EnTag
Enhanced Tagging for Discovery

Koraljka Golub, Catherine Jones, Marianne Lykke Nielsen, Brian Matthews, Jim Moon, Bartłomiej Puzoń, Douglas Tudhope

JISC Repository Search Infrastructure Projects Meeting, 20 March 2009
Project context

• Partners
  – UKOLN
  – University of Glamorgan
  – STFC (former CCLRC)
  – Intute

  – Non-funded
    • OCLC Office of Research, USA
    • Danish Royal School of Library and Information Science

• Period: 1 Sep 2007 -- 31 Oct 2008
Background

- Controlled vocabularies help improve information retrieval and discovery
  - But, costly to index with, especially the amount of digital documents

- Social tagging holds the promise of reducing indexing costs
  - But, tags entirely uncontrolled
  - Still, may identify new trends and hot concepts
Purpose

- Investigate the combination of controlled and folksonomy approaches to support resource discovery in repositories and digital collections

- Aim
  Investigate whether use of an established controlled vocabulary can help improve social tagging for better resource discovery
Overall approach

• Main focus:
  free tagging with no instructions
  versus
  tagging using a combined system and guidance for users

• Two demonstrators
  – Intute digital collection http://www.intute.ac.uk demos
    • Major development
    • Tagging by reader
    • DDC
  – STFC repository http://epubs.cclrc.ac.uk/
    • Complementary development
    • Tagging by author
    • ACM Computing Classification Scheme
### Intute demonstrator: simple tagging

**EnTag Project ST**

<table>
<thead>
<tr>
<th>Taggers</th>
<th>All Alex's Tags</th>
<th>My Tags For This Document</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anniv878, Basewell, Cicinho, Chikern, Clare, Dij</td>
<td>journal, academic paper, achieve of european integration, afghanistan, Alphabetical Listing, annual report, Archive of European Integration, archives, ARENA</td>
<td></td>
</tr>
<tr>
<td>Donn1l, Emoff, Evis Gautam, Gemma, Gop131, Gix365, Jay, John, Katy, Kihns, Laurad67, Liam</td>
<td>Logic, Louise, Matt, Maudry, Max, Megan11, Meng, Mike, Mikes, Most, Noel, Palermo, Pemai, Pogono, Randvar, Ricco, Rikann</td>
<td></td>
</tr>
<tr>
<td>Shannon, Squelch, Tato, The75, Tom, Ur1981</td>
<td>archives, ATTITUDES TO EUROPEAN INTEGRATION, bibliography, blog, Book, Review, briefing</td>
<td></td>
</tr>
</tbody>
</table>

**Document Details**

**Title:** 50th anniversary of the 1956 Hungarian revolution: Canadian Embassy  
**Description:** This site has been created by the Canadian Embassy in Hungary to commemorate the 50th anniversary of the 1956 Hungarian revolution. In addition to a listing of events, it also provides free access to some interesting full text publications about Canadian involvement in the events. These include booklets about Canadian diplomacy. There are also links to other websites from Canadian organisations which include eye-witness accounts.
Intute demonstrator: enhanced tagging

EnTag Project ST

Current Location: Document Tagging Page

Current User:

Taggers:
- Adamecok, Adziya, Alex, Alicho, Alonso
- Amy878, Boswell, Cebich, Ckern, Clare, Dju
- Donnelly, Emoff, Eve, Guinta, Germina, Gip131
- Sla365, Jay, John, Kay, Krish, Laura62, Lian
- Logic, Louise, Matt, Maury, Max, Megan11
- Meng, Mike, Mikel, Most, Noel, Palermo
- Pema, Prason, Ronald, Ryan, Ekanm
- Shannon, Sapers, Tane, Tine, Tom, Una81

All Alex's Tags:
- Academic Papers
- advocacy
- Afghan War, 2001
- Armed Forces - Political activity
- Arms control
- Arms control - Europe
- art
- Asia for Educators
- bibliography
- British foreign policy
- Buddhism
- China - Tibet
- buffer zone
- Canada
- Carnegie Endowment
- casualties
- figures
- cease fire
- China - Politics and

My Tags For This Document:
- Disarmament and arms control
- human rights
- podcast
- Sino-American Summit
- Sino-US specialists
- Taiwan - 20th century
- Tiananmen Square
- United States - Foreign relations - China

Document Details:

Title: 1998 Sino-Soviet Summit by Robert S. Ross
URL: http://www.asiasociety.org/publications/sino_soviet_summit.html

Description: This site contains the full-text of the paper by Robert Ross which was originally published by the Asia Society, a non-profit educational institution, in June 1998. It examines US-China relations in the period after the Tiananmen Square massacre in 1989, focusing on such issues as human rights, international trade, nuclear weapons and arms control. Consideration is then turned to what the 1998 Sino-Soviet Summit hopes to achieve in these areas. An additional feature of the article is an appendix listing contact addresses for specialists on US-China relations.

Automatically suggested matches:
- Soviet Union - Foreign relations
- African American women legislators-biography
- Eastern Europe-political...
- Communist Party of China
- Relations of federal to state, regional, provincial &
- United States foreign relations - Eastern Europe...
- End of Political Science
- 1945-1998

Explore hierarchy around the selected context:
- International relations
- Soviet Union - Foreign relations

Selected/edit relevant tags:
- Eastern Europe - foreign relations
- Europe
- Eastern - Foreign relations
- Soviet Union
- Europe, Eastern - Politics and government
- 1945-1989
- Foreign relations - Eastern Europe
- Foreign relations - Russia
- Foreign relations - Russia - Foreign relations
- 1894-1917
- Russia (Federation)
- Foreign relations - Russia (Federation)
- Foreign relations - Russia (Federation)
Grey literature, institutional repositories and the organisational context

Abstract: A wide variety of grey literature is produced during the work carried out at a large multidisciplinary scientific research organisation. This paper examines how the grey literature may be managed and the benefits that result. Trends in Technical Reports, which have always been an important medium for transmission of knowledge, are examined and the use of an institutional repository is advocated for the future of the medium. Other kinds of grey literature produced in research projects are also described, and again the value of repositories is stressed.

Current thesaurus path:
TOP

Top terms:
- computer applications
- computer systems organization
- computing methodologies
- computing malware
- data
- general literature
- hardware
- information systems
- macromedia application

Add:
Search thesaurus

freight

-controlled tags
- strategic information systems
Intute demonstrator user study
Objectives

- Test: comparison of simple and enhanced systems
  - Indexing behaviour
    - Use (tags selected, clouds consulted, order of consultation, source of tag)
    - Consistency and quality analysis of tagging
  - Retrieval
    - Degree of match between user and system terminology
Pre-study questionnaire

• 28 participants
  – Equal distribution of gender
  – Majority solid subject experience
  – Majority experienced Web users
  – Majority without Intute use
  – Half with tagging experience before but little tagging
  – A third familiar with DDC
Task example

Task 1: Simple Tagger, “European integration”

Imagine that as part of one of your courses, you are asked to write a four-page essay on the **topic of European integration**, as a joint project in groups of four. The essay should critically discuss existing theories about the creation of the European Union and its institutions. Your lecturer has instructed you to look for resources in the EnTag system. Since you will be working together with three other students, you should tag the documents you retrieve with tags that would be useful to you but would also enable other students to find those documents in EnTag and understand from your tags what the documents are about.

Go to the EnTag login page*, choose **Task 1** and **Simple Tagger Log In** and in the "Search for Documents" box enter these words: **European integration**.

Then, tag the **first 15 retrieved documents**. Do only the ones you can open - if a URL is unavailable move on to the next document in the Results.

Tagging each document should on average take between **5 and 10 minutes**. Please describe as many aspects and topics as you think appropriate for the task. Remember to open the URL, but you do not need to follow further internal links within a Web site. If the document is very long, focus on its abstract, introduction, conclusion, headings and table of contents.
Test setting and data collection

• Test setting
  – 28 UK students in political science
  – 60 documents, covering 4 topics of relevance for the students
    • 2 controlled tasks
    • 2 free tasks
    • Rotation
      – Instructions and training documents

• Data collection
  – Logging
  – Pre- and post-questionnaires
<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Enhanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags in total</td>
<td>4022</td>
<td>3546</td>
<td>7568</td>
</tr>
<tr>
<td>Controlled task</td>
<td>2025</td>
<td>1688</td>
<td>3713</td>
</tr>
<tr>
<td>Free task</td>
<td>1997</td>
<td>1858</td>
<td>3855</td>
</tr>
<tr>
<td>Tags per document (controlled)</td>
<td>avg. 49 (41 docs)</td>
<td>avg. 32 (53 docs)</td>
<td>avg. 41 (94 docs)</td>
</tr>
<tr>
<td>Tags per document (free)</td>
<td>avg. 5 (374 docs)</td>
<td>avg. 5 (377 docs)</td>
<td>avg. 5 (751 docs)</td>
</tr>
<tr>
<td>Tags per tagger (controlled)</td>
<td>avg. 72</td>
<td>avg. 63</td>
<td>total avg. 135</td>
</tr>
<tr>
<td>Tags per tagger (free)</td>
<td>avg. 74</td>
<td>avg. 69</td>
<td>total avg. 143</td>
</tr>
</tbody>
</table>
Choosing a tag

<table>
<thead>
<tr>
<th>Activity</th>
<th>Simple</th>
<th>Enhanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typing Own Tag</td>
<td>3656</td>
<td>2525</td>
<td>6181</td>
</tr>
<tr>
<td>Main Tag Cloud</td>
<td>94</td>
<td>88</td>
<td>182</td>
</tr>
<tr>
<td>Own Tag</td>
<td>0</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Certain Tagger’s Tag</td>
<td>272</td>
<td>303</td>
<td>575</td>
</tr>
<tr>
<td>Dewey Tag</td>
<td></td>
<td>598</td>
<td>598</td>
</tr>
<tr>
<td>Total</td>
<td>4022</td>
<td>3546</td>
<td>7568</td>
</tr>
</tbody>
</table>

Activity Simple Enhanced Total
---
Typing Own Tag 3656 90.9% 2525 71.2% 6181 81.7%
Main Tag Cloud 94 2.3% 88 2.5% 182 2.4%
Own Tag 0 0% 32 0.9% 32 0.4%
Certain Tagger’s Tag 272 6.8% 303 8.5% 575 7.6%
Dewey Tag           598 16.9% 598 n/a
Total 4022 100% 3546 100% 7568 100%
## Browsing for tags

<table>
<thead>
<tr>
<th>Activity</th>
<th>Simple</th>
<th>Enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Tag Cloud Clicked</td>
<td>16.8%</td>
<td>5.1%</td>
</tr>
<tr>
<td>Own Tag Clicked</td>
<td>10.3%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Certain Tagger's Name Clicked</td>
<td>18.9%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Certain Tagger's Tag Clicked</td>
<td>54.0%</td>
<td>14.5%</td>
</tr>
<tr>
<td>Dewey Suggest Button Clicked</td>
<td></td>
<td>28.9%</td>
</tr>
<tr>
<td>Dewey TreeView Clicked (1st pane)</td>
<td></td>
<td>13.7%</td>
</tr>
<tr>
<td>Dewey Hierarchy Clicked (2nd pane)</td>
<td></td>
<td>3.2%</td>
</tr>
<tr>
<td>Dewey Suggested Tags Clicked (3rd pane)</td>
<td></td>
<td>28.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>
Number of documents’ records where tags found

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Enhanced</th>
<th>In Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags in Title</td>
<td>200</td>
<td>174</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>48.3%</td>
<td>41.8%</td>
<td>44.9%</td>
</tr>
<tr>
<td>Tags in URL</td>
<td>76</td>
<td>72</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>18.4%</td>
<td>17.3%</td>
<td>17.6%</td>
</tr>
<tr>
<td>Tags in Description</td>
<td>262</td>
<td>232</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td>63.3%</td>
<td>55.8%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Tagged Documents in</td>
<td>414</td>
<td>416</td>
<td>780</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Post-task questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity with tasks</td>
<td>majorit familiarity or very familiar</td>
<td></td>
</tr>
<tr>
<td>Easy to choose tags</td>
<td>majorit easy or very easy</td>
<td></td>
</tr>
<tr>
<td>Satisfaction with tags assigned</td>
<td>majorit satisfied or very satisfied</td>
<td></td>
</tr>
<tr>
<td>Certainty that tags assigned correctly</td>
<td>majorit certain or very certain</td>
<td></td>
</tr>
<tr>
<td>Main Tag Cloud</td>
<td></td>
<td>helpful to half, unhelpful to half</td>
</tr>
<tr>
<td>Clickable Names of Others</td>
<td>helpful to half, unhelpful to half</td>
<td>unhelpful to majority</td>
</tr>
<tr>
<td>Listing of Own Tags</td>
<td>helpful to majority</td>
<td></td>
</tr>
<tr>
<td>Dewey Tree Disambiguation</td>
<td>n/a</td>
<td>helpful to majority</td>
</tr>
<tr>
<td>Dewey Hierarchy</td>
<td>n/a</td>
<td>helpful to half, unhelpful to half</td>
</tr>
<tr>
<td>Dewey/LCSH suggestions</td>
<td>n/a</td>
<td>helpful to majority</td>
</tr>
</tbody>
</table>
Post-study questionnaire

- Majority enjoyed the study
- Majority thought it extremely or very easy to learn and use Simple Tagger
- Majority thought it somewhat or very easy to learn and use Enhanced Tagger
- Majority think a similar system would be useful in real life
Conclusions
Outcomes

- Two demonstrators (tagger as author, tagger as reader)
- Two user studies and recommendations
  - Potential for KOS-based tagging suggestions, various lessons

Overlaps STAR project and W3C SKOS standards effort
- SKOS-based services and browser widgets

- Outputs
  ALISS08, NKOS DC08 Session,
  ISKO-UK09, IFLA-preConf09, JCDL09
Recommendations (broad)

- Both simple and enhanced tagging provide additional entry points. Some evidence KOS-based suggestions provided access points beyond literal text.

- Both Simple and Enhanced Demonstrators usable with little prior training. However (particularly) Enhanced interface experienced as complex.

- Global tag cloud little used and had mixed response in questionnaires.

- Enhanced suggestions sometimes useful but sometimes wide of the mark and not helpful. Evidence of support for auto-suggestions, provided high quality and user oriented.

- While users appreciated the ‘direct’ suggestions and made some use of the disambiguation interface element, they did not browse the hierarchy much. Further work explore when/how browsing useful (tagging behaviour studies) – needs to be more integrated with suggestions.
Recommendations

- Filter/process automatic suggestions, more selective, more user-oriented.
- Test interface cluttered by design. Operational interface needs to be less cluttered with interaction streamlined.
- Auto-completion part of streamlined interface and browsing options more integrated.
- Unclear whether displaying all global tags useful for retrieval purposes. Consider Document and Community-of-use tag clouds (or filters) as well as visualisation/personalisation options.
- Integrate KOS with other bases for automatic suggestions (e.g., auto-classification, IE, co-occurrence).
- Develop provision for other vocabularies, develop import facility and underlying vocabulary web services.
- Develop SKOS based tagging service (building on STAR project and W3C SKOS work).
Further research

- Qualitative analysis of user tags (including topic facets)
- Further investigate user tagging styles and motivation for tagging
- Retrieval test with performance measures
- Extended (longitudinal) evaluation in live settings
- Investigate different KOS and subject domains
- Potential to augment KOS entry vocabularies from user tags
- Retrospective system ‘improvement’ of user tags (consistency)
- Possibility to integrate non-subject (non-topical) tags might prove useful for general retrieval purposes, in some contexts. E.g., genre or utility of a document

http://www.ukoln.ac.uk/projects/enhanced-tagging/

www.ukoln.ac.uk
EnTag: More info

- Live demonstrator and final reports at

http://www.ukoln.ac.uk/projects/enhanced-tagging/