



# On Supporting Complex Relationships and Knowledge Discovery in Semantic Web

**Amit Sheth**

**Large Scale Distributed Information Systems (LSDIS) Lab  
University Of Georgia**

**Founder/CEO, Taalee, Inc → CTO, Voquette, Inc.**

**August 2001**



# Interoperable Information System

*(a personal perspective)*

**Semantics** (Ontology, Context, Relationships, KB)

Generation III  
(information  
brokering)  
1997...

**InfoQuilt**  
**ADEPT**  
**VideoAnywhere**

Semantic Web, some DL-II projects,  
Taalee Semantic Engine, Oingo

**Metadata** (Domain model)

Generation II  
(mediators)  
1990s

**VisualHarness**  
**InfoHarness**

InfoSleuth, KMed, DL-I projects  
Infoscopes, HERMES, SIMS,  
Garlic, TSIMMIS, Harvest, RUFUS, ...

**Data** (Schema, "semantic data modeling")

Generation I  
(federated DB/  
multidatabases)  
1980s

**Mermaid**  
**DDTS**

Multibase, MRDSM, ADDS,  
IISS, Omnibase, ...



# Semantics:

## The Next Step in the Web's Evolution

---

***“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



Unstructured Text



Feeds



Websites



Enterprise Data

Content Understanding Applications

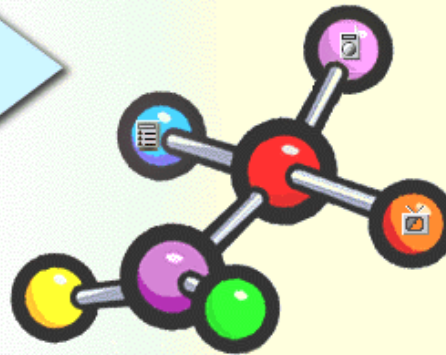
Classifiers

Catalogers

Relationship Builder

Metadata Enhancers

## Intelligent Content



Semantic Web



Content Metabase

## Content Engines

Search

Directory

Personalization

Content Integration

Targeting

Alert

Semantic Cluster

Interactive Marketing & E-Commerce



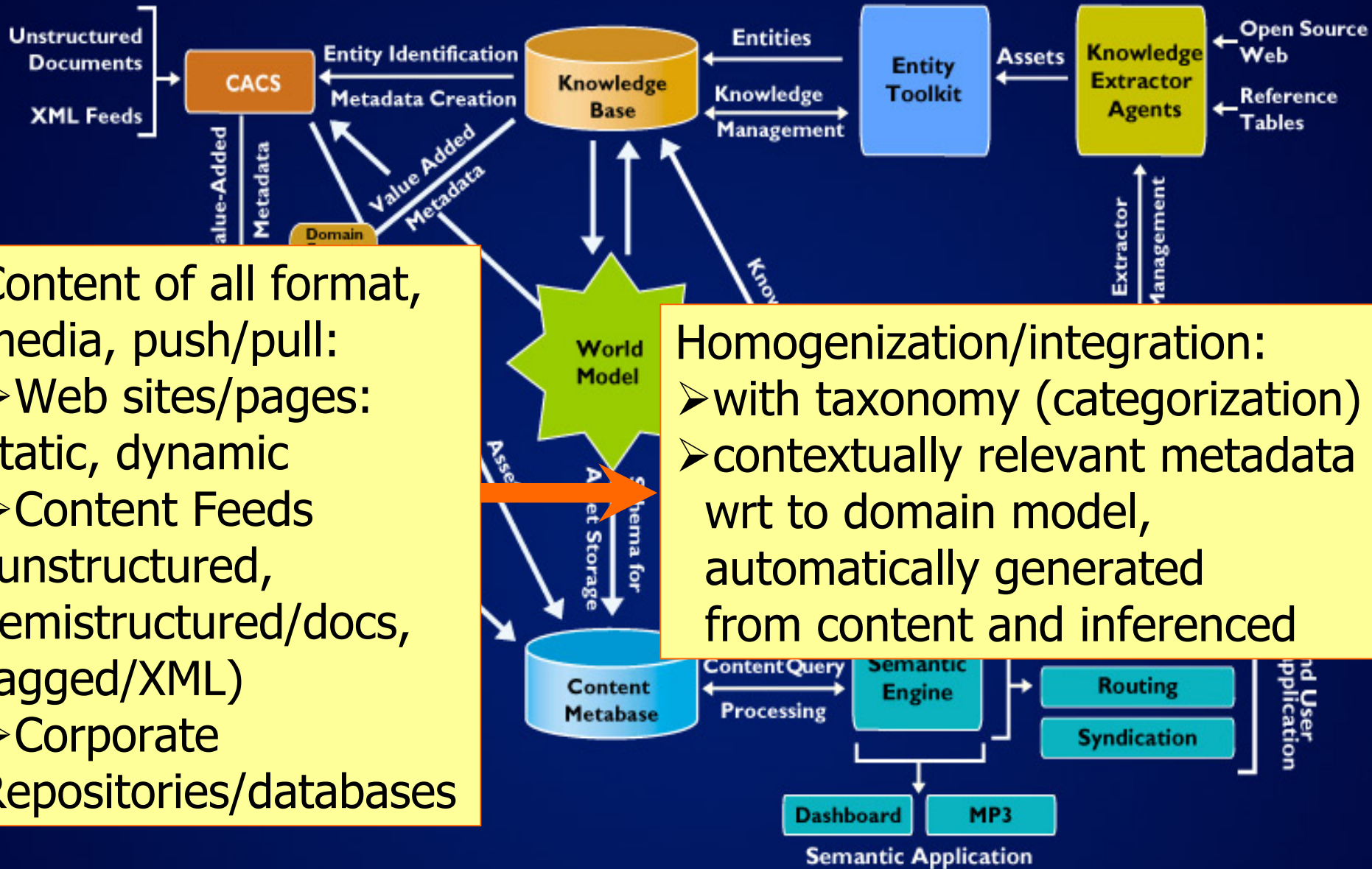
Third Party Applications

Content Management

Syndication

Targeting

# Taalee/Voquette Semantic Platform Architecture



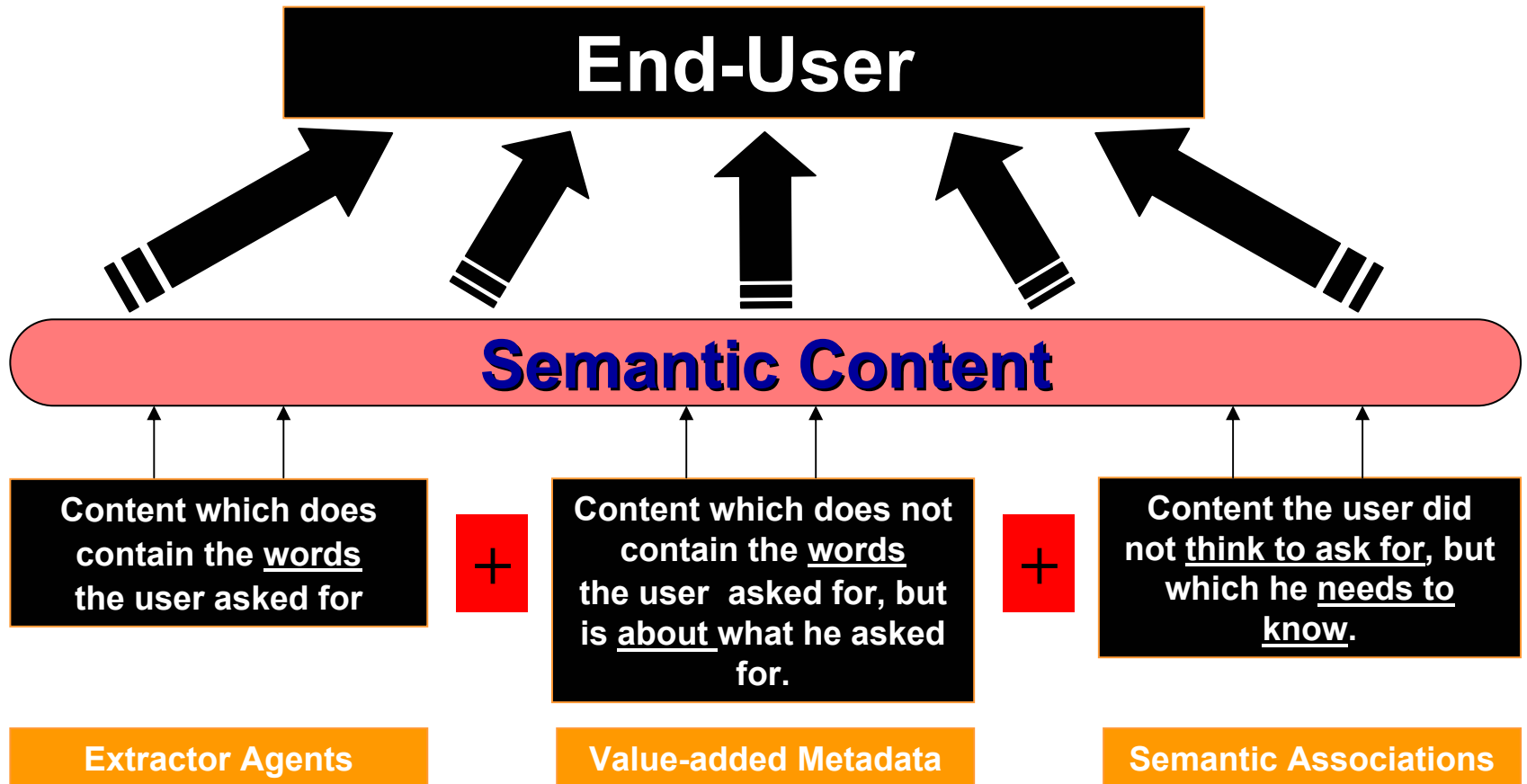
Content of all format, media, push/pull:

- Web sites/pages: static, dynamic
- Content Feeds (unstructured, semistructured/docs, tagged/XML)
- Corporate Repositories/databases

Homogenization/integration:

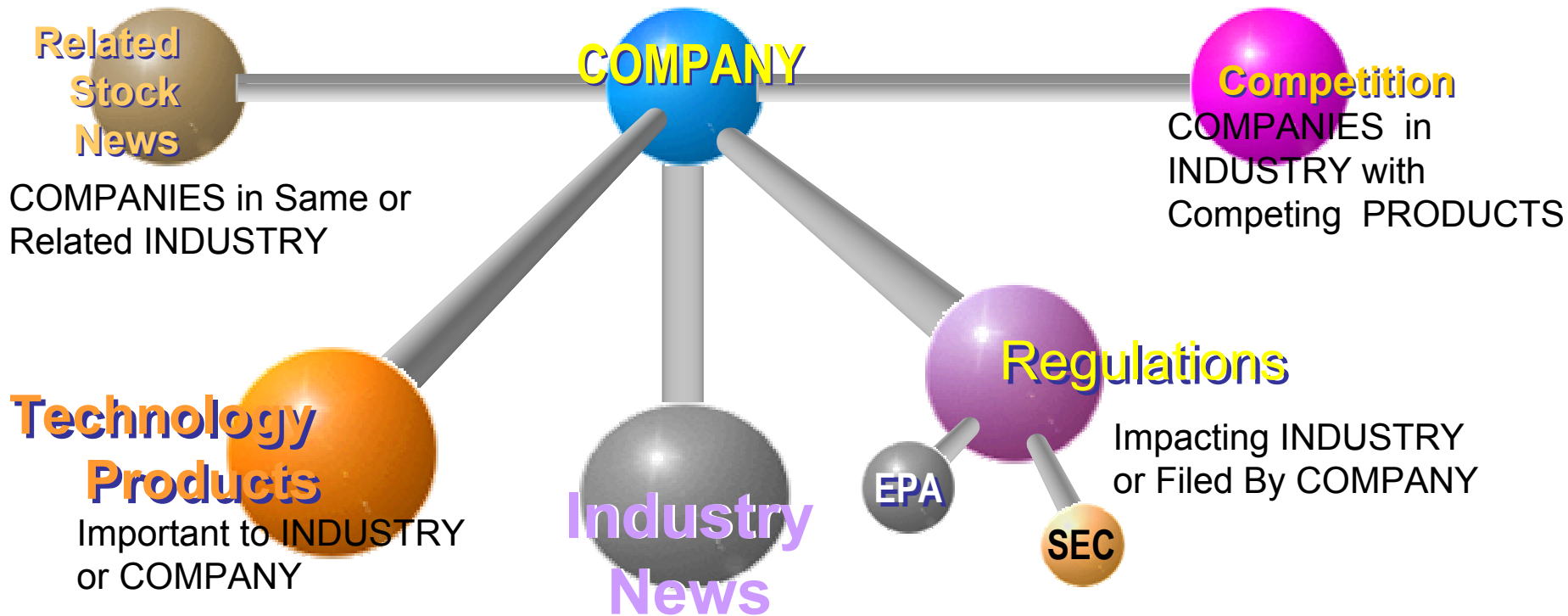
- with taxonomy (categorization)
- contextually relevant metadata wrt to domain model, automatically generated from content and inferenced

# Semantic Content



# Semantic Associations using Simple Relationships

supported by Taalee Semantic Engine

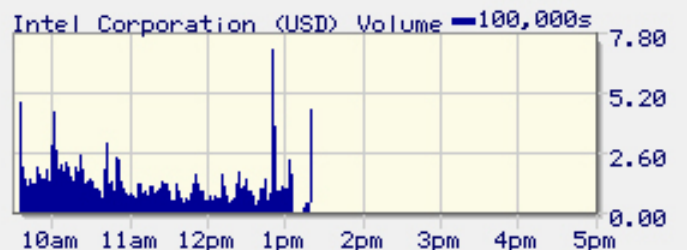


# Intel Corporation

07/03 4:57 PT

Symbol	Change	Price	Volume	Equity Indices	
INTC	+0.81	30.06	38,618,100	NASDAQ	2084.79
				DOW	10539.06
				S&P 500	1215.68

[View more charts...](#)



[View competitors...](#)

[Listen to audio programs...](#)

## Resources for Intel Corporation

- [TheStreet.com](#) Key Stats | Income | Cash Flow | Broker Rating
- [EXCAR's DataView](#) SEC Financials
- [Market Guide](#) Snapshot | Analyst Corner | Company Profile
- [sage](#) Message Boards
- [Zacks](#) Analyst Recommendations
- [MarketGauge](#) Price and volume action

## 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

- [vWebsense 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 Competitor News

- [vWebsense 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

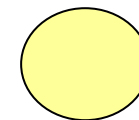
- [Techs rise modestly](#) 07/02/2001 CNNFN
- [European bourses mixed](#) 06/29/2001 CNNFN
- [Markets Up On Microsoft News](#) 06/28/2001 CBS



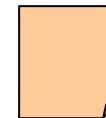
# Semantic Web Research today

---

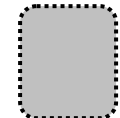
- Resource Modeling/Markup with ontology
  - RDFS, OIL, **DAML(+OIL)**
  - service modeling ?
- Inferencing/Reasoning: Description Logic (subsumption), frame based, theorem proving ...
- Applications: Search
- Query Language: ?
- Learning, discovery, : ??



Agent



Resource



Markup



---

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).**



# Complex Relationships

---

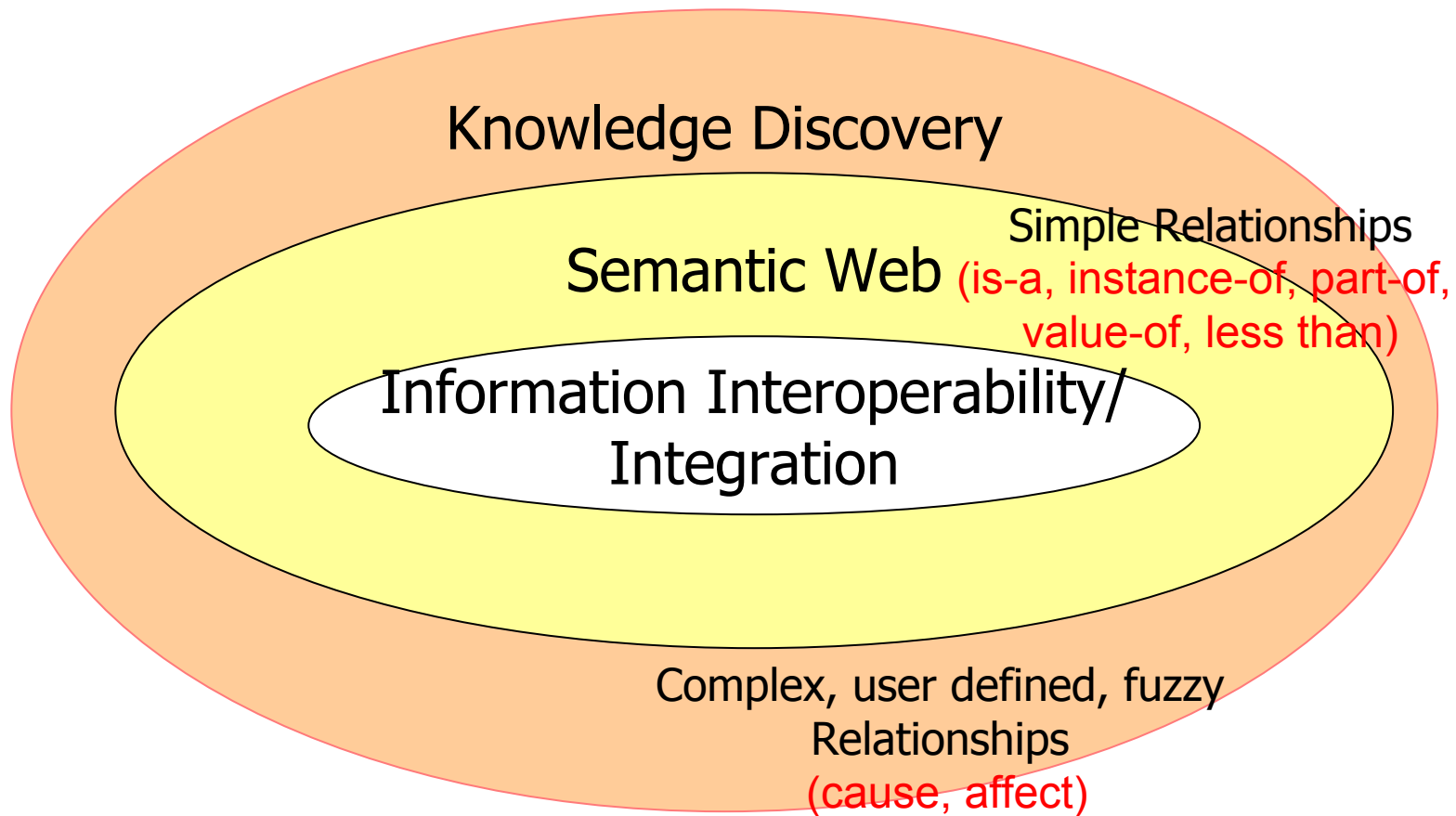
- Arise in several contexts, especially involving multiple ontologies (hence mappings)
  - information interoperability where related resources subscribe to different but related ontologies
  - information requestor and resource modelers choose to use different ontologies
  - information requests to 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.

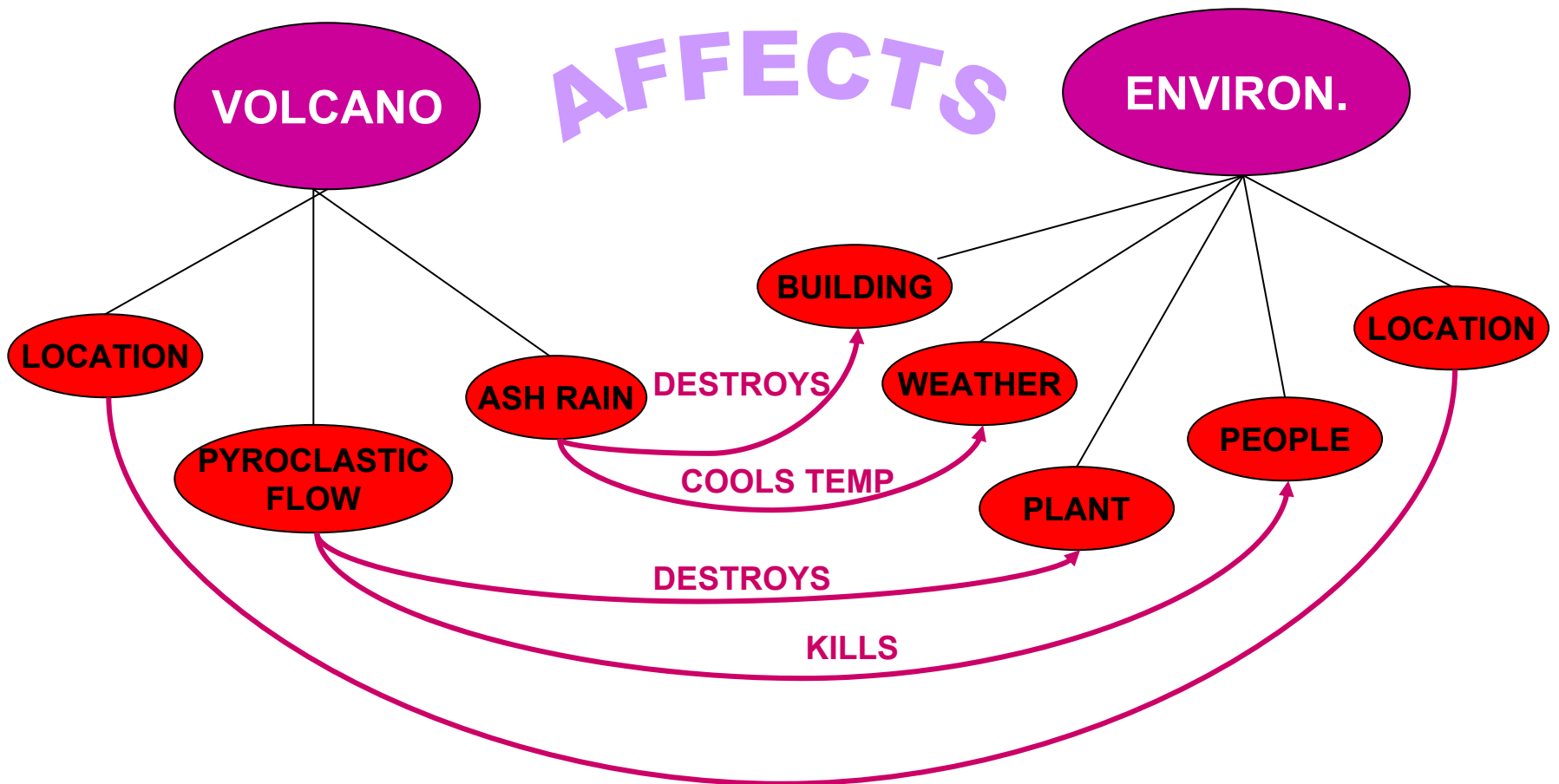


# Semantic Web And Beyond...

---



# Complex Relationships and Knowledge discovery





# 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)



# Inter-Ontological Relationships

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**

```
<= dateDifference( NuclearTest.eventDate,  
                    Earthquake.eventDate ) < 30  
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)*

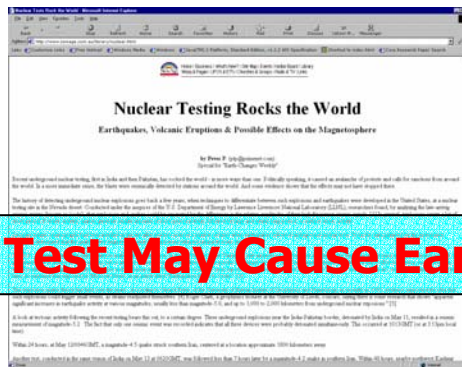
*“Find all earthquakes with epicenter less than 5000 mile from the location at latitude 60.790 North and longitude 97.570 East and find all tsunamis that they might have caused”*

Next - KD using ISapces

# Knowledge Discovery - Example



Earthquake Sources



Nuclear Test Sources

**Nuclear Test May Cause Earthquakes**

Is it really true?

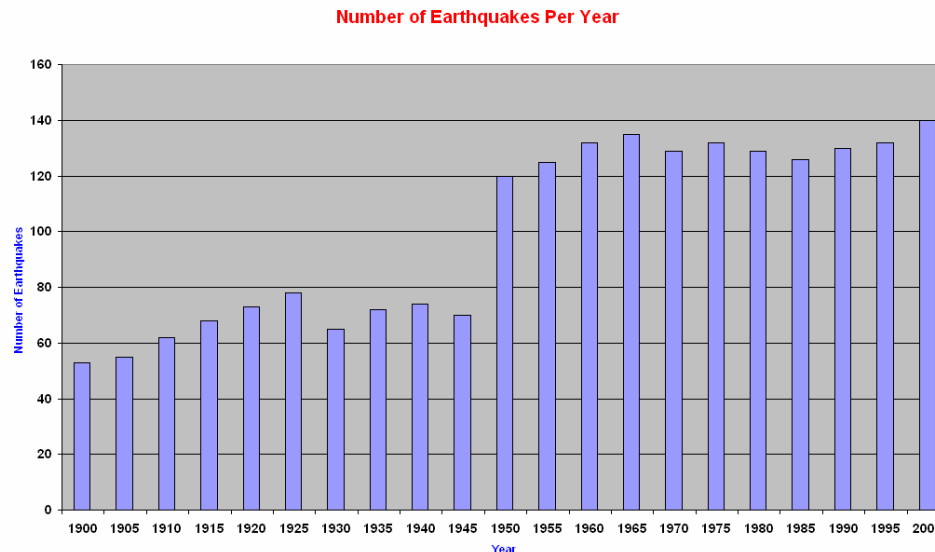
**Complex Relationship:**  
How do you model this?

# 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

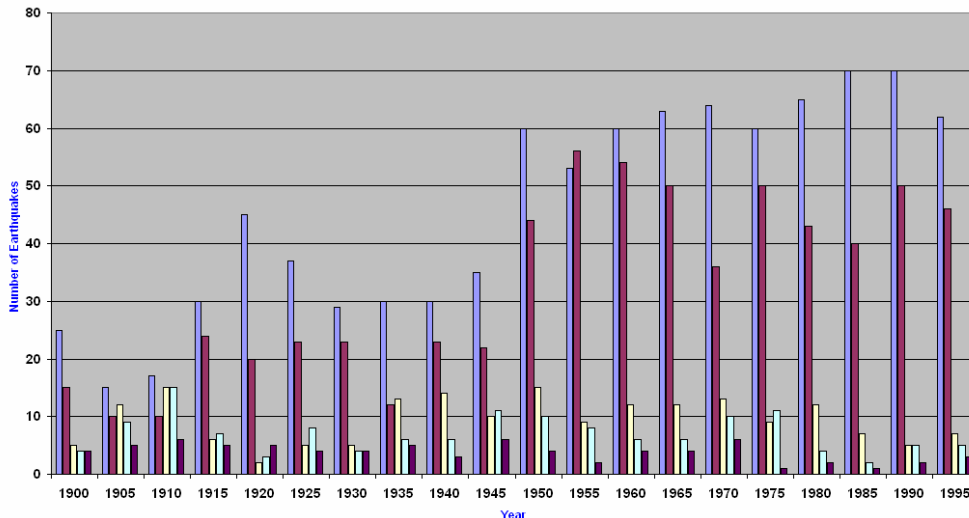


Increase in number of earthquakes since 1945

# Knowledge Discovery — exploring relationship...

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 Per Year By Magnitudes




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

**Result 2**

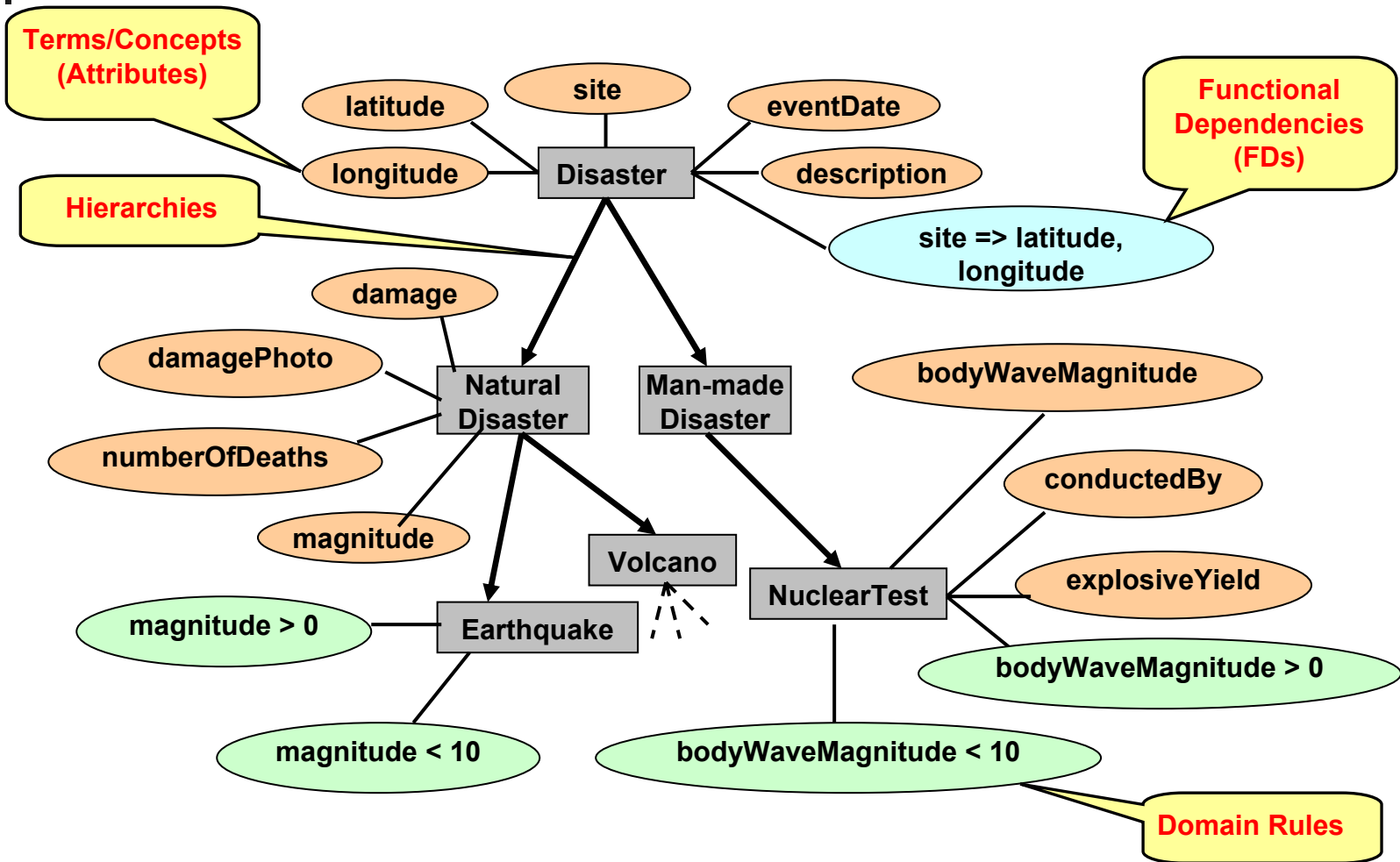
Earthquake.magnitude	5.5
Earthquake.region	Egypt, Cairo, Israel, Elat
NuclearTest.testSite	Pakistan, Chagai test site
NuclearTest.bodyWaveMagnitude	4.7
NuclearTest.explosiveYield	N/A
NuclearTest.conductedBy	Pakistan
distance (Earthquake.latitude ,NuclearTest.latitude ,Earthquake.longitude ,NuclearTest.longitude)	4427.121
diffInDays (NuclearTest.eventDate ,Earthquake.eventDate)	0
Earthquake.eventDate	1998, May, 28
Earthquake.description	One person injured at Cairo. Felt in the Cairo area. Also felt at Elat, Israel.
Multimedia (Earthquake.image)	

**Result 3**

Earthquake.magnitude	5.6
Earthquake.region	China, South Xinjiang
NuclearTest.testSite	India, Rajasthan, Pokaran test site
NuclearTest.bodyWaveMagnitude	4.7
NuclearTest.explosiveYield	N/A
NuclearTest.conductedBy	India
distance (Earthquake.latitude ,NuclearTest.latitude ,Earthquake.longitude ,NuclearTest.longitude)	5286.8604
diffInDays (NuclearTest.eventDate ,Earthquake.eventDate)	17
Earthquake.eventDate	1998, May, 28
Earthquake.description	Twenty-eight people injured, more than 2,000 buildings destroyed, an additional 3,000 damaged and 5,000 head of cattle killed in southwestern Xinjiang.
Multimedia (Earthquake.image)	

KB

# Ontologies



# Knowledge Builder

The screenshot displays the Knowledge Builder application with two main windows. The primary window is for editing the 'Earthquake' ontology, and a secondary window shows a tree view of the system's knowledge.

**Knowledge Builder - Earthquake Ontology Configuration**

- Ontology Name:** Earthquake
- Description (Do not use newline):** destruction at Kodiak, Cordova, and Seward and cause America, particularly at Crescent City, California. Some Liquefaction, which happens when loosely packed, wa strong shaking, causes major damage during earthqu liquefaction of the soils and debris used to fill in a lago sliding of the ground surface in the Marina district in Sa cause more destruction than the earthquakes themsel landslides devastated the Turnagain Heights residenti Anchorage.
- Parent Ontology:** <None>
- Keywords:** hazard, disaster, quake, seismic
- Buttons:** Add / Edit, Keyword, Remove Keyword, Save All

**Knowledge Tree - Knowledge of the system represented as a tree**

- NuclearTest** (Ontology)
  - testSite (Attribute)
  - bodyWaveMagnitude (Attribute)
  - explosiveYield (Attribute)
  - eventDate (Attribute)
  - testType (Attribute)
  - conductedBy (Attribute)
  - longitude (Attribute)
  - latitude (Attribute)
  - FD (testSite) -> (longitude)(latitude) (Functional Dependency)
  - Rule bodyWaveMagnitude GREATER THAN 0 (Rule)
  - Rule bodyWaveMagnitude LESSTHAN 10 (Rule)
- TestSitesDB** (Resource)
- NuclearTestsDB** (Resource)
  - testSite (Attribute)
  - explosiveYield (Attribute)
  - bodyWaveMagnitude (Attribute)
  - testType (Attribute)
  - eventDate (Attribute)
  - conductedBy (Attribute)
  - DC bodyWaveMagnitude GREATER THAN 3 (Data Characteristic Rule)
- NuclearTestCausesEarthquake** (Relationship)
- Vegetation** (Ontology)

**Legend:**

- O** Ontology
- a** Attribute
- FD** Functional Dependency
- R** Relationship
- P** Condition
- BP** Binding Pattern
- DC** Data Characteristic Rule
- LC** Local Completeness
- S** Resource
- Rule** Domain Rule

**Buttons:** Back

# IScape Builder

The screenshot shows the IScape Builder application window. At the top, there is a title bar with the text "IScape Builder" and standard window control buttons. Below the title bar is a navigation bar with tabs for "Step1", "Step2", "Step3", "Step4", "Step5", and "Run". The main area of the window is titled "Specify all the ontologies and relationships to be used in IScape".

At the top of the main area, there are five buttons: "New IScape", "Open IScape", "Save IScape", "Clear IScape", and "Delete IScape".

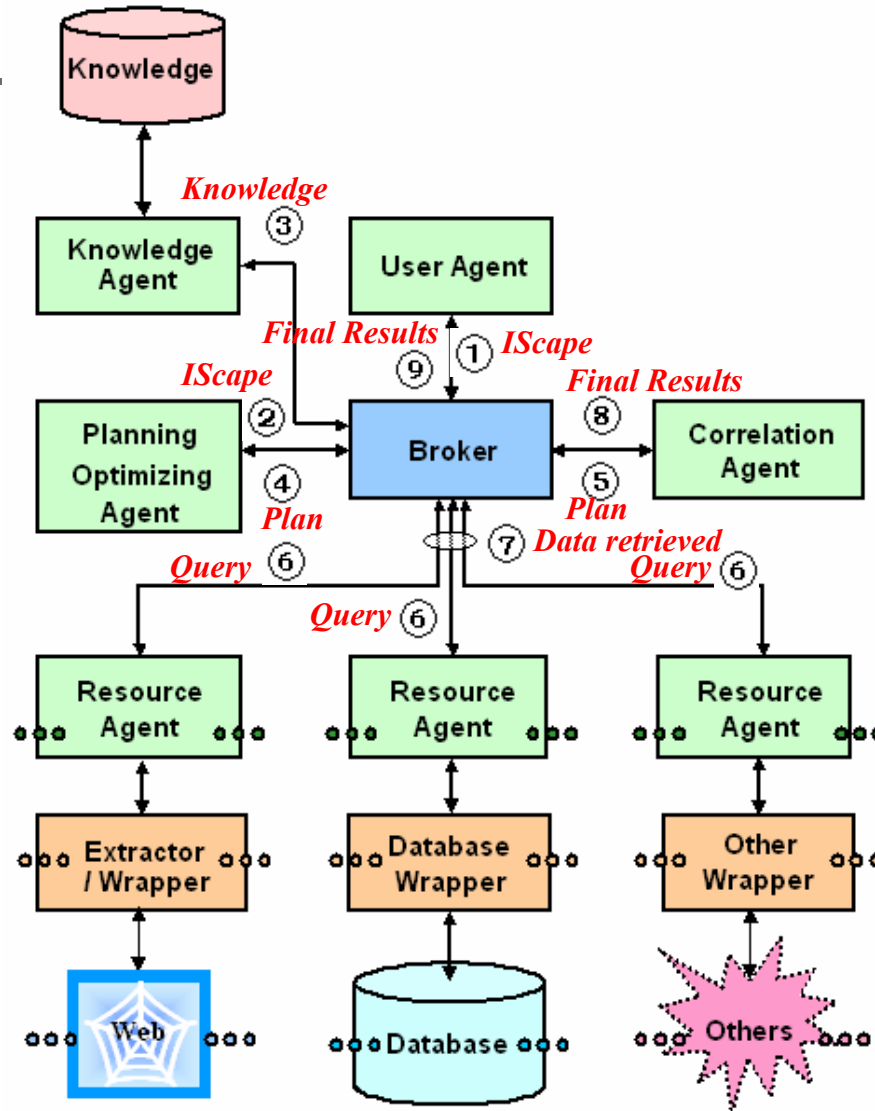
Below the buttons, there are two text input fields:

- Statement:** "Study the effect of nuclear tests on earthquakes"
- Description:** "List earthquakes caused just after a nuclear test and in a nearby location"

The interface is divided into four sections for selecting ontologies and relationships:

- Ontologies:** A list box containing "Tsunami", "Earthquake", "NuclearTest", "Vegetation", and "DailyWeather". A "Select" button is positioned to the right of this list.
- Ontologies in IScape:** A list box containing "Earthquake" and "NuclearTest". A "Remove" button is positioned to the right of this list.
- Relationships:** A list box containing "NuclearTestCausesEarth", "EarthquakeCausesTsun", and "EarthquakeAffectsEnviron". A "Select" button is positioned to the right of this list.
- Relationships in IScape:** A list box containing "NuclearTestCausesEarth". A "Remove" button is positioned to the right of this list.

# IScape Execution





# 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

# IScape 1

*Resource*

```
NuclearTestsDB(  
  testSite, explosiveYield, waveMagnitude,  
  testType, eventDate, conductedBy,  
  [dc] waveMagnitude > 3 );
```

*Resource*

```
SignificantEarthquakesDB(  
  eventDate, description, region,  
  magnitude, latitude, longitude,  
  numberOfDeaths, damagePhoto,  
  [dc] eventDate > "January 1, 1970" );
```

USGS site

<http://sun00781.dn.net/nuke/hew/Library/Catalog>

*Resource*

```
NuclearTestSites(  
  testSite, latitude, longitude );
```

*Ontology*

*Ontology*

```
NuclearTest(  
  testSite, explosiveYield, waveMagnitude, testType,  
  eventDate, conductedBy, latitude, longitude,  
  waveMagnitude > 0, waveMagnitude < 10,  
  testSite -> latitude longitude );
```

```
Earthquake(  
  eventDate, description, region,  
  magnitude, latitude, longitude,  
  numberOfDeaths, damagePhoto,  
  magnitude > 0 );
```

*Relationship*

**NuclearTest Causes Earthquake**

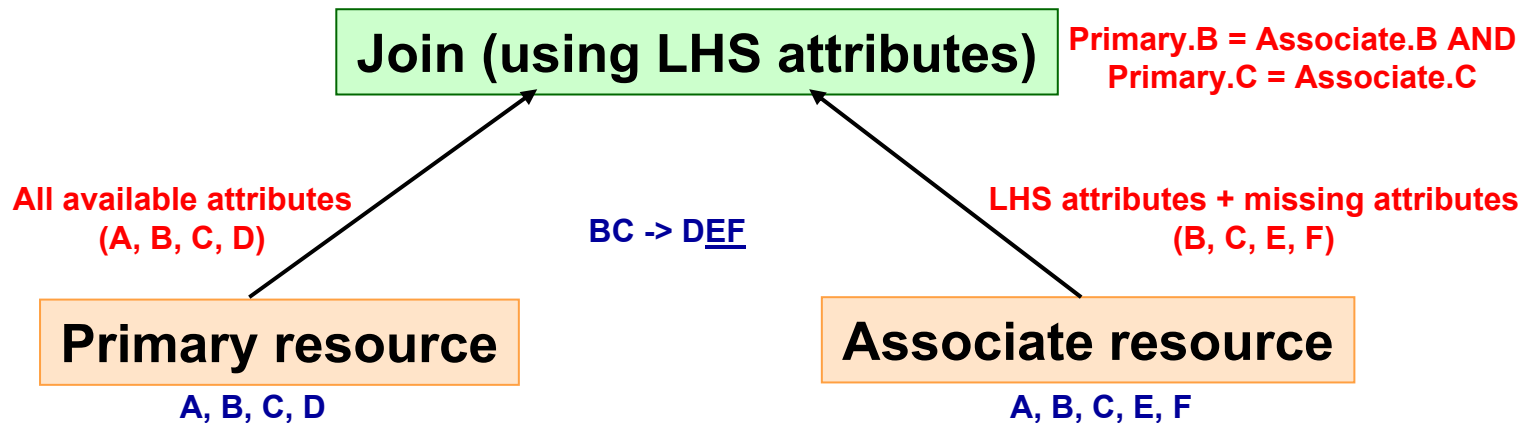
```
<= dateDifference( NuclearTest.eventDate, Earthquake.eventDate ) < 30  
  AND distance( NuclearTest.latitude, NuclearTest.longitude,  
               Earthquake.latitude, Earthquake.longitude ) < 10000
```

*IScape*

***“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)
  - $\langle \text{attribute} \rangle^+ \rightarrow \langle \text{missing attributes} \rangle \langle \text{attribute} \rangle^*$
- Couple with *associate resource*



# IScape Processing Monitor (IScape 1)

IScape Processing Monitor

IScape 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 du

Processing Log :

Message Id	Time Stamp	Message From	Brief Message
0	20:29:14.484	User Agent	Started processing
1	20:29:14.671	Broker Agent	Started processing
2	20:29:15.046	Planning Agent	Applied Domain Rules to IScape's constraint
3	20:29:15.109	Planning Agent	Checking IScape's constraint and relationshi...
4	20:29:15.187	Planning Agent	Selecting sources for Earthquake
5	20:29:15.234	Planning Agent	Selecting sources for Nuclear Test
6	20:29:15.453	Planning Agent	Plan created by the planner. Returning it to Br...
7	20:29:15.578	Broker Agent	Received plan from planner
8	20:29:15.843	CorrelationAgent	Executing IScape
9	20:29:16.390	TestSitesDB Resource Agent	Queried TestSitesDB
10	20:29:16.359	SignificantEarthquakesDB Resource Agent	Queried SignificantEarthquakesDB
11	20:29:16.500	Correlation Agent	Computed Union
12	20:29:16.718	NuclearTestsDB Resource Agent	Queried NuclearTestsDB
13	20:29:16.843	CorrelationAgent	SelectNodeProcessor done processing
14	20:29:16.984	Correlation Agent	Evaluated Join
15	20:29:17.046	Correlation Agent	Computed Union
16	20:29:17.250	Correlation Agent	Evaluated functions on a relation.
17	20:29:17.359	CorrelationAgent	SelectNodeProcessor done processing
18	20:29:17.750	Correlation Agent	Evaluated Relationship on a set of relations
19	20:29:17.765	Correlation Agent	Evaluated projection on a relation
20	20:29:17.796	Broker Agent	Received IScape results from Correlation Ag...
21	20:29:17.937	User Agent	Returning final result to client

Detailed Log Message Show Associated Data



## 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>*

Modeling Features	
User-Defined Functions	Users can provide their own implementation of function based on their own requirements
Complex Predicates/ Relationships	<ul style="list-style-type: none"> <li>- Relationships can be defined as complex expressions based on user-defined and/or existing functions, relational and logical operators</li> <li>- This means that properties of relationships beyond <i>sub-typing</i> and <i>part-of</i> can be modeled explicitly e.g. <i>Causes</i></li> </ul>
Iscape	<ul style="list-style-type: none"> <li>- Ability to define complex information requests utilizing existing domain models, user-defined functions and complex relationships.</li> <li>- Ability to re-use and re-configure such requests</li> </ul>
Mapping Rules	<ul style="list-style-type: none"> <li>- Enables users and domain experts to model mappings between concepts in different ontologies</li> <li>- This allows for data integration analogous to multi-database joins but more complex</li> </ul>
Resource Modeling	<ul style="list-style-type: none"> <li>- Captures description of <i>actual</i> content as opposed to <i>potential</i> content which helps in better resource selection</li> </ul>

Execution Features	
Iscape Consistency Checks	- Checks if Iscape is meaningful with respect to domain models chosen. Support for dynamic information resources
Resource Selection	- More specific because of dc and lc rules - More complete because of FDs.
Binding Patterns	- Support for dynamic information resources