Relationship Web: Spinning the Semantic Web from Trailblazing to Complex Hypothesis Evaluation

Amit P. Sheth
Wright State University - Main Campus, amit.sheth@wright.edu

Follow this and additional works at: http://corescholar.libraries.wright.edu/knoesis
Part of the Bioinformatics Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, OS and Networks Commons, and the Science and Technology Studies Commons

Repository Citation

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu.
Relationship Web:
Spinning the Semantic Web from Trailblazing to Complex Hypothesis Evaluation

August 2007

Amit Sheth
Kno.e.sis Center, Wright State University, Dayton, OH

This talk also represents work of several members of Kno.e.sis team, esp. Cartic Ramakrishnan. http://knoesis.wright.edu
Not data (search), but integration, analysis and insight, leading to decisions and discovery.
“An object by itself is intensely uninteresting”.

Grady Booch, Object Oriented Design with Applications, 1991

Changing the paradigm from document centric to relationship centric view of information
Data captured per year = 1 exabyte \((10^{18})\)  

(Eric Neumann, Science, 2005)

Multiple formats: Structured, unstructured, semi-structured

Multimodal: text, image, a/v, sensor, scientific/engineering

Thematic, Spatial, Temporal

Enterprise to Globally Distributed
Is There A Silver Bullet?

Moving from Syntax/Structure to Semantics
Semantics: Meaning & Use of Data

Semantic Web: Labeling data on the Web so both humans and machines can use them more effectively

i.e., Formal, machine processable description ⇒ more automation;
emerging standards/technologies

(RDF, OWL, Rules, ...)

Knowledge Enabled Information and Services Science
How?

**Ontology**: Agreement with Common Vocabulary & Domain Knowledge

**Semantic Annotation**: metadata (manual & automatic metadata extraction)

**Reasoning**: semantics enabled search, integration, analysis, mining, discovery
Biomedicine is one of the most popular domains in which lots of ontologies have been developed and are in use. See: 
http://obo.sourceforge.net/browse.html

Clinical/medical domain is also a popular domain for ontology development and applications: 
http://www.openclinical.org/ontologies.html
is a focused ontology for the description of glycomics
models the biosynthesis, metabolism, and biological relevance of complex glycans
models complex carbohydrates as sets of simpler structures that are connected with rich relationships
An ontology for structure and function of Glycopeptides
Published through the National Center for Biomedical Ontology (NCBO)

More at: http://knoesis.wright.edu/research/bioinformatics/
ProPreO ontology

An ontology for capturing process and lifecycle information related to proteomic experiments

Two aspects of glycoproteomics:

*What is it?* → identification

*How much of it is there?* → quantification

Heterogeneity in *data generation process*, *instrumental parameters*, and *formats*

Need *data* and *process provenance* → ontology-mediated provenance

Hence, ProPreO models both the glycoproteomics experimental process and attendant data

Approx 500 classes, 3million+ instances

Published through the National Center for Biomedical Ontology (NCBO) and Open Biomedical Ontologies (OBO)

More info. [On Knowledge Representation in Life Sciences at Kno.e.sis](http://www.knoesis.wrightstate.edu)
GNT-I attaches GlcNAc at position 2:

$$\text{UDP-N-acetyl-D-glucosamine + } \alpha-D\text{-Man}1,3(\beta1-R1)\beta-D\text{-Man}1,3(\beta1-R2) \rightarrow \text{UDP + N-Acetyl-\beta-D-glucosaminyl-1,2-\alpha-D\text{-Man}1,3(\beta1-R1)\beta-D\text{-Man}1,3(\beta1-R2)}$$

GNT-V attaches GlcNAc at position 6:

$$\text{UDP-N-acetyl-D-glucosamine + } \text{G00020} \rightarrow \text{UDP + G00021}$$
Pathway visualization tool by M. Eavenson and M. Janik, LSDIS Lab, Univ. of Georgia
Abundance of this glycan in three experiments
Semagix Freedom for building ontology-driven information system
Information Extraction for Metadata Creation

WWW, Enterprise Repositories

Create/extract as much (semantics) metadata automatically as possible

EXTRACTORS

METADATA
Blue-chip bonanza continues

Dow above 9,000 as HP, Home Depot lead advance; Microsoft upgrade helps techs.

August 22, 2002; 11:44 AM EDT

By Alexandra Twin, CNN/Money Staff Writer

New York (CNN/Money) - An upgrade of software leader Microsoft and strength in blue chips including Hewlett-Packard and Home Depot were among the factors pushing stocks higher at midday Thursday, with the Dow Jones industrial average spending time above the 9,000 level.

Around 11:40 a.m. ET, the Dow Jones industrial average gained 65.06 to 9,022.09, continuing a more than 1,300-point resurgence since July 23. The Nasdaq composite gained 9.12 to 1,418.37.

The Standard & Poor's 500 index rose 9.61 to 958.97.

Hewlett-Packard ( HPQ: up $0.33 to $15.03, Research, Estimates) said a report shows its share of the printer market grew in the second quarter, although another report showed that its share of the computer server market declined in Europe, the Middle East and Africa.

Home Depot ( HD: up $1.07 to $33.75, Research, Estimates) was up for the third straight day after topping fiscal second-quarter earnings estimates on Tuesday.

Tech stocks managed a turnaround. Software continued to rise after Salomon Smith Barney upgraded No. 1 software maker Microsoft ( MSFT: up $0.55 to $52.83, Research, Estimates) to "outperform" from "neutral" and raised its price target to $59 from $56. Business software makers Oracle ( ORCL: up $0.18 to $10.94, Research, Estimates), PeopleSoft ( PSFT: up $1.17 to $20.67, Research, Estimates) and BEA Systems ( BEAS: up $0.28 to $7.12, Research, Estimates) all rose in tandem.
Today, the Food and Drug Administration (FDA) is announcing an important change to voluntarily withdraw Bextra from the market. Pfizer has agreed to further discussions with the agency.

Reports of serious and potentially life-threatening skin reactions, including deaths, in patients using Bextra. The risk of these reactions in individual patients is unpredictable, occurring in patients with and without a prior history of sulfa allergy.

Date created: April 7, 2005
N-Glycosylation Process (NGP)

**Cell Culture**
- Extract

**Glycoprotein Fraction**
- Proteolysis

**Glycopeptides Fraction**
- Separation technique I

**Glycopeptides Fraction**
- PNGase

**Peptide Fraction**
- Separation technique II

**Peptide Fraction**
- Mass spectrometry

**N-dimensional array**
- Signal integration
  - Data correlation

**Glycopeptide identification and quantification**
- Peptide list
  - Peptide identification
  - Peptide fraction

**Signal reduction**
- Data reduction
  - Binning

**ms data**
- ms peaklist

**ms/ms data**
- ms/ms peaklist

**Data reduction**
- Data correlation
ISiS – Integrated Semantic Information and Knowledge System

Semantic Web Process to incorporate provenance

- Biological Sample Analysis by MS/MS
- Raw Data to Standard Format
- Data Pre-process
- DB Search (Mascot/Sequest)
- Results Post-process (ProValt)

Semantic Annotation Applications

- Raw Data
- Standard Format Data
- Filtered Data
- Search Results
- Final Output

Storage

Biological Information
### Mass Spectrometry (MS) Data

<table>
<thead>
<tr>
<th>Parent Ion m/z</th>
<th>Fragment Ion m/z</th>
<th>Fragment Ion Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>830.9570</td>
<td>580.2985</td>
<td>0.3592</td>
</tr>
<tr>
<td></td>
<td>688.3214</td>
<td>0.2526</td>
</tr>
<tr>
<td></td>
<td>779.4759</td>
<td>38.4939</td>
</tr>
<tr>
<td></td>
<td>784.3607</td>
<td>21.7736</td>
</tr>
<tr>
<td></td>
<td>1543.7476</td>
<td>1.3822</td>
</tr>
<tr>
<td></td>
<td>1544.7595</td>
<td>2.9977</td>
</tr>
<tr>
<td></td>
<td>1562.8113</td>
<td>37.4790</td>
</tr>
<tr>
<td></td>
<td>1660.7776</td>
<td>476.5043</td>
</tr>
</tbody>
</table>

**Parent Ion Charge:** 2

**Parent Ion m/z:**

**Fragment Ion m/z:**

**Fragment Ion Abundance:**

**MS/MS Peaklist Data:**

**Semantic Extraction/Annotation of Experimental Data**

**ProPreO: Ontology-mediated provenance**

**Mass Spectrometry (MS)**

**Knowledge Enabled Information and Services Science**
• **Evaluate the specific effects of changing a biological parameter:** Retrieve *abundance* data for a given *protein* expressed by three different *cell types* of a specific *organism*.

• **Retrieve raw data supporting a structural assignment:** Find all the *raw ms data files* that contain the *spectrum* of a given *peptide sequence* having a specific *modification* and *charge state*.

• **Detect errors:** Find and compare all *peptide* lists identified in *Mascot output files* obtained using a similar *organism*, cell-type, *sample preparation protocol*, and mass spectrometry conditions.

A Web Service Must Be Invoked

ProPreO concepts highlighted in red
Example of Relevant Subgraph Discovery based on evidence
Anecdotal Example

UNDISCOVERED PUBLIC KNOWLEDGE
Discovering connections hidden in text
The Priory of Sion was supposedly led by a Grand Master or Neufchatel.

<table>
<thead>
<tr>
<th>No.</th>
<th>Member</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ugo de Blanchefort</td>
<td>1159-1159</td>
</tr>
<tr>
<td>2</td>
<td>Bernard de Tremblay</td>
<td>1185-1198</td>
</tr>
<tr>
<td>3</td>
<td>Guillaume de Chandelles</td>
<td>1159-1159</td>
</tr>
<tr>
<td>4</td>
<td>Évrard de Niv,Y</td>
<td>1164-1164</td>
</tr>
<tr>
<td>5</td>
<td>Audrey de Montfort</td>
<td>1155-1160</td>
</tr>
<tr>
<td>6</td>
<td>Bernard de Blanchefort</td>
<td>1166-1169</td>
</tr>
<tr>
<td>7</td>
<td>Philippe de Mail</td>
<td>1159-1170</td>
</tr>
<tr>
<td>8</td>
<td>Etienne de Saint-Amand</td>
<td>1170-1180</td>
</tr>
<tr>
<td>9</td>
<td>Amaud de Tencur</td>
<td>1181-1194</td>
</tr>
<tr>
<td>10</td>
<td>Girard de Bouteville</td>
<td>1184-1188</td>
</tr>
<tr>
<td>11</td>
<td>Jean de Gisors</td>
<td>1186-1226</td>
</tr>
<tr>
<td>12</td>
<td>Marie de Saint-Claire</td>
<td>1220-1228</td>
</tr>
<tr>
<td>13</td>
<td>Guillaume de Gisors</td>
<td>1306-1307</td>
</tr>
<tr>
<td>14</td>
<td>Édouard de Bag</td>
<td>1307-1320</td>
</tr>
<tr>
<td>15</td>
<td>Jean de Bar</td>
<td>1336-1361</td>
</tr>
<tr>
<td>16</td>
<td>Jean de Saint-Clair</td>
<td>1351-1396</td>
</tr>
<tr>
<td>17</td>
<td>Blanche d'Évreux</td>
<td>1360-1370</td>
</tr>
<tr>
<td>18</td>
<td>Nicole de Flavigny</td>
<td>1300-1346</td>
</tr>
<tr>
<td>19</td>
<td>René d'Anjou</td>
<td>1414-1460</td>
</tr>
<tr>
<td>20</td>
<td>Isadore de Bar</td>
<td>1460-1483</td>
</tr>
<tr>
<td>21</td>
<td>Charles Filastri AKA Portelet</td>
<td>1460-1510</td>
</tr>
<tr>
<td>22</td>
<td>Jeanne de Vinci</td>
<td>1510-1519</td>
</tr>
<tr>
<td>23</td>
<td>Charles III (Duke of Brion)</td>
<td>1519-1527</td>
</tr>
<tr>
<td>24</td>
<td>François de Guise</td>
<td>1557-1559</td>
</tr>
<tr>
<td>25</td>
<td>Michel de Noircy AKA Nostradamus</td>
<td>1558-1568</td>
</tr>
<tr>
<td>26</td>
<td>Gard de Languedoc &amp; Nicolas Flammel</td>
<td>1660-1676</td>
</tr>
<tr>
<td>27</td>
<td>Louis de Lorraine</td>
<td>1575-1595</td>
</tr>
<tr>
<td>28</td>
<td>Robert Fludd</td>
<td>1565-1627</td>
</tr>
<tr>
<td>29</td>
<td>Johann Valentin Andreae</td>
<td>1637-1664</td>
</tr>
<tr>
<td>30</td>
<td>Robert Boyle</td>
<td>1664-1721</td>
</tr>
<tr>
<td>31</td>
<td>Isaac Newton</td>
<td>1671-1727</td>
</tr>
<tr>
<td>32</td>
<td>Charles Radclyffe</td>
<td>1727-1746</td>
</tr>
<tr>
<td>33</td>
<td>Charles de Larroque</td>
<td>1747-1749</td>
</tr>
<tr>
<td>34</td>
<td>Maximilian de Lorraine</td>
<td>1700-1701</td>
</tr>
<tr>
<td>35</td>
<td>Charles Nodier</td>
<td>1801-1845</td>
</tr>
<tr>
<td>36</td>
<td>Peter Hooke</td>
<td>1814-1836</td>
</tr>
<tr>
<td>37</td>
<td>Claude Octavian</td>
<td>1885-1915</td>
</tr>
<tr>
<td>38</td>
<td>Jean Cocteau</td>
<td>1919-1963</td>
</tr>
</tbody>
</table>

A second List of the Grand Masters of the Priory of Sion that included the names of Roger Patrice Pèlerin and Thomas Plantard appeared in 1999, but it should not be confused with the...
Schema-Driven Extraction of Relationships from Biomedical Text

An excessive endogenous or exogenous stimulation by estrogen induces adenomatous hyperplasia of the endometrium.

SS-Tagger (University of Tokyo)
SS-Parser (University of Tokyo)

- Entities (MeSH terms) in sentences occur in modified forms
  - "adenomatous" modifies "hyperplasia"
  - An excessive endogenous or exogenous stimulation modifies estrogen
- Entities can also occur as composites of 2 or more other entities
  - "adenomatous hyperplasia" and "endometrium" occur as "adenomatous hyperplasia of the endometrium"
An excessive endogenous or exogenous stimulation by estrogen induces adenomatous hyperplasia of the endometrium.
An excessive endogenous or exogenous stimulation induces modified_entity1, which hasModifier adenomatous hyperplasia. This modified_entity1 hasPart composite_entity1, which hasPart endometrium.

- **Modifiers**
- **Modified entities**
- **Composite Entities**
Now possible – Extracting relationships between MeSH terms from PubMed
Once you have Semantic Web Data

- **migraine** (D008881)
  - caused_by
  - hasPart
- **me_3142**
  - by_a_primary_abnormality_of_platelet_behavior
- **platelet** (D001792)
  - hasPart
- **me_2286**
  - _13%_ and _17%_ adp_and_collagen_induced_platelet_aggregation
  - stimulated
- **collagen** (D003094)
  - hasPart
- **magnesium** (D008274)
  - hasPart
  - stimulated
**QUESTION 1:** Given an RDF graph without weights can we use domain knowledge to compute the strength of connection between any two entities?

**QUESTION 2:** Can we then compute the most “relevant” connections for a given pair of entities?

**QUESTION 3:** How many such connections can there be? Will this lead to a combinatorial explosion? Can the notion of relevance help?
Overview

• Problem: Discovering relevant connections between entities
  – All Paths problem is NP-Complete
  – Most informative paths are not necessarily the shortest paths

• Possible Solution: Heuristics-based Approach*
  – Find a smart, systematic way to weight the edges of the RDF graph so that the most important paths will have highest weight
  – Adopt algorithms for weighted graphs
    • Model graph as an electrical circuit† with weight representing conductance and find paths with highest current flow – i.e. top-k

---

* Cartic Ramakrishnan, William Milnor, Matthew Perry, Amit Sheth. "Discovering Informative Connection Subgraphs in Multi-relational Graphs", SIGKDD Explorations Special Issue on Link Mining, Volume 7, Issue 2, December 2005

† Christos Faloutsos, Kevin S. McCurley, Andrew Tomkins: Fast discovery of connection subgraphs. KDD 2004: 118-127
What is a good path with respect to *knowledge discovery*?

- Uses more specific classes and relationships
  - e.g. Employee vs. Assistant Professor

- Uses rarer facts
  - Analogous to information gain

- Involves unexpected connections
  - e.g. connects entities from different domains
Class and Property Specificity (CS, PS)

- More specific classes and properties convey more information

- Specificity of property $p_i$:
  - $d(p_i)$ is the depth of $p_i$
  - $d(p_{iH})$ is the depth of the property hierarchy

\[
\mu(p_i) = \frac{d(p_i)}{d(p_{iH})}
\]

- Specificity of class $c_j$:
  - $d(c_j)$ is the depth of $c_j$
  - $d(c_{jH'})$ is the depth of the class hierarchy

\[
\mu(c_j) = \frac{d(c_j)}{d(c_{jH'})}
\]

- Node is weighted and this weight is propagated to edges incident to the node
Instance Participation Selectivity (ISP)

- Rare facts are more informative than frequent facts
- Define a *type* of an statement RDF \(<s,p,o>\)
  - Triple \(\pi = <C_i,p_j,C_k>\)
    - \(\text{typeOf}(s) = C_i\)
    - \(\text{typeOf}(o) = C_k\)
- \(|\pi| = \text{number of statements of type } \pi \text{ in an RDF instance base}\)
- *ISP* for a statement: \(\sigma_{\pi} = 1/|\pi|\)
• \( \pi = \langle \text{Person}, \text{lives_in}, \text{City} \rangle \)

• \( \pi' = \langle \text{Person}, \text{council_member_of}, \text{City} \rangle \)

• \( \sigma_{\pi} = \frac{1}{(k-m)} \) and \( \sigma_{\pi'} = \frac{1}{m} \), and if \( k-m > m \) then \( \sigma_{\pi'} > \sigma_{\pi} \)
Span Heuristic (SPAN)

- RDF allows Multiple classification of entities
  - Possibly classified in different schemas
  - Tie different schemas together
- Refraction is Indicative of anomalous paths
- SPAN favors *refracting* paths
  - Give extra weight to multi-classified nodes and propagate it to the incident edges
Going Further

• What if we are not just interested in knowledge discovery style searches?

• Can we provide a mechanism to adjust relevance measures with respect to users’ needs?
  – Conventional Search vs. Discovery Search

Yes! ... SemRank*

High Information Gain
High Refraction Count
High S-Match

Low Information Gain
Low Refraction Count
High S-Match

Adjustable search mode
Blazing Semantic Trails in Biomedical Literature

“The physician, puzzled by her patient's reactions, strikes the trail established in studying an earlier similar case, and runs rapidly through analogous case histories, with side references to the classics for the pertinent anatomy and histology. The chemist, struggling with the synthesis of an organic compound, has all the chemical literature before him in his laboratory, with trails following the analogies of compounds, and side trails to their physical and chemical behavior.” [V. Bush, As We May Think. The Atlantic Monthly, 1945. 176(1): p. 101-108.]
PMID-15886201

Identification of the substrates for BRCA1-dependent ubiquitination activity is important for understanding how mutation of BRCA1 is associated with loss of tumor suppression activity. The currently identified substrates include histone proteins, p53, Pisciotti amino protein D2, and centrosomal proteins including NPM1 and (gamma)-tubulin (24, 51–53). Among these, only the modification of (gamma)-tubulin by BRCA1/BARD1 has been shown to affect the biology of breast cells. It has been shown that failure to ubiquitinate (gamma)-tubulin results in centrosome amplification (24). The BRCA1/BARD1 proteins are known to regulate multiple processes in the cell, including transcription, DNA repair, and centrosome dynamics (5, 25–29). Although the ubiquitination of (gamma)-tubulin may in part explain the BRCA1-dependent regulation of centrosome dynamics, it was unclear whether the BRCA1-dependent ubiquitination activity also regulates the transcription and DNA repair function of BRCA1.

We had proposed that the BRCA1-dependent ubiquitination activity may function in DNA repair by modification of RNAPII that transcribes DNA near a lesion (14, 15). This proposed role for BRCA1 in transcription-coupled repair could be important following UV damage or double strand breaks. One prediction of this model was that BRCA1/BARD1 ubiquitination activity would be targeted to the elongating, hyperphosphorylated form of RNAPII. Actively transcribing RNAPII is phosphorylated on Ser5 proximal to the promoter and on Ser2 further downstream (23). Thus, the principal form of RNAPII that elongates through a gene is the Ser2p form, which we now show is a substrate for BRCA1/BARD1. The model that BRCA1-dependent ubiquitination directly links transcription elongation to repair is thus not supported. Instead, we found that Ser5 phosphorylation of RNAPII is a generalized response to UV irradiation, and BRCA1-dependent ubiquitination modifies the RNAPII. It has been observed that transcriptionally engaged RNAPII does become phosphorylated on Ser5 by the action of extracellular signal-regulated kinases 1 and 2 (69). The data are most consistent with a model whereby DNA damage causes phosphorylation of a subpopulation of RNAPII, followed by ubiquitination by BRCA1/BARD1 and subsequent degradation at the proteasome.

In these experiments we found that overexpression of BRCA1 in cells could stimulate the damage-induced ubiquitination of RNAPII. When we inhibited BRCA1 expression by transfection of short interfering RNA specific for BRCA1, we did not observe a decrease in ubiquitination of RNAPII. We interpret these results to indicate that one or more other ubiquitin ligases can execute this function. Several other factors have been implicated in the ubiquitination of RNAPII, including Cdc20.

PMID-10037099

In the mouse two-stage skin carcinogenesis model, tumor promotion is a distinct, rate-limiting step that determines the formation of premalignant tumors. As discussed above, the role of tumor promoters in human cancer is more complex because human exposure tends to involve sporadic low doses of complex mixtures of carcinogens, co-carcinogens, and tumor-promoting agents. Nonetheless, studies of rodent tumor models of liver, bladder, colon, and breast cancers—and analyses of human tumor formation—suggest that processes analogous to tumor promotion by TPA on the mouse skin are a common feature of carcinogenesis (1). Thus, epigenetic changes in cell signaling, such as altered growth factor production and receptor expression, and elevated synthesis of inflammatory and mitogenic factors, such as cytokines and eicosanoids, are key targets for inhibiting tumor promotion.

Tumor Progression

As noted earlier, tumor progression involves the accumulation of additional genetic alterations in an initiated cell clone and generally gives a growth advantage to the progressing clone. In progression, a focal lesion consisting of a population of initiated and promoted cells ultimately becomes an invasive malignant tumor. One frequently observed genetic alteration that appears to contribute to malignant progression is mutation in the p53 [also known as TP53] tumor suppressor gene (62). The p53 gene product is a transcription factor that regulates the expression of a number of DNA-damage and cell cycle-regulatory genes and genes regulating apoptosis. By enhancing transcription of these critical genes, p53 regulates the cellular response to DNA damage (63). p53 also plays a role in maintaining genomic stability (64). Genomic instability, a hallmark of spontaneous malignant progression, is characterized by sequential chromosomal aberrations, such as duplications, deletions, and loss of
The p53 gene product is a transcription factor that regulates the expression of a number of DNA-damage and cell cycle-regulatory genes and genes regulating apoptosis. The data are most consistent with a model whereby DNA damage causes phosphorylation of a subpopulation of RNAII, followed by ubiquitination by BRCA1/BARD1 and subsequent degradation at the proteasome.
p53 gene

isa

transcription factor

regulates → DNA-damage → causes → phosphorylation

(b)
Semantic Trails can be built over a Web of Semantic (Meta)Data extracted (manually, semi-automatically and automatically) and gleaned from

- **Structured data** (e.g., NCBI databases)
- **Semi-structured data** (e.g., XML based and semantic metadata standards for domain specific data representations and exchanges)
- **Unstructured data** (e.g., Pubmed and other biomedical literature)

and

- **Various modalities** (experimental data, medical images, etc.)
Semantic Metadata can be extracted from unstructured (eg, biomedical literature), semi-structured (eg, some of the Web content), structured (eg, databases) data and data of various modalities (eg, sensor data, biomedical experimental data). Focusing on the relationships and the web of their interconnections over entities and facts (knowledge) implicit in data leads to a Relationship Web.

Relationship Web takes you away from “which document” could have information I need, to “what’s in the resources” that gives me the insight and knowledge I need for decision making.

Demonstration of Semantic Trailblazing using a Semantic Browser

This application demonstrating use of ontology-supported relationship extraction (represented in RDF) and their traversal in context (as deemed relevant by the scientists), linking parts of knowledge represented in one biomedical document (currently a sentence in an abstract in Pubmed) to parts of knowledge represented in another document.

This is a prototype and lot more work remains to be done to build a robust system that can support Semantic Trailblazing. For more information:


Applications

“Everything's connected, all along the line. Cause and effect. That's the beauty of it. Our job is to trace the connections and reveal them.”

Jack in Terry Gilliam’s 1985 film - “Brazil”
Ahmed Yaseer:
- Appears on Watchlist ‘FBI’
- Works for Company ‘WorldCom’
- Member of organization ‘Hamas’
Example of Fraud prevention application used in financial services

Scores the entity based on the content and entity relationships
Creating “logical web” through Media Independent Metadata based Correlation

Model for Logical Correlation using Ontological Terms and Metadata

Framework for Representing MREFs

Serialization (one implementation choice)

Abstract: The rates of growth of 29 hepatic metastases from 15 patients with primary colorectal carcinoma were studied using serial computed tomography (CT). Eleven metastases were found by the surgeon at laparotomy (overt metastases); the remaining 18 were not evident to the surgeon at laparotomy, but were detected by CT scan during the immediate postoperative period postoperative (occult metastases). An estimate of tumour volume doubling time was obtained from a semi-logarithmic plot of tumour cell number against time. The mean doubling time for the overt metastases compared with 86 days for the occult metastases. The time of surgery was estimated to have growth curve assuming that the corresponding age of the overt metastases is 2.3 +/- 0.4 years.
Hypothesis driven retrieval of Scientific Text
More about the Relationship Web

Relationship Web takes you away from “which document” could have information I need, to “what’s in the resources” that gives me the insight and knowledge I need for decision making.

World class research center - coupled with dataOhio for tech transfer and commercialization

Core expertise in

- data management: integration, mining, analytics, visualization
- distributed computing: services/grid computing
- Semantic Web
- Bioinformatics, etc.

With domain/application expertise in Government, Industry, Biomedicine

Member of World Wide Web Consortium and extensive industry relationships
Core expertise in

- data management: integration, mining, analytics, visualization
- distributed computing: services/grid computing
- Semantic Web
- Bioinformatics, etc.

With domain/application expertise in Government, Industry, Biomedicine

W3C member, extensive industry relationships

http://knoesis.wright.edu