# PMHR 2011
# The First Workshop on Personalised Multilingual Hypertext Retrieval

Maristella Agosti
University of Padua, Italy
*agosti@dei.unipd.it*

Ernesto William De Luca
Berlin Institute of Technology, Germany
*Ernesto.DeLuca@dai-labor.de*

Séamus Lawless
Trinity College Dublin, Ireland
*seamus.lawless@scss.tcd.ie*

Johannes Leveling
Dublin City University, Ireland
*jleveling@computing.dcu.ie*

## 1 Introduction

The *First Workshop on Personalised Multilingual Hypertext Retrieval* (PMHR 2011)[1] was held on the 6th June 2011 in conjunction with the 22nd ACM conference on Hypertext and Hypermedia in Eindhoven, The Netherlands[2]. The aims of PMHR 2011 were to promote collaboration and the exchange of ideas between researchers working on hypertext and adaptive hypertexts, cross-lingual information retrieval, personalised search, personalisation for the web and hypertexts, and recommender systems. The workshop had a specific focus on research in user modelling and multilingual personalisation for hypertext retrieval.

## 2 Background and Motivation

Search engines have traditionally followed "a one size fits all" paradigm and returned the same results for all users. They do not adapt to the user, the domain, or the search context. Thus, the search process and the number and type of results returned are not tailored to the individual user or his search situation. Personalised hypertext retrieval is concerned with adapting the search process to the user's needs. This includes adapting the system, the query-document similarity metrics, the search results, and their presentation to an individual user. The personalisation process can be based on models of the user, the domain, and the search context, but no standard representation or resources have evolved to-date.

It can often be the case that non-native English speakers suffer limited or restricted online experiences as typically the majority of web content is still authored in English. Machine translated

[1] http://www.dai-labor.de/pmhr2011/
[2] http://www.ht2011.org/

versions of content may be generated for some languages, but this is not always the case. This results in the user being greatly restricted in the content collections across which they can search.

The papers accepted for the PMHR 2011 workshop start to explore the use of multilingual hypertext retrieval technologies and adaptive personalisation techniques to enable end-users to write queries in their native language, but receive results collated from content collections in a variety of languages, all tailored for consumption by that individual.

The PMHR 2011 workshop aimed at soliciting cooperation between researchers working in many different areas, such as hypertext and adaptive hypertexts, cross-lingual information retrieval, personalised search, personalisation for web and hypertexts, recommender systems and human-computer interaction.

# 3 Goals and Objectives

Personalised multilingual hypertext retrieval has to be viewed as an interdisciplinary research task, as it can include ideas from a variety of different research disciplines and subject domains. The PMHR 2011 workshop brought together researchers from across these domains with the main goals of:

- developing a roadmap to follow which identifies the different research challenges that need to be faced in order to envisage, design and implement personalised multilingual hypertext retrieval systems and tools;
- identifying a process for the evaluation of PMHR systems and the correct metrics to use; and
- defining the hypertext collections available for use in the evaluation of PMHR systems.

# 4 Invited Presentations

To set the scene two keynote presenters were invited to address challenges related to the personalisation of the web and the experimental evaluation of hypertexts.

The first presentation was given by Prof. Vincent Wade of Trinity College Dublin, Ireland, on *Challenges for the Multi-Lingual, Multi-Dimensional Personalised Web*. The presentation addressed the personalised web and identified the many dimensions, extending far beyond the appropriateness of content selection, that need to be addressed concurrently by any system attempting to deliver personalised web experiences. Among these dimension are tasks and activities, cultural preferences, language and social interaction. Adaptive Hypermedia and Adaptive Web research examines the personalised delivery and exploration of content; however, this content has traditionally been closed-corpus in nature, with less research having been conducted on open corpus content. From a commercial perspective, web adaptivity has been more focused on adaptive content retrieval rather than adaptive content composition. The key challenges identified by this talk include: integrated open corpus and service personalisation, cultural adaptivity, dialogue and simulation-based personalisation and the power of the crowd.

The second presentation was given by Dr. Nicola Ferro of University of Padua, Italy, on *Experimental Evaluation and Hypertexts*. The presentation addressed open issues concerning how experimental evaluation and hypertext could be better benefit each other. On the one hand, it is time for experimental evaluation to explicitly take into consideration the hypertextual nature of resources when assessing performance based on retrieved items and to not only consider systems as black-boxes that internally exploit the exiting hypertext. On the other hand, experimental evaluation produces huge amounts of scientific data that would be better understood and interpreted if they were enriched with links to each other, to other resources, and to user-generated content, such as annotations explaining them.

# 5 The Papers Accepted and Presented at the Workshop

The main issues discussed during the workshop were introduced by the papers accepted and presented at the workshop. The papers are briefly described below.

- Ghorab et al. [3] which investigated the representation, creation, use, and evaluation of user models for personalised information retrieval in a multilingual environment. The authors argued that their proposed representation of a user model, which is created from the user's search history, would be more suitable for personalised multilingual information retrieval. In addition, they outlined two algorithms for query adaptation based on the proposed multilingual user model.

- Lops et al. [5] presented a language-independent content-based recommender system which is based on cross-language user profiles. They aim to reduce the effects of polysemy and synonymy by moving from word-based representation towards concept-based representation, taking the meaning of words into account. The recommender system builds on word sense disambiguation and the concept inventory of MultiWordNet. Experimental results in the domain of movie recommendation have shown the effectiveness of the approach.

- Levacher et al. [4] described a method for evaluating adaptive content retrieval, modification, and delivery. They focused on text segmentation for adaptive hypermedia systems and described challenges in evaluating different text slicing techniques.

- Ganguly et al. [2] proposed to create log information in a controlled topic creation process, where information about different query formulations, retrieved documents, and viewed documents is logged. The log data will be used in an evaluation task which investigates aspects of how to improve personalised and collaborative information retrieval. One subtask is concerned with predicting the search category of a query based on a user's search history. The second subtask involves tuning methods to address individual user needs based on information about user groups. The proposed methodology will be employed for a task at the FIRE 2011 evaluation campaign.

- Agosti et al. [1] reported results from experiments using linking annotations which express relationships between images in a digital archive. The annotations are represented as a hypergraph and exploited to personalise result presentation for different users.

# 6 Conclusions: Research Roadmap Topics

In general, an information retrieval system tries to find and retrieve relevant documents related to a user query, where both the documents and the query are written or composed in the same language. Dealing with a multilingual document collection brings up new challenges. Being able to read a document in a foreign language does not always automatically imply that the user can formulate appropriate queries in that language as well. When handling documents in languages they are less confident with, users find cross-language text retrieval particularly useful as they can express their information needs more effectively in their mother tongue [6].

Concentrating on the Web which is, by far, the largest data repository in the world and considering the number of web pages in different languages, a significant amount of work is necessary for the construction of reasonable multilingual information retrieval (MLIR) tools that can handle multilingual resources, especially web collections. Zhou et al. [7] analyse web-based MLIR systems and explain that it is not clear whether or how a generic approach for multilingual web retrieval systems that includes different languages can be developed. While the web includes many linguistic resources, as of yet no data mining process exists that can improve the performance of multilingual web retrieval systems. It is also not clear whether machine translation can really help in a multilingual task. However, traditional MLIR techniques cannot be directly adopted to web applications. There are different aspects to be considered if we want to apply MLIR techniques to web documents. The workshop discussion sessions debated these different aspects and challenges that constitute the research roadmap for personalised multilingual hypertext retrieval have been recognised. These challenges include:

- multi-linguality: facilitating translation of queries, content, and metadata;
- intelligent content: moving annotation towards making content structurally rich and semantically aware, discoverable, reusable,reconfigurable and adaptable;
- multi-dimensionality: supporting multiple dimensions for adaptivity, allowing different sets of dimensions for different application areas; enabling multi-dimensional evaluation; and
- move beyond adaptive content retrieval and adaptive recommendation to adaptive creation: dynamic content repurposing and generation for reuse.

Many of these challenges are part of the research agenda of the CULTURA[3] and CNGL[4] projects.

# Acknowledgements

The work of Maristella Agosti and Séamus Lawless has been partially supported by the CULTURA project, as part of the Seventh Framework Programme of the European Commission, Area "Digital Libraries and Digital Preservation" (ICT-2009.4.1), grant agreement no. 269973.
The work of Séamus Lawless and Johannes Leveling has been partially supported by Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation.

[3]CULTURA Project - `http://www.cultura-strep.eu/`
[4]Centre for Next Generation Localisation - `http://www.cngl.ie/`

# References


[1] Maristella Agosti, Giordana Mariani Canova, Nicola Orio, and Chiara Ponchia. Methods of personalising a collection of images using linking annotations. In *The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011)*. 2011.

[2] Debasis Ganguly, Johannes Leveling, Wei Li, and Gareth Jones. Towards evaluation of personalized and collaborative information retrieval. In *The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011)*. 2011.

[3] M. Rami Ghorab, Dong Zhou, Ben Steichen, and Vincent Wade. Towards multilingual user models for personalized multilingual information retrieval. In *The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011)*. 2011.

[4] Killian Levacher, Seamus Lawless, and Vincent Wade. A proposal for the evaluation of adaptive content retrieval, modification and delivery. In *The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011)*. 2011.

[5] Pasquale Lops, Cataldo Musto, Fedelucio Narducci, Marco De Gemmis, Pierapolo Basile, and Giovanni Semeraro. Learning semantic content-based profiles for cross-language recommendations. In *The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011)*. 2011.

[6] Douglas W. Oard. Alternative approaches for cross-language text retrieval. In *AAAI Symposium on Cross-Language Text and Speech Retrieval.*, American Association for Artificial Intelligence, 1997.

[7] Yilu Zhou, Jialun Qin, Hsinchun Chen, and Jay F. Nunamaker. Multilingual web retrieval: An experiment on a multilingual business intelligence portal. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 1*, page 43a, Hawaii, 2005.