Outline

- Web development in China
- Semantic Web communities in China
- Semantic Web projects in China
  - IODT from IBM Research China
  - Falcon from Southeast University
  - APEX from Shanghai Jiaotong University
  - KEG from Tsinghua University
  - DartGrid from Zhejiang University
- Summary
Web Development In China

- As revealed by the 21st China Internet Development Survey released in Jan. 7, 2008, the registered Internet users in China has reached to 210,000,000, only less than US right now.
Semantic Web Communities (1/2)

- Two major google groups:
  - AI and Web: http://groups.google.com/group/ai-and-web-in-china
  - Central community of Semantic Web in China http://groups.google.com/group/ChinaSemanticWeb
- Annual China Semantic Web Symposium started last year.
- The first Asian semantic web conference was held in Beijing in 2006.
One major student-developer-oriented community: W3China.org
- It was created in 2003 by many students
- Its original intention was actually to provide an open forum for Semantic Web technology.
- As of April, 2008, the total number of registered user is nearly 180,000.
- The number of active members in semantic-web-related bbs/blog is over 3,000, mainly consisting of students and developers from over 100 institutions or universities.
Major achievement

- Increase in SW publications in major SW events
- One best paper award in the in-use track at ISWC2006 and one best student paper nominee at ISWC2007.
- First places in several ontology matching and mapping contests
Semantic Web Projects
Toolkit Development and Applications
The main focus

- Semantic Web based modeling
- Semantic Web data management including high-performance semantic data storage, reasoning, query.
- Meta-data management for enterprise applications.

Main toolkits developed:

- IODT: Integrated Ontology Development Toolkit
- SOR: Scalable Ontology Repository
Integrated Ontology Development Toolkit

- IODT is an ontology toolkit for storage, manipulation, query, and inference of ontologies and corresponding instances.
SOR: Scalable Ontology Repository

- One of the major component of IODT is an OWL ontology repository, named Scalable Ontology Repository (SOR),
- It is a high-performance OWL storage, inference, and query system based on RDBMS.
- It supports DLP (Description Logic Program), a subset of OWL DL, and the SPARQL language.
- Also it supports datalog rule inference in its new version which will be released at IBM alphaworks soon.

http://www.alphaworks.ibm.com/tech/semanticstk
IBM is using IODT in several products and projects, and a government project with different specific application purposes.

<table>
<thead>
<tr>
<th>Product name / project description</th>
<th>sector</th>
<th>purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Websphere Service Registry and Repository 6.0</td>
<td>industry technology</td>
<td>service classification</td>
</tr>
<tr>
<td>Service Oriented Asset Repository</td>
<td>industry solution</td>
<td>asset classification</td>
</tr>
<tr>
<td>A government project</td>
<td>public sector</td>
<td>service classification</td>
</tr>
</tbody>
</table>

Applications
Main focus:
- Ontology matching and ontology alignment.
- Semantic search engine.

Major tools developed:
- Falcon-AO: a practical ontology matching system
  http://iws.seu.edu.cn/projects/matching/
- Falcons: a semantic web search engine
  http://iws.seu.edu.cn/services/falcons/
Falcon-AO: A Practical Ontology Matching System

- Falcon-AO is an automatic ontology matching system for matching ontologies written in RDF(S) and OWL.
- Contains a set of novel matcher: V-Doc, GMO, PBM.
- Has won first place in major ontology matching contest, and recognized by the community as one of the best ontology matching systems
Falcons: A Semantic Web Search Engine

- **Key features:**
  - Recommending ontologies for concept search
  - Recommending types for object search
  - Entity summarization for browsing

- **Data Sets:**
  As of today, Falcons has discovered more than 7 million well-formed RDF documents containing 250 million RDF statements. 4,400 ontologies have been identified. 30 million semantic web entities have been indexed, in which 2 million are concepts.

- http://iws.seu.edu.cn/services/falcons/
APEX from Shanghai Jiaotong University

- Main focus:
  - Ontology exaction
  - Semantic search
  - Semantic web user interaction

- Semantic tools developed:
  EachWiki, Orient, PORE, HS Explorer, CE², Semplore, Q2S emantic, PANTO, ReSE, CE²-Pedia
people related to "World Wide Web" science fiction novel related to "Mars" institutions that Turing Award winners work in
Find specifications about "SVG" whose author's name is "Capin"
KEG from Tsinghua University

Main Focus

- Semantic search
- Semantic annotation
- Ontology matching

Semantic Tools developed:

- ArnetMiner: Academic Researcher Social Network
  - http://www.arnetminer.org/
- Meta Data Editor and Annotator
- RiMOM: Ontology matching tool
It is a semantic search engine aiming at providing comprehensive analysis and mining for academic community.

- **social information extraction**: academic social information is extracted and integrated to build a researcher network based on FOAF.

- **expertise ranking**: propose a probabilistic random walk model to rank person, conference, and paper simultaneously.

- **social association finding**: finding connection between people.

- **hot-topic mining**: find hot sub-topics and their trends in a given research field.
Semantic Annotation

- Develop three different approaches in automatic semantic annotation
  - Rule induction
  - Unbalanced classifier
  - Sequence labeling
DartGrid from Zhejiang University

◆ Main focus:
  ◆ Bridging the gap between relational database and semantic web
  ◆ Data mining over web of data.
  ◆ Focus on life science application domain, particularly for traditional Chinese medicine (TCM)
  ◆ Semantic tools developed:
    ◆ DartMapper: visualized SQL2RDF mapping tool.
    ◆ DartQuery: Form-based interactive SPARQL query builder.
    ◆ DartSearch: Semantic search and navigation based on ontology
  ◆ One paper from the group won the best paper award in ISWC2006's In-Use track.
Visualized SQL2RDF Mapper

- Visualized mapper for defining mappings from relational schema to RDF ontology.
- A new browser-host mapper provides online mapping manipulation supporting XMLP.
- Provides fast linked data publishing and deployment for both manually defining the mapping and direct D2R transformation.
Semantic Data Query and Analysis

- Ontology-based Query and Search
  - Query form for user to construct SPARQL queries.
  - Ontology-based automatically generating query form.
  - Ontology-enabled semantic navigation.
- Semantic Graph Mining:
  - Enable relational mining over web of data
Application: Health Care and Life Science
Make a connection between TCM and modern medicine
TCM Ontology Effort

- Collaborative online ontology engineering toolkit
- Has supported around 200 people from 17 institutes to develop the TCM ontology and language systems together.
- The ontology has reached around 20,000 class descriptions and 100,000 instances.
Summary

- A lot of energetic players on the ground.
- A bunch of fancy semantic toolkits at disposal.
- A variety of several real-life applications.

And
- A lot of promise when we look into the future...
Acknowledgment

◆ Presentation is a result of many participants' contribution:
  ◆ Guotong Xie, IBM Research China
  ◆ Li Ma, IBM Research China
  ◆ Wei Hu, Southeast University
  ◆ Haofen Wang, Shanghai Jiaotong University
  ◆ Jie Tang, Tsinghua University
  ◆ Yimin Wang, Karlsruhe University, Gemerny
  ◆ Collin Xu, W3China.org
◆ Particular thanks to Pro. Yu Yong, and Pro. Qu Yuzhong for their very kind supports.
Thanks for your attention!
Enjoy your stay in China!