## EnTag Enhanced Tagging for Discovery

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

JISC Repository Search Infrastructure Projects Meeting, 20 March 2009





**UKOLN** is supported by:







BATH

www.bath.ac.uk





- 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









- 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 <a href="http://www.intute.ac.uk">http://www.intute.ac.uk</a> demos
    - Major development
    - Tagging by reader
    - DDC
  - STFC repository <a href="http://epubs.cclrc.ac.uk/">http://epubs.cclrc.ac.uk/</a>
    - Complementary development
    - Tagging by author
    - ACM Computing Classification Scheme



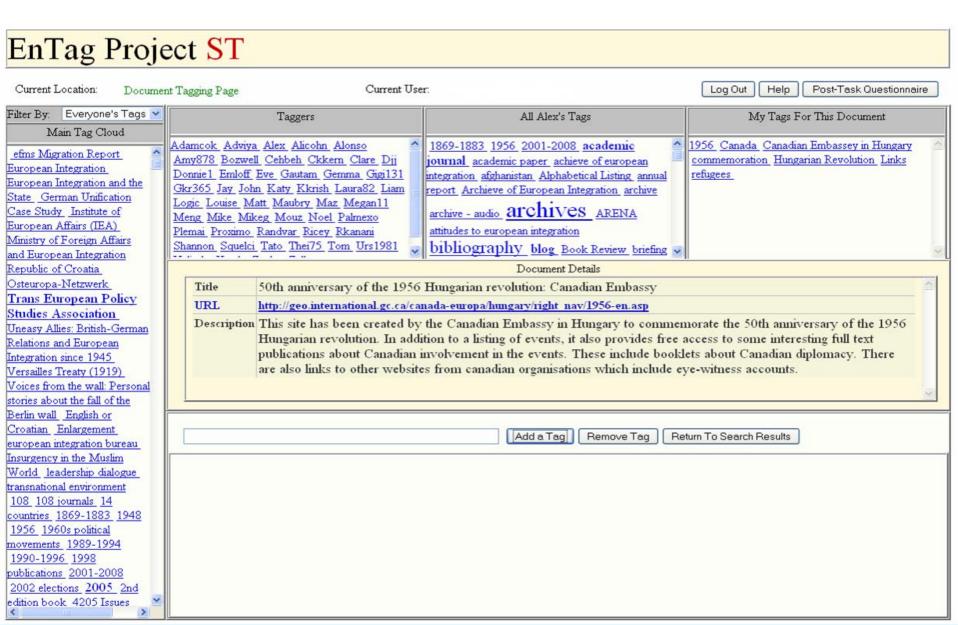


### Intute demonstrator: searching

#### EnTag Project ST Current Location: Current User Loa Out Help Post-Task Questionnaire Document Search and Browsing Page Everyone's Tags V Filter By: Taggers Search For Documents Search In Main Tag Cloud Tags Adamcok Adviya Alex Alicohn Alonso ✓ Titles efms Migration Report Enter your search terms Amy878 Bozwell Cehbeh Ckkern Clare European Integration ✓ Descriptions Dij Donnie1 Emloff Eve Gautam Gemma European Integration and the Gigi131 Gkr365 Jay John Katy Kkrish State German Unification Laura82 Liam Logic Louise Matt Maubry Start Search Case Study Institute of Maz Megan11 Meng Mike Mikeg Mouz Displaying Documents Tagged by: Alex European Affairs (IEA) Noel Palmexo Plemai Proximo Randvar Ministry of Foreign Affairs Ricey Rkanani Shannon Squelci Tato and European Integration Republic of Croatia Results Osteuropa-Netzwerk Add Title (Click to Open) Description Tags Trans European Policy Tags Studies Association 50th anniversary of the 1956 This site has been created by the Canadian 1956, Canada, Canadian Embassey in Hungary, commemoration, Uneasy Allies: British-German Hungarian revolution: Canadian Embassy in Hungary to commemorate the Tag Hungarian Revolution, Links, refugees Relations and European Embassy 50th anniversary... Integration since 1945 The Canadian Defence and Foreign Affairs archives, Canada, conferences, courses, defence, foreign affairs, Canadian Defence and Foreign Versailles Treaty (1919) Institute (CDFAI)is an independent journalism, media, newsletter, obama, polls, Relations - US -Tag Affairs Institute Voices from the wall: Personal research institute wh Canada, research, submissions, turkey stories about the fall of the Canadian Dimension is a left wing Canadian Berlin wall English or academic journal, blog, Canada, Canadian Dimension, left-wing Tag journal that examines world events from a Canadian Dimension Croatian Enlargement persepective, NDP - New Democratic Party of Canada Socialist pe.. european integration bureau This site offers access to a searchable Insurgency in the Muslim Canadian Illustrated News 1869-1883, archives, Canada, Canadian Illustrated News, database of almost 4,000 images and Tag World leadership dialogue 1869-1883 Library and Archives Canada cartoons taken from t... transnational environment This site provides access to the full-text of archives, Canada, Canadian Pamphlets and Broadsides Online 108 108 journals 14 Canadian Pamphlets and over 500 pre-1930s Canadian pamphlets Collection, database, pre-1930, Thomas Fisher Rare Book Tag countries 1869-1883 1948 Broadsides and broadside Library, University of Toronto 1956 1960s political Canadian Parliamentary movements 1989-1994 Canadian Parliamentary Review publishes academic journal, Book Review, Canada, Canadian Parliamentary Review: Revue Parlementaire information on Canadian legislation and the Review, Canadian Politics - Domestic, Commonwealth Taq 1990-1996 1998 Parliamentary Association, Legislative Reports publications 2001-2008 Canadienne activities of ... 2002 elections 2005 2nd 12345

edition book 4205 Issues

## Intute demonstrator: simple tagging



## Intute demonstrator: enhanced tagging

#### EnTag Project ST







Logged user: cmg45 Active thesaurus: acm Perspective: Tag						
TAG CLOUD		Grey literature, institutional repositories and 🖺	FREE TAGS			
ePubs physical sciences and engineering repositories strategic information systems planning research outputs	(3)	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	Business Process Data Curation Grey Literature Institutional Repository Metadata Repositories Research Lifecycle Technical reports			
Business Process communication Data Curation earth and atmospheric sciences Grey Literature Institutional Repository Metadata Repositories repostories Research Lifecycle	(1) (1) (1) (1) (1) (1) (1)	literature produced in research projects are also described, and again the  Current thesaurus path:  TOP  Top terms:	Add:  CONTROLLED TAGS  Strategic information systems			
Show all	(1)	# general literature # hardware # information systems # carm V OK  Search thesauri				



www.ukoln.ac.uk

n.ac.uk www.bath.ac.uk



## Intute demonstrator user study









- 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









- 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

# PREFY CO.

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







PREP VOLV	1000

	Simple	Enhanced	Total
Tags in total	4022	3546	7568
Controlled task	2025	1688	3713
Free task	1997	1858	3855
Tags per document (controlled)	avg. 49 (41 docs)	avg. 32 (53 docs)	avg. 41 (94 docs)
Tags per document (free)	avg. 5 (374 docs)	avg. 5 (377 docs)	avg. 5 (751 docs)
Tags per tagger (controlled)	avg. 72	avg. 63	total avg. 135
Tags per tagger (free)	avg. 74	avg. 69	total avg. 143









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%







Activity	Simple	Enhanced
Main Tag Cloud Clicked	16.8%	5.1%
Own Tag Clicked	10.3%	2.7%
Certain Tagger's Name Clicked	18.9%	3.7%
Certain Tagger's Tag Clicked	54.0%	14.5%
Dewey Suggest Button Clicked		28.9%
Dewey TreeView Clicked (1st pane)		13.7%
Dewey Hierarchy Clicked (2nd pane)		3.2%
Dewey Suggested Tags Clicked (3rd pane)		28.2%
	100.0%	100.0%











	Simple		Enhanced		In Both	
Tags in Title	200	48.3%	174	41.8%	350	44.9%
Tags in URL	76	18.4%	72	17.3%	137	17.6%
Tags in Description	262	63.3%	232	55.8%	468	60.0%
Tagged Documents in Total	414		416		780	







## Post-task questionnaire

	Simple	Enhanced		
Familiarity with tasks	majority familiar or very familiar			
Easy to choose tags	majority easy or very easy			
Satisfaction with tags assigned	majority satisfied or very satisfied			
Certaity that tags assigned correctly	majority certain or very certain			
Main Tag Cloud	helpful to half, unhelpful to half			
Clickable Names of Others	helpful to half, unhelpful to half	unhelpful to majority		
Listing of Own Tags	helpful to majority			
Dewey Tree Disambiguation	n/a	helpful to majority		
Dewey Hierarchy	n/a	helpful to half, unhelpful to half		
Dewey/LCSH suggestions	n/a	helpful to majority		







- 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









- 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









www.bath.ac.uk

- 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.

www.ukoln.ac.uk





- 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/personalistion 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).

BATH www.bath.ac.uk

www.ukoln.ac.uk





- 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/







Live demonstrator and final reports at

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



