Building semantic knowledge organization systems for interdisciplinary research

-----An example of wetland remote sensing

Asian NKOS workshop, Seoul, Dec, 9th, 2015
Outline

1. Introduction: background, aim
2. Experimental design: workflow
3. Results and discussions
4. Conclusions and future work
1. Introduction

- **Background**

  **Interdisciplinary and emerging field:**
  - brain Science
  - quantum communications
  - wetland

  How to use
1. Introduction

● **Aims**

develop the methods and workflows in building semantic knowledge organization system (KOS), in support of interdisciplinary scientific research

● **Experimental Field**

  – Wetland
    • remote sensing
    • Biosphere
    • Atmosphere
    • Lithosphere
    • …
2. Workflow

- Documents
- Doc processing
- Terms & frequency
- Mapping to STKOS concepts
- STKOS metathesaurus
- KOS
  - Model Construction
  - Selecting concept
  - Analyze semantic type & relations
  - Indexing
- Query expansion
- List of concepts, relations
- Search and ranking
2. Workflow

● Data source

  – Terms from current available KOS
    • STKOS metathesaurus: 199 KOS (150 thesauri, 3 classifications, 37 glossaries, etc.)
    • Remote sensing lexicon, Science publishing, 1990

  – Terms extracted from literature
    • Web of Science Database(1900-2014), Topic field = “remote sensing & wetland”, bibliographic records

● Methods

  – Semantic matching
  – Domain experts participation
2. Workflow

Knowledge collecting

- **Pre-processing / cleaning:** Thomson Data Analyzer (TDA)
  - Stop word removal, analysis, clustering

- **Semantic match to STKOS:**
  - Normalization, phrase analysis, word sense analysis, structure analysis
  - term list

- **term extraction**
  - Terms selected over a specific threshold: term frequency (TF)
  - \[ \text{Threshold}_{\text{term}} = \text{mean (TF + standard deviation TF)}, \text{where high TF, medium TF, low TF} \]
2. workflow

- **domain expert participation**
  - Reconstruct and reuse model, *Remote sensing lexicon*
  - Remove general terms, such as “classification, case study”
  - Select compound term
  - Category, Seed term, **doubly-anchored patterns**
2. Workflow

- **Evaluation**
  - Annotation and index by human and computer
  - Precision, recall, how correct is the KOS learned
3. Results and discussions

- Match STKOS result

<table>
<thead>
<tr>
<th>frequency</th>
<th>term</th>
<th>Match STKOS</th>
<th>percent</th>
<th>Match + Standard deviation</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4271</td>
<td>707</td>
<td>14.98%</td>
<td>707</td>
<td>14.98%</td>
</tr>
<tr>
<td>2</td>
<td>548</td>
<td>178</td>
<td>32.48%</td>
<td>178</td>
<td>32.48%</td>
</tr>
<tr>
<td>3</td>
<td>196</td>
<td>85</td>
<td>43.37%</td>
<td>87</td>
<td>44.39%</td>
</tr>
<tr>
<td>4</td>
<td>106</td>
<td>57</td>
<td>53.77%</td>
<td>58</td>
<td>54.72%</td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>41</td>
<td>63.07%</td>
<td>43</td>
<td>66.15%</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>21</td>
<td>52.5%</td>
<td>22</td>
<td>55%</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>19</td>
<td>67.86%</td>
<td>19</td>
<td>67.86%</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>17</td>
<td>68%</td>
<td>18</td>
<td>72%</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
<td>14</td>
<td>73.68%</td>
<td>14</td>
<td>73.68%</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>5</td>
<td>45.45%</td>
<td>5</td>
<td>45.45%</td>
</tr>
<tr>
<td>11-20</td>
<td>47</td>
<td>30</td>
<td>63.83%</td>
<td>31</td>
<td>65.96%</td>
</tr>
<tr>
<td>21-30</td>
<td>9</td>
<td>9</td>
<td>100%</td>
<td>9</td>
<td>100%</td>
</tr>
<tr>
<td>31-40</td>
<td>8</td>
<td>7</td>
<td>87.5%</td>
<td>7</td>
<td>87.5%</td>
</tr>
<tr>
<td>41-50</td>
<td>6</td>
<td>5</td>
<td>83.3%</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>5</td>
<td>5</td>
<td>100%</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>total</td>
<td>5834</td>
<td>1200</td>
<td>20.57%</td>
<td>1209</td>
<td>20.72%</td>
</tr>
</tbody>
</table>
3. Results and discussions

- **Source of STKOS matching result**

![Source of matching STKOS](image)
3. Results and discussions

KOS model of Wetland remote sensing

- **General**
  - RS classification
  - Resolution

- **RS base**
  - Physics
  - Math
  - Geoscience
  - Information science

- **RS technology**
  - Remote Sensor
  - platform
  - ground station
  - image processing
  - Photography
  - Interpretation

- **Wetland RS application**
3 Results and discussions

- **Semantic KOS of Wetland RS**
  - Classification: 4 primary, 12 secondary
  - Concept: 209
  - Term: 409
  - Relationship: hierarchical, associative, equivalence
  - Semantic type and relation: equipment, product, concept?
  
Relation: Equipment A is a kind of B
3. Results and discussions

Parent Term
- satellite

Preferred Term
  Environmental Satellite
Non-preferred Term
  Earth observation satellite
  ENVISAT
Chinese: 环境卫星
Semantic Type: Equipment

Children Terms
- IKONOS
- QuickBird
- RapidEye
- SPOT
- WorldView
- GeoEye
- Landsat satellites

Semantic relationships

- Equipment
  • Sensors
  • Thematic mapper
  • Imaging spectrometer

- Product
  • Image
  • Data

- Application
  • Land hydrology
  • Ecological community
  • Crop
3. Results and discussions

Evaluation

Index term, index term number, index semantic type
- Does index term represent the content in wetland RS
- Need to add some terms
- Is the article belong to wetland RS?

<table>
<thead>
<tr>
<th>Index term Number</th>
<th>Article Number</th>
<th>Evaluation by domain experts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td>0</td>
<td>81</td>
<td>31</td>
</tr>
<tr>
<td>1</td>
<td>102</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>110</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>135</td>
<td>126</td>
</tr>
<tr>
<td>5</td>
<td>126</td>
<td>124</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>148</td>
</tr>
<tr>
<td>&gt;7</td>
<td>765</td>
<td>765</td>
</tr>
</tbody>
</table>
4. Conclusions and future work

- **Conclusions**
  - Methods of interdisciplinary KOS construction
  - Cover the content of wetland remote sensing
  - Use in wetland DB
  - Improve retrieval efficiency and discover

- **Future work**
  - Semantic type revise
  - continue in other field of wetland
Our Team

Liu Zheng\textsuperscript{1}, Sun Tan\textsuperscript{2}, Sheng Chunlei\textsuperscript{3}, Liu Xiumin\textsuperscript{1}

1 National Science Library, Chinese Academy of Sciences
2 Agricultural information Institute, Chinese Academy of Agricultural Sciences
3 Northeast institute of Geography and Agroecology, Chinese Academy of Sciences

Wetland experts

Lv Xiangguo, Song Kaishan, Wang Zongmin

Key laboratory of wetland ecology and environment, Northeast institute of Geography and Agroecology, CAS
Thanks

Any questions?