WebScripter

Tools for Building a Semantic Web from the Ground Up

May 26, 2004

Dr. Robert MacGregor
USC/Information Sciences Institute

macgregor@isi.edu
End-user Knowledge Engineering

- Semantic experts are scarce, knowledge acquisition is expensive and inexact, semantic expert unavailable after initial session(s).
- End-users are the real experts – need to insure that their expertise is fully-exploited

→ Need tools that help end-users to evolve ontologies while working with data
Grassroots Approach

- New breed of tools that can function without centralized ontologies and alignments
- Carrots instead of sticks – reward users for:
  - creating vocabulary
  - adopting each others’ vocabularies
  - creating alignment catalogs
  - adopting each others’ alignments
  - annotations
Key End-user Activities

- **Authoring**
  - eNotes to self, others, annotations

- **Markup (authoring + advertisement)**
  - of maps, text documents, Web pages

- **Visualization**
  - of annotated maps, reports, etc.

- **Retrieving**
  - eDocuments, eNotes, communications, annotations

- **Organizing**

⇒ Ontologies should be a side-effect of each of these activities
CHIME: GeoSpatial Visualization

Powered by OWL
Grassroots Spatio-temporal Queries

- Choose Topics (or Entities, Event Threads, ...)
- Select Integrity Level
- "Focus Box" settings
- Identify Spatial Region
- Position Time Slider
- Identify Spatial Region
- “Focus Box” settings
- Select Integrity Level
- Choose Topics (or Entities, Event Threads, ...)
- Position Time Slider
HiKE Semantic Desktop

Triples all the way down

Notes taken yesterday
Report Generation

- How do I print out a section of a HiKE network?

- WebScripter can’t handle nested/hierarchic data

  ➔ inspired development of VIVID
VIVID: Visual Variable-Depth Information Displays

- Semantic Web data is
  - frequently non-uniform
  - often deeply nested
  - not easily viewed or browsed using conventional tools

- "Navigating a semantic web is like looking at data through a soda straw." – Joe Rockmore
How to Display SWeb data

Problems

- Hierarchic displays waste most of the screen real estate
- Conventional tabular displays space-efficient, but handle only “uniform” data
Sigmod Record Example of Low-Density Display
### Hierarchic Header

Data warehousing and OLAP for decision support

Foundations of Statistical Natural Language Processing - by C. Manning et al

Communication Efficient Distributed Mining of Association Rules

Performance Evaluation of a New Distributed Deadlock Detection Algorithm

A New Join Algorithm

Online Association Rule Mining

Constructing the Next 100 Database Management Systems

Left-deep vs. bushy trees: an analysis of strategy spaces and its implications for query optimization

Template-based wrappers in the TSIMMIS system

Lore: A Database Management System for Semistructured Data

The Impact of Database Research on Industrial Products (Panel Summary)

---

### Tabular Data

<table>
<thead>
<tr>
<th>Title</th>
<th>Article Code</th>
<th>作者</th>
<th>Volume</th>
<th>Issue</th>
<th>On-line Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data warehousing and OLAP for decision support</td>
<td>262686</td>
<td>Saroj Chaudhuri, Umeshwar Dayal</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Foundations of Statistical Natural Language Processing - by C. Manning et al</td>
<td>313194</td>
<td>Reviewer: G. Weikum</td>
<td>00</td>
<td>2</td>
<td>false</td>
</tr>
<tr>
<td>Communication Efficient Distributed Mining of Association Rules</td>
<td>302071</td>
<td>Assaf Schuster</td>
<td>00</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Performance Evaluation of a New Distributed Deadlock Detection Algorithm</td>
<td>233211</td>
<td>Chim-fu Yeung, Kam-yiu Lam, Sheung-lun Hung</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>A New Join Algorithm</td>
<td>234223</td>
<td>Arnold Charles Meltzer, Dong Keun Shin</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Online Association Rule Mining</td>
<td>282030</td>
<td>Christian Hidber</td>
<td>00</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Constructing the Next 100 Database Management Systems</td>
<td>231196</td>
<td>Andreas GopPERT, Klaus R. Dittrich</td>
<td>00</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Left-deep vs. bushy trees: an analysis of strategy spaces and its implications for query optimization</td>
<td>202253</td>
<td>Yannis E. Ioannidis, Younkyung Cha Kang</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Template-based wrappers in the TSIMMIS system</td>
<td>262694</td>
<td>Ramana Yemeni, Marcus Breunig, Svetlana Nestorov, Joachim Hammer, Vasilis Vassalos, Héctor García-Molina</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>Lore: A Database Management System for Semistructured Data</td>
<td>263363</td>
<td>Serge Abiteboul, Roy Goldman, Jennifer Widom, Jason McHugh, Dallan Quass</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
<tr>
<td>The Impact of Database Research on Industrial Products (Panel Summary)</td>
<td>233214</td>
<td>Daniel H. Fishman, Michael Stonebraker, David B. Lomet, Joseph A. Buhler</td>
<td>01</td>
<td>2</td>
<td>true</td>
</tr>
</tbody>
</table>
How to Manage Header Information (CIA Fact Book dataset)

Full header is ~18 feet wide
Header Formatting Tool

Header Formatter is ~2.5 feet tall
Pivoting

View dataset from ‘Author’ perspective

Pivot on ‘Title’
### View dataset from ‘Title’ perspective

<table>
<thead>
<tr>
<th>articleCode</th>
<th>title</th>
<th>authors</th>
<th>volume</th>
<th>issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>282139</td>
<td>The WASA2 Object-Oriented Workflow Management System</td>
<td>Mathias Weske Gottfried Vossen</td>
<td>02</td>
<td>28</td>
</tr>
<tr>
<td>192193</td>
<td>ACTA: a framework for specifying and reasoning about transaction structure and behavior</td>
<td>Panayiotis K. Chrysanthis, Krithi Ramamritham</td>
<td>01</td>
<td>19</td>
</tr>
<tr>
<td>294022</td>
<td>Cache Invalidation Scheme for Mobile Computing Systems with Real-time Data</td>
<td>E. Chan, K.-Y. Lam, H. Leung, J. Yuen</td>
<td>02</td>
<td>29</td>
</tr>
<tr>
<td>292061</td>
<td>Self-Organizing Data Sharing Communities with SAGRES</td>
<td>Ewa Jasliowska, Zachary G. Ives, Jing Su, Alon Y. Levy, Qiong Chen, Shiori Betzler, Wai Tak Theodora Yeung, Rachel Pottinger, Stefan Saroiu</td>
<td>08</td>
<td>29</td>
</tr>
</tbody>
</table>
Users easily generate complex queries

WHERE (?article, rdf:type, cf:Article) AND
(?article, cf:title_data, ?td) AND
(?td, cf:title, ?title) AND
(?article, cf:authors, ?authors) AND
(?authors, cf:author_data, ?ad) AND
(?ad, cf:author, ?author) AND
(?articles, cf:article, ?article) AND
(?issue, cf:articles, ?articles) AND
(?issue, cf:number, ?number) AND
(?issue, cf:volume, 22)
USING cf For <http://www.isi.edu/webscripter/CIAFactBook#>
Using Queries to Build Vocabulary

- Retrieve cargo ships with length between 70 and 150
- Name query “MidSizedCargoShip”
- Automatically positioned in class hierarchy

MidSizedCargoShip = Ship and length > 70 and length < 150

How can I do this in OWL or SWRL?
VIVID would like to be tightly-coupled to an app
- needs high bandwidth to browse knowledge base

Problem: No standard API for interfacing RDF/OWL tools

The plumbing to do this is not in place
Transition to SemWebCentral (cont.)

- For now, use Jena2 as bridge between app and VIVID
Semantic Engineering Workbench

For SemWebCentral, extract these components

- Web Documents
- N-Dimensional Visualization
- CHIME Console
- n-Dimensional Abstraction
- Protege Memory
- RDF Connection
- Jena Triple Store
- MySQL Relational DBMS
- RDF (Data + Metadata)
- RDF (Data + Metadata)
- XML Files
- Excel Spreadsheets
- RDF Files
- N3 Files
- OWL Files

WebScripter

Open Source & Web Database

VIVID Reports

Protege Memory

For SemWebCentral, extract these components

Semantic Engineering Workbench
Tasks for Remainder of ‘04

- Round-out first VIVID release
  - Printing/Report Generation
  - Scale-up
- First release to SemWebCentral
- Additional features
  - Updating
  - Bookmarking/User Profiles
- Second release to SemWebCentral
END OF TALK ONE
Ontoagents
(Stefan Decker, Siegfried Handschuh,
Steffen Staab, Rudi Studer)

TRIPLE: http://triple.semanticweb.org (ISI)

- **Goals/Status:**
  - Inference Engine API
  - Grid Reasoning Service (Grid Matchmaker)
  - Data integration in NSF Argos project (integrating transport data)
  - Many users and contributors: WU Vienna, University of Hannover, Fraunhofer Berlin, TU Berlin, ISI, DFKI, DERI,…

- **Work in Progress**
  - Re-Integrate work from outside contributors
  - "Decker Problem": Specifying Input/Output of Web Services – using this for validation
  - Implementing OWL semantics (AMAP)

- **SemWebCentral**
  - Currently on SourceForge: Date: After DAML Meeting, end of June 2004
Ontoagents

(Stefan Decker, Siegfried Handschuh, Steffen Staab, Rudi Studer)

Annotation/Crawling (AIFB, Karlsruhe):

- **Goals/Status:**
  - OWL compatible Annotation Tool (CREAM/OntoMat)
  - OWL-Crawler (standalone and OntoMat-plugin)
  - Annotation/Meta-Ontology
  - Extended Information-Extraction (The Self-annotating Web)

- **Work in Progress**
  - Annotation & Authoring Web Pages
  - Ontology Engineering & Annotation
  - Date: Oct 2004

- **SemWebCentral**
  - OWL-Crawler (project approved for release)
  - Annotation-Tool
  - Date: After DAML Meeting, June 2004
BACKUPS
For more information

- Dr Robert MacGregor
  - University of Southern California/Information Sciences Institute
  - 310-448-8423
  - macgregor@isi.edu
VIVID Challenges

- **Solved:**
  - Data compaction
  - How to manage very large headers
  - Pivoting
  - Queries

- **Current Focus:**
  - Data formatting
  - Displaying recursive structures
  - Printing/Report Generation

- **Next:**
  - Scaling to very large datasets
  - Updating
  - Bookmarks, User-Profiles
  - Inverse links
  - Annotation (of rows and cells)
  - ...

DAML PI Meeting, May 2004
HiKE: Semantic Desktop

- Simple, friendly Semantic Web authoring tool
  - Take notes, annotate, organize documents

- Semantic Web interrelates desktop entities
  - Drag desktop objects into hierarchy
  - Attach attribute values anywhere
  - Semantic annotations enable precise searches
Empirical Evolution

1) Select dataset
2) Enhance VIVID to handle dataset
3) Repeat

- CIA Factbook
  - Handle very wide fanout
- Sigmod Record
  - Pivoting
- SATURN dataset
  - Recursion
What We Need From Definition Language

- Something that supports rule-based (if-and-only-if) definitions
  (recall the ‘MidSizedShip’ example)

- Precedents:
  - Loom ‘satisfies’
  - KIF ‘kappa’