Enhancing social tagging with a knowledge organization system

Brian Matthews, K. Golub, C. Jones, J. Moon, M. L. Nielsen, B. Puzoń, D. Tudhope
EnTag: Enhancing Social Tagging for Discovery

K. Golub, C. Jones, B. Matthews, J. Moon, M. L. Nielsen, B. Puzoń, D. Tudhope

JCDL 2009, Austin, Texas
EnTag Project

Enhanced tagging for discovery

• JISC funded project
• Partners
  - UKOLN – Koralyka Golub
  - University of Glamorgan – Doug Tudhope, Jim Moon
  - STFC – Brian Matthews, Cathy Jones, Bartek Puzon
  - Intute
  - Non-funded
    • OCLC Office of Research, USA
    • Danish Royal School of Library and Information Science - Marianne Lykke Nielsen

Period: 1 Sep 2007 -- 30 Sep 2008

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

- UK JISC
- One year
- Six institutional partners

- [http://www.ukoln.ac.uk/projects/enhanced-tagging/](http://www.ukoln.ac.uk/projects/enhanced-tagging/)
EnTag Background

• Controlled vocabularies
  - Improve information retrieval and discovery
  - But, costly to index with, especially the amount of digital documents
  - Require subject and classification experts

• Social tagging
  - Holds the promise of reducing indexing costs
  - Uses terms describing how people see the resource
  - Serendipity
  - But, tags uncontrolled,
    • missed associations
    • Relating different views
    • Highly personal (“me”, “important”),
    • Quality and ranking
    • Depth of term

• Aim to investigate
  whether use of an established controlled vocabulary can help move social tagging beyond personal bookmarking to aid resource discovery
Rationale

• Controlled vocabularies
  – For information retrieval
  – Natural language control
  – Expensive

• Social tagging
  – Personal purposes
  – Lack natural language control
  – Free
Objectives

• Social tagging only *versus* combined with controlled vocabularies
  - Indexing aspects
  - Influence on retrieval

• Two contexts
  - Intute subject gateway
  - STFC repository
Intute study
Intute is a free online service providing you with a database of hand selected Web resources for education and research.

Search our database of selected Internet resources

Search

Subject area(s)

All subjects

Arts and Humanities

Health and Life Sciences

Science, Engineering and Technology

Social Sciences

Advanced search Subject A-Z New resources Help

Browse

Browse by subject

Login to MyIntute

Email:

Password:

New user? Register

Intute Blog

Follow Intute on Twitter

Some Small Changes to Intute Economics

Spotlight for May

>> Go to Blog

Behind the Headlines

The Mouse at 40

The computer mouse has celebrated its 40th anniversary.

>> Archive
<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Slavery in New York</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>This online exhibition has been created by the New York Historical Society. It covers the social and political history of slavery in New York state from the 1600s until 1827 when it was made illegal. The site includes a history of slavery in the region and its connections with the Atlantic slave trade, accounts of personal experiences and digitised images of contemporary maps, prints and documents. The site also has a section for teachers with selected classroom activities and worksheets. Topics covered include: the slave trade and emancipation movement.</td>
</tr>
<tr>
<td><strong>Keywords - controlled</strong></td>
<td>slave trade; slavery; slaves</td>
</tr>
<tr>
<td><strong>Keywords - uncontrolled</strong></td>
<td>African-Americans; New York</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Papers/reports/articles/texts</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://www.slaveryinnewyork.org/">http://www.slaveryinnewyork.org/</a> [English]</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>Slavery</td>
</tr>
<tr>
<td><strong>Country of origin</strong></td>
<td>United States</td>
</tr>
<tr>
<td><strong>Date added</strong></td>
<td>2006-01-09</td>
</tr>
</tbody>
</table>
Intute demonstrator

- 11,042 stripped records
- Politics
- Free tagging or DDC / LCSH / Relative Index
- MS .NET application
- Searching, simple and enhanced interfaces
Intute demonstrator: enhanced tagging

EnTag Project SST

Document Details
Title: Arts Council England
URL: http://www.artscouncil.org.uk/
Description: The Arts Council for England is the national funding body for the arts in England. It acts to promote the arts and distribute public money from central government and income generated from the National Lottery. Its website provides information on the history, purpose and current projects of the body. It also includes information on Arts funding in the UK, how to apply for public events and forthcoming events. The latter includes a database of carnivals.

Related Headings
- Political Science (Politics and Government)
  - Comparative government
- Local government
- National government
- Government

Dewey Hierarchy
- Political Science
  - Comparative government
  - Historical, Geographic, Persons Treatment
  - Local government
- Philosophy and theory
- Policy making
- Political ideologies

Suggestions
- Comparative government
- Government
- Historical, Geographic, Persons Treatment
- Local government
- Philosophy and theory
- Policy making
- Political ideologies
Intute demonstrator: searching

EnTag Project ST

Search For Documents

Enter your search terms

slavery new york

Displaying Search Results: slavery new york

<table>
<thead>
<tr>
<th>Title (Click to Open)</th>
<th>Description</th>
<th>Tags</th>
<th>Add Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York divided: slavery and the Civil war</td>
<td>This site provides free access to an online exhibition created by New-York Historical Society wh...</td>
<td>Antislavery movements - United States, Slave trade, slavery, Slavery and emancipation</td>
<td>Teg</td>
</tr>
<tr>
<td>Slavery in New York</td>
<td>This online exhibition has been created by the New York Historical Society. It covers the social...</td>
<td>Antislavery movements - United States, Civil and political rights of nondominant groups, Minorities, Slavery, Slavery and emancipation</td>
<td>Teg</td>
</tr>
<tr>
<td>Last we Forget the Triumph Over Slavery</td>
<td>This is an online exhibition created by the Schomburg Centre for Research in Black Culture, New Y...</td>
<td>Anti-Slavery International is a pressure group which campaigns against contemporary slavery world...</td>
<td>Teg</td>
</tr>
<tr>
<td>Anti-Slavery International (ASID)</td>
<td>Anti-Slavery International is a pressure group which campaigns against contemporary slavery world...</td>
<td></td>
<td>Teg</td>
</tr>
<tr>
<td>Abolish Anti-Slavery Portal</td>
<td>The Abolish portal is a project of the American Anti-Slavery Group which monitors and publicises...</td>
<td></td>
<td>Teg</td>
</tr>
<tr>
<td>Studies in the World History of Slavery, Abolition and Emancipation</td>
<td>Studies in the World History of Slavery, Abolition and Emancipation is an electronic journal prod...</td>
<td></td>
<td>Teg</td>
</tr>
</tbody>
</table>
EnTag Project ST

Title: Slavery in New York
URL: http://www.slaveryinnnewyork.org/
Description: This online exhibition has been created by the New York Historical Society. It covers the social and political history of slavery in New York state from the 1600s until 1827 when it was made illegal. The site includes a history of slavery in the region and its connections with the Atlantic slave trade, accounts of personal experiences and digitised images of contemporary maps, prints and documents. The site also has a section for teachers with selected classroom activities and worksheets. Topics covered include: the slave trade and emancipation movement.

Intute demonstrator: tagging interfaces

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aboriginal Australian local government</td>
<td></td>
</tr>
<tr>
<td>Aboriginal Australian local government</td>
<td></td>
</tr>
<tr>
<td>Abolition movements - United States</td>
<td></td>
</tr>
<tr>
<td>Australian Civil and political rights of non-dominant groups</td>
<td></td>
</tr>
</tbody>
</table>
# Intute demonstrator: Enhanced interface

## Automatically suggested matches

Find appropriate context(s)

- Slavery and emancipation
- Emancipation
- End of Political Science
- Discriminatory practices and slavery
- Extension of slavery
- Slavery
- American fiction-1830-1861...
- American poetry-1830-1861...

## Explore hierarchy around the selected context

Explore Slavery and emancipation
- Emancipation
- Political science

## Select/edit relevant tags

- Amistad (Schooner)
- Abolition of slavery
- Emancipation
- Political science
- Slave insurrections
- Slave trade
- Slavery
- Slavery - political science
- Slavery and emancipation
Intute user study 1

• Questions
  – Choice of tag
  – Retrieval implications

• Participants
  – 28 UK politics students
  – Little tagging experience
Intute user study 2

• Data collection
  - Logging
  - Three questionnaires

• Four tagging tasks
  - Two controlled, two free
  - Tag 15 documents in each task
Intute user study 3

• Hypothetical group project scenario

• Instructions
  – 5 to 10 min per document
  – Open document but focus
  – Try consider enhanced suggestions where appropriate
STFC repository study
Complementary study

- 10 authors depositors
- ACM classification scheme
- Questions
  - Purpose and use of tags
  - Value of controlled vocabulary
  - User interface
Measurement of the Kinetic Energy and Lattice Constant in hcp Solid Helium at Temperatures 0.07 - 0.4K

Author: M A Adams (CCLRC), J Mayers (CCLRC), O Kirichek (CCLRC), R B E Down (CCLRC)
OAI Item Identifier: oai:epubs.cclrc.ac.uk:nwork/3697

Current thesaurus path: electric field effects > electric breakdown

Narrower terms
- avalanche breakdown
- discharges (electric)
- semiconductor device breakdown
- trees (electrical)
- vacuum breakdown
- Zener effect

Related terms
- arcs (electric)
- corona
- electric strength
- electrical faults
- fault currents

Go to the top Add keyword
Approach

• Separate demonstrator

• Supervised sessions
  - Logging
  - Task worksheet
    - Tagging own papers
    - Tag cloud, own tags, controlled vocabulary
Intute study results
Number of tags

- 7,568 tags in total
- 278 tags per person
- 94 + 751 documents tagged (controlled + free task)

- More in simple
- More in free, esp. in enhanced
Tag selection

• **Simple interface:**
  - 91% freely assigned

• **Enhanced interface:**
  - 71% freely assigned
  - 17% controlled tags

• **Other features (both interfaces):**
  - 8% other taggers’ tags
  - 2% main tag cloud
  - < 1% own tag
Browsing for tags

• Simple interface
  - 73% others’ tags
  - main tag cloud and own tag less (17% and 10%)

• Enhanced interface
  - 74% controlled vocabulary
  - 18% others’ tags
  - DDC captions problem
Retrieval implications 1

- **Versus search engine**
  - All tags: new access points for 36% documents
  - Controlled tags: new access points for 69% documents

- **Versus subject indexing**
  - > 90% tags new access points
Retrieval implications 2

• Search terms
  – More in tags than in (un)controlled keywords (up to 3x)

• Free *versus* controlled tags
  – Analysis under way
  – Hypothesis: if pre-coordinated, more facets
Post-questionnaires 1

- **Post-task**
  - Familiar / easy / satisfied / certain
  - Useful: own tags, DDC disambiguation pane, DDC suggestions
  - Not useful: main tag cloud, others’ names, hierarchical DDC pane
Post-questionnaires 2

• Post-study
  – Easy to learn and useful in real life
  – Simple
    + Simplicity, speed, freedom of choice
    – No suggestions
  – Enhanced
    + Suggestions
    – Inappropriate suggestions, cluttered interface, many steps
STFC study

- 6 tags per document
- 67% free tags

- Value of consistency
- Usable in real life
- Is it the right controlled vocabulary
Some observations

• Users are willing to add tags – using a controlled vocab more than free text
  - By and large they understand why it's useful
    • Good search terms = good retrieval
  - But they need help
    • Automation, suggestions, good interfaces
    • Support for different styles of interaction
  - Also need flexible and targeted controlled vocabularies need to be supplied
  - “Web 2.0” features need to be thought through very carefully
    • “tag clouds” not a success
    • Need much better structuring and presentation
    • Integrated

• Interaction between tag clouds and structured vocab needs further investigation
  - Develop a flexible user focussed vocab from tags
  - “structured folksonomy”
Conclusions

• Controlled vocabulary suggestions valued if appropriate

• More access points potential

• Value of consistency for IR
Further research

- Qualitative analysis (scheduled)
- Enhancements
  - Controlled vocabulary
  - Auto suggestions
  - Interface
- Motivation for tagging