

EASTER (Evaluating Automated Subject Tools for Enhancing Retrieval)

<http://www.ukoln.ac.uk/projects/easter/>

EASTER is an 18-month JISC project funded under the Information Environment Programme 2009-11. It started in April this year and involves eight institutional partners, with UKOLN acting as the lead, University of Glamorgan, Intute, City University London and Dagobert Soergel as major partners, and Royal School Library and Information Science, University College London, and OCLC Office of Research acting as non-funded supporting partners.

The purpose of the project is to test and evaluate existing tools for automated subject metadata generation in order to better understand what is possible and what the limitations of current solutions are, and to make subsequent recommendations for services employing subject metadata in the JISC community.

The project is concerned both with the creation and with the enrichment of subject metadata using existing automated tools. Subject metadata are most important in resource discovery, yet most expensive to produce manually. In addition, they are much more difficult to generate automatically especially in comparison to formal metadata such as file type, title, etc. Also, due to the high cost of evaluation, automated subject metadata tools are rarely tested in live environments of use. There is a huge challenge facing UK HE digital collections, institutional repositories, and aggregators of institutional repository content, as to how to provide high quality subject metadata for increasing numbers of digital information at reasonable costs.

The information centre chosen as a test-bed for this project is Intute. Selected tools for automated subject metadata generation will be tested in two contexts: by Intute cataloguers in the cataloguing workflow and by end-users who search for information in Intute as part of their research, learning, and information management.

The project will first develop a methodology for evaluating the tools. The methodology will then be implemented in the above contexts. First, we will create a 'gold standard' using a consensus of cataloguers and then we will evaluate all tools against it. The best tool(s) for the purposes of Intute will be implemented into a demonstrator that will feed its results into the cataloguing workflow. This will be evaluated. Furthermore, we will conduct a task-based end-user retrieval study to determine the contribution of automatically assigned terms and manually assigned terms, each alone and in combination, to retrieval success (retrieving relevant documents) and failure (missing relevant documents and retrieving irrelevant documents).