天津大学
Tianjin University

CCF 2013 开放课题汇报
课题编号：CCF2013-02-02

基于Linked Data的中文学术期刊语义知识组织与服务关键技术研究

课题负责人：王鑫
wangx@tju.edu.cn
天津大学 计算机科学与技术学院

2014.12.09 @ NLPCC 2014

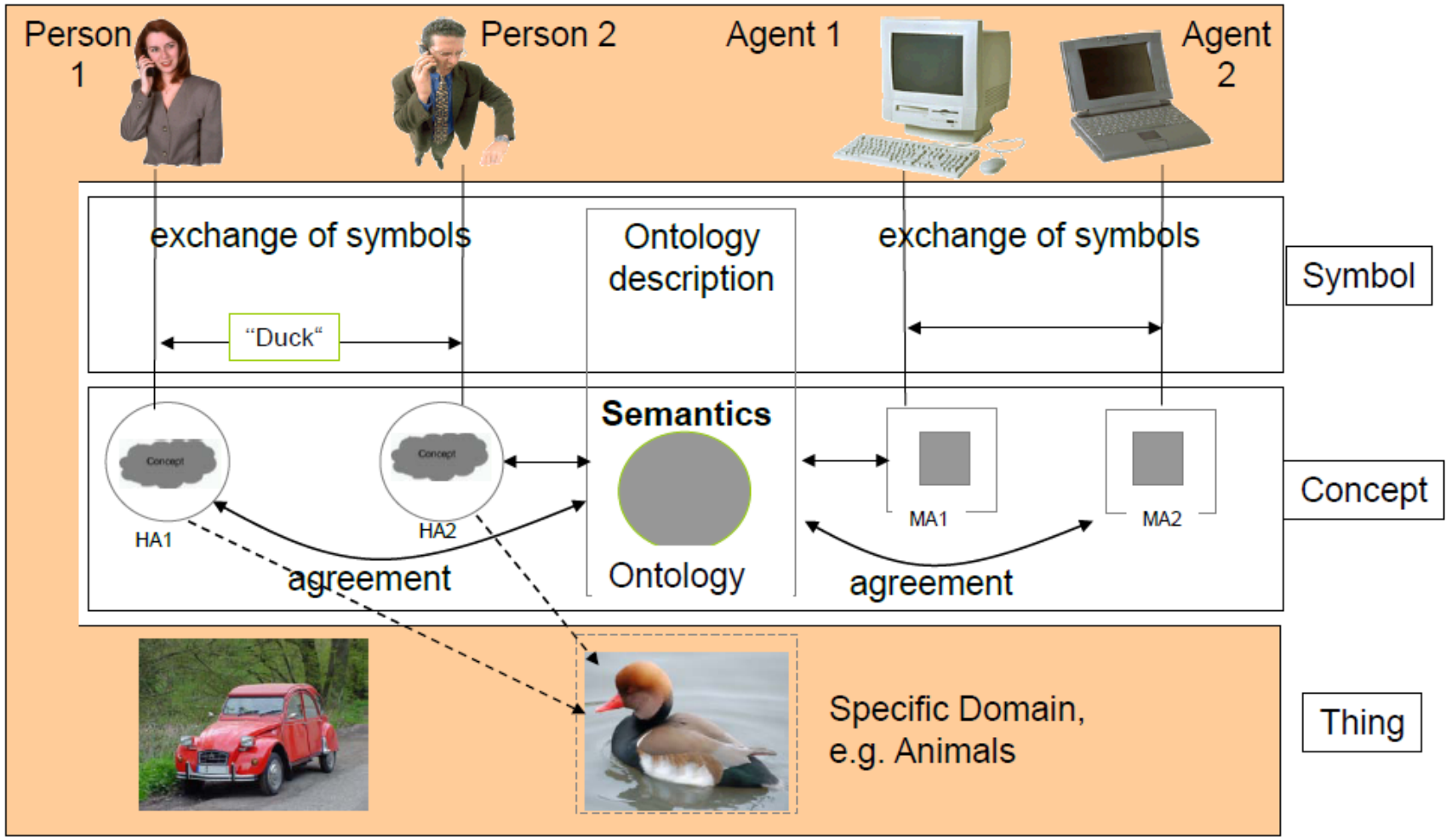
一、课题背景

二、中文学术期刊本体与知识库

三、中文学术期刊语义知识服务

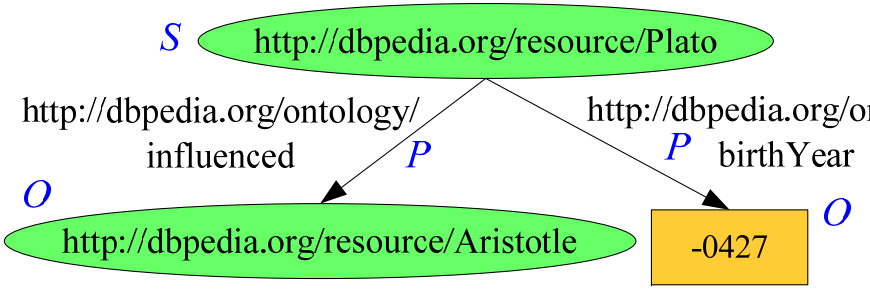
四、原型系统与功能演示

语义万维网(Semantic Web)的基本思想



Linked Data：语义大数据

- RDF：资源描述框架

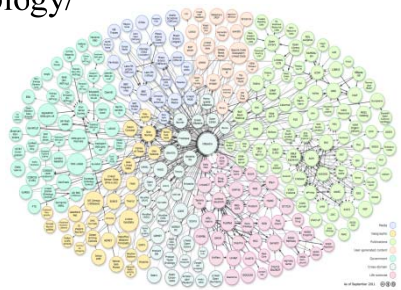


Cumulative numbers

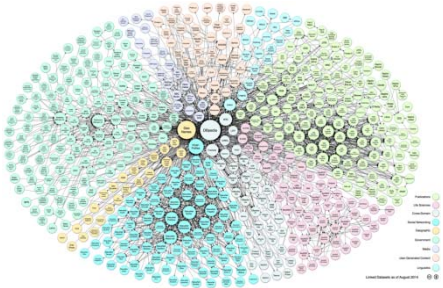
2122 datasets

61,976,332,795 triples from

- Linked Data

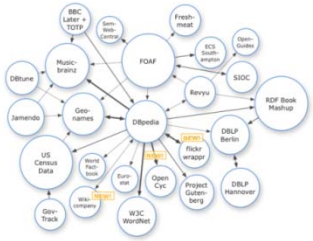


2011.09

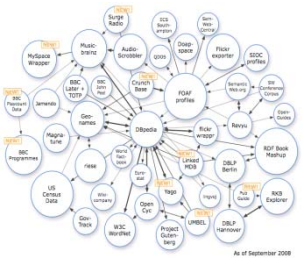


2014.08

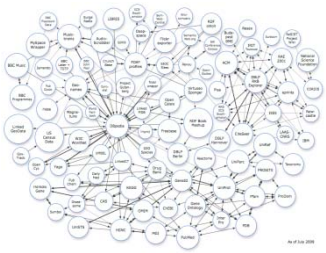
百科知识
地理信息
生物医学
媒体出版
电子政务
……



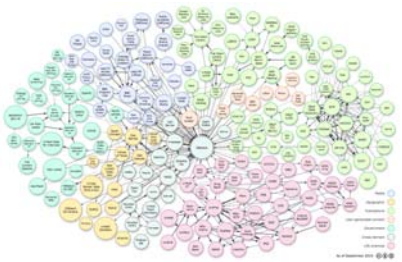
2007.10



2008.09



2009.07

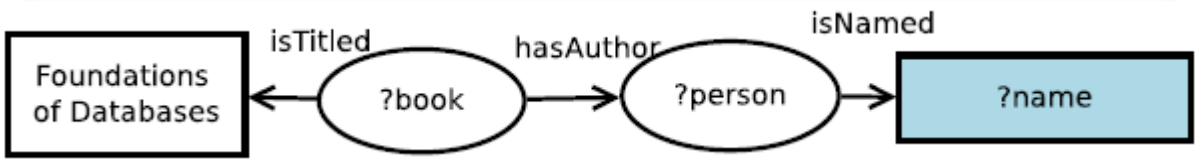


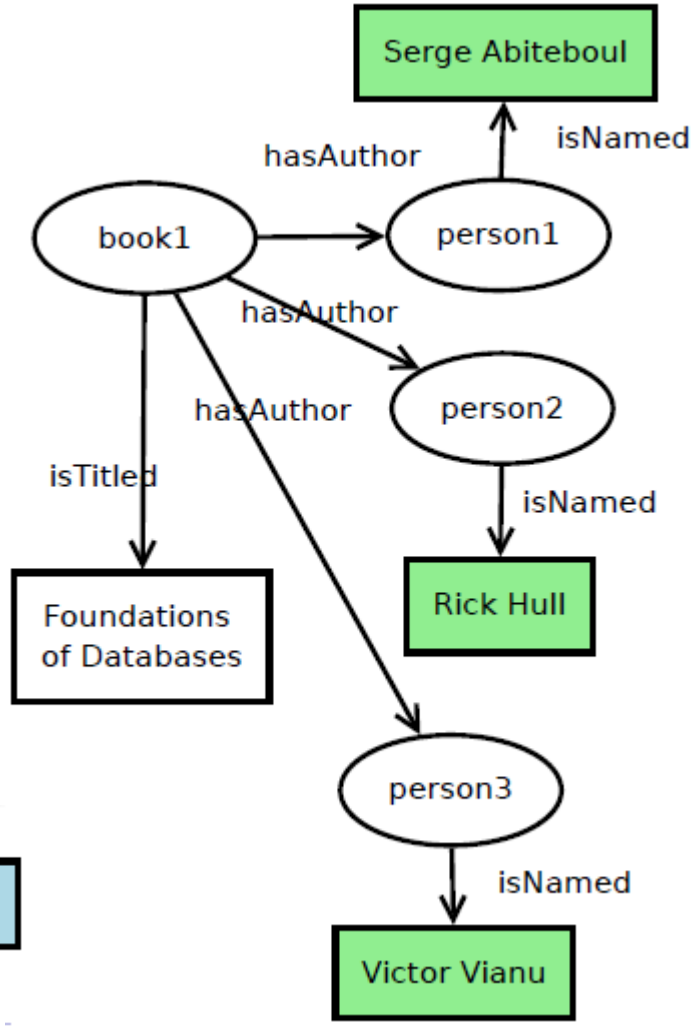
2010.09

RDF查询语言：SPARQL

举例

```
SELECT ?name
WHERE {
  ?book isTitled "Foundations of Databases".
  ?book hasAuthor ?person .
  ?person isNamed ?name .
}
```





Linked Data与数字出版

- Linked Data四原则
  1. 用URI指代实体
  2. 用HTTP URI使实体可引用，可被人或Agent查看
  3. 查看URI时提供关于实体的有用信息
  4. 在Web上发布数据时包含实体到其他相关实体的链接

——Tim Berners-Lee

- Linked Data数字出版
  - 知识库、服务、共享

nature.com linked data

DBpedia

dblp computer science bibliography

crossref.org

万方数据 WANFANG DATA 知识服务平台

维普网 仓储式在线出版平台

cnki 中国知网 cnki.net

课题研究内容

1. 基于Linked Data的中文学术期刊与文献知识库构建

2. 基于Linked Data知识库的Web知识服务API构建

3. 基于Linked Data的中文学术期刊与文献语义服务平台

一、课题背景

二、中文学术期刊本体与知识库

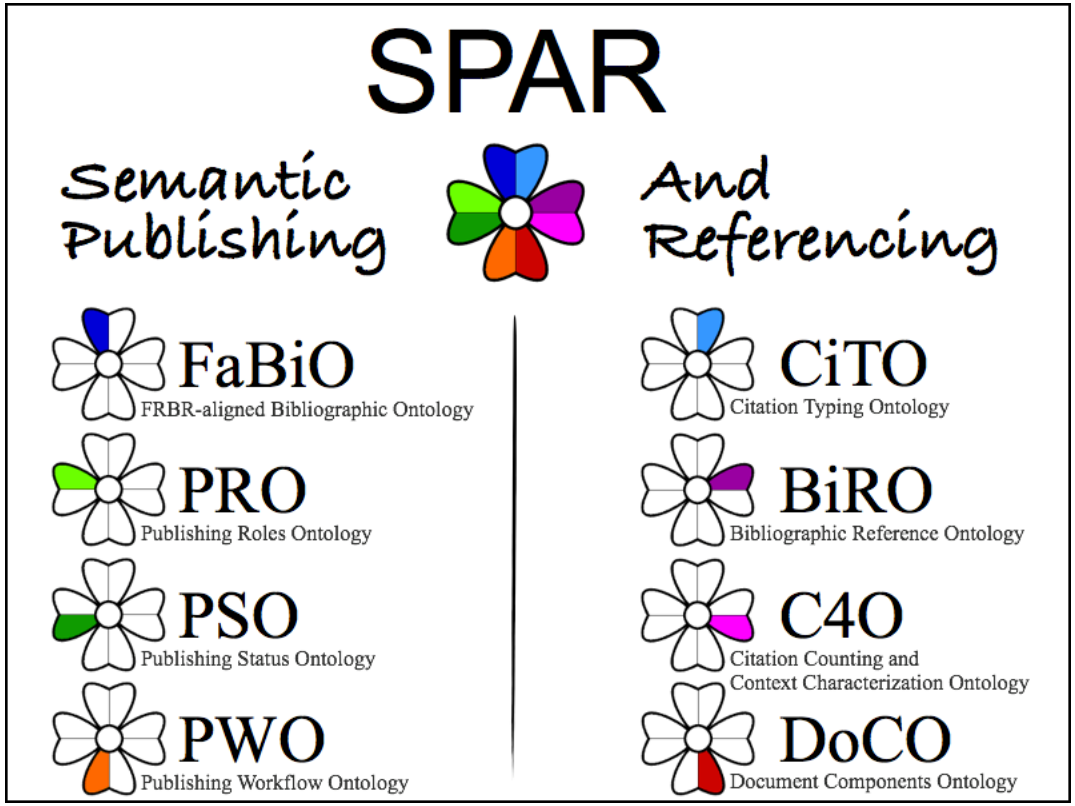
三、中文学术期刊语义知识服务

四、原型系统与功能演示

➢ 中文学术期刊本体构建

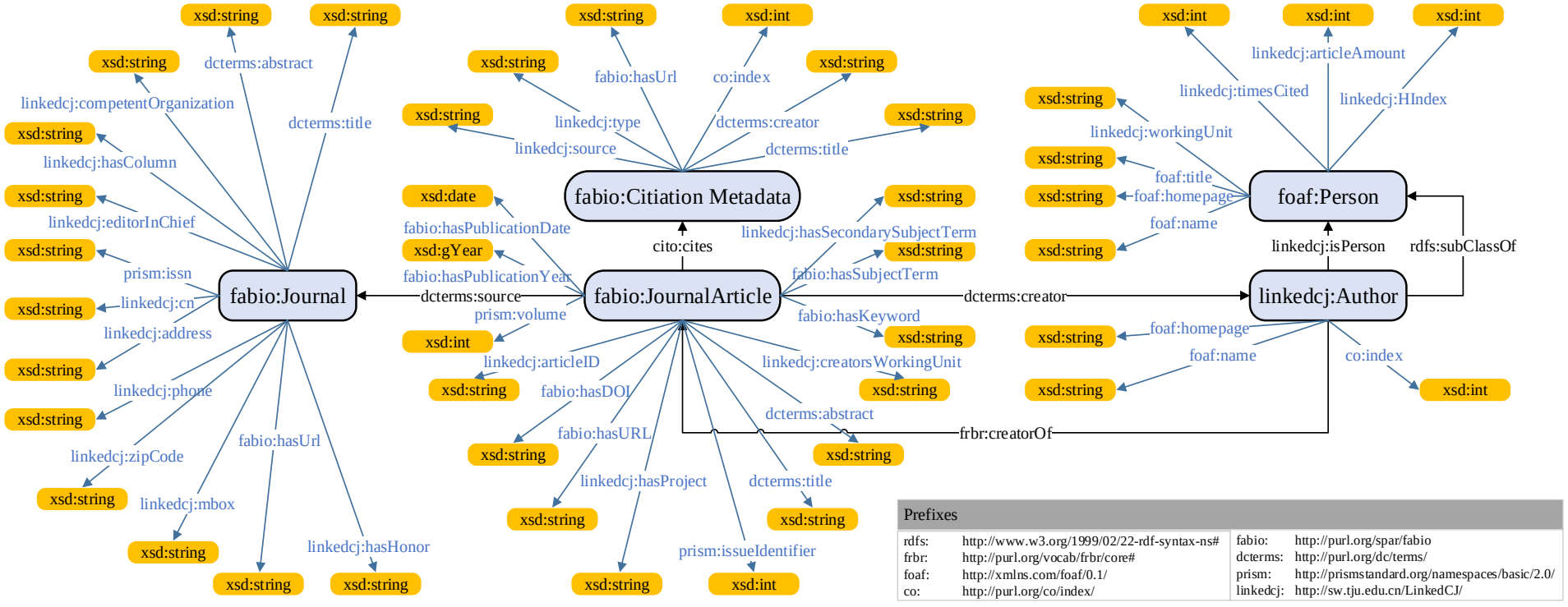
□ 原则：尽量复用国际上已有的语义出版本体

- FaBio
- CiTO
- Dublin Core
- FOAF
- PRISM
- …



中文学术期刊本体构建

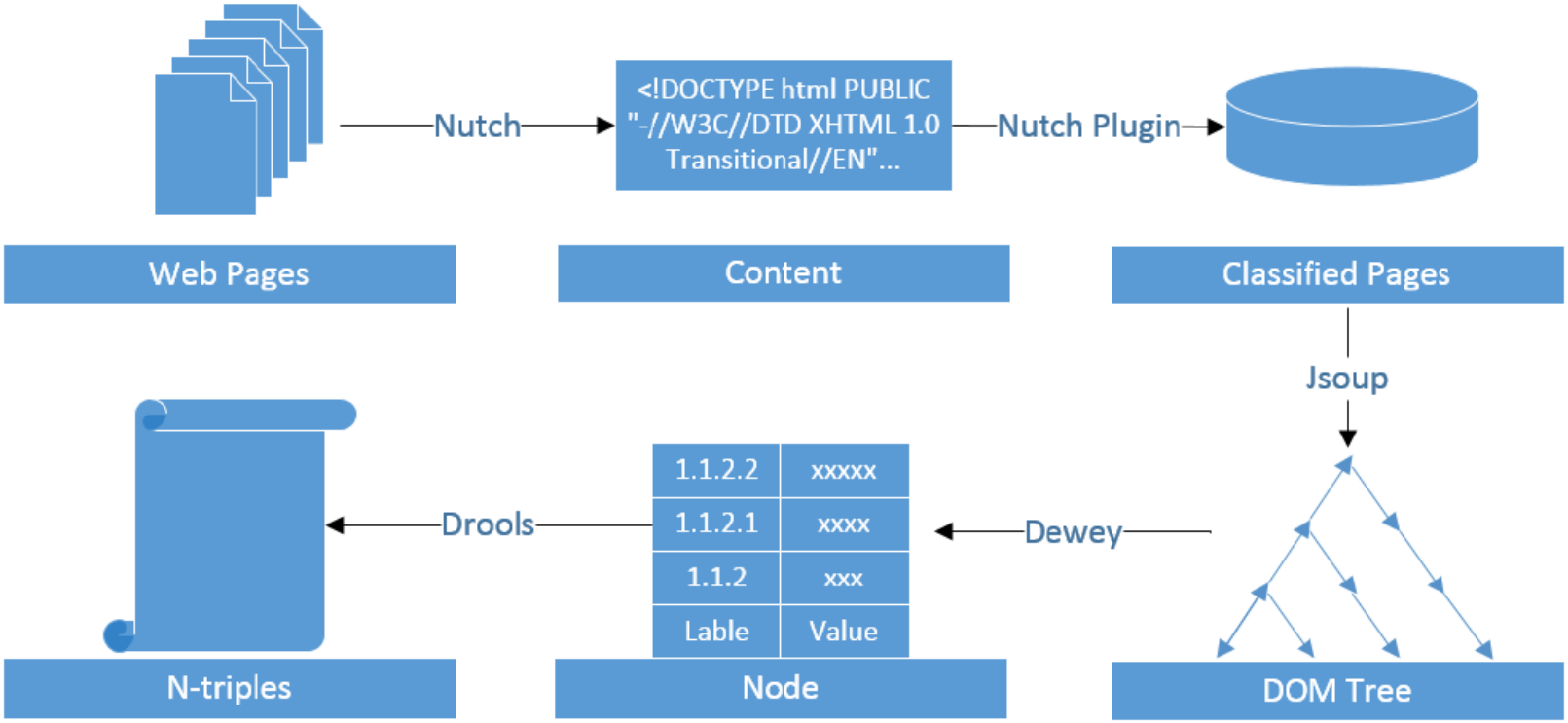
LinkedCJ本体：融入中文学术期刊特点



Prefixes			
rdfs:	http://www.w3.org/1999/02/22-rdf-syntax-ns#	fabio:	http://purl.org/spar/fabio
frbr:	http://purl.org/vocab/frbr/core#	dcterms:	http://purl.org/dc/terms/
foaf:	http://xmlns.com/foaf/0.1/	prism:	http://prismstandard.org/namespaces/basic/2.0/
co:	http://purl.org/co/index/	linkedcj:	http://sw.tju.edu.cn/LinkedCJ/

中文学术期刊知识库构建

知识库生成

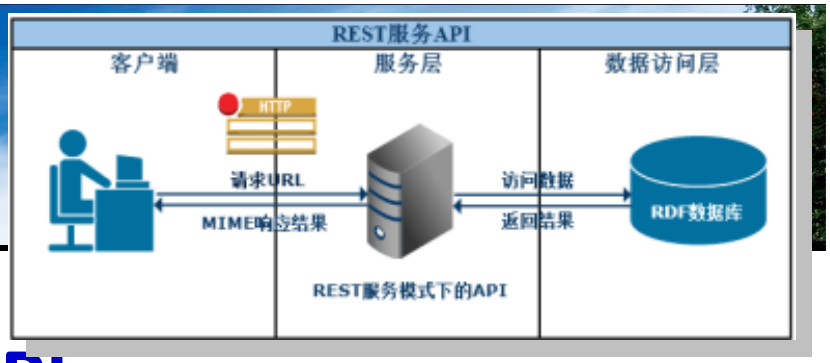


一、课题背景

二、中文学术期刊本体与知识库

三、中文学术期刊语义知识服务

四、原型系统与功能演示



➢ 中文学术期刊知识服务API

□ 基于REST服务的知识服务API

Linked Data Platform

方法名	方法描述	资源 URI
queryArticleByDoi()	根据某篇文献的 Doi，查询并获取该文献资源。	http://localhost:8080/linkedcj/rest/journals/article?doi=
queryPersonByAuthorId()	根据某位作者的 ID，查询并获取该作者的个人信息资源。	http://localhost:8080/linkedcj/rest/journals/person?authorId=
queryPersonItem()	根据某个（些）人的姓名（同名不同人），查询并获取这个（些）人的姓名、作者 ID、学历和工作单位信息资源。	http://localhost:8080/linkedcj/rest/journals/personItem?name=
queryAuthor()	根据某篇文献的其中一个创作者的 ID，查询并获取这个创作者的姓名、主页等信息资源。	http://localhost:8080/linkedcj/rest/journals/author?creatorId=
queryJournal()	根据某个期刊的 ID，查询并获取这个期刊的刊名、主管单位、主办单位、主编等信息资源。	http://localhost:8080/linkedcj/rest/journals/journal?journalId=
queryJournalByname()	根据某个期刊的刊名，查询并获取这个期刊的刊名、主管单位、主办单位、主编等信息资源。	http://localhost:8080/linkedcj/rest/journals/journal?name=
queryCitation()	根据某篇文献的某个引文 ID，查询并获取这个引文的标题、作者、引文类型、引文来源等信息资源。	http://localhost:8080/linkedcj/rest/journals/citation?citeId=
queryResult()	根据用户输入的 SPARQL 语句，查询并获取作者所需的信息资源。	http://localhost:8080/linkedcj/rest/journals/sparql?sparqlstr=
queryTitleDoiByKeyword()	根据用户输入字段，查询并获取含有这些字段的文献标题和 Doi。	http://localhost:8080/linkedcj/rest/journals/title?keyword=

中文学术期刊知识服务平台

- Linked Data 语义链接

中文学术期刊知识服务平台

SPARQL查询服务

SPARQL查询

```
select ?title?source where { ?article
<http://prismstandard.org/namespaces/basic/
2.0/doi>
"http://dx.doi.org/10.3724/SP.J.1016.2014.00
621"; <http://purl.org/dc/terms/title> ?title;
<http://purl.org/dc/terms/source> ?source}
```

查询示例

1. 如何查询DOI为"http://dx.doi.org/10.3724/SP.J.1016.2014.00621"的文章的题目和来源?

2. 如何查询文献《一种多变量决策树方法研究》的作者和他的工作单位?

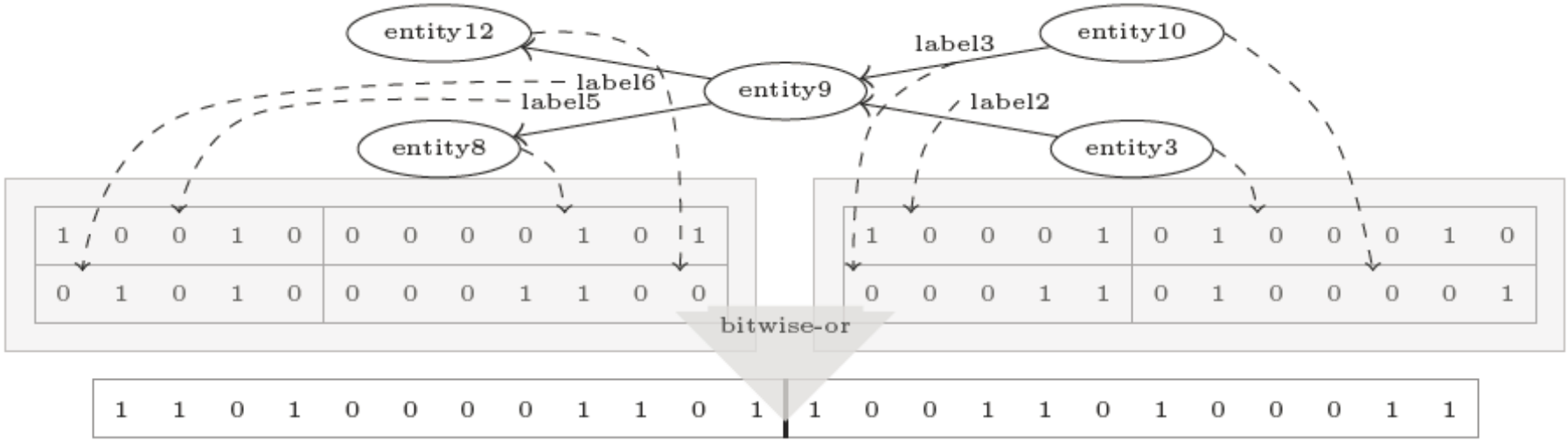
3. 如何查询期刊《计算机科学》的主编和主办单位?

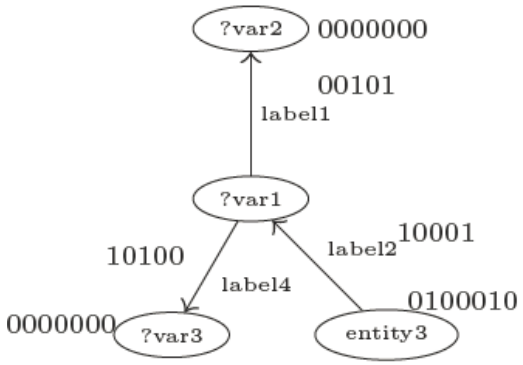
查询结果

title	source
基于样本的大规模人群快速创作	http://purl.org/spar/fabio/Journal/l-jsjxb

➢ 大规模RDF图上的子图匹配查询

□ 星形查询的高效执行：SPARQL关键操作





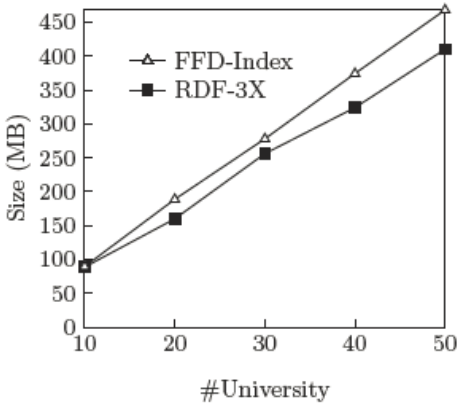
(a) Q_1^* and Bitstrings of Lables

$$\frac{\underline{10101\ 0000000}}{fp_{out}(Q_1^*)} \quad \frac{\underline{10001\ 0100010}}{fp_{in}(Q_1^*)}$$

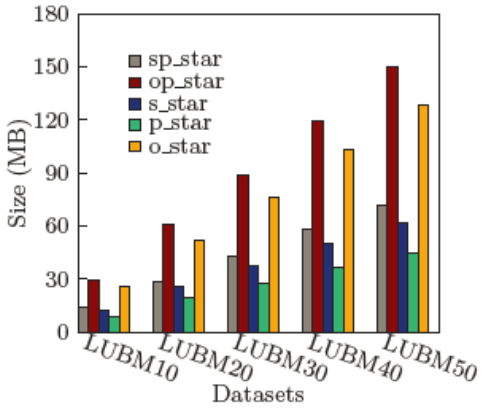
(b) $fp(Q_1^*)$

➢大规模RDF图上的子图匹配查询

□ 星形查询的高效执行： SPARQL关键操作



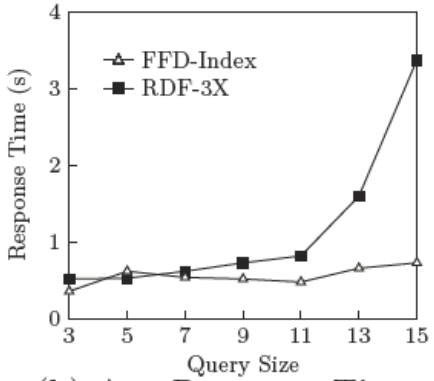
(a) Index Size



(b) FFD-Index Size

	# BEF. VLD.	# AFT. VLD.
Q_3	542.6	67.4
Q_5	477.3	45.8
Q_7	237.0	21.1
Q_9	97.3	12.7
Q_{11}	104.4	4.2
Q_{13}	85.5	1.6
Q_{15}	57.2	1.8

(a) Filtering Ability



(b) Avg Response Time

Fig. 8: Performance on YAGO

一、课题背景

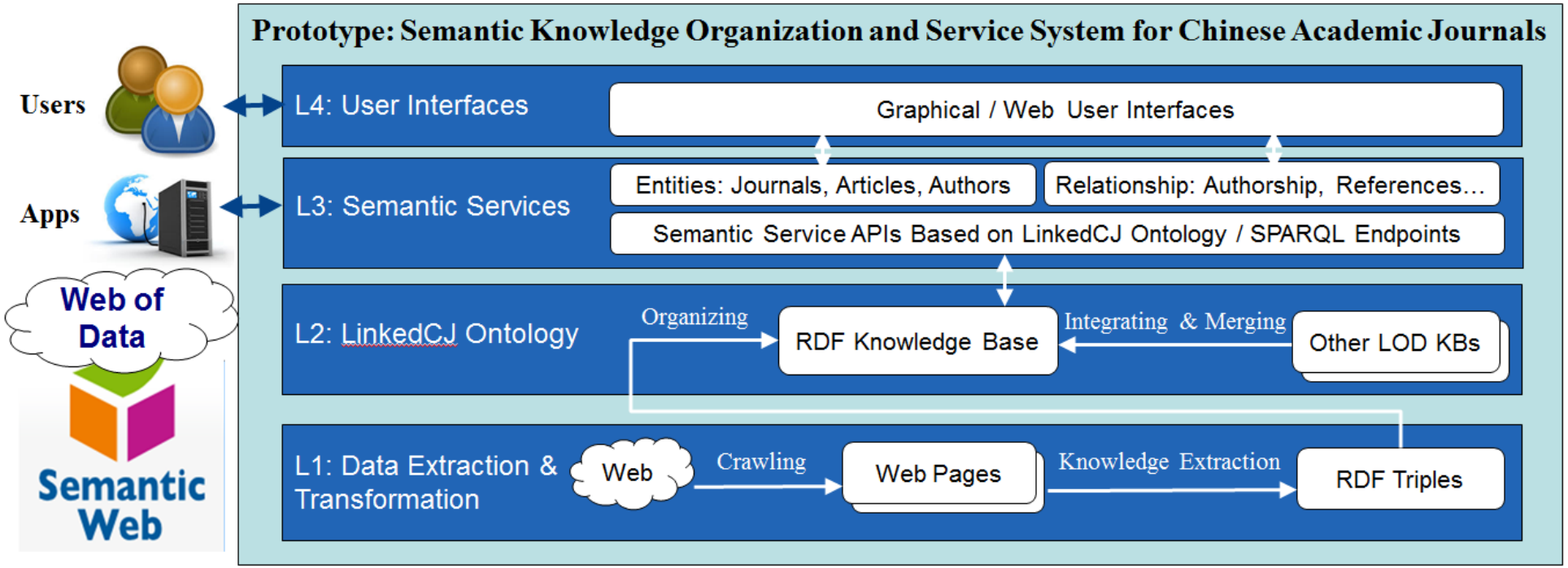
二、中文学术期刊本体与知识库

三、中文学术期刊语义知识服务

四、原型系统与功能演示

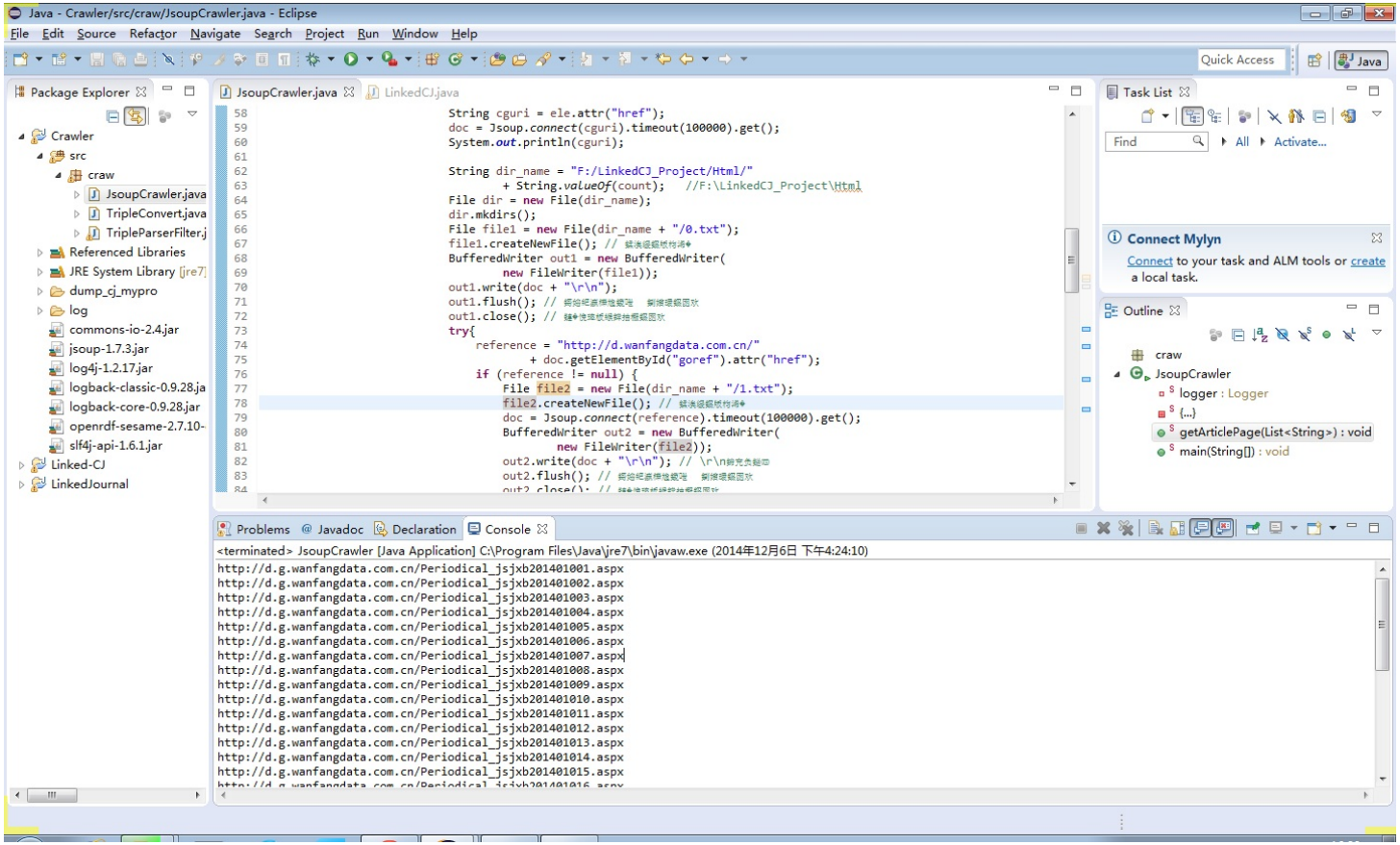
➢ 中文学术期刊语义知识组织与服务系统

□系统架构



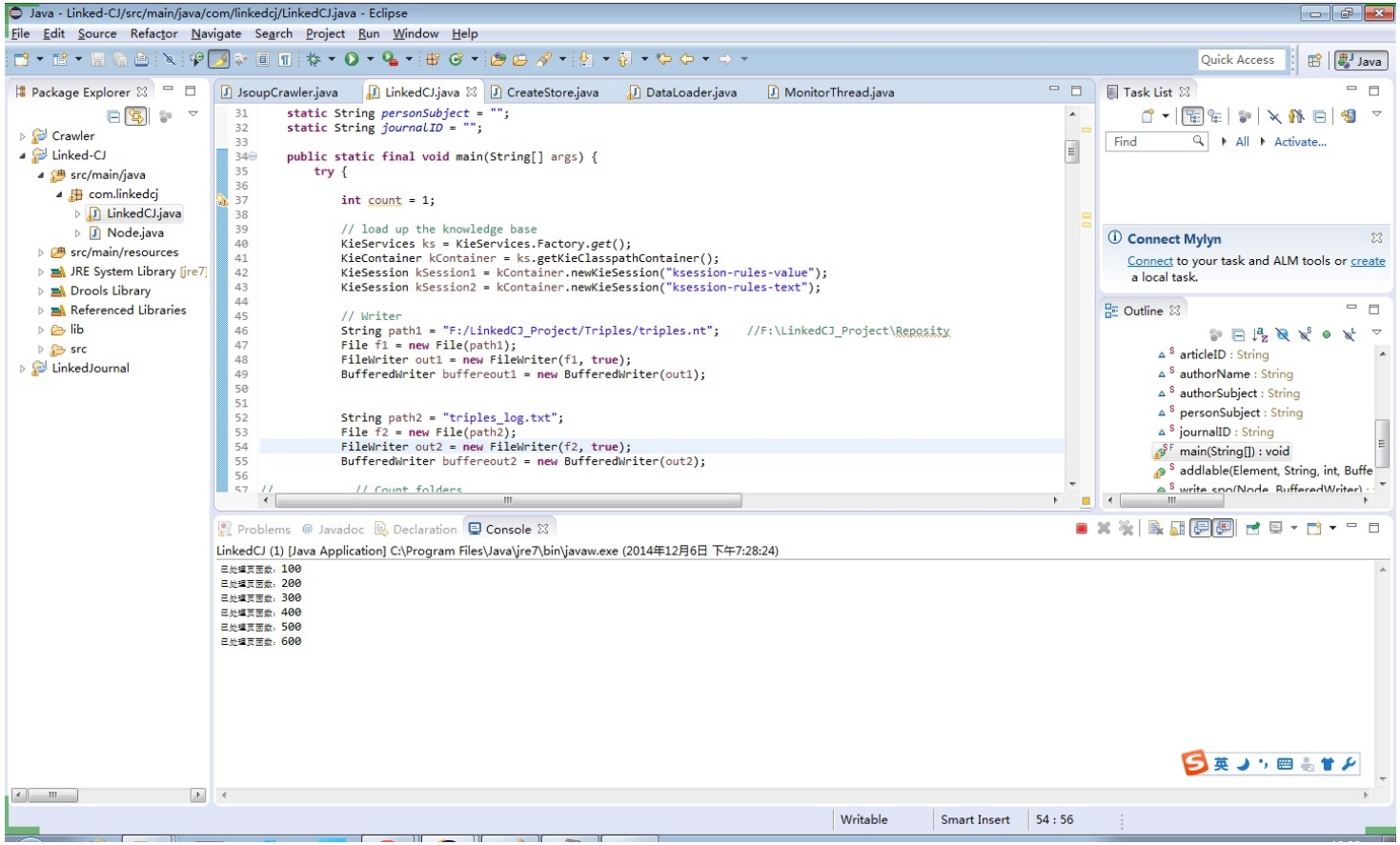
原型系统演示

演示1：Web页面爬取



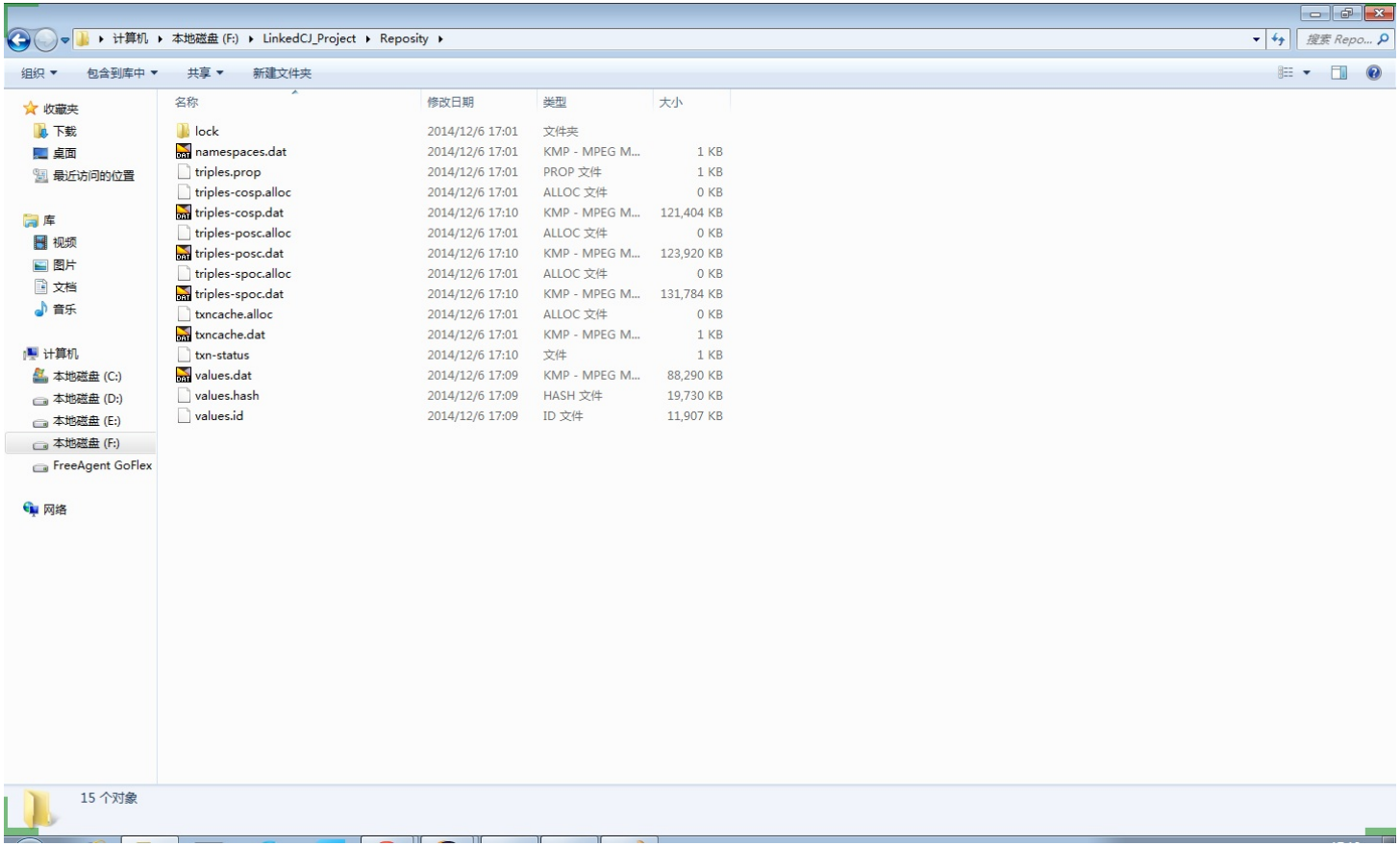
原型系统演示

演示2：知识抽取



原型系统演示

演示3：知识库装载



➢ 原型系统演示

□ 演示4：服务平台

课题资助发表的论文

	名称	期刊/会议
1	Xu Peng, Xin Wang, and Haofen Wang. LinkedCJ: A Knowledge Base of Chinese Academic Journals Based on Linked Data. In Proceedings of the 4th Joint International Semantic Technology Conference (JIST 2014), poster, 2014.	*JIST 2014*
2	Xuedong Lyu, Xin Wang, Yuan-Fang Li, and Zhiyong Feng. FFD-Index: An Efficient Indexing Scheme for Star Subgraph Matching on Large RDF Graphs. Submitted to DASFAA 2015.	*DASFAA 2015* (CCF B)
3	张啸野, 王鑫, 徐鹏, 冯志勇. 基于Linked Data的中文学术期刊语义知识组织与服务. 计算机工程. (待发表)	核心期刊
4	魏亚洲, 王鑫, 冯志勇, 饶国政. S-Index: 一种面向大规模 RDF 数据的高效率语义索引方案. 武汉大学学报（理学版）（待发表）（第8届中国语义Web和Web科学会议推荐）	核心期刊

总结

- 基于Linked Data的中文学术期刊语义知识组织与服务关键技术研究

谢谢大家
请提问题

2014.12.09@NLPCC2014