Enhancing social tagging with a knowledge organization system

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

Science & Technology Facilities Council
e-Science

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

UKOLN is supported by:

Science & Technology Facilities Council
e-Science

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/

Science & Technology Facilities Council
e-Science

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

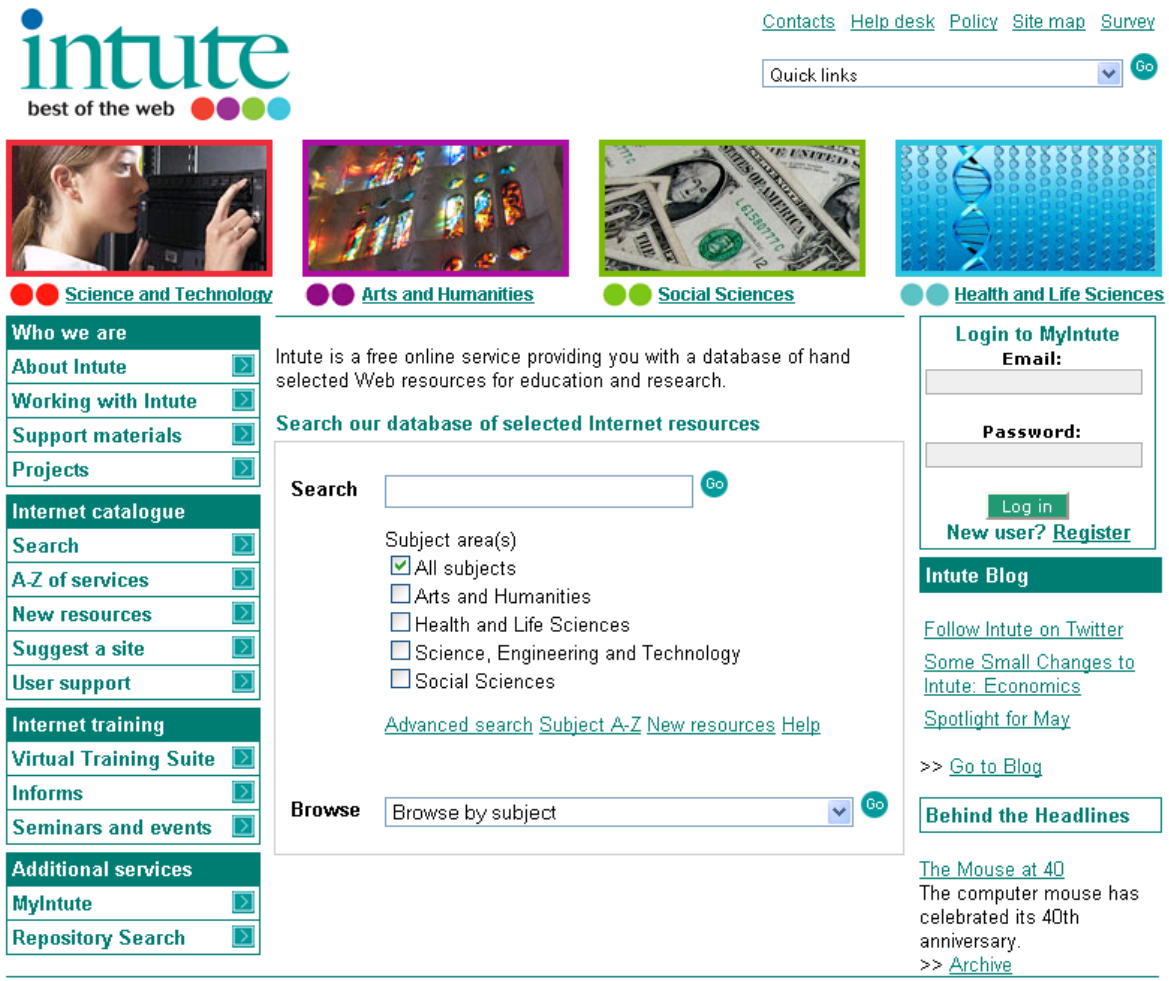
- Controlled vocabularies
  - For information retrieval
  - Natural language control
  - Expensive
- Social tagging
  - Personal purposes
  - Lack natural language control
  - Free

Objectives

- Social tagging only *versus* combined with controlled vocabularies
  - Indexing aspects
  - Influence on retrieval
- Two contexts
  - Intute subject gateway
  - STFC repository

Intute study

Intute http://www.intute.ac.uk



Intute metadata

Title	**Slavery in New York**
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.
Keywords - controlled	slave trade; slavery; slaves
Keywords - uncontrolled	African-Americans; New York
Type	Papers/reports/articles/texts
URL	http://www.slaveryinnewyork.org/ [English]
Classification	Slavery
Country of origin	United States
Date added	2006-01-09

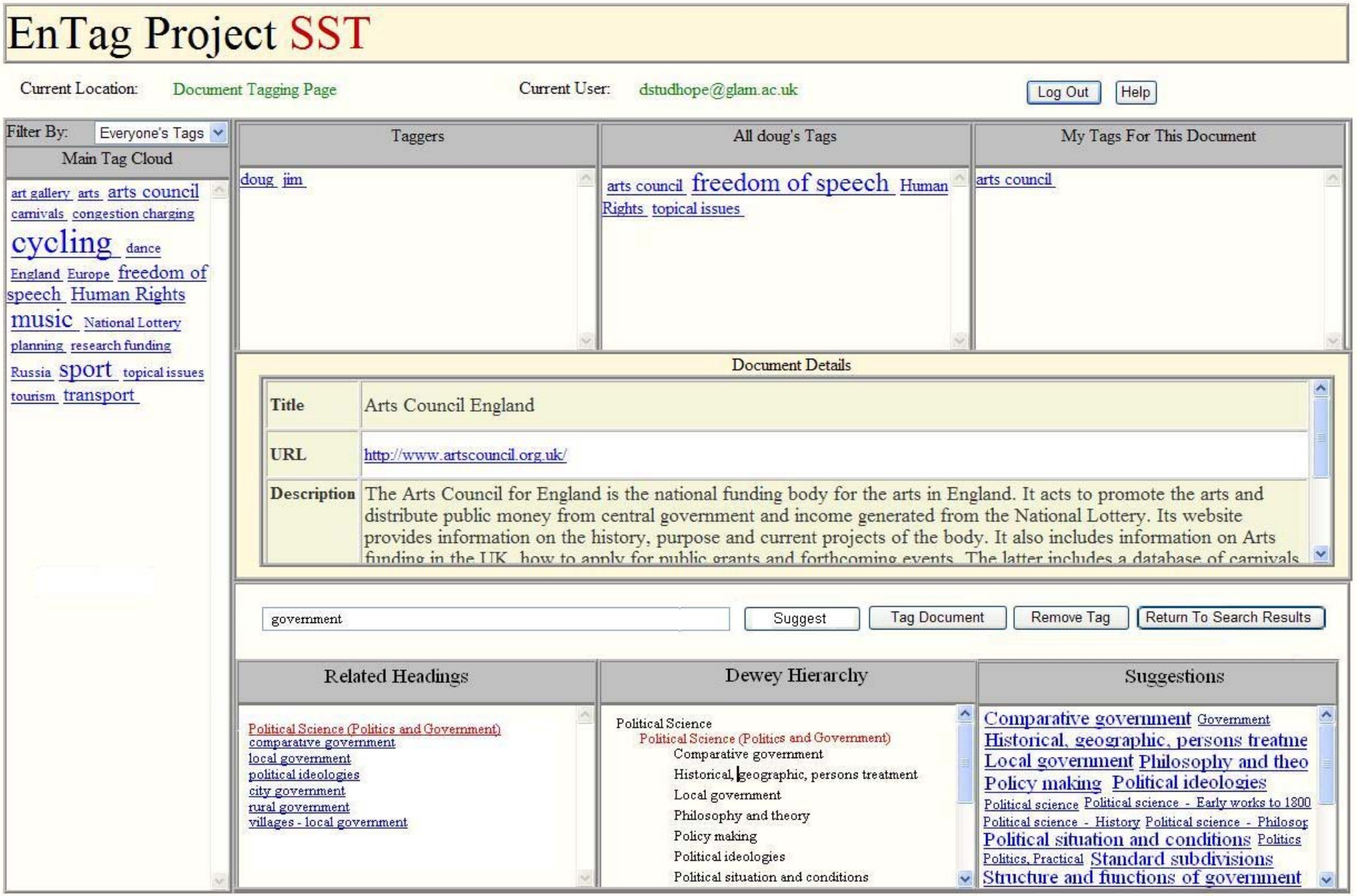
..

Intute demonstrator

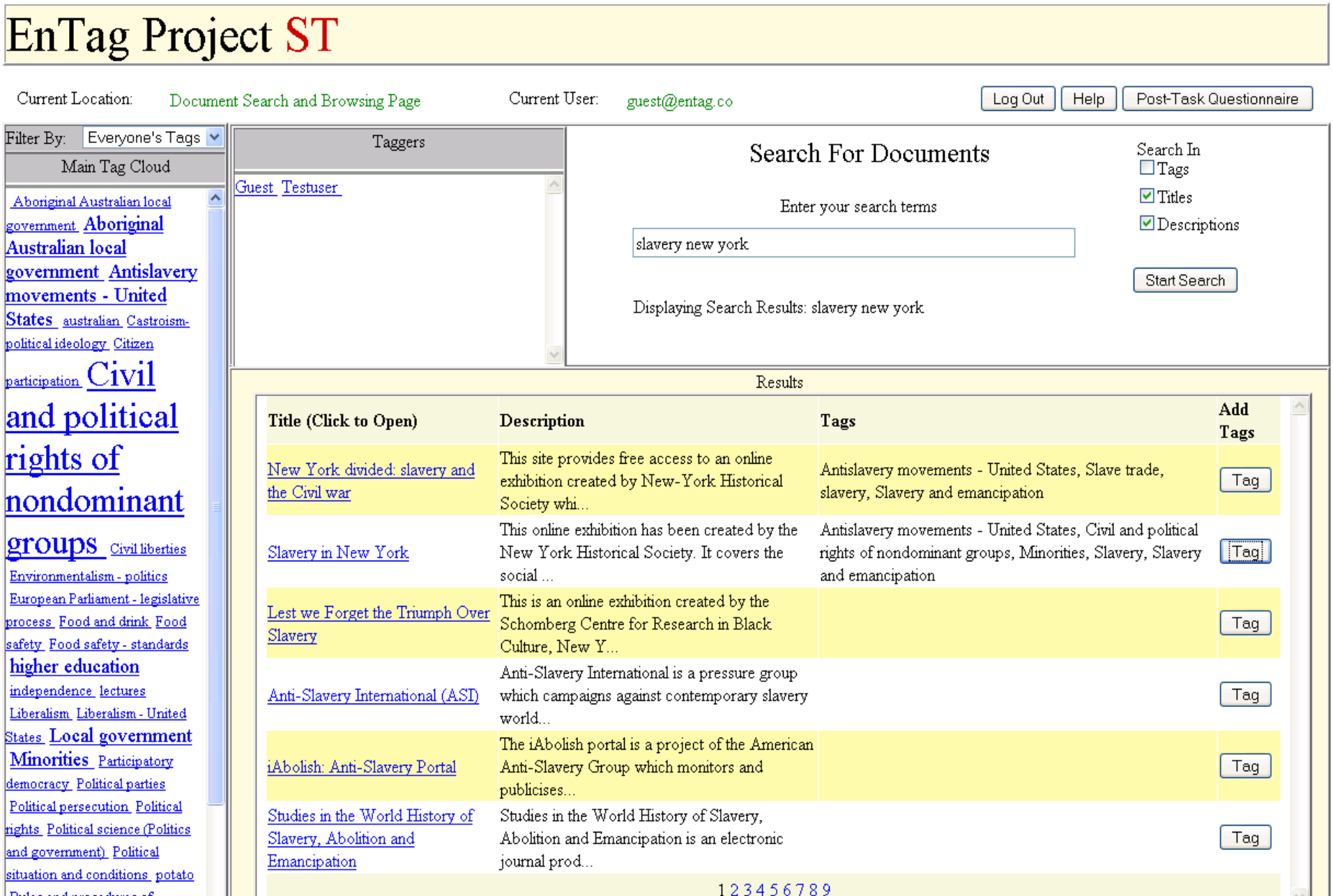
- 11,042 stripped records
- Politics
- Free tagging *or*
  DDC / LCSH / Relative Index

- MS .NET application
- Searching, simple and enhanced interfaces

Intute demonstrator: enhanced tagging



Intute demonstrator: searching



Intute demonstrator tagging interfaces

Intute demonstrator

EnTag Project ST

Current Location: Document Tagging Page Current User: guest@entag.co

Log Out | Help | Post-Task Questionnaire

Filter By: Everyone's Tags

Main Tag Cloud

Aboriginal Australian local government **Aboriginal Australian local government** **Antislavery movements - United States** australian Castroism-political ideology Citizen participation **Civil and political rights of nondominant groups** Civil liberties Environmentalism - politics European Parliament - legislative process Food and drink Food safety Food safety - standards **higher education** independence lectures Liberalism Liberalism - United States **Local government** **Minorities** Participatory democracy Political parties Political persecution Political rights Political science (Politics and government) Political situation and conditions potato

Taggers	All Guest's Tags	My Tags For This Document
Guest Testuser	Aboriginal Australian local government **Aboriginal Australian local government** **Antislavery movements - United States** australian **Civil and political rights of nondominant groups** European	Antislavery movements - United States Civil and political rights of nondominant groups Minorities Slavery Slavery and emancipation

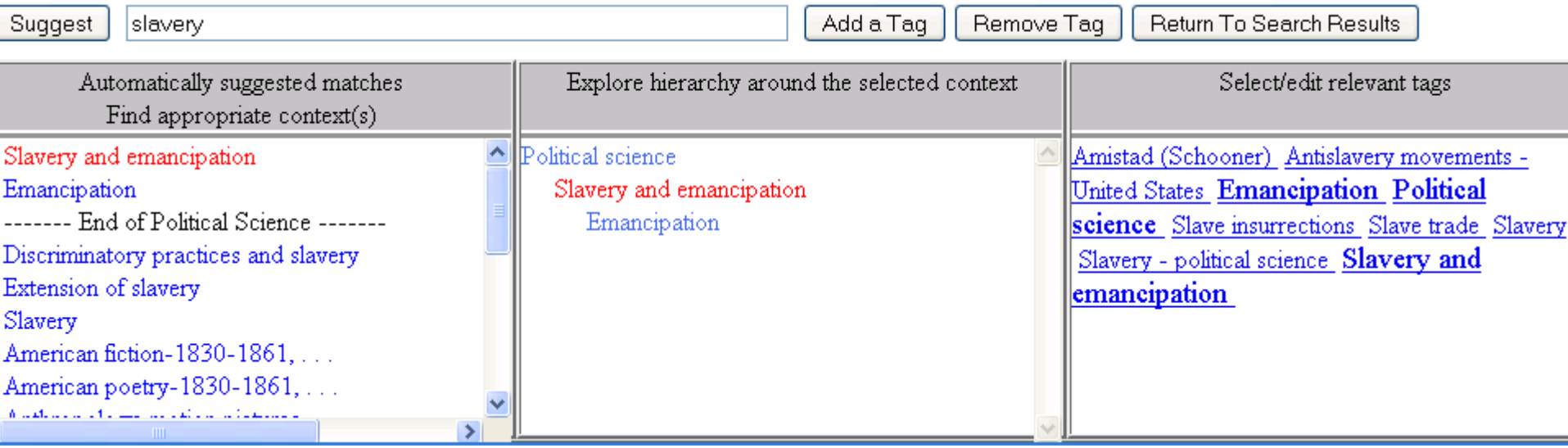
Document Details

Title	Slavery in New York
URL	http://www.slaveryinnewyork.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.

Suggest | slavery | Add a Tag | Remove Tag | Return To Search Results

Automatically suggested matches Find appropriate context(s)	Explore hierarchy around the selected context	Select/edit relevant tags
Slavery and emancipation Emancipation ------- End of Political Science ------- Discriminatory practices and slavery Extension of slavery Slavery American fiction-1830-1861, . . . American poetry-1830-1861, . . .	Political science Slavery and emancipation Emancipation	Amistad (Schooner) Antislavery movements - United States **Emancipation** **Political science** Slave insurrections Slave trade Slavery Slavery - political science **Slavery and emancipation**

Intute demonstrator: Enhanced interface



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

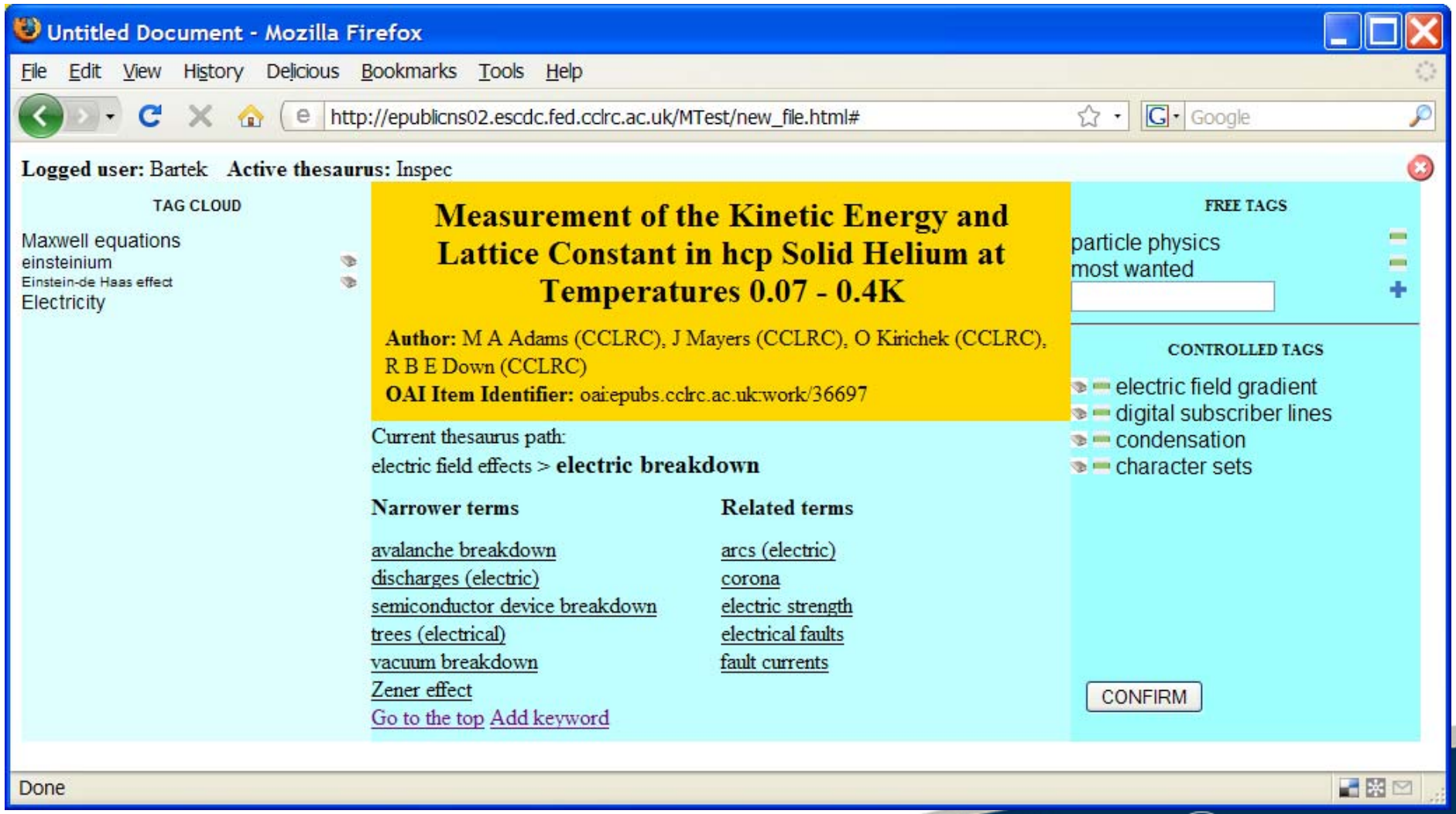
STFC repository study

Science & Technology Facilities Council
e-Science

Complementary study

- 10 authors depositors
- ACM classification scheme
- Questions
  - Purpose and use of tags
  - Value of controlled vocabulary
  - User interface

STFC demonstrator



Science & Technology Facilities Council
e-Science

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 its 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 vocabs 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