

# Personalised Multilingual Hypertext Retrieval: An Overview

Maristella Agosti<sup>1</sup>, Ernesto William De Luca<sup>2</sup>, Séamus Lawless<sup>3</sup>, and  
Johannes Leveling<sup>4</sup>

<sup>1</sup> Department of Information Engineering – University of Padua  
[maristella.agosti@unipd.it](mailto:maristella.agosti@unipd.it)

<sup>2</sup> DAI Lab, Berlin Institute of Technology, Germany  
[ernesto.deluca@dai-labor.de](mailto:ernesto.deluca@dai-labor.de)

<sup>3</sup> Knowledge and Data Engineering Group, Trinity College Dublin, Ireland  
[seamus.lawless@scss.tcd.ie](mailto:seamus.lawless@scss.tcd.ie)

<sup>4</sup> School of Computing, Dublin City University, Ireland  
[johannes.leveling@computing.dcu.ie](mailto:johannes.leveling@computing.dcu.ie)

**Abstract.** The aims of the workshop on *Personalised Multilingual Hypertext Retrieval* (PMHR) are twofold: to set the scene in this challenging area, allowing the different communities engaged in related research topics to meet and to determine a program of actions to undertake; to devise a strategy for the evaluation of PMHR systems, which should define the collection of resources to use to evaluate such systems together with the evaluation metrics to use.

The workshop results will be of use in the design of personalised tools that can help end-users fully benefit from the use of distributed multilingual hypertext content.

## 1 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 their 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 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 the end-user.

To reach useful results, it is necessary to solicit 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, and recommender systems.

## **2 Goals and Objectives**

Personalised multilingual hypertext retrieval has to be viewed as an interdisciplinary research task, as it can include ideas from adaptive hypertext, user modelling for adaptive hypertext, personalisation for Web and hypertext, human-computer interaction, and interactive authoring environments.

The main goals of the workshop are: to develop a roadmap to follow which identifies the different research challenges that need to be faced in the design and implementation of personalised multilingual hypertext retrieval systems and tools of use for the end-user; to develop a strategy for the evaluation of PMHR systems, in addition, to define the collections of hypertext documents to use in the evaluation of such systems together with the metrics to use.

## **3 Accepted Workshop Papers**

In total, the PMHR workshop received seven submissions, of which five were accepted for presentation.

Ghorab, Zhou et al. [1] investigate the representation, creation, use, and evaluation of user models for personalised IR in a multilingual environment. They argue that their proposed representation of a user model, which is created from the user search history, would be more suitable for personalised multilingual information retrieval. In addition, they outline two algorithms for query adaptation based on the proposed multilingual user model.

Lops, Musto et al. [2] present 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 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 show the effectiveness of this approach.

Levacher, Lawless et al. [3] present a proposed methodology for the evaluation of adaptive content retrieval, modification and delivery. Content slicing systems are introduced and the objectives and challenges involved in the evaluation of such systems are detailed. The paper focuses on content slicing for adaptive hypermedia (AH). Some of the limitations of traditional AH systems are addressed

by providing the adaptive retrieval of open corpus resources, tailored to suit the content requirements of specific AH systems. These on-demand tailored content resources are called slices. A methodology for the evaluation of such slicing systems is presented along with a proposed experimental implementation.

Ganguly, Leveling et al. [4] propose 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 IR. 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, Canova et al. [5] report results for 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 of the images for different users.

As one of the major goals of the PMHR workshop is to identify the collection of resources for personalised multilingual hypertext retrieval, the workshop proceedings also include two notes about ongoing research activities. The first note is about the CULTURA project, in which community-aware adaptivity for digital humanities is explored. The second note describes log file resources used in the LogCLEF evaluation initiative.

## Acknowledgements

The work of Maristella Agosti and Samus Lawless has been partially supported by the CULTURA project<sup>5</sup>, 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.

This work has also been supported by the Science Foundation of Ireland (grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (<http://www.cngl.ie/>) at Dublin City University and Trinity College Dublin.

## References

1. Ghorab, M.R., Zhou, D., Steichen, B., Wade, V.: Towards multilingual user models for personalized multilingual information retrieval. In: The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011). (2011)
2. Lops, P., Musto, C., Narducci, F., Gemmis, M.D., Basile, P., Semeraro, G.: Learning semantic content-based profiles for cross-language recommendations. In: The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011). (2011)

---

<sup>5</sup> CULTURA Project Website, URL: <http://www.cultura-strep.eu/>

3. Levacher, K., Lawless, S., Wade, V.: A proposal for the evaluation of adaptive content retrieval, modification and delivery. In: The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011). (2011)
4. Ganguly, D., Leveling, J., Li, W., Jones, G.: Towards evaluation of personalized and collaborative information retrieval. In: The First Workshop on Personalised Multilingual Hypertext Retrieval (PMHR 2011). (2011)
5. Agosti, M., Canova, G.M., Orio, N., Ponchia, C.: Methods of personalising a collection of images using linking annotations. (2011)