

Exploration, Navigation and Retrieval of Information in Cultural Heritage ENRICH 2013

Séamus Lawless
Owen Conlan
Knowledge and Data
Engineering Group
Trinity College Dublin
Ireland
+353 1 896 1765
{seamus.lawless,
owen.conlan}@scss.tcd.ie

Maristella Agosti
Department of Information
Engineering
University of Padua
Italy
+39 049 827 7650
agosti@dei.unipd.it

Paul Clough
Information School
University of Sheffield
United Kingdom
+44 114 222 2664
p.d.clough@sheffield.ac.uk

ABSTRACT

The Exploration, Navigation and Retrieval of Information in Cultural Heritage Workshop (ENRICH 2013) offers a forum to 1) discuss the challenges and opportunities in Information Retrieval research in the area of Cultural Heritage; 2) encourage collaboration between researchers engaged in work in this specialist area of Information Retrieval, and to foster the formation of a research community; and 3) identify a set of actions which the community should undertake to progress the research agenda.

The workshop will foster a new stream of Information Retrieval research and support the design of search tools that can help end-users fully exploit the wonderful Cultural Heritage material that is available across the globe.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Search Process*

Keywords

Cultural Heritage, Digital Humanities, Information Retrieval

1. OVERVIEW

A key challenge facing the curators and providers of digital cultural heritage worldwide is to instigate, increase and enhance engagement with their collections. To achieve this, a fundamental change in the way these artefacts can be discovered, explored and contributed to by users and communities is required. Cultural heritage artefacts are digital representations of primary resources: manuscript collections, paintings, books, photographs etc. The text-based resources are often innately “noisy”, contain non-standard spelling, poor punctuation and obsolete grammar and word forms. Image-based resources often have limited associated metadata which describes the resources and their content. In addition, the information needs and tasks of cultural heritage users are often complex and diverse, evolving through information exploration and discovery. This presents a specific set of challenges to traditional Information Retrieval (IR) techniques and approaches.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '13, July 28–August 1, 2013, Dublin, Ireland.

ACM 978-1-4503-2034-4/13/07.

This workshop investigates the enhanced retrieval of, and interaction with, cultural heritage collections. We are interested in investigating innovative forms of personalised, multi-lingual IR, which can include:

- Content-aware retrieval which responds to the entities and relationships contained within the artefacts and across collections.
- Personalised IR, exploration and presentation which responds to models of user and contextual intent.
- Community-aware retrieval which responds to wider community activity, interest, contribution and experience.

Such new forms of enhanced IR require rigorous evaluation and validation using appropriate metrics, contrasting digital cultural heritage collections and diverse users and communities. This workshop promotes such evaluation, taking into account the specific requirements of the domain. The workshop also explores the use of multilingual, multimodal and personalised IR techniques and technologies to enable end-users to search in their native language, but receive results collated from content collections in numerous languages, all tailored for their consumption. The nature of cultural heritage resources means that content analysis in support of IR is of specific interest. This includes the automated normalisation of historical texts, the use of Natural Language Processing (NLP) for entity extraction and metadata generation.

The ENRICH workshop promotes the exchange of ideas between researchers working on the theory and foundations of IR, cross and multi-lingual IR, personalised search and recommender systems. There are numerous research areas that can support such improved retrieval and exploration in the area of cultural heritage. Relevant topics include, but are not limited to: Multilingual semantic search; Context-aware and semantic recommender systems; Adaptation engines and algorithms for personalised multilingual IR; User modelling and adaptation; Content personalisation and personalised result presentation; Domain modelling; External knowledge resources for IR; Evaluation methodologies and metrics for personalised multilingual IR; Information Extraction, Data Mining and Natural Language Processing; and Social Network Analysis.