Semantic Visualization for Subject Authority Data of Chinese Classified Thesaurus

Wei Fan
Sichuan University

Shuqing Bu
National Library of the China

Qing Zou
Lakehead University

Outline

I. Background
   - Chinese Classified Thesaurus (CCT)
   - Open and Linked Data Environment

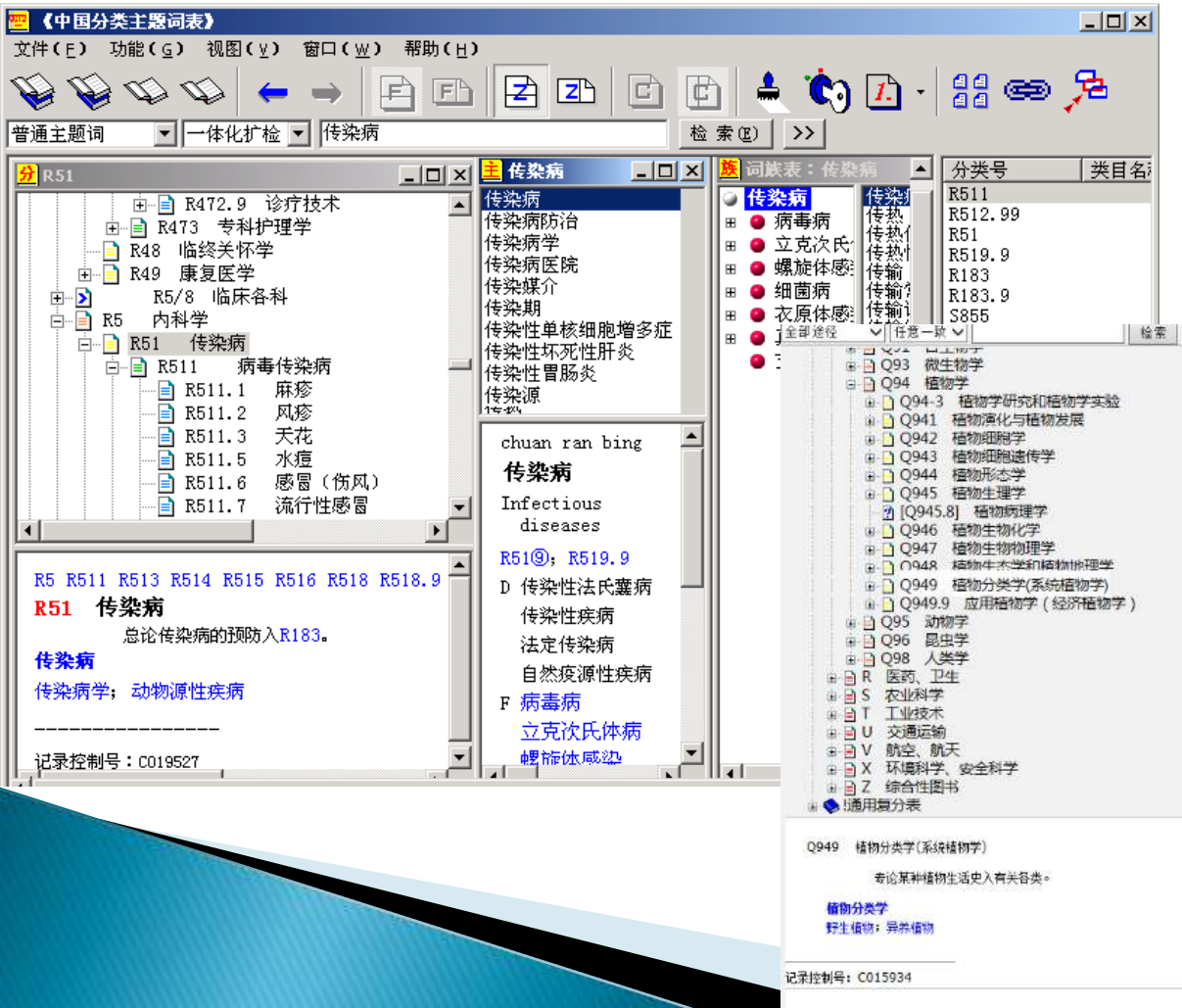
II. SKOS Modelling for CCT
   - Subject Authority Data modelling
   - Integration Structure Considerations

III. Semantic Visualization
   - Design Architecture and Implementation
   - Visualization Interfaces

IV. Conclusion

I. Background - *CCT Introduction*

Chinese Classified Thesaurus is integrated from Chinese Library Classification (CLC) and Chinese Thesaurus (CT).



Electronic Version ←

Web Version ↓

I. Background - *Some Practical Points*

- CCT is designed for traditional cataloguers' workflows
- Its complicated knowledge structure and relation mappings between CLC and CT are hidden to non-expert users

---

- A relatively isolated system with use limited to the library field (eg. OPAC search and annotation)
- Lack of capacity for open linking and communication with external web applications

I. Background - *Seizing Open Linked Data Chance*

- Linked Data provide a feasible technical mechanism for publishing open data (Heath & Bizer, 2011)

- Terminology Services (TS) have brought KOS's applications to the level of Web Services which means that TS "can be m2m or interactive, user-facing services and can be applied at all stages of the search process" (Tudhope, Koch & Heery, 2006)

CCT could play an important part in structuring and inter-linking Semantic Web data.

I. Background - *What can we do in this paper*

Show our approach that how to transform CCT into linked data and supporting it with an interactive visualization interface.

- Discuss a basic semantic modelling for subject authority data. While, some integration issues are discussed.
- Design and implement a visualization demo system on an existing terminology service platform.

2. SKOS Modelling for CCT

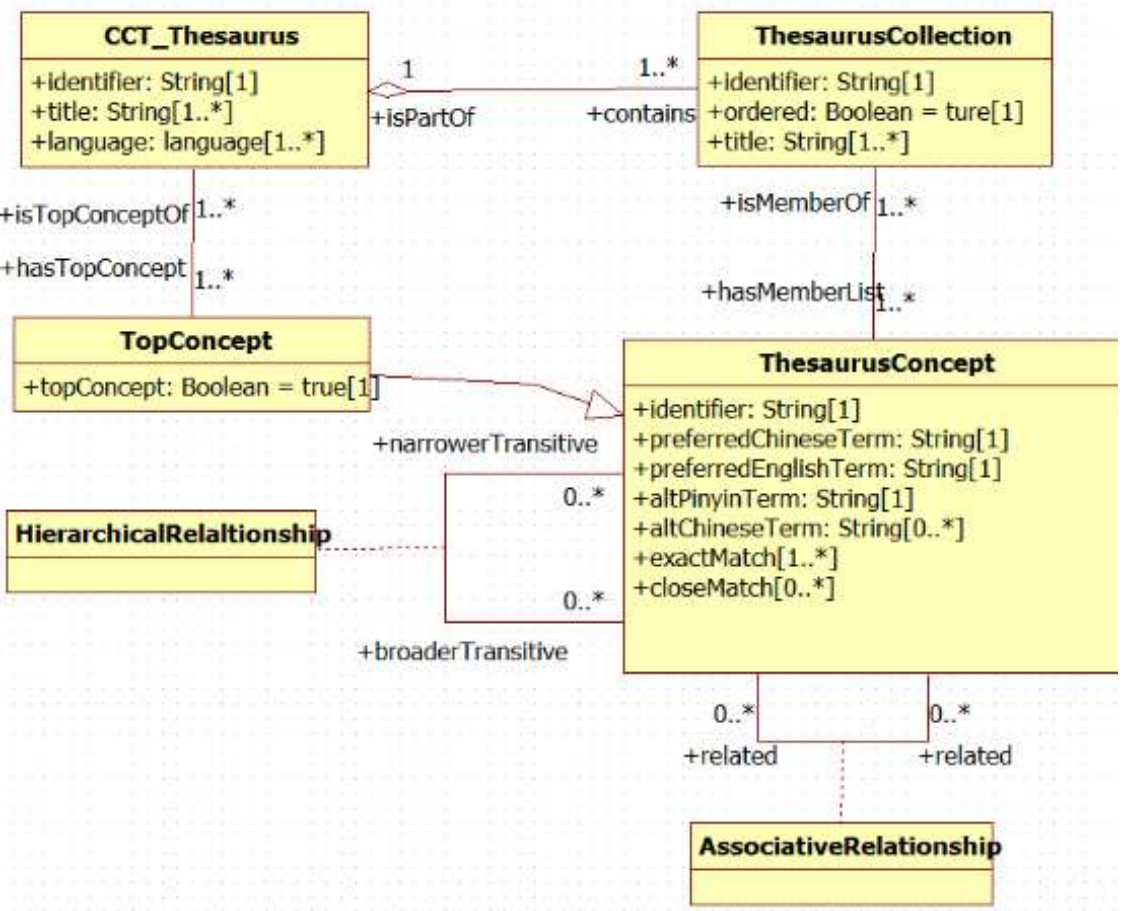
Existed Data Format

- China Machine-Readable Cataloguing Formats (CNMARC) for subject authority data.
- China Library Classification Machine-Readable Cataloguing Formats (CLCMARC) which are based on Universal MARC (UNIMARC) Format for classification data containing 22 main-classes, 52,992 sub-classes;

With complex integration considerations

- Starting with subject authority data (Thesaurus Part)
- Express semantic relationships progressively by carefully following the development of both web technology and vocabulary standards.

2. SKOS Modelling for CCT – *Our Approach*



TopConcept itself is not only a ThesaurusConcept, but also has additional features in a specific domain. Thus, TopConcept could be a generalization of ThesaurusConcept as its children class.

SKOS broader/narrower transitive properties are selected for representing the semantic relationships in the hierarchical structures.

2. SKOS Modelling for CCT – *Our Approach*

more than 100,000 subject authority entries have been converted from CNMARC into SKOS. Subject authority data have mainly included preferred terms, non-preferred terms and coordinated terms.

```
<skos:Concept rdf:about="http://cct.nlc.gov.cn/Subject/S095502#concept">
  <skos:inScheme rdf:resource="http://cct.nlc.gov.cn/Subject#conceptScheme"/>
  <skos:topConceptOf rdf:resource="http://cct.nlc.gov.cn/Subject#conceptScheme"/>
  <skos:prefLabel xml:lang="zh">植物</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Plants</skos:prefLabel>
  <skos:altLabel xml:lang="zh-pinyin">zhi wu</skos:prefLabel>
  <skos:narrowerTransitive rdf:resource="http://cct.nlc.gov.cn/Subject/S006361#concept"/>
  <skos:narrowerTransitive rdf:resource="http://cct.nlc.gov.cn/Subject/S012991#concept"/>
 ......
  <skos:related rdf:resource="http://cct.nlc.gov.cn/Subject/S002198#concept"/>
  <skos:related rdf:resource="http://cct.nlc.gov.cn/Subject/S107282#concept"/>
......
 <skos:exactMatch rdf:resource="http://cct.nlc.gov.cn/Classification/C015667#concept"/>
  <skos:notation>Q94</skos:notation>  <!-- temporary use for plain literal of Notation -->
</skos:Concept>
```

2. SKOS Modelling for CCT – *subject-notation issue*

In the subject-classification table of CCT, one subject concept can have one or more corresponding notations.

skos:notation property only shows what the class notation is but does not indicate the specific relationships among these notations.

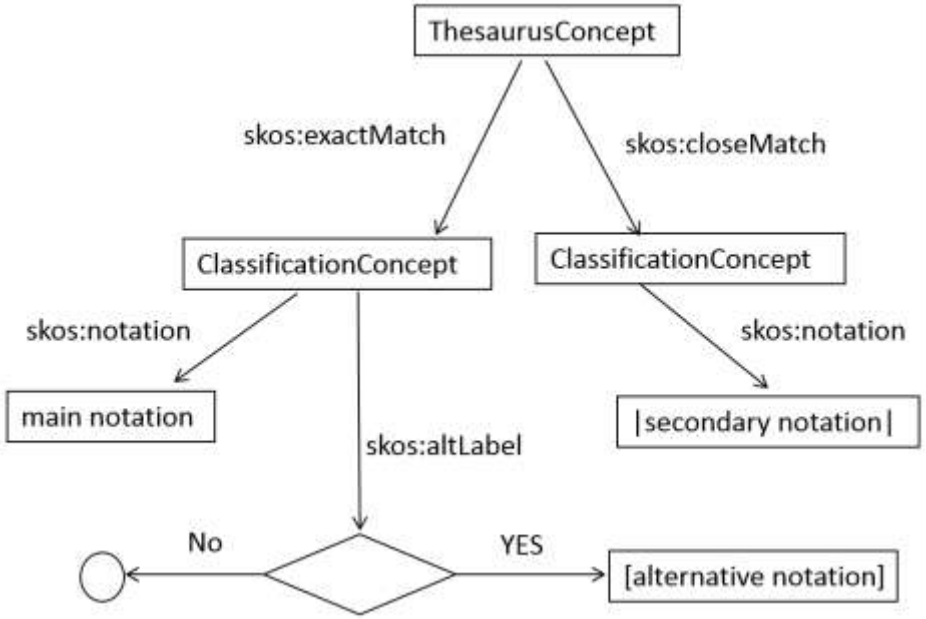
- Main notation which indicates the main discipline aspect of a concept
- Secondary notation which indicates the related aspect of a concept with two “|” marks.
- Alternative notation which is generated from the relationships between CLC classes marked by the symbols “[” and “]”.

海藻 Seaweed Q949.2；|R282.71|；[R931.71]	Q949.2 藻类 R282.71 植物药	[R931.71] 植物药 宜入R282.71

2. SKOS Modelling for CCT – *Mapping with subject-notation*

With this mapping approach, the classification scheme skeleton of CCT is constructed by subject-notation mapping. Since the classification part of CCT is derived from CLC, 22 main classes were taken from the major categories of CLC as top concepts. In each main class, a hierarchy can be built by using notations from subject authority data.

- The first two types of notations are subject and class mappings.
- The third types of notations can be automatically derived from the classes and the mappings among them.



Partly generate category browsing interface

<table>
<tr><td colspan="7">Chinese Classified Thesaurus</td></tr>
<tr><td></td><td colspan="2">Subject Authority Data
http://cct.nlc.gov.cn/Subject#conceptScheme</td><td colspan="4">Classification Skeleton</td></tr>
<tr><td>Identifier
(URI)</td><td colspan="2">http://cct.nlc.gov.cn/Subject/Sxxxxxx
(Control Number)</td><td colspan="4">http://cct.nlc.gov.cn/Classification/Cxxxxxx
(Control Number)</td></tr>
<tr><td></td><td>D(代)
Y(用)</td><td>skos:altLabel
(Plain literals)</td><td></td><td colspan="2"></td></tr>
<tr><td></td><td>S(属)
F(分)</td><td>skos:broaderTransitive
skos:narrowerTransitive</td><td></td><td colspan="2"></td></tr>
<tr><td></td><td>C(参)</td><td>skos:related</td><td></td><td colspan="2"></td></tr>
<tr><td></td><td>Z(族)</td><td>skos:topConceptOf
skos:hasTopConcept</td><td></td><td colspan="2"></td></tr>
<tr><td rowspan="3"></td><td rowspan="3">Notation</td><td rowspan="3">skos:notation
(Plain literals)</td><td rowspan="3">Subject Notation Mapping</td><td>main notation</td><td>skos:exactMatch</td></tr>
<tr><td>secondary notation</td><td>skos:closeMatch</td></tr>
<tr><td>alternative notation</td><td>skos:altLabel</td></tr>
<tr><td></td><td>Collection</td><td>skos:Collection</td><td></td><td></td><td></td></tr>
<tr><td>Identifier
(URI)</td><td colspan="2">http://cct.nlc.gov.cn/XXXX#OrderedCollection
(Personal names, Corporate names, Geographic names, Title names and etc.)</td><td></td><td></td><td></td></tr>
</table>

3. Semantic Visualization – *Related tools*

Existed visualization approaches are not entirely suitable for controlled vocabularies for two reasons.

•OWL visualization tools are designed for ontologies without consideration for the requirements of thesauri and classification schemes

•closely related to specific tools and some visualization are generated in local environments.

3. Semantic Visualization – *Loosely Coupled Strategy*

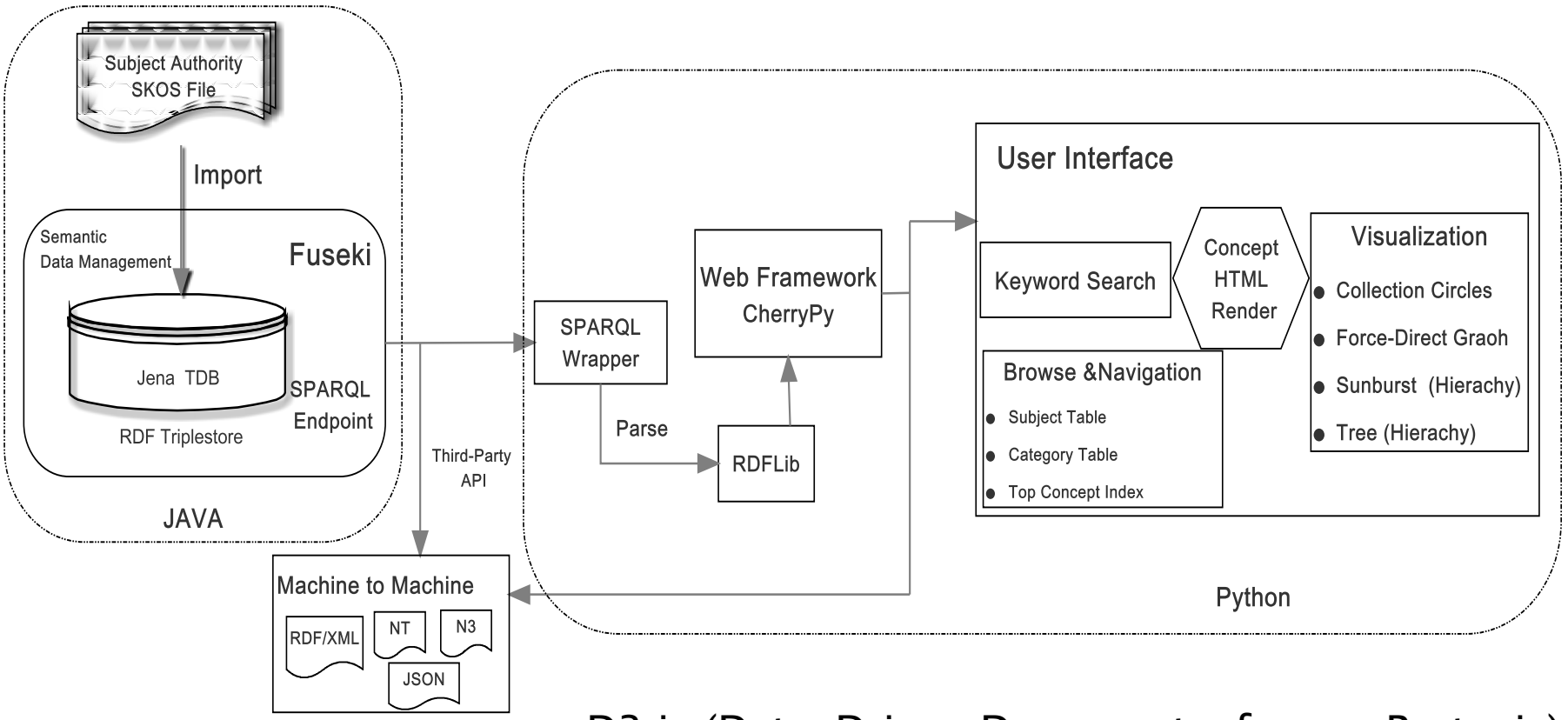
From the perspective of terminology web service, data and their representation are loosely coupled.

Browser/Server (B/S) model with two major advantages:

- no specific tool requires installation;
- users could take any modern web browser to explore KOS in an interactive manner.

web related visualization technology was selected not only for visualizing SKOS data, but also for supporting web access.

3. Semantic Visualization – *Technology Architecture*



D3.js (Data-Driven Documents, former Protovis)

Visualization

FrontPage of CCT

Visualization

Visualization

Concept Page

Halophytes

盐土植物

yan tu zhi wu

其他词汇形式

- xi yan zhi wu
- 喜盐植物

上位类- 传递

- 植物

相关概念

- 耐盐性

类号

- Q949.4

所属词表

- 中国分类主题词表

其他形式

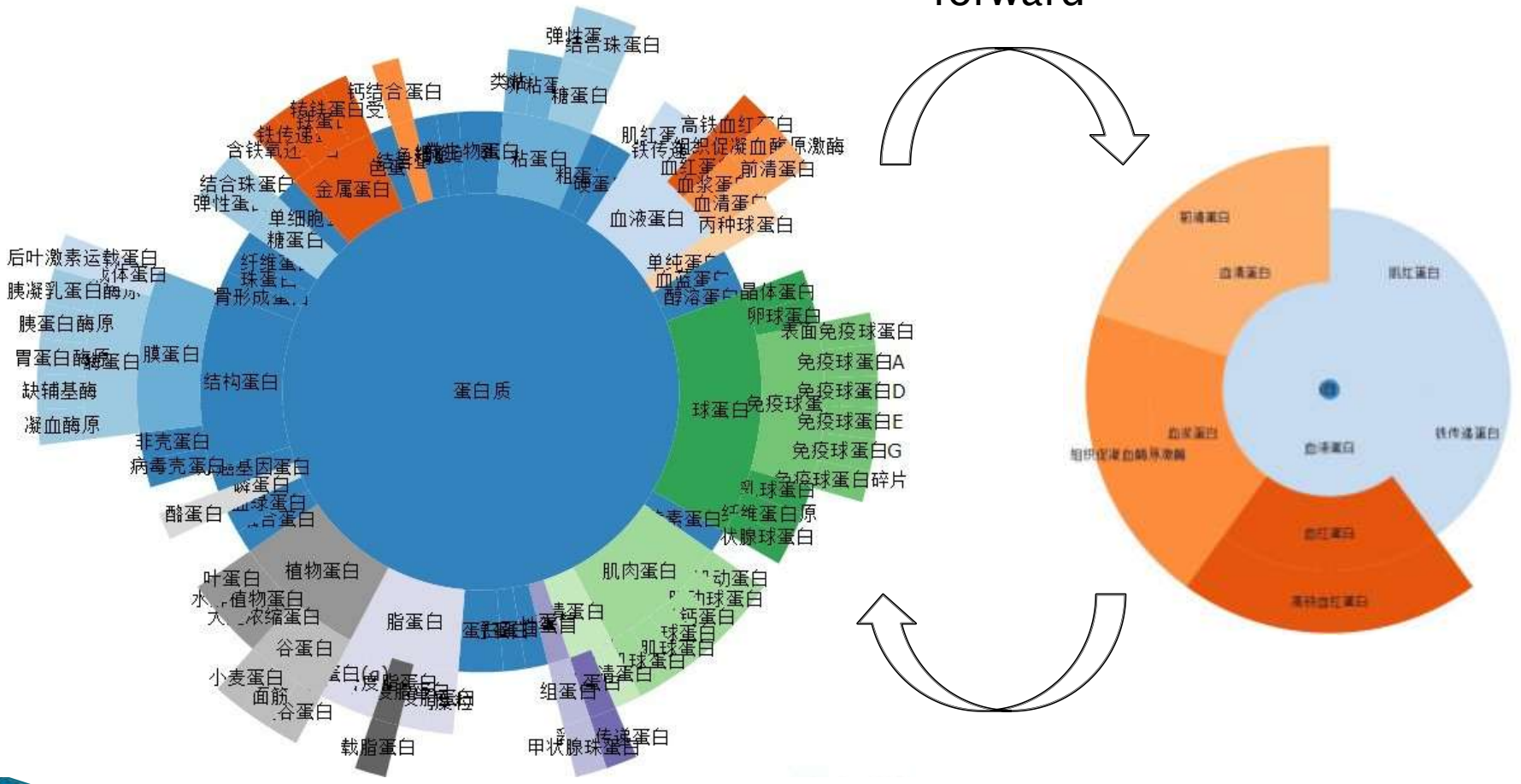
- RDF/XML
- N3
- NT
- Dot

- Purple: centre node with preferred labels.
- Green: alternative labels.
- Yellow: class notation(s).
- Blue: direct broader concept(s).
- Red: related concept(s).

Visualization Sunburst

Hierarchy

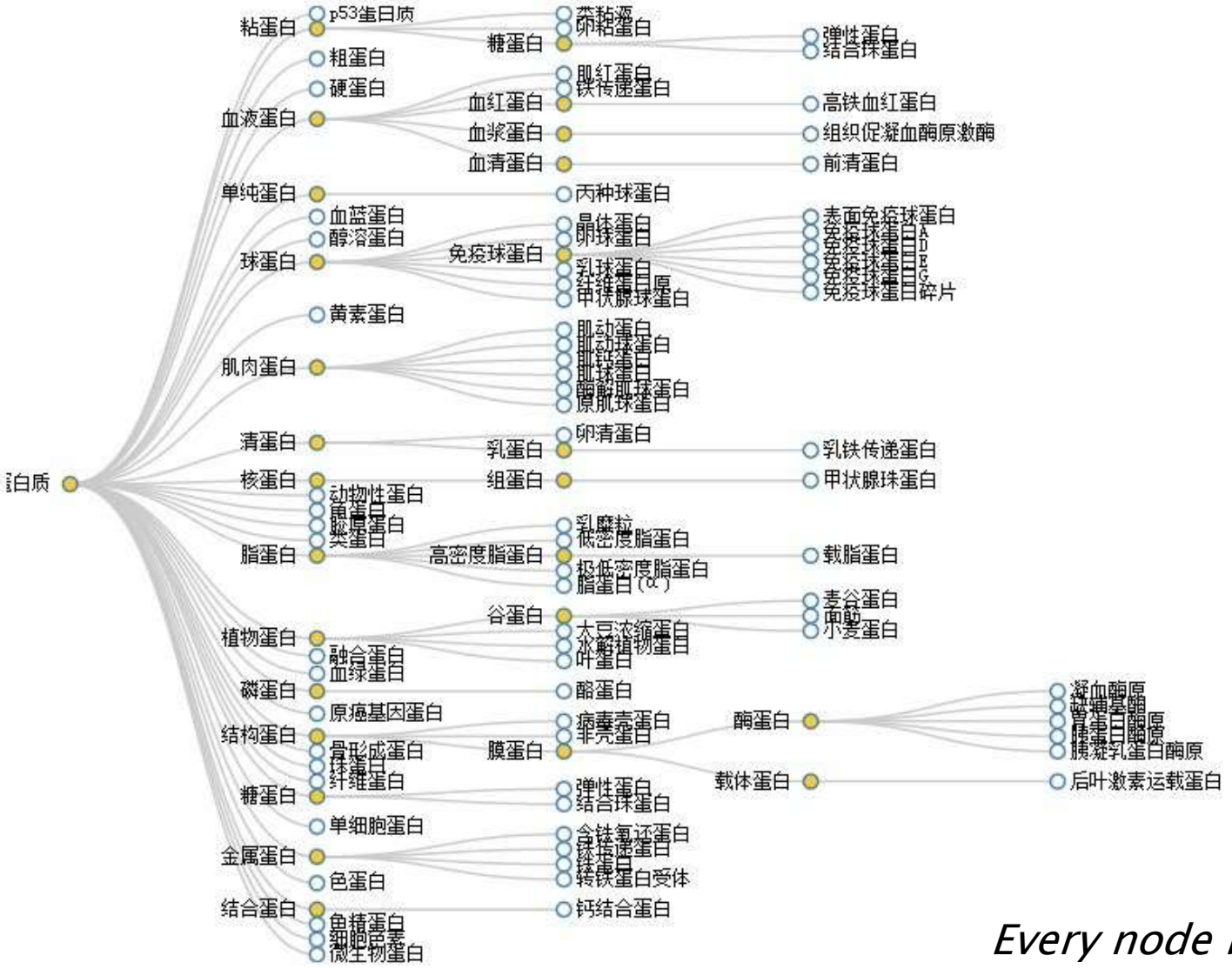
forward



backward

Visualization Tree

Hierarchy



Every node is clickable

Visualization *Subject A-Z Index*

主题浏览表 - a: 共计 2038

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z others

00001 A-D转换器
00002 A调单簧管
00003 A-稳定
00004 A型花岗岩
00005 A型特殊星
00006 A-B-O血型
00007 A.G.S.工业公司
00008 A.S.吕士那克滑石公司(法国)
00009 AACR2
00010 AACR
00011 AACR2R
00012 AAD法
00013 AAJ
00014 ABC分析法
00015 ABJ反常

Visualization *Top Concept A-Z Index*

族首词索引 - a: 共计 42

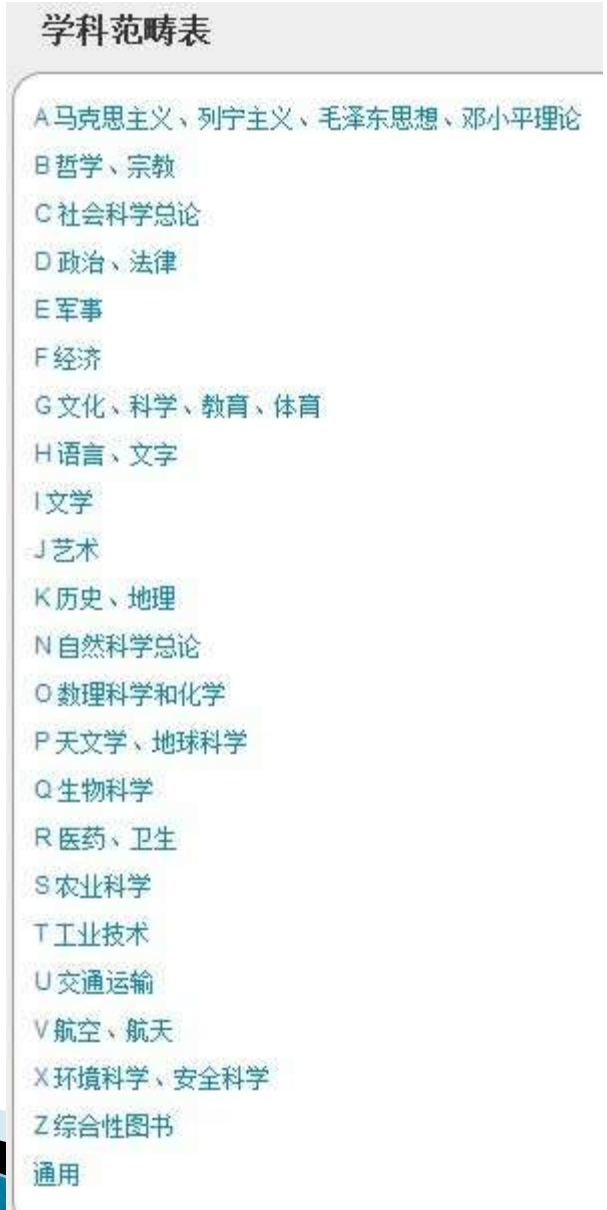
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z others

00001	阿尔巴尼亚	Albania
00002	阿尔衮琴人	Algonkins
00003	阿尔及利亚	Algeria
00004	阿尔泰语系	Altaic family
00005	阿富汗	Afghanistan
00006	阿根廷	Argentian
00007	阿拉伯联合酋长国	United Arab Emirates
00008	阿拉伯人	Arabs
00009	阿拉伯石油输出国组织	Arab Organization of Petroleum Exporting Countries
00010	阿曼	Oman
00011	阿塞拜疆	Azerbaijan
00012	锕系化合物	Actinide series compounds
00013	爱尔兰	Ireland
00014	爱国卫生运动	Patriotic health movement
00015	埃及	Egypt
00016	艾美尔属	Eimeria
00017	埃塞俄比亚	Ethiopia

Visualization

CLC Main Class

Category table



General Auxiliary

Conclusion *Next Steps*

- The class notation issue may be more complicated and needs to be further explored.
- Inner mapping visualization of classification scheme from current subject notation.
- Cross mapping visualization with other vocabularies, such as UDC and DDC which have already published vocabulary data sets. - Interoperability

Conclusion *To Future*

- A starting point for exposing and sharing CCT.
- Re-engineering CCT represents a shift from traditional vocabulary editing and the displaying of patterns to broader data-intensive and technology-driven developments

From an isolated KOS tool to a Chinese vocabulary hub in the open linked data environment.

Acknowledge

- Collaboration with The Editorial Office of the Chinese Library Classification
- Supported by State Commission of Science Technology of China (Grant No. 2009FY220400)

National Library of China

Thanks

Q & A