

Overview of the PENG project

Gabriella Pasi and Gloria Bordogna

Abstract— In this paper a synthetic overview of the main aims, characteristics and innovations of the PENG project is presented. PENG . “PERSONALISED NEWS content programminG” is a Specific TARGETED RESEARCH PROJECT (IST-004597) funded within the Sixth Program Framework of the European Research Area.

Index Terms— Information Filtering, Distributed Information Retrieval, Personalized User Profiles

I. INTRODUCTION

THE main objective of the PENG project was to define an innovative technological solution to the personalised multimedia news access, composition and presentation, with an emphasis on personalised filtering, retrieval and composition of multimedia news. Indeed, the proposed system aims to collect news from both news-feeds and specialised archives in a personalised way, so as to provide media professionals with a fully customisable environment.

This is performed by pushing personalised news towards the user, and by allowing her or him to expand a selected topic by searching for additional information or editing the final news through a personalized presentation approach. This involves the integration of personalised filtering and distributed information retrieval, producing information that constitute a first but very important aid to a journalist's writing activity (the initial target user for PENG). The target users of PENG are classified according to a bi-dimensional schema defined in terms of their level of interest in the news, and their topical interests. Possible user targets include information-intensive workers, students of communication faculties, journalists with specific topical interests etc.

An important characteristic that is ensured by the system is the flexibility in modelling the user's topical interests and context. This means modelling the capability to be tolerant to

This work was supported by the E.C. through the Specific Targeted Research Project named PENG (Personalized NEWS content programminG). This project (IST-004597) funded within the Sixth Framework Programme, Priority 2, Information Society Technology, Thematic Priority: Cross Media Content for Leisure and Entertainment.

Gabriella Pasi is with the Università degli Studi di Milano Bicocca (DISCO), Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy, , (corresponding author phone: +390264487847; e-mail: pasi@disco.unimib.it).

Gloria Bordogna is with the National Council of Research (IDPA-CNR), Via Pasubio 5, 24044 Dalmine, Italy, (e-mail: gloria.bordogna@idpa.cnr.it).

the vagueness and uncertainty in both the user-system interaction, and adaptive in the learning of users' changing preferences over time. In particular the PENG prototypal system supports both the filtering of news based on a new multi-criteria decision approach, and the cataloguing of news into overlapping topics. These characteristics are not met by the current systems commonly used by journalists to carry out their habitual tasks, and their replacement with the PENG system will greatly innovate the current way of making news. The PENG project was mainly addressed to news professionals, such as journalists and editors, with the view of extending the use of the defined system to more general users in the future. In this context, with the term news we refer to any kind of news, including information regarding leisure and entertainment.

This PENG prototype is conceived as a personal assistant, supporting journalists in all stages of the news lifecycle. Information (text, images, and videos) is gathered from different sources (including the Web) using a combination of push and pull technologies and is presented to the user in a personalised way.

II. PENG SYSTEM'S ARCHITECTURE

The main functionalities of the PENG system are separated into three main phases: a push phase, a pull phase and a presentation phase.

In the push phase a filtering system was developed, by means of which a first selection of news are selected from newswires and other news archives. This filtering is based on a dynamic user profile including the personal user's trust in the information sources.

In the pull phase a user query and the user profile are used to retrieve further and more specific information from both the same sources used in the pull phase, and also from additional sources automatically selected in relation to the content of the query and the user profile. A distributed information retrieval approach is used, where the query can be automatically generated from user feedback on the information presented by the push phase.

The presentation phase uses multi-document and multi-media visualisation to present the results from the push and pull phases to the user. This takes into account the trust a user places in the information sources. The results visualized by the system can be personalised not only to the user information need but also to visualisation preferences and the subjective interpretation of the users trust in sources of information.

Figure 1, below, presents a high level view of the PENG

architecture, with the modules corresponding to the three main phases highlighted in gray.

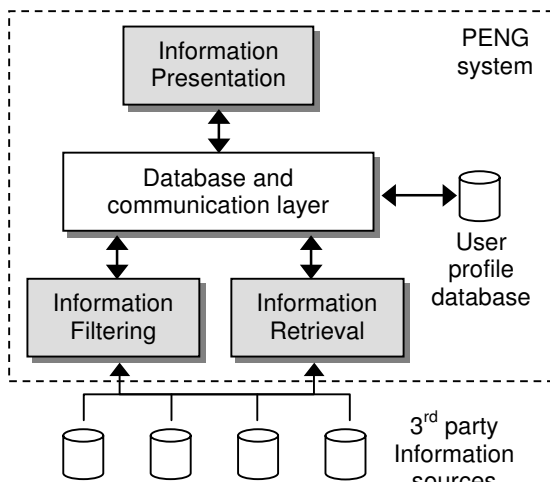


Figure 1: a sketch of the PENG system architecture

The user accesses the system locally through an interface provided by the presentation module. Each of the three main modules communicate via an intermediary layer that also manages access to the common databases required by the system, the most important of which is the user profile database (the other databases are not shown for clarity). This database and communication layer is composed of a user profile manager (which manages the user profile database) and a common database manager (which manages other common databases and coordinates the communication between the modules). In the PENG system, a user profile contains the data relating to a single user: personal information (such as name, email etc), information preferences (what information is relevant to the user, from where), presentation preferences (how this information is to be displayed) and interaction history (the history of the user's interaction with the PENG system). Since a user may be interested in numerous different subjects, the information and presentation preferences are split into a set of different user interests. Each interest is personal to the user to which it belongs, and plays an important part in the filtering module (which is intended to not only filter documents to users, but to adjust user interest) and information retrieval modules (as providing a context in which a search can take place).

Importantly, the profile stores the degree to which a user trusts different information sources ('trust scores'), information hypothesised as being important in news gathering and filtering. Trust scores are conceived as indications of the potential reliability of the information sources to a specific user (or category of users) with respect to a given topical area.

PENG has the potential to greatly contribute to the continuing development of filtering and retrieval systems, for the benefit of the journalists, and ultimately for all users of news services. Professionals, such as journalists or editors, can tune the contribution of the distinct sources to their information gathering, filtering and editing tasks. This is achieved by specifying queries expressing constraints on the

multimedia and time-dependent content of the news so as to focus on a particular event. This enables the tuning of a personalised gathering and presenting of news that expresses an individual's view and opinion on an event, a condition for journalism that has become predominant in recent years and a very important condition for a personalised presentation of news to the general user. In fact, while this can greatly reduce the time needed for a journalist to consult the distinct sources and to report on a given topic of interest, it also enables the presentation of news tailored to a specific user's interests.

The automatic classification of the news into thematic clusters represented by sets of keywords can be coupled by successively using PENG to yield personalised presentations of up to date topics. This can help in drafting a personalised multimedia newspaper and can thus be a powerful tool for the editorial staff of a journal.

REFERENCES

- [1] M. Agosti, F. Crestani and G. Pasi eds., "Lectures on Information Retrieval", Springer-Verlag, 2001.
- [2] G. Bordogna, G.Pasi, R. Yager, "Soft Approaches to information Retrieval on the WEB", *Int. Journal of Approximate Reasoning*, 34, 105-120, (2003).
- [3] G. Bordogna, G.Pasi, "Personalised indexing and retrieval of heterogeneous structured documents", *Information Retrieval Journal*, in press (2004).
- [4] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. netes, M. Sartin, "Combining content-based and collaborative filters in an online newspaper", ACM sigir workshop on recommender systems Aug. 19, Berkeley. (1999).
- [5] F. Crestani and G. Pasi, editors. "Soft Computing in Information Retrieval: Techniques and Applications." Physica-Verlag (Springer-Verlag), Heidelberg, Germany, 2000.
- [6] F. Kilander, "A brief comparison of News filtering Software", <http://www.glue.umd.edu/enec/medlab/filter/filter.html>.
- [7] D. Moraru, L. Besacier, P. Mulhem and G. Quénot, "CLIPS-IMAG at TREC-11 : Experiments in Video Retrieval", *11th Text Retrieval Conference*, Gaithersburg, MD, USA, 19-22 November, 2002.
- [8] P. Mulhem, J. Gensel and H. Martin, "Adaptive Video Summarization", in Handbook on Video Databases, CRC Press, to appear, 2003
- [9] G. Pasi, "Modelling users' preferences in systems for information access", *International Journal of Intelligent Systems*, 18, 793-808, (2003).
- [10] G. Amato and U. Straccia (1999) "User Profile Modeling and Applications to Digital Libraries", 3rd European Conference on Digital Libraries, ECDL99, Paris, France, September 22-24, LNCS 1696.