Interoperable Information System
(a personal perspective)

<table>
<thead>
<tr>
<th>Generation III</th>
<th>Information System</th>
</tr>
</thead>
<tbody>
<tr>
<td>(information brokering)</td>
<td>Semantic Web, some DL-II projects, Taalee Semantic Engine, Oingo</td>
</tr>
<tr>
<td>1997...</td>
<td>InfoQuilt, ADEPT, VideoAnywhere</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generation II</th>
<th>Metadata System</th>
</tr>
</thead>
<tbody>
<tr>
<td>(mediators)</td>
<td>InfoSleuth, KMed, DL-I projects, Infoscopes, HERMES, SIMS, Garlic, TSIMMIS, Harvest, RUFUS, ...</td>
</tr>
<tr>
<td>1990s</td>
<td>VisualHarness, InfoHarness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generation I</th>
<th>Data System</th>
</tr>
</thead>
<tbody>
<tr>
<td>(federated DB/ multidatabases)</td>
<td>Multibase, MRDSM, ADDS, IISS, Omnibase, ...</td>
</tr>
<tr>
<td>1980s</td>
<td>Mermaid, DDTS</td>
</tr>
</tbody>
</table>

Semantics (Ontology, Context, Relationships, KB)
Metadata (Domain model)
Data (Schema, “semantic data modeling”)
“The Web of data (and connections) with meaning in the sense that a computer program can learn enough about what the data means to process it. . . . Imagine what computers can understand when there is a vast tangle of interconnected terms and data that can automatically be followed.” (Tim Berners-Lee, Weaving the Web, 1999)

A personal definition
**Semantic Web:** The concept that Web-accessible content can be organized and utilized semantically, rather than through syntactic and structural methods.
Semantics:

- “meaning or relationship of meanings, or relating to meaning” (Webster)
- is concerned with the relationship between the linguistic symbols and their meaning [Kromrey 1994]
- meaning and use of data (Information System)
Important Building Blocks

- Semantic Domain Modeling
- Semantic Inter-Domain Relationship Modeling
- Resource Characteristics Modeling
- Complex Operations
- Semantic Information Request Modeling
- Knowledge Discovery Paradigm
Outline

- Commercial SOA of Semantic Web
- Research SOA
- Research for realizing full potential of Semantic Web - complex relationships
Taalee Semantic Engine™

Content Aggregation and Enhancement
- Content Understanding Applications
- Classifiers
- Catalogers
- Relationship Builder
- Metadata Enhancers

Intelligent Content
- WorldModel™
- Semantic Web
- Content Metabase

Content Engines
- Search
- Directory
- Personalization
- Content Integration
- Targeting
- Alert
- Semantic Cluster
- Interactive Marketing & E-Commerce

Third Party Applications
- Content Management
- Syndication
- Targeting
Content of all format, media, push/pull:
- Web sites/pages: static, dynamic
- Content Feeds (unstructured, semistructured/docs, tagged/XML)
- Corporate Repositories/databases

Homogenization/integration:
- with taxonomy (categorization)
- contextually relevant metadata wrt to domain model, automatically generated from content and inferenced
Semantic Content

End-User

Semantic Content

Content which does not contain the words the user asked for

Extractor Agents

Content which does not contain the words the user asked for, but is about what he asked for.

Value-added Metadata

Content the user did not think to ask for, but which he needs to know.

Semantic Associations
Semantic Associations using Simple Relationships supported by Taalee Semantic Engine

- RELATED STOCK NEWS: COMPANIES in Same or Related INDUSTRY
- TECHNOLOGY PRODUCTS: Important to INDUSTRY or COMPANY
- INDUSTRY NEWS: Impacting INDUSTRY or COMPANY
- REGULATIONS: Impacting INDUSTRY or Filed By COMPANY
- COMPETITION: COMPANIES in INDUSTRY with Competing PRODUCTS

COMPANY
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Change</th>
<th>Price</th>
<th>Volume</th>
<th>NASDAQ</th>
<th>Dow</th>
<th>S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTC</td>
<td>+0.81</td>
<td>30.06</td>
<td>38,618,100</td>
<td>2084.79</td>
<td>10593.06</td>
<td>1215.68</td>
</tr>
</tbody>
</table>

### Analysis News
- J.P. Morgan Boosts Microsoft to 'Buy' 06/29/2001 BusinessWeek
- Microsoft Corp. 06/29/2001 ON24
- Microsoft Corp. 06/28/2001 ON24

### Company News
- Websense Inc. Presents Corporate Highlights at The... 06/18/2001 COMTEX
- London shares still on offer midday, techs and ad... 06/15/2001 CNNFN
- THE WASHINGTON POST: NINTENDO'S AIM IS BIGGER 06/12/2001 COMTEX

### Market Commentary News
- Techs rise modestly 07/02/2001 CNNFN
- European bourses mixed 06/29/2001 CNNFN
- Markets Up On Microsoft News 06/28/2001 CBS

### Mergers & Acquisitions
- Landis acquires QUAYONE 07/02/2001 COMTEX
- THE TIMES: MICROSOFT STILL FACES BATTLE 07/01/2001 COMTEX
- BOSTON GLOBE: STOCKS BOUNCE ON RULING 06/29/2001 COMTEX

### Industry and Competition News
- Websense Inc. Presents Corporate Highlights at The... 06/18/2001 COMTEX
- London shares still on offer midday, techs and ad... 06/15/2001 CNNFN
- THE WASHINGTON POST: NINTENDO'S AIM IS BIGGER 06/12/2001 COMTEX

### Company News
- Techs rise modestly 07/02/2001 CNNFN
- European bourses mixed 06/29/2001 CNNFN
- Markets Up On Microsoft News 06/28/2001 CBS

© Taalee Inc
Semantic Web Research today

- Resource Modeling/Markup with ontology
  - RDFS, OIL, DAML(+OIL)
  - service modeling?
- Inferencing/Reasoning: Description Logic (subsumption), frame based, theorem proving ...
- Applications: Search
- Query Language: ?
- Learning, discovery, : ??
Classical “crisp” logic, inferencing..
Same fate as Q/A systems?
Some comments on State of the Art

- “the ontology cited in the query matches one of the ontologies in the list of the namespaces of the web page” -- what if they have different ontologies (likely)
- “DAML-M would be a service that for a given ontology and a set of classes and properties returns mappings to other ontologies and their classes and properties that are declared to be equivalent.” Equivalence is not sufficient... it quickly gets more complicated.
- “There might be a “Name-Match” service available that given a name like “Jim” returns “James” and “J.”. Highly context sensitive..so no generalized mappings are possible! Search engine cannot decide if name-match can be used (e.g., for legal purpose).

Quotes from Denker et al at SWWS, 2001
Complex Relationships

- Arise in several contexts, especially involving multiple ontologies (hence mappings)
  - information interoperability where related resources subscribe to different but related ontologies
  - information requestor and resource modelers choose to use different ontologies
  - information requests to support analysis, knowledge discovery, decision making, learning that requires linking multiple domains with different ontologies

Developing all encompassing, unified ontology is not shown to be practical. Preexisting classifications/metadata standards/taxonomies are hard to ignore.
Semantic Web And Beyond...

- Simple Relationships
  - (is-a, instance-of, part-of, value-of, less than)

- Complex, user defined, fuzzy relationships
  - (cause, affect)

- Information Interoperability/Integration

- Semantic Web

- Knowledge Discovery
Complex Relationships and Knowledge discovery

- VOLCANO
  - LOCATION
  - ASH RAIN
  - PYROCLASTIC FLOW

- ENVIRON.
  - BUILDING
  - WEATHER
  - PEOPLE
  - PLANT

AFFECTS

- DESTROYS
- COOLS TEMP
- DESTROYS
- KILLS

LOCATION

- DESTROYS
InfoQuilt System Core capabilities

- Ability to handle heterogeneous, static or dynamic content – wrappers & extractors, with resource modeling (completeness, data characteristics, binding patterns)
- Information Extraction: Semi-Automatically or Automatically create domain-specific or contextually relevant metadata
- Domain modeling with complex (user defined, inter-ontology) relationships, domain rules and FD
- User defined Functions (esp. for fuzzy/approximate matching) and Simulation
- Post processing result analysis (e.g., chart creator)
A nuclear test could have *caused* an earthquake if the earthquake occurred *some time after* the nuclear test was conducted and *in a nearby region*.

**NuclearTest Causes Earthquake**

\[
\text{<= dateDifference( NuclearTest.eventDate, Earthquake.eventDate ) < 30}
\]
\[
\text{AND distance( NuclearTest.latitude, NuclearTest.longitude, Earthquake.latitude, Earthquake.longitude ) < 10000}
\]
Iscape

- A computing paradigm that allows users to query and analyze the data available from a diverse autonomous sources, gain better understanding of the domains and their interactions as well as discover and study relationships.
Iscape ...a simple example

- user’s request
  - for semantically related information (regardless of all forms of heterogeneity)
  - specified in terms of components of knowledge base (domain model, relationships, functions, simulations)

“Find all earthquakes with epicenter less than 5000 mile from the location at latitude 60.790 North and longitude 97.570 East and find all tsunamis that they might have caused”

Next - KD using ISacpes
Knowledge Discovery - Example

Earthquake Sources

Nuclear Test Sources

Nuclear Test May Cause Earthquakes

Is it really true?

Complex Relationship:
How do you model this?
Knowledge Discovery - Example

When was the first recorded nuclear test conducted?

1950

Find the total number of earthquakes with a magnitude 5.8 or higher on the Richter scale per year starting from 1900

Increase in number of earthquakes since 1945
For each group of earthquakes with magnitudes in the ranges 5.8-6, 6-7, 7-8, 8-9, and >9 on the Richter scale per year starting from 1900, find number of earthquakes

Number of earthquakes with magnitude > 7 almost constant. So nuclear tests probably only cause earthquakes with magnitude < 7
Knowledge Discovery - Example...

Find nuclear tests and earthquakes that may have occurred as a result of the test
Ontologies

Terms/Concepts (Attributes)
- site
- eventDate
- description
- site => latitude, longitude
- damage
- damagePhoto
- numberOfDeaths
- magnitude
- magnitude > 0
- magnitude < 10
- bodyWaveMagnitude
- bodyWaveMagnitude < 10
- bodyWaveMagnitude > 0
- conductedBy
- explosiveYield

Hierarchies
- Natural Disaster
- Man-made Disaster
- Volcano
- NuclearTest
- Earthquake

Functional Dependencies (FDs)
- site => latitude, longitude

Domain Rules

Functional Dependencies (FDs)
- bodyWaveMagnitude
- conductedBy
- explosiveYield
- bodyWaveMagnitude > 0
- bodyWaveMagnitude < 10
Knowledge Builder

Knowledge Tree

- Nuclear Test
  - testSite
  - bodyWaveMagnitude
  - explosionYield
  - eventDate
  - eventType
  - conductedBy
  - longitude
  - latitude
  - F1 (testSite) -> (longitude)(latitude)
  - Rule bodyWaveMagnitude GREATER THAN 0
  - Rule bodyWaveMagnitude LESS THAN 10
- TestSitesDB
  - testSite
  - explosiveYield
  - bodyWaveMagnitude
  - eventType
  - eventDate
  - conductedBy
  - DC bodyWaveMagnitude GREATER THAN 3
- Nuclear Test Causes Earthquake

- Ontology
- Attribute
- DC Data Characteristic
- Relationship
- Condition
- Resource
- Local Completeness
- Functional Dependency
- Binding Pattern
IScape Execution

1. Broke
2. Planning Optimize Agent
3. Knowledge Agent
4. Query
5. Data retrieved
6. Query
7. Plan
8. Final Results
9. IScape

Final Results

Knowledge

Extract/Wrapper

Web

Database

Other Wrapper

Other
Planning and Optimization

- IScapes are specified in terms of ontologies
- Source selection
- Execution plans that are *executable*
- Plans that retrieve more complete information
- Integrate data from sources
- Optimization using domain and resource characteristics
“Find all nuclear tests conducted by India or Pakistan after January 1, 1995 with seismic body wave magnitude > 4.5 and find all earthquakes that could have been caused due to these tests.”
Missing Attributes

- Use functional dependencies (FD)
  - \(<\text{attribute}>^+ \rightarrow <\text{missing attributes}><\text{attribute}>^*\)
- Couple with *associate resource*

Join (using LHS attributes)

Primary resource
- A, B, C, D

Associate resource
- A, B, C, E, F

Primary.B = Associate.B AND Primary.C = Associate.C

BC \rightarrow DEF

LHS attributes + missing attributes
  - (B, C, E, F)
IScape Processing Monitor (IScape 1)

<table>
<thead>
<tr>
<th>Message Id</th>
<th>Time Stamp</th>
<th>Message From</th>
<th>Brief Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20:29:14.484</td>
<td>User Agent</td>
<td>Started processing</td>
</tr>
<tr>
<td>1</td>
<td>20:29:14.671</td>
<td>Broker Agent</td>
<td>Started processing</td>
</tr>
<tr>
<td>2</td>
<td>20:29:15.046</td>
<td>Planning Agent</td>
<td>Applied Domain Rules to IScape's constraint</td>
</tr>
<tr>
<td>3</td>
<td>20:29:15.189</td>
<td>Planning Agent</td>
<td>Checking IScape's constraint and relationship</td>
</tr>
<tr>
<td>4</td>
<td>20:29:15.187</td>
<td>Planning Agent</td>
<td>Selecting sources for Earthquake</td>
</tr>
<tr>
<td>5</td>
<td>20:29:15.234</td>
<td>Planning Agent</td>
<td>Selecting sources for Nuclear Test</td>
</tr>
<tr>
<td>6</td>
<td>20:29:16.453</td>
<td>Planning Agent</td>
<td>Plan created by the planner, Returning it to Bro...</td>
</tr>
<tr>
<td>7</td>
<td>20:29:15.578</td>
<td>Broker Agent</td>
<td>Received plan from planner</td>
</tr>
<tr>
<td>8</td>
<td>20:29:15.814</td>
<td>CorrelationAgent</td>
<td>Executing IScape</td>
</tr>
<tr>
<td>9</td>
<td>20:29:16.390</td>
<td>TestSitesDB Resource Agent</td>
<td>Queried TestSitesDB</td>
</tr>
<tr>
<td>11</td>
<td>20:29:16.500</td>
<td>Correlation Agent</td>
<td>Computed Union</td>
</tr>
<tr>
<td>13</td>
<td>20:29:16.843</td>
<td>CorrelationAgent</td>
<td>SelectNodeProcessor done processing</td>
</tr>
<tr>
<td>14</td>
<td>20:29:16.904</td>
<td>Correlation Agent</td>
<td>Evaluated Join</td>
</tr>
<tr>
<td>16</td>
<td>20:29:17.250</td>
<td>Correlation Agent</td>
<td>SelectNodeProcessor done processing</td>
</tr>
<tr>
<td>17</td>
<td>20:29:17.359</td>
<td>CorrelationAgent</td>
<td>Evaluated Relationship on a set of relations</td>
</tr>
<tr>
<td>18</td>
<td>20:29:17.750</td>
<td>Correlation Agent</td>
<td>Evaluated projection on a relation</td>
</tr>
<tr>
<td>19</td>
<td>20:29:17.765</td>
<td>Correlation Agent</td>
<td>Received IScape results from Correlation Ag...</td>
</tr>
<tr>
<td>20</td>
<td>20:29:17.906</td>
<td>Broker Agent</td>
<td>Returning final result to client</td>
</tr>
<tr>
<td>21</td>
<td>20:29:17.957</td>
<td>User Agent</td>
<td></td>
</tr>
</tbody>
</table>
Related Work

- Features of InfoQuilt not supported by any other systems
  - Ability to assist in learning about domains and complex inter-domain relationships
  - Support for use of functions and simulations to post-process
  - Support for complex relationships and constraints that can use functions as special operators
  - Powerful semantic query interface (IScapes)
Related Work

- **SIMS/ARIADNE/InfoMaster**
  - Mediator specialized to one domain
  - No local completeness information about sources

- **OBSERVER**
  - Limited to basic relationships
  - Resource models are not as rich
Related Work...

- **TSIMMIS**
  - Mediators defined using MSL
  - Adding or removing sources is difficult
  - Query-centric (uses pre-defined query templates)
  - Can answer a restricted set of queries

- **Information Manifold**
  - No domain rules, FDs
  - Capability records cannot model query capability limitations precisely
Future

- SemWeb activities (DAML+OIL)
- Heuristic, Inexact/Fuzzy vs Formal/Classical Reasoning (DL??)
- P2P - more dynamic/volatile/heterogeneous
Thank You!

http://lsdis.cs.uga.edu
<table>
<thead>
<tr>
<th>Modeling Features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>User-Defined Functions</td>
<td>Users can provide their own implementation of function based on their own requirements</td>
</tr>
<tr>
<td>Complex Predicates/</td>
<td>Relationships can be defined as complex expressions based on user-defined and/or existing functions, relational and logical operators.</td>
</tr>
<tr>
<td>Relationships</td>
<td>This means that properties of relationships beyond sub-typing and part-of can be modeled explicitly e.g. Causes</td>
</tr>
<tr>
<td>Iscape</td>
<td>Ability to define complex information requests utilizing existing domain models, user-defined functions and complex relationships. Ability to re-use and re-configure such requests</td>
</tr>
<tr>
<td>Mapping Rules</td>
<td>Enables users and domain experts to model mappings between concepts in different ontologies.</td>
</tr>
<tr>
<td></td>
<td>This allows for data integration analogous to multi-database joins but more complex</td>
</tr>
<tr>
<td>Resource Modeling</td>
<td>Captures description of actual content as opposed to potential content which helps in better resource selection</td>
</tr>
<tr>
<td>Execution Features</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>Iscape Consistency Checks</td>
<td>- Checks if Iscape is meaningful with respect to domain models chosen. Support for dynamic information resources</td>
</tr>
<tr>
<td>Resource Selection</td>
<td>- More specific because of dc and lc rules</td>
</tr>
<tr>
<td></td>
<td>- More complete because of FDs.</td>
</tr>
<tr>
<td>Binding Patterns</td>
<td>- Support for dynamic information resources</td>
</tr>
</tbody>
</table>