

EASTER: Evaluating Automated Subject Tools for Enhancing Retrieval

Koraljka Golub

AMG-UC WS, 5th June 2009, Bath



UKOLN is supported by:  



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Overview

- 1 Apr 2009 – 30 Sep 2010
- Partners
 - 4 funded partners
 - 1 consulting expert
 - 3 non-funded supporting partners



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Purpose & rationale

- Main purpose
 - Evaluate tools for automated **subject metadata** generation
- Rationale
 - Manual subject metadata expensive
 - Difficult to automatically generate
 - High cost of evaluation



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Aims & objectives

- 2 types of subject metadata
 - Controlled subject terms
 - Non-controlled subject keyphrases
<http://www.ukoln.ac.uk/projects/easter/demonstration/>
- 2 contexts
 - Intute cataloguers
 - Intute end-users



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Approach

- Devise evaluation methodology
- Develop 'gold standard'
- Evaluate applying 'gold standard'
- Evaluate through integration in cataloguing workflow
- Evaluate through retrieval by end users



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Candidate tools

- Controlled terms
 - Machine learning: iVia, KEA, Temis, TextGarden
 - String matching: KnowLib, Scorpion
- Key phrases
 - KEA, Temis, TerMine, TextGarden



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Data collection

- Intute
 - Social Sciences (IBSS and HASSET)
 - Health and Life Sciences (CABI)
 - Arts (AAT & GNT)



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Evaluation framework 1

- 'Gold standard'
 - A sample of documents
 - 2 cataloguers
 - Focus group of 3 end users
- Apply a variety of measures



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Evaluation framework 2

- Cataloguing workflow integration
 - Integration demonstrator
 - Cataloguers' assessment of automatically assigned terms
- End-user retrieval test
 - Search tasks
 - Contribution of automatically and manually assigned terms to retrieval success and failure



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management



Dissemination

- Web site
 - <http://www.ukoln.ac.uk/projects/easter>
- Blog
 - <http://blogs.ukoln.ac.uk/easter/>



www.ukoln.ac.uk

UKOLN A centre of expertise in digital information management

