Semantic Web Development in China

W3C Semantic Web

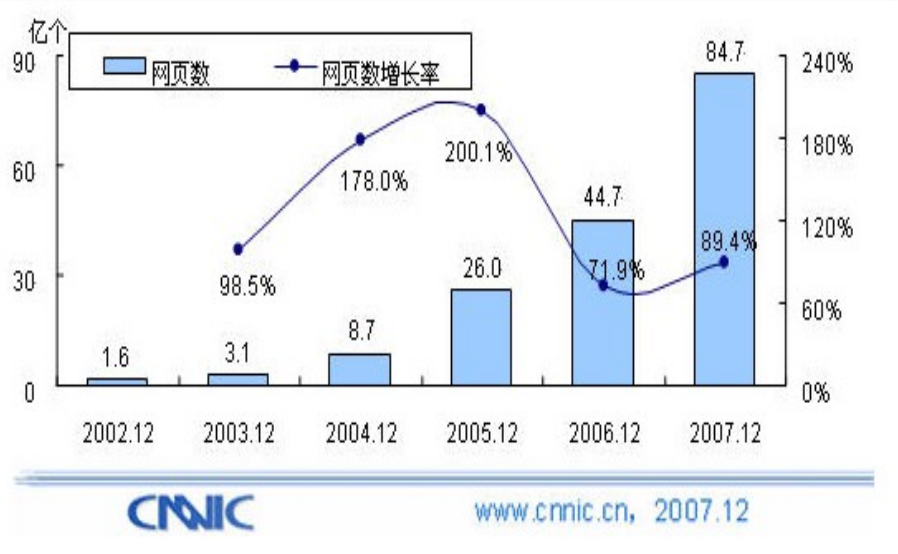
Beijing 2008
One World, One Web
17th International World Wide Web Conference

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)

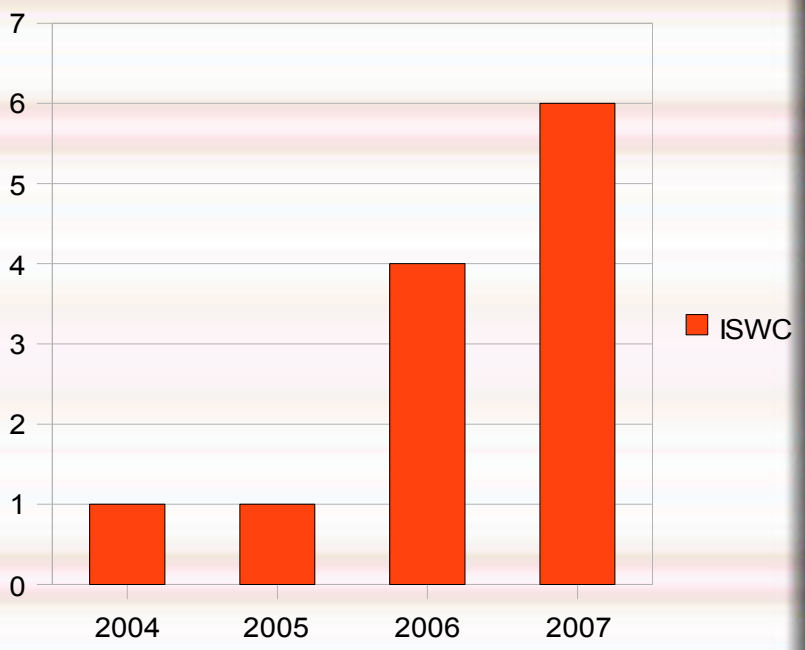
- Two major google groups:
  - AI and Web:
    - http://groups.google.com/group/ai-and-web-in-china
  - Central community of Semantic Web in China
    - http://groups.google.com/group/ChinaSemanticWeb
- Annual China Semantic Web Symposium started last year.
- The first Asian semantic web conference was hold in Beijing in 2006.

Semantic Web Communities (2/2)

- 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



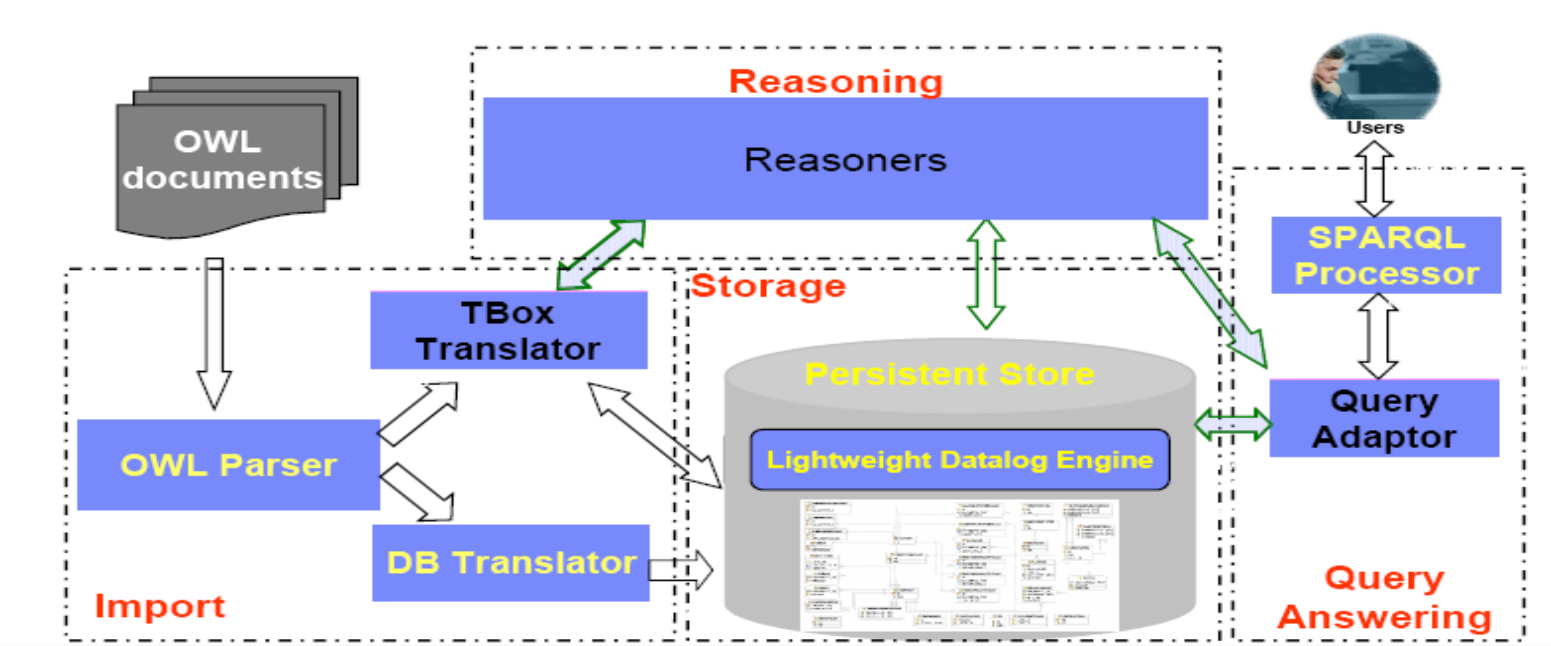
Semantic Web Projects
Toolkit Development and Applications

IODT from IBM Research China

- **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

Applications

- IBM is using IODT in several products and projects, and a government project with different specific application purposes.

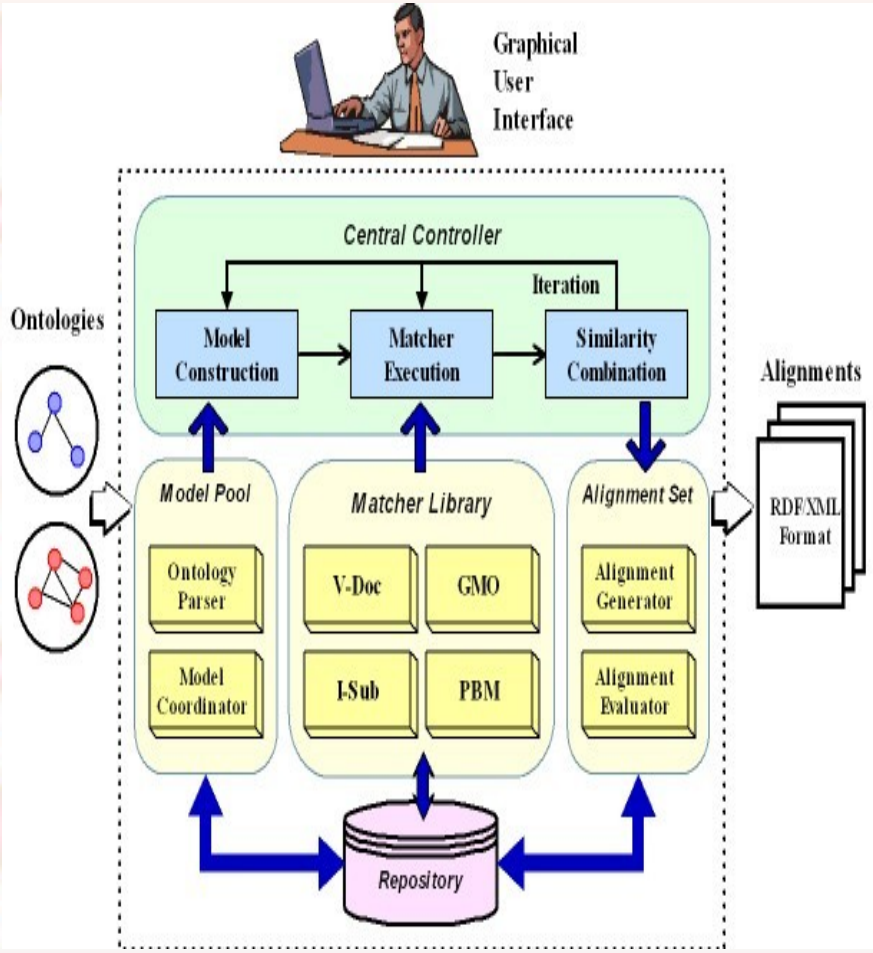
Product name / project description	sector	purpose
Websphere Service Registry and Repository 6.0	industry technology	service classification
Service Oriented Asset Repository	industry solution	asset classification
A government project	public sector	service classification

Falcon from Southeast University

- 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/

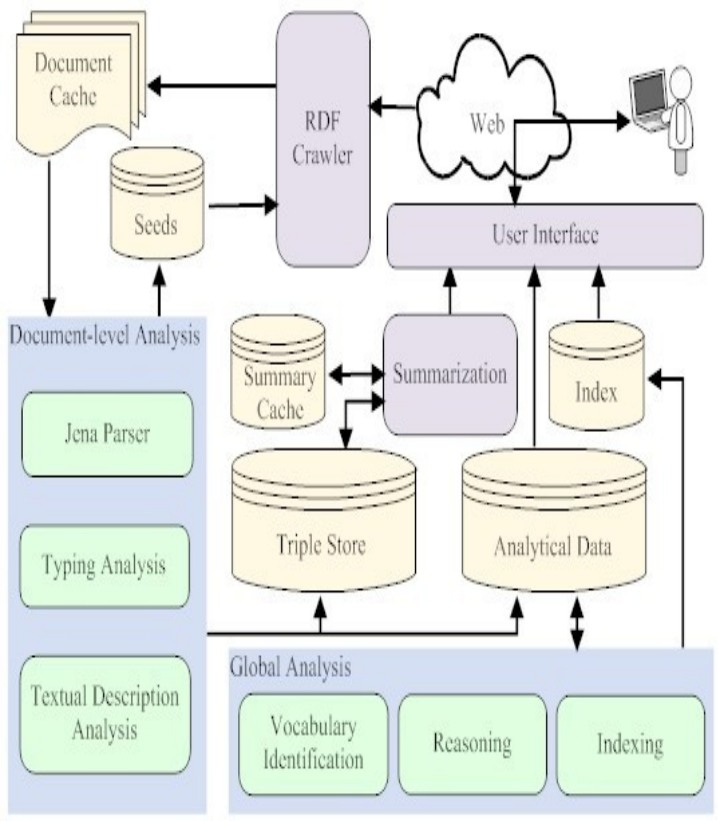
Falcon-AO: A Practical Ontology Matching System

- 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



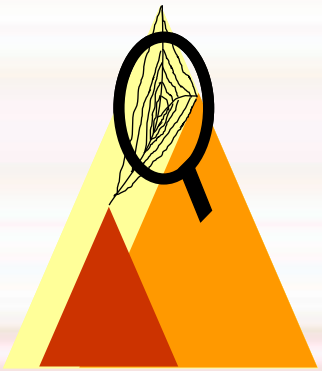
Falcons: A Semantic Web Search Engine

- 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, Q2Semantic, PANTO, RelSE, CE2-Pedia



RelSE

http://relse.apexlab.org

RELATIONAL SEARCH FOR WIKIPEDIA Powered by Semplore

WANNA SURF WIKIPEDIA EFFECTIVELY?

Facing the colossal Wikipedia
How can I find the information I want?
Ask GOOGLE for help?

MAYBE WE CAN MAKE A DIFFERENCE

Find action films starring a Hong Kong martial actor

"action film"
"Hong Kong"
martial actor
site:
en.wikipedia.org

Google

Hong Kong action cinema
Martial arts film
Max Mok Siu-Chung
...

WHERE'RE THE FILMS???

Traditional Search

Relational Search

"action film"
starring
Hong Kong actors
"martial"

RelSE

Righting Wrongs
Eastern Condors
Shaolin Chamber of Death
...

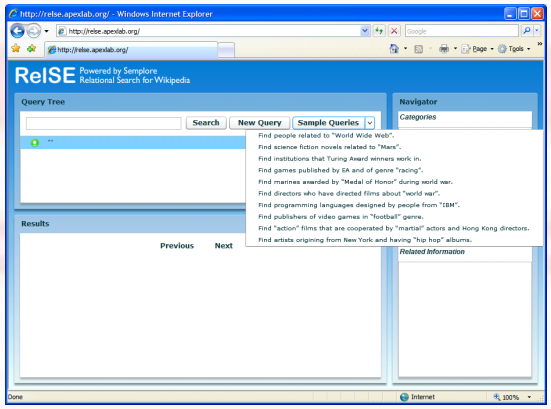
THAT'S WHAT I WANT!

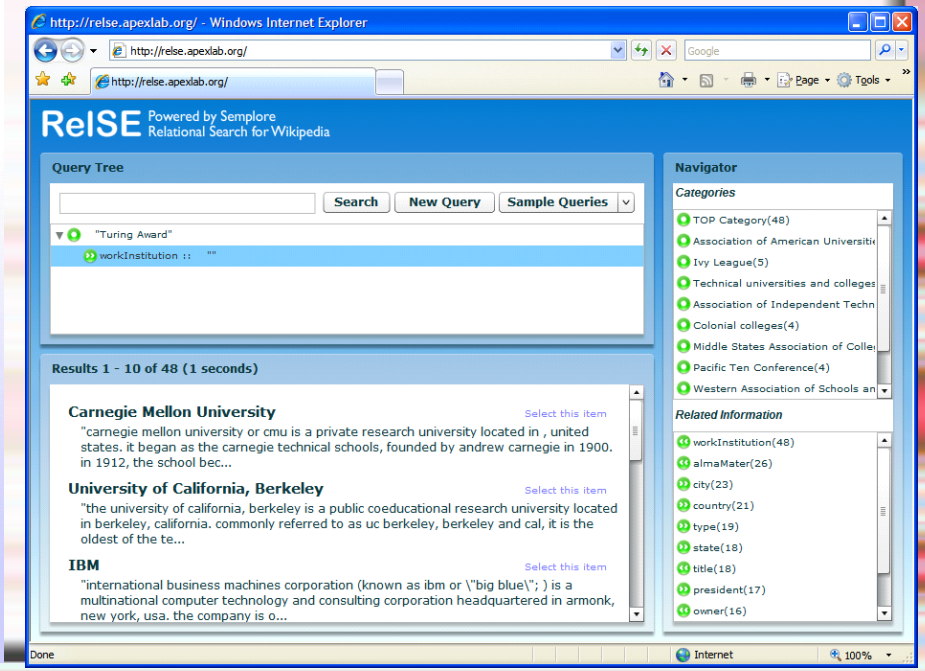
Relational Data + Relational Query + Relational Ranking
Relational Search is what you want!

APEX 数据和知识管理实验室
Apex Data & Knowledge Management Lab

people related to "World Wide Web"
science fiction novel related to "Mars"
institutions that Turing Award winners work in

……





Query to Semantic

Try it now! http://q2semantic.apexlab.org/

How can I find the specification about "SVG" created by "Capin"?

```
PREFIX tap: <http://tap.stanford.edu/tap#>
SELECT ?spec
WHERE {
  ?spec tap:hasAuthor ?person.
  ?spec tap:label "SVG".
  ?person tap:name "Capin".
}
```

Comfused with complex query schema when using semantic search engine?

Let Q2Semantic help you.

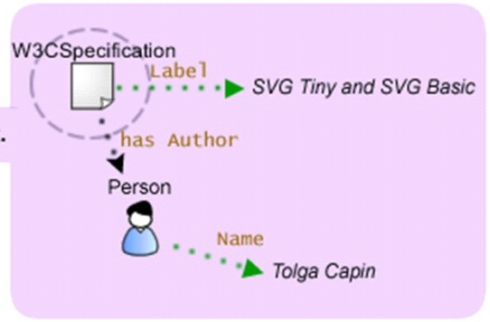
SVG

Capin

Search

Just input your keywords.

Get your formal query by a click.

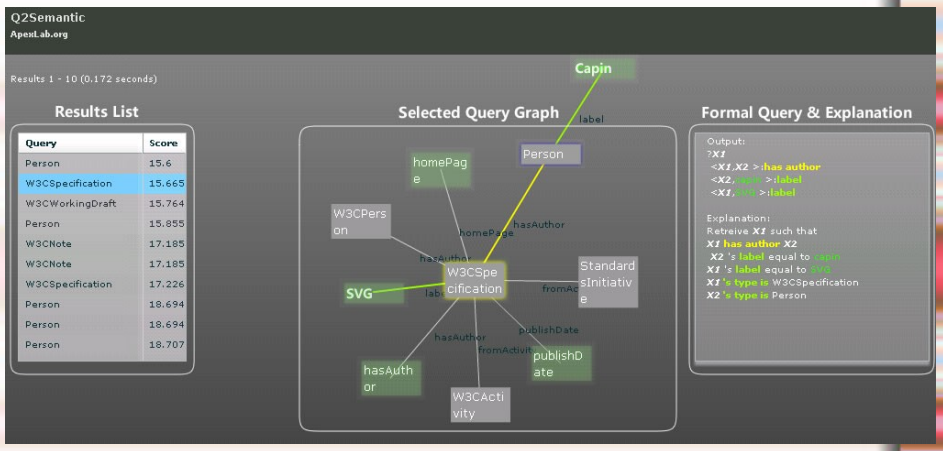


Advantage:

* Adapt Wiki-thesaurus to enrich terms and distinguish relevant resources.
* Clustered resource graph is used to speed up the query construction.
* Well-defined ranking scheme is involved in the top-k search process

APEX 数据和知识管理实验室
Apex Data & Knowledge Management Lab

Find specifications about “SVG” whose author‘s name is “Capin "



Query	Score
Person	15.6
W3CSpecification	15.665
W3CWorkingDraft	15.764
Person	15.855
W3CNote	17.185
W3CNote	17.185
W3CSpecification	17.226
Person	18.694
Person	18.694
Person	18.707

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

Jie Tang, Jing Zhang, Limin Yao, Duo Zhang, Mingcai Hong, and Juanzi Li
Knowledge Engineering Group, DCST, Tsinghua University
{tangjie, zhangjing, ylm, zhangduo, hmc}@keg.cs.tsinghua.edu.cn

Arnetminer
Http://www.arnetminer.org

Association Finding

Expert Finding

Person Search

Survey Pa

Search

Sub-topi

Search

Hot-topi

- 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

Social Network Extraction

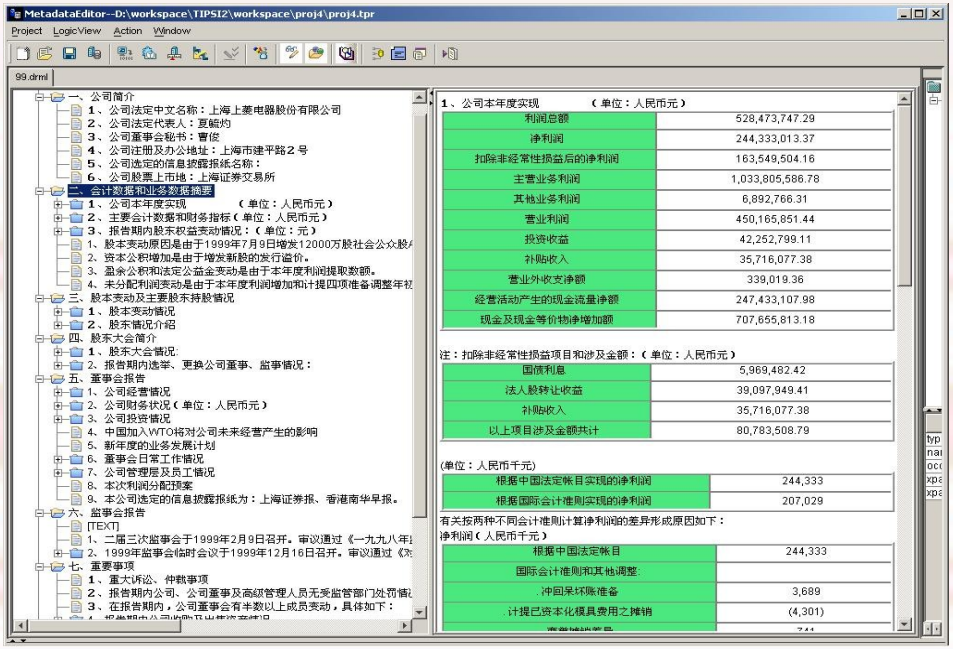
Name Disambiguation

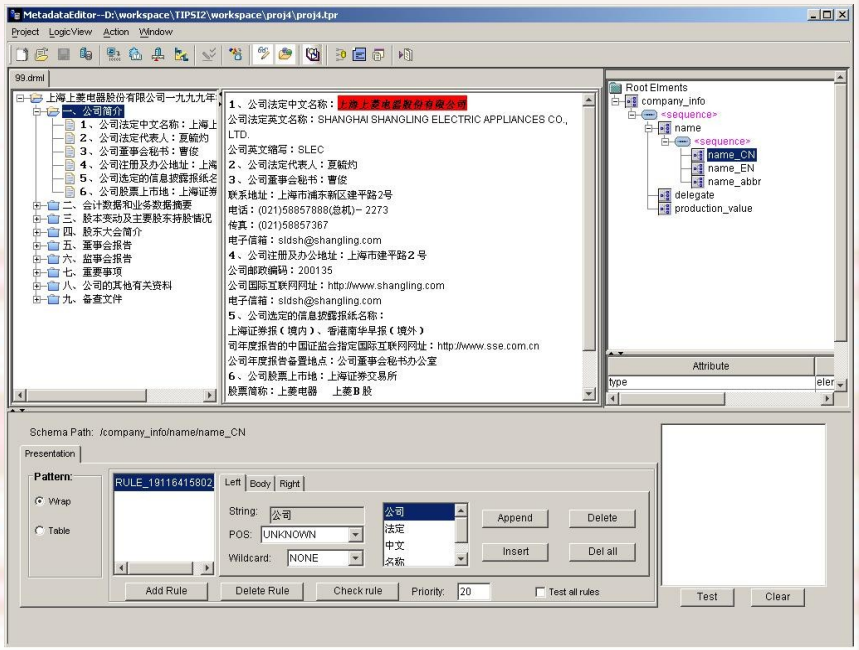
Integration

Profiling

Semantic Annotation

- Develop three different approaches in automatic semantic annotation
  - Rule induction
  - Unbalanced classifier
  - Sequence labeling



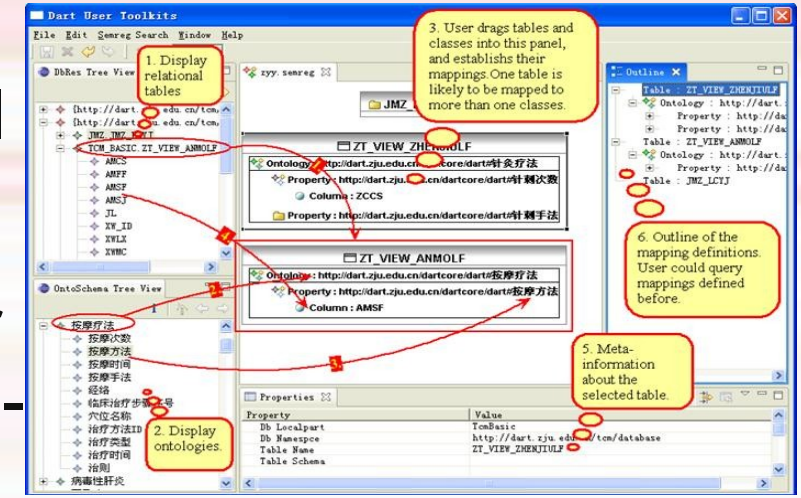


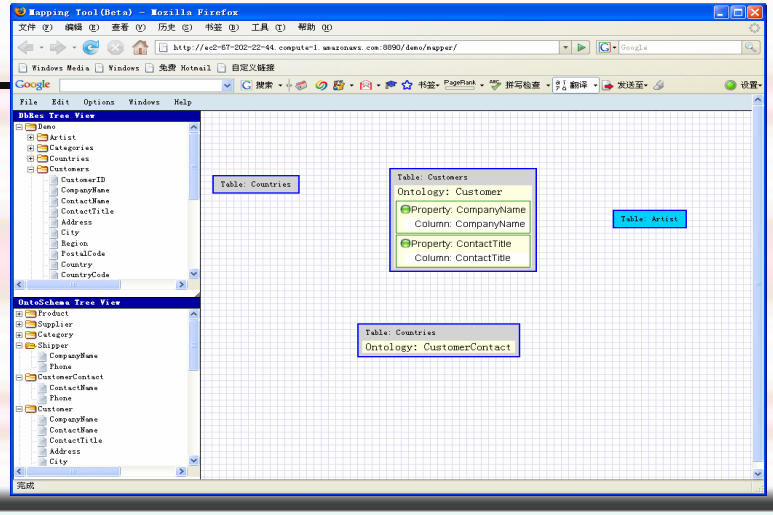
DartGrid from Zhejiang University

- 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: visulized 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

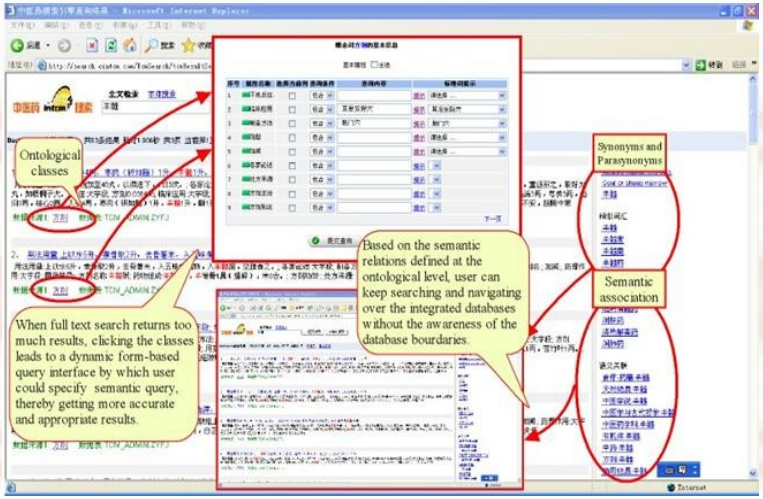
- Visualized mapper for defining mappings from relational schema to RDF ontology.
- A new browser-host mapper provides online mapping manipulation supporting XMLP.
- Provides fast linked data publishing and deployment for both manually defining the mapping and direct D2R transformation.

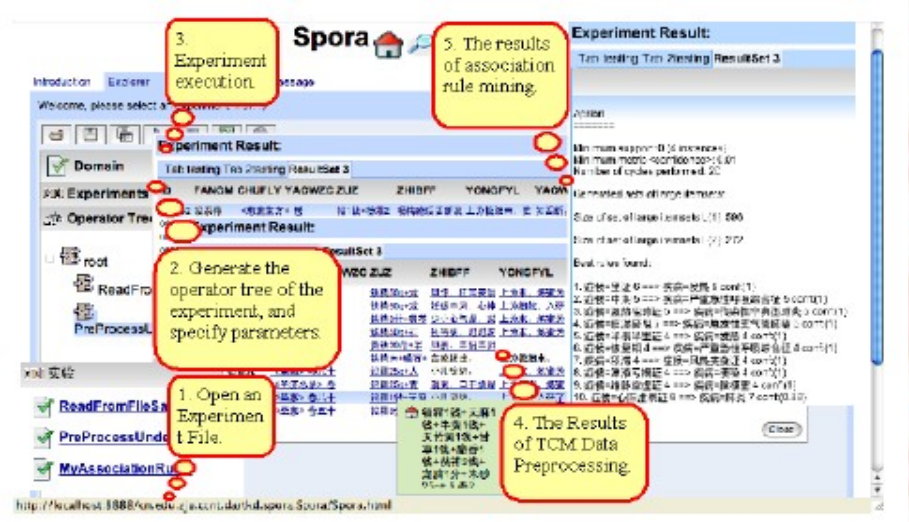




Semantic Data Query and Analysis

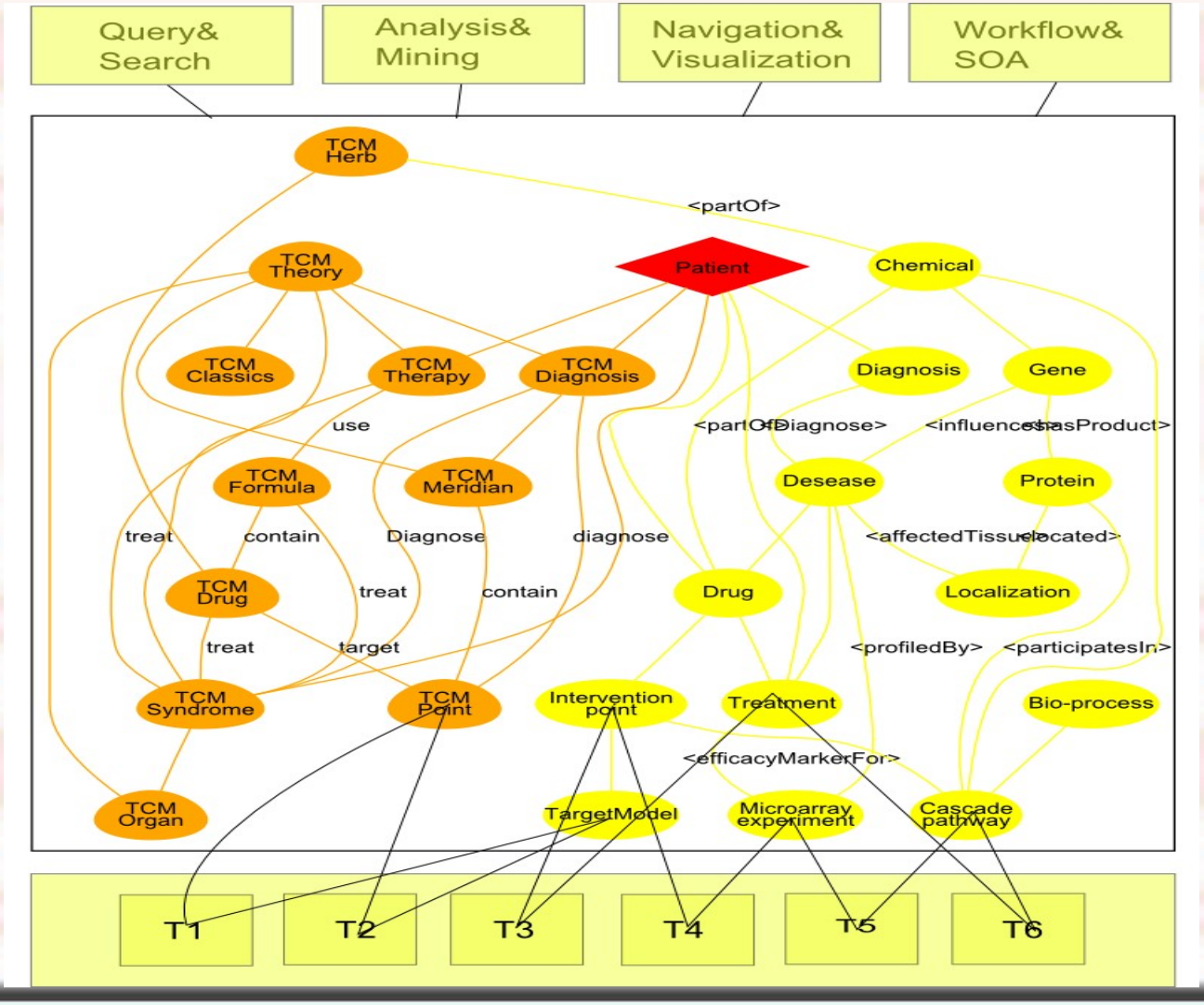
- **Ontology-based Query and Search**
  - Query form for user to construct SPARQL queries.
  - Ontology-based automatically generating query form.
  - Ontology-enabled semantic navigation.
- Semantic Graph Mining:
  - Enable relational mining over web of data





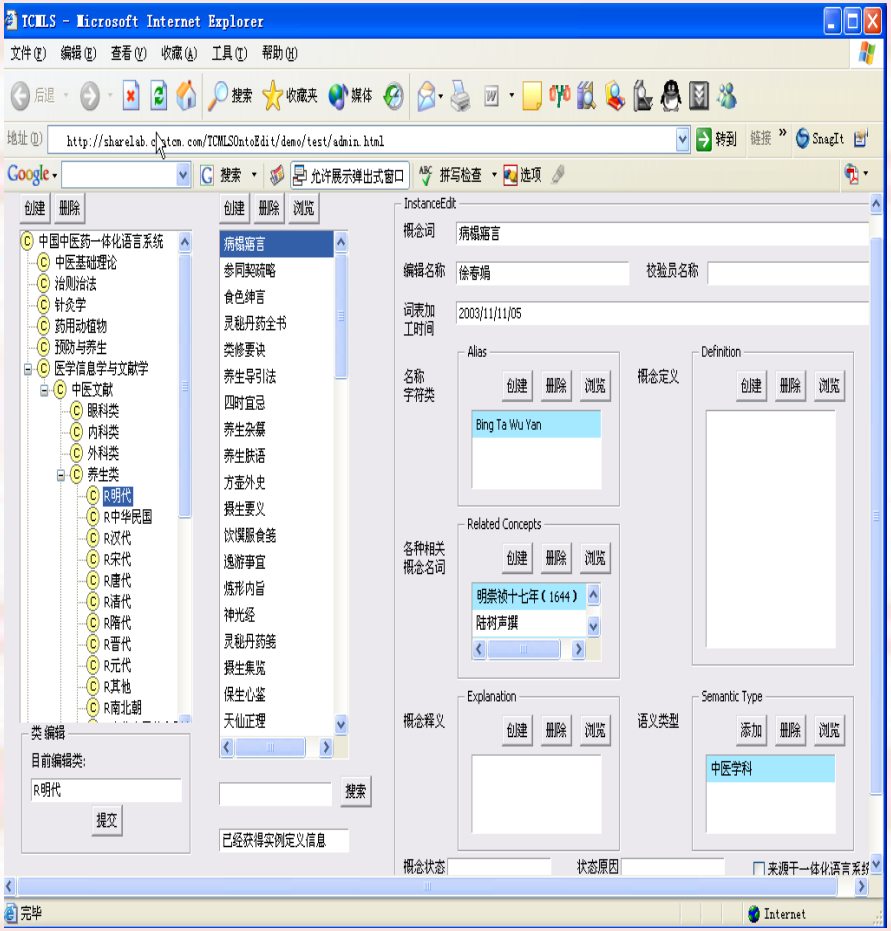
Application: Health Care and Life Science

Make a connection between TCM and modern medicine



TCM Ontology Effort

- Collaborative on line 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, Karlsuhe 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!