would permit more accurate and efficient Web searches, which are among the most important Web-based activities.”

— A personal definition

Semantic Web: The concept that Web-accessible content can be organized semantically, rather than though syntactic and structural methods.
Example of searching on the semantic Web

1. User initiates query “Find an expert on XML” to DAML agent.

2. Student Web page contains markup describing university course “Advanced Web Development” that covers XML.

3. University course catalog page has markup that lists course “Advanced Web Development” as Web202.

4. University Web page profiling Professor Lila Smith has markup listing courses she teaches, including Web202.

5. Agent returns Professor Smith as potential XML expert.

Sites using DAML will enable agents to understand content in a Web page and use it intelligently with data from other pages.

Source: http://www.zdnet.com/pcweek/stories/jumps/0,4270,2432946,00.html
Information brokering is an architecture that guides creation and management of information systems and semantic-level solutions to serve a variety of information stakeholders (participants), including providers, facilitators, consumers, and the business involved in creating, enhancing and using of information.
Graduate students in a College of Geography have a final project in which a case of study is proposed. In the case, they are supposed to help a City Council in making decisions over the planning of a new landfill. This is a hands-on learning exercise through the interaction with a Digital Earth and the starting point would be to find the best location for the landfill*.

* This scenario comes in support of one of the suggestions for Digital Earth scenarios sampled by the "First Inter-Agency Digital Earth Working Group, an effort on behalf of NASA's inter-agency Digital Earth Program.
An example scenario of learning on the Web

A high level information request would be:

Find a landfill site for a new landfill near the source of the wastes.

The earthquakes’ impacts must be evaluated.

A first cut refinement leads us to the following information request:

Find a proper soil in sites not subject to flooding or high groundwater levels for a new landfill near the industrial zone.

Liquefaction phenomenon cannot occur.
An example scenario of learning on the Web

Adding on-the-fly user constraints while processing the information request:

*Retrieve satellite images in 12-meter resolution or higher,*

*looking for soils with permeability rate < 10 (silty clay loam)*

*for a new landfill*

*whose distance from the city industrial park is less than 5km.*

*Using the images’ coordinates, forecast seismic activity up to moderate magnitude (5 - 5.9, Richter scale) in the pointed areas.*

- domain specific metadata;
- correlation among multiple ontologies;
- return results in multiple media (in this case, images and a simulation)
An example scenario of learning on the Web

Partial sample ontologies for semantic information brokering:

- LAND (SITE)
  - LANDFILL SITE
  - CULTIVATED AREA
  - GREENLAND AREA
  - LAND BANK
- ZONING
- WASTE DISPOSAL
- SOLID
  - LANDFILL
  - RESOURCE REC.
  - RECYCLING
    - shredding
    - magnetic separation
    - screening
    - washing
- SEWAGE
- COMERCIAL
- INDUSTRIAL
- RURAL
- RESIDENTIAL
- MILITARY
- AGRICULTURAL
- RECREATIONAL
- LANDFILL SITE
- NATURAL DISASTER
  - TSUNAMI
  - VOLCANO
  - FIRE
  - FLOOD
  - STORM
  - EARTHQUAKE
  - LANDSLIDE

causes
causes
causes
Potential locations for a future shopping mall identified by all regions having a population greater than 5000, and area greater than 50 sq. ft. having an urban land cover and moderate relief. <A MRER ATTRIBUTES(population > 5000; area > 50; region-type = 'block'; land-cover = 'urban'; relief = 'moderate') can be viewed here</A>

domain specific metadata: terms chosen from domain specific ontologies

=> media-independent relationships between domain specific metadata: population, area, land cover, relief

=> correlation between image and structured data at a higher domain specific level as opposed to physical “link-chasing” in the WWW
Generation II

(a metadata classification: the information pyramid)

- User
  - Ontologies
    - Classifications
    - Domain Models
  - Domain Specific Metadata
    - Domain Independent (structural) Metadata
      - Area, population (Census), land-cover, relief (GIS), metadata concept descriptions from ontologies
    - Direct Content Based Metadata
      - (inverted lists, document vectors, WAIS, Glimpse, LSI)
    - Content Dependent Metadata
      - (size, max colors, rows, columns...)
    - Content Independent Metadata
      - (creation-date, location, type-of-sensor...)
- Data (Heterogeneous Types/Media)

METADATA STANDARDS
- General Purpose: 
  - Dublin Core, MCF
- Domain/industry specific: 
  - Geographic (FGDC, UDK, …), Library (MARC, …)

Move in this direction to tackle information overload!!
An example scenario of learning on the Web

A sample result (depending on information providers) could be:

OrbView-4’s stereo imaging capacity providing 3-D terrain images

Hyperspectral data will be valuable for identifying material types

⇒ The students now have the information requested for helping the City Council in the planning of the new landfill
Agents for inter-ontology interoperation

In our starting example scenario:

Integrating aerial-videography with the process of land use classification is identified as an exciting example of SEMANTIC INFORMATION BROKERING, leading to the gathering of ground-truth satellite image interpretation and post-classification land-cover mappings.
Agents for media independent correlations

**Metadata Agent**
- Information content
- Metadata extraction
- Metadata exportation
- Information requests

**Correlation Agent**
- Control strategy
- User constraints
- GUI construction
- Final answer (objects satisfying constraints)

**Provider Agent**
- Metadata extraction

**Seismicity (SQL)**
- UCSB’s Institute for Crustal Studies Earthquake Collection

**Land (SQL)**
- USGS Global Land Info System

**Image Features (IP routines)**
- NASA’s EOS Earth Images and Data

**Satellite image:**
- Landfill site:
- Seismic forecast:

**User Interface**
- Metadata extraction
- Metadata exportation
- Information requests
- GUI construction
- User constraints
- Final answer (objects satisfying constraints)
Semantic Information Brokering

A taxonomy for heterogeneity and interoperability

INFORMATION HETEROGENEITY

Semantic Heterogeneity
Structural, Representational/Schematic Heterogeneity
Syntactic, Format Heterogeneity

SYSTEM HETEROGENEITY

Information System Heterogeneity
Digital Media Repository Management Systems
Database Management Systems, (heterogeneity of DBMSs, data models, system capabilities such as concurrency control and recovery)

Platform Heterogeneity
Operating System (heterogeneity of file system, naming, file types, operation, transaction support, IPC)
Hardware/System (heterogeneity of instruction set, data representation/coding)
Semantic Information Brokering

**Impacts on information overload**

**Heterogeneity:**

- differences in structure;
- differences in query languages;
- semi-structured or unstructured data;
- media heterogeneity;
- terminology (and language) heterogeneity;
- contextual heterogeneity
Heterogeneity... is a Babel Tower!!

SEMANTIC HETEROGENEITY

metadata

ontologies

contexts

Heterogeneity... is a Babel Tower!!

SEMANTIC INTEROPERABILITY
Semantic Information Brokering

Impacts on information overload

- Globalization: challenge of the scale
  - information resource discovery;
  - modeling of information content;
  - querying of information content
    - information focusing
    - information correlation

Much more than returning documents with potentially relevant information.
Semantic Information Brokering

Stakeholders and components

- **CONSUMERS**
- **FACILITATORS**
- **PROVIDERS**

- **SEMANTICS**
- **STRUCTURE**
- **SYNTAX**
- **SYSTEM**

- **METADATA**
- **DATA**
- **TERMINOLOGY**
Semantic Information Brokering

Handling the information overload: stakeholders and beneficiaries in information brokering

key player on the GII, who help to match the consumers' REQUESTS with the INFORMATION exported by the providers

key player on the GII, which USE information

key player on the GII, which EXPORT information
Semantic Information Brokering

Stakeholder perspective

INFORMATION CONSUMERS
- Corporations
- Universities
- People
- Programs
- Government

User Query

INFORMATION/DATA OVERLOAD

INFORMATION PROVIDERS
- Newswires
- Corporations
- Research Labs

INFORMATION SYSTEMS
- Information System
- Data Repository

INFORMATION BROKERING
- Information Request
- Information Request
- Information Request

Corporations
Universities
People
Programs
Government
Semantic Information Brokering

Handling the information overload

Information facilitators enable brokering between the information consumer and provider that may be defined as:

- arbitration between consumers and providers for resolving the information impedance, the differing worldviews of the consumers and providers;

- dynamic re-interpretation of information requests for determination of relevant information services and products;

- dynamic creation and composition of information products after suitable assembly or correlation of information components available from the various providers, or other value addition activities.
Semantic Information Brokering

Levels of information brokering

- semantic information brokering
- metadata-based brokering
- data brokering standards-based interoperability
Handling the information overload

The role of facilitators on the GII:

- design and construction of multimedia views using metadata descriptions from standardized terminologies available to the consumers;

- maintenance and storage of associations between the digital data exported by the providers and the metadata descriptions designed by it;

- bridge between providers and consumers

  - use of domain specific ontologies to characterize vocabularies and terminological relationships in the interoperation, reducing information overload
Services the facilitators can offer:

- provide a collection of standard terminologies captured as domain specific ontologies to the consumers, used by the providers to construct metadata descriptions;

- support the interoperation across multiple domain specific ontologies;

- support the transformation of information requests across different ontologies, minimizing the loss of information.

Key objective of the approach: enable semantic interoperability, reducing the problem of knowing the structure and semantics of data in the huge number of information sources.
Semantic Information Brokering

Appropriateness of multi-agents systems

.extensions to agent functionalities to support semantic information brokering:

- capture, view and interrogate the semantics of the underlying (multimedia) data;
- request information without regards to the underlying data type, structure, format or media in which the information may be represented or stored;
- support interoperation across multiple standard terminologies or domain specific ontologies.
Multi-agents architecture (personal perspective)
Some volcanoes are more active than others, and a few are in a state of permanent eruption, at least for the geological present. Volcanoes may become *quiescent* (dormant) for months or years. The danger to life posed by active volcanoes is not limited to eruption of molten rock or showers of ash and cinders. Mudflows that melt ice and snow on the volcano's flanks are equally hazardous.*

* Encarta® 98 Desk Encyclopedia © & 1996-97 Microsoft Corporation. All rights reserved.
A sample information request:

Find information on volcanoes and also find how these volcanoes affect/cause landslides and tsunamis.

Some of the ontologies involved in processing this information request are:

- Ontology for GIS Datasets;
- Ontology for Natural Disasters;
- Ontology for Volcanoes;
- Ontology for Landslides;
- Ontology for Tsunamis.
Realizing Semantic Information Brokering and Semantic Web *in summary*

<table>
<thead>
<tr>
<th>Visual, Scientific/Eng.</th>
<th>Knowledge</th>
<th>Semantic</th>
<th>Knowledge Mgmt., Information Brokering, Cooperative IS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-structured</td>
<td>Metadata</td>
<td>Structural, Schematic</td>
<td>Mediator, Federated IS</td>
</tr>
<tr>
<td>Text Structured Databases</td>
<td>Data</td>
<td>Syntax, System</td>
<td>Federated DB</td>
</tr>
</tbody>
</table>

**Visual, Scientific/Engineering**
- Visual, Scientific/Engineering
- Knowledge
- Semantic
- Knowledge Mgmt., Information Brokering, Cooperative IS
- Semi-structured
- Metadata
- Structural, Schematic
- Mediator, Federated IS
- Text Structured Databases
- Data
- Syntax, System
- Federated DB
Realizing Semantic Information Brokering and Semantic Web in conclusion

- From Procedures, Objects, Components to Agents we have a nice abstraction of computation. Now let’s apply them to address semantic-level issues.

- Semantic Information Brokering gives a framework for enabling complex decision making and learning involving heterogeneous digital media on the GII, which is much more than current e-commerce involving products and services.

- Semantic Web is a basis of handling information overload.
A few references (with Paolo). Also,

http://www.semanticweb.org

“DAML could take search to a new level”
http://www.zdnet.com/pcweek/stories/news/0,4153,2432538,00.html

R. Hellman, “A Semantic Approach Adds Meaning to the Web”,

“Humankind has not woven the web of life.
We are but one thread within it.
Whatever we do to the web, we do to ourselves.
All things connect.”
– Chief Seattle, 1854

amit@taalee.com – http://www.taalee.com
amit@cs.uga.edu – http://lsdis.cs.uga.edu