

## A blended learning approach in the news domain

**Bilal Zaka, Christian Safran, and Frank Kappe**

Institute for Information Systems and Computer Media, Graz University of Technology, Austria

A multimodal approach for acquiring personalized news contents, with customized aggregators, collaborative & personal content filters is presented. The proposed framework is based on the effectiveness of combined use of taxonomies and folksonomies. This platform is an effort towards extending web 2.0 concepts of information selection for various modes of user interaction. Traditional classification (taxonomies) uses explicit relations and predefined concepts about a subject and while being more authoritative these are more confined and may require effort to maintain in data environment of diverse nature. In contrast to that folksonomies are developed using rather uncontrolled relations and concepts about some subject and can be updated and developed by users in real time. Metadata for news classification are semantically enhanced using online linguistic resources. While dealing with over communicated daily news contents one can benefit from taxonomy based content filters to get only information of interest, along with the possibility of finding and sharing information of interest with rather hidden or deep relation which might not be available otherwise. The blended information classification and filtration technique presented in this paper can prove to be very effective in multimodal systems where system's response to user inputs has a lower level of certainty.

**Keywords:** knowledge management; personalization; cross-media publishing

### 1. Introduction

The ever growing knowledge base of the internet offers convenience of time and place independent news access about virtually any topic. Progression in technology also resulted in a variety of devices being used to access this information. Different ways to access information have always been developed and experimented with, as varying environmental and physical conditions require more appropriate interfaces. Ubiquitous computing environments with this much information load force an individual to have efficient personalization of available information. This requirement is the basis of the proposed framework, which offers a central platform to aggregate news contents of choice and present user adaptive view of information. This proposed system uses content filters, and group behaviours to filter out news for individuals. Semantic inferences are added to strengthen the use of filtration techniques in modes of interaction where communication with the system is rather uncertain. In his book "Being Digital" N. Negroponte wrote almost 11 years ago "True personalization is now upon us. It's not just a matter of selecting relish over mustard once. The post-information age is about acquaintance over time: machines' understanding individuals with the same degree of subtlety (or more than) we can expect from other human beings" [1]. The information systems since then have evolved; many offer nifty filtration and recommendation techniques for individuals based on profiles, social or group information and behaviour knowledge. One widely experienced example is amazon.com<sup>1</sup>, which intelligently keeps track of user interest behaviour, selections, and present similar products and items that were viewed by like minded peers. The same filtering and recommendation trends are being experimented in news and information industry. More the 10,000 newspapers are available online as per listing of an online newspaper directory [2]. One can observe many online information portals offer some level of customization or personalization options. These options normally include setting appearance, weather localization, defining rule based filters for desired view of information.

<sup>1</sup> <http://www.amazon.com>

A research study in Finish newspaper industry highlights the facts that although more and more newspapers are becoming online, these digital products or services usually have ambiguous business models. They are mainly taken as experience goods or marketing tool. The online papers intensively apply customization which may not be a direct predictor of business success but it has positive affect to accumulated customer base and thus revenue. Thus customization is seen as very effective if not only business model by many online newspapers [3]. However there are many critics of “Daily Me” approach, Sunstein in his book *republic.com* states that excessive personalization or customization limits a person’s exposure to topics and point of view in a narrow domain. A broader or looser sphere of interest may expose valuable contents which one would not have selected in advance [4]. This situation led us to an experiment approach of blended filters to effectively convey news in multimodal environment. Information noise is filtered using rule based filters; semantic inferences from rule based filters, behaviour tracking and collaborative filters are added to have efficient recommender system and extend news delivery.

## 2. Blending personalization and socialization

Importance of personalization for online news industry and to provide a thorough coverage of all aspects of certain concept, the proposed news delivery framework uses a blending of

- i. rule based content filters: applied to predefined news category classifications, terms in these structures are defined using explicit relation to one another (taxonomies)
- ii. social or collaborative filters: applied to user defined classification approach of any term, the tags created by users may or not have any explicit relation, these structures are created and consumed by like minded groups (folksonomies)
- iii. inference filters: derived from semantic aspects of user defined content filters, and behaviour knowledge.

Rule based personalization is easiest to comprehend and implement, as in this approach the user and the system know ahead of time about filter conditions and the information to select. This approach can be a bit strict and may act as an information blinder in case of diverse and varying information environments such as news. This factor can be effectively dealt with a social or collaborative filtering approach; these filters are dependent on selection criteria and behaviour data of a like minded group. This gives users a feeling of interactivity with system, sharing thoughts with other people and voicing their opinion which in turn makes an online news delivery platform more attractive. The social aspects of web communities or so called Web 2.0 is making interesting transformations in existing web, this change is not merely technological but represents a mind shift [5]. In this environment the user is leveraged with blogs, wikis, social tagging and sharing tools for an enthusiastic participation. A complete social paradigm for the news domain however can not be completely practical; limited use of Wikinews<sup>2</sup> as compared to Wikipedia shows the limitation which online collaborative news publishing face. The synthesis of collaborative news keeping the contents to generally acceptable quality is a time consuming process and depends on collective user participation. This is not feasible in news domain where old news is considered no news. This blending of personal and social filters is supported with semantic and behaviour inferences. Both rule based and social filters lacks real-time adaptive personalization. This type of learning approach is added by determining user interest by examining behaviour and inferring further recommendations which is useful in certain modes of user interaction where user input to the system is difficult or probabilistic. The semantic inferences of filters from term to concept require a rich ontology of every news category, proposed system initially uses external linguistic resources (Wortschatz<sup>3</sup>) to acquire concept vocabulary of any term. Later on this conceptual vocabulary will be used to enrich the local news ontology of the system. This method may lead to certain undesired results but presenting recommenda-

<sup>2</sup> <http://en.wikinews.org>

<sup>3</sup> <http://wortschatz.uni-leipzig.de/>

tions often help users to develop a complete understanding of subject. We can see that personalization schemes discussed above covering premeditated, social and spontaneous customization have both pros and cons. We can also observe that the advantages one offers can be used to minimize the negative factors of other, thus a blended approach is best suited for news and information delivery platform.

### 3. System architecture

Online news platforms are considered very important in future journalism because the internet today is used extensively and its access interfaces are expanding beyond conventional computers. Network connected mobile device and traditional telephones linked to voice applications are increasing online information system influence. In order to extend access to the broadest possible community of users, information systems should be both adaptable, i.e., modifiable by the user, as well as adaptive, that is observe user behaviour and learn to optimally meet user requirements [6]. Our system is an effort to link online news systems to next generation information systems. The proposed system comprises of the following components.

*Harvesting:* This is an agent based module where news entities are fetched from numerous sources. The harvesting includes web mining news portals, media extraction from multimedia repositories (podcast, videocast, webcasts), RSS feeds, NNTP servers etc. Indexing information is then relayed to information preprocessing module.

*Information preprocessing:* This module extracts information about broader categories to classify the news, tokenize the keywords from news descriptors, and gather location, date, time, media type, author, publisher information to enrich metadata in the information repository.

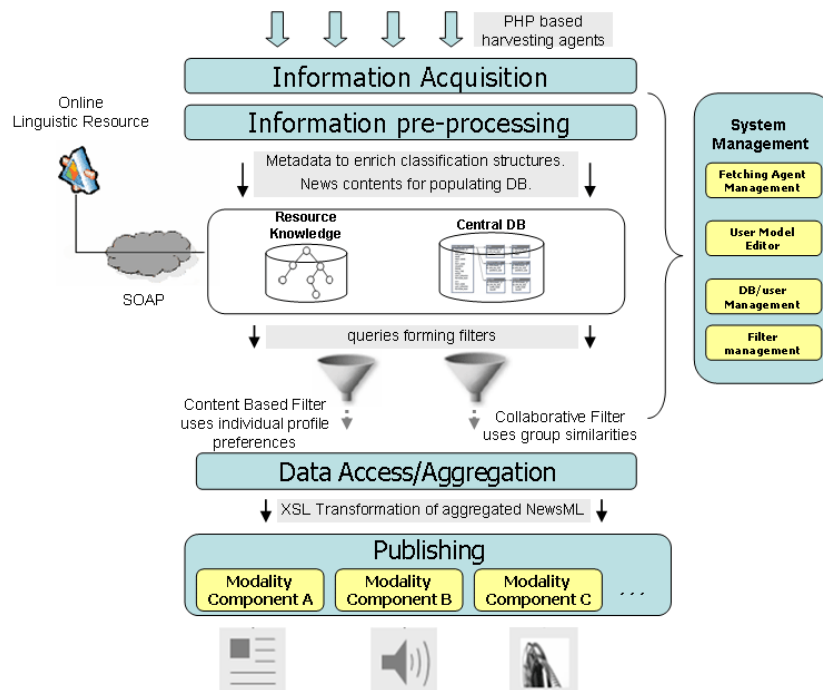


Fig. 1 News delivery platform architecture

*Filters:* User are given options to define keyword based filters, these filters can be made active or made available to selected groups. The system also gives option to create or join user groups; a group utilizes keyword filters of member users. The collaborative filters are applied to the system taxonomy and user defined metadata (tags). System users are also given option to activate semantic inferences of keyword filters and make use of behavior based recommendations.

*Data aggregation:* This module forms a personalized and generic information dataset based on individual personalization, social affiliation and modality component constraint. The standardized integration of news entities from several sources to form a seamless newscast refers to phenomena known as mashup and considered to be a beacon of future oriented internet applications.

*Publishing:* This module takes the aggregated data and passes it through appropriate transformer depending on the interaction modality and restraints of the user interface device. It also provides means of feedback mechanism to (i) give behavior data for recommendation inferences (ii) provide means of adding information tags to news entities.

*System Management:* System users have access to web based management consoles that offers registration, profile building options, selection and customization of harvesting agents. It gives user ability to view, modify, create and delete personal filters; join, create and edit groups of common interest.

Figure 1 shows the architecture of the system. A partial implementation of the proposed architecture is tested with news data from affiliated news portals. The system aggregates the contents in standardized NewsML<sup>4</sup> format which offers the ease of data interchange for multimodal publishing. XSL transforms are used to convert data into various forms of XHTML for desktop, mobile or e-ink<sup>5</sup> device browsers. NewsML data is also converted into VoiceXML<sup>6</sup> for vocal access to system via voice over IP software or standard telephone.

#### 4. Conclusion and future work

The natural instinct of humans to get what they want, when they want it, how they want it, to influence, and seek more depth are the driving factors behind this news framework. In our initial implementation we tested the combined use of various customization techniques to enhance the personalized view of news. This enhancement includes addition of multiplicity in view through social and semantic means. Cross media publishing capabilities of this framework add to the extensibility of online news. A promising start has been made with the initiation of Personalized Interactive News Cast; several techniques for metadata extraction, keyword selection, term to concept mapping, automated taxonomy building for the news domain and user modes of interaction have been tested and applied. In order to better evaluate the potential of the approach, more extensive testing with user and group interaction with system is planned. In order to have better concept oriented filtration, development of extensive news domain ontology is in progress. A wide range of user interfaces are required to be tested and integrated with the delivery system. Online news being more interactive, user adaptive and serving a larger audience than conventional media seems to be the futuristic choice of news publishing. Our work aims to determine the most extensible and useful news publishing platform.

**Acknowledgements** The research presented in this paper is partly funded by the Styria Medien AG through a funded professorship for innovative Media Technologies. The voice browsing platform by Loquendo provided for testing and research is gratefully acknowledged.

<sup>4</sup> <http://www.newsml.org>

<sup>5</sup> <http://www.eink.com>

<sup>6</sup> <http://www.w3.org/TR/voicexml21>

---

## References

- [1] Negroponte, N.: Being Digital, Chapter 13: The post-Information Age, 1995
- [2] Directory of online newspapers and magazines, <http://www.onlinenewspapers.com/>
- [3] Markku Sääksjärvi: Customization as a business model for online newspapers, 16th Bled eCommerce Conference eTransformation Bled, Slovenia, June 9 – 11, 2003
- [4] Cass R. Sunstein: Republic.com, Princeton University Press, 2001
- [5] Josef Kolbitsch, Hermann Maurer: The Transformation of the Web: How Emerging Communities Shape the Information we Consume; *Journal of Universal Computer Science*, vol. 12, no. 2 (2006), 187-213
- [6] Mark T. Maybury: Universal multimedia information access. *Universal Access in the Information Society* 2(2): 96-104 (2003)