An overview of trends in personalized content retrieval

Pogacnik, Matevz; Tasic F., Jurij; Tomazic, Saso

Abstract— ***The topic of this paper is an overview of the state of the art of personalized content retrieval and the analysis of future trends in this domain. The paper first discusses the main problems and issues related to content descriptions, which are important for content identification and selection process. The second part of the paper presents some of the most interesting topics related to personalized content retrieval such as usage of semantics, user interaction mechanisms and standardization of user models. The paper concludes with the presentation of implications that the future development might have on personalized systems.***

Index Terms—*personalization, user modeling, ontology, metadata, user-system interaction, MPEG-7 User Profile*

1. Introduction

THE amount of available content increases daily, providing us with textual documents, images, videos, music, etc. Corpora of digital(ized) content items, access to virtual museums, research publications, news (Usenet), product and service catalogues and many more have found their way into the public, mostly through the Internet. Additionally, other distribution mechanisms such as peer-to-peer networks and digital television with its distribution channels (Digital Video Broadcasting - DVB) are becoming more and more widespread. With the ascent of peer-to-peer systems, user devices are becoming sources of information, which will increase the number and heterogeneity of available content even further. Today, searching for particular information (document, image…) usually results in a vast number of hits, with a high amount of irrelevant ones. It is unlikely that hundreds of millions of users are so similar in their interests that one approach to information search fits all needs. Information and content retrieval can be more effective if individual users' interests and preferences are taken into account. Such an approach to information and content retrieval is usually called personalized content retrieval, sometimes also referred to as user modeling.

The focus of this paper is the state of the art of personalized content search and the future development trends in this field.

Manuscript received April 18, 2005.

M. Pogacnik is with the Faculty of electrical engineering, University of Ljubljana, Slovenia (e-mail: matevz.pogacnik@fe.uni-lj.si).

J. F. Tasic is with the Faculty of electrical engineering, University of Ljubljana, Slovenia (e-mail: jurij.tasic@fe.uni-lj.si).

S. Tomazic is with the Faculty of electrical engineering, University of Ljubljana, Slovenia (e-mail: saso.tomazic@fe.uni-lj.si).

2. Different content types, their indexing and descriptions

In order to understand the possibilities offered by systems for personalized content search, we should first take a brief look at existing content types and their descriptions. Content descriptions represent a basis for the identification and distinction of content items, therefore their understanding is very important.

2.1 Types of content descriptions

Content descriptions differ significantly and can be categorized based on a number of criteria.

The first categorization of content descriptions is very basic and is based on the properties of different content items. Namely, content descriptions can be either textual or non-textual. The first category is usually referred to as high-level descriptions, the second as low-level descriptions. The difference comes from the fact that images and video differ significantly from conventional data items (textual documents). This is because the extractable features are not high-level concepts like word lists (word vectors) [4], which enable relatively direct matching of search queries against data item representation. This is not the case for low-level descriptions, therefore they will be only briefly mentioned in this paper. Basic image features are the so-called low-level features, which contain information about colors, textures, shapes etc. [15], [16], [17]. Videos contain the same low-level features and have additional features like position, direction and speed of motion [1]. We should stress that, at this moment, no metadata standard seems to be widely used for image descriptions. An exception might be the MPEG-7 standard [1], even though its widespread use is also questionable. Today, text-based search techniques are still the most direct, accurate, and efficient methods for finding images and video. Text annotations can be obtained either by manual effort, embedded text, or from hyperlinked documents containing images (Google, Yahoo Ditto). The problem is that for truly accurate textual descriptions, manual indexing is required. The classification tree of high-level and low-level descriptors is presented in Figure 1.

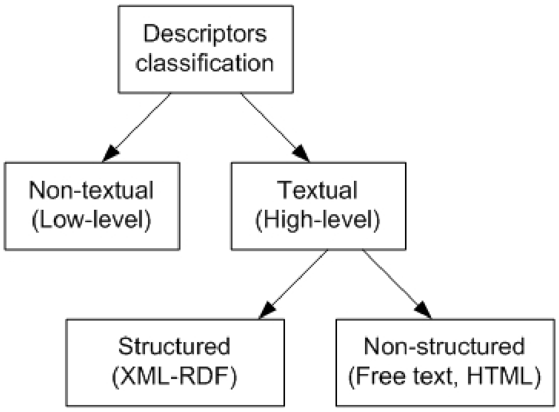

FIGURE 1: CLASSIFICATION TREE OF (MULTIMEDIA) CONTENT DESCRIPTORS

The second important categorization splits the high-level descriptions into two main categories: structured and unstructured content descriptions. The first group represents content descriptions that follow a certain metadata specification, which provides a set of rules for the proper use of metadata elements. Examples of metadata standards are TV Anytime [2] – used for TV program and video descriptions, Dublin Core [3] - a general metadata standard, ID3 – used for describing MP3 audio files, etc. Typical metadata elements used in these standards are title, genre, album, synopsis, author, lists of actors, directors, etc. In the second category are unstructured descriptions, mostly belonging to textual documents. In such cases the descriptions usually do not even exist in an explicit way, but are later generated by automatic indexing methods.

There are a number of ways to index (assign context terms to) textual documents. The automatic methods can be based on either single term or multiple terms indexing [4][5]. With single term indexing, the documents are most often represented with a standardized set of terms, named word vector. Each vector entry belongs to a predefined term and contains a weight value. The weight value is directly proportional to the frequency with which its term appears in a document and inversely proportional to the number of documents in which the term appears [4], [18]. Single terms are not ideal for indexing as their meanings out of context are often ambiguous. Term phrases on the other hand have more discriminative power. These approaches include statistical methods, probabilistic methods and linguistic methods [14]. Statistical methods use term frequency within a document as information about term relevance. Probabilistic methods generate complex index terms based on term dependence information. In practice only certain dependant term-pairs are considered as relevant. Linguistic methods are used to enhance the statistical methods by assigning syntactic class indicators to terms (verb, noun, adjective, etc.). Identification of most relevant syntactic units results in the selection of relevant phrase elements.

Opposed to automatic indexing methods are manual methods. When compared to automatic methods, they are extremely time consuming, but the generated indexes are much more exact. Moreover, some concept terms can be used to describe content even though they do not appear in the content, etc.

It is evident that content descriptions vary in a number of ways and that it is difficult to make a universal content retrieval search engine. This task is even more tedious if it is supposed to provide search results matching user's preferences. Since high-level descriptions contain keywords, concepts and semantic annotations, they are much more appropriate for annotation of a user's preferences. Therefore, the majority of personalization approaches are based on high-level (textual) descriptions.

.

3. PERSONALIZED CONTENT RETRIEVAL

3.1 Background

According to [6] the first user modeling approaches appeared in the late 1970's. At the time, they were general and modular in structure, containing explicit assumptions about users' preferences. Over time they became more topic oriented, focusing on commercial applications. The complex reasoning capabilities and assumptions of generic user modeling systems were gradually replaced by simpler approaches. On the other hand, the introduction of the client-server architecture in these systems has enabled the possibility of comparing users among each other [6].

3.2 Basic user modeling elements

The user interaction mechanisms provide basic observation mechanisms, based on which conclusions about user preferences can be made. Usually these mechanisms are implemented in the framework of the graphical user interface (GUI). User interaction mechanisms are related to the question of the type of user feedback. Users can be either asked to explicitly evaluate the suitability of a particular content item, or the system has to make implicit conclusions about the suitability of content, based on content usage (select, ignore, delete, record etc.). A combination of both approaches is also used. One would normally expect better results from the explicit feedback approach, since implicit feedback systems have to make decisions relying on incomplete and uncertain information. However, some authors [28] report improved results in the domain of television programmes using implicit feedback obtained through the analysis of the Personal Video Recorder (PVR) usage history.

User models store information about user preferences. They can have a special structure (decision trees, hierarchical structure, keyword vectors etc.) or can contain only lists of content items selected/rejected by the user. The first standardized approach to user modeling in the multimedia (MM) domain is presented in Section 4.

Content selection algorithms can be standalone computational procedures (e.g. similarity calculations) or can be a part of the user model structure (e.g. decision trees). Today we can speak about a number of recommendation approaches [9]: collaborative recommenders, content-based recommenders, demographic recommenders, utility-based recommenders, knowledge-based recommenders and also group-oriented recommenders. The main two approaches are content-based filtering (CBF) [18],[19],[20], [21], [22] and collaborative filtering (CF) [23], [24], [25], [28]. The difference between the two is in the process of identification of suitable content for the user. When using the CBF, the suitability of particular content for the user is estimated through direct comparison of content description (meta-data) and the model of the active user. When using the CF, first the similarity between users is estimated, and then content liked by users 'similar' to the active user is recommended. The advantages of collaborative systems are the independence of the representation of content items and the capability of recommending content items from cross-genre niches (music, movies, books...). The drawbacks are problems of new users and new content items. New users have not yet rated enough content items, while new content items have not yet been rated by a sufficient number of users. The content-based approach similarly suffers from the problem of new users and is not capable of cross-genre recommendations. Both approaches also need large historical data sets in order to provide quality recommendations. The mentioned approaches make recommendations to individual users, while another approach, which has come up recently, makes recommendations for groups of users (viewers) [8]. They all have their advantages and drawbacks and are not universally suitable for all fields of usage. Therefore, they are often combined into hybrids (see [9] for an overview).

4. *THE TRENDS IN THE DOMAIN OF PERSONALIZED CONTENT RETRIEVAL*

It is clear that personalized content recommender systems have a future and that they will soon appear in most of the content retrieval domains. Regardless of the content types, the personalization approaches used today have similar usage scenarios. Users browse content listings, rate content items implicitly or explicitly and get content suggestions. The drawbacks of today's systems are context unaware keyword approaches, relatively primitive means of interaction with systems resulting in poor implicit feedback, unawareness of a user's mood, etc. At this point, the question arises: what improvements can we expect in the future from personalized content recommended systems?

In the following subsections we mostly focus on the issues of user/system interaction and the seamless acquisition of information about user’s preferences and moods. Later we address the issue of contextual understanding of content and the existing approaches. The section concludes with a presentation of the MPEG-7 user modeling standard [28]proposed by the MPEG initiative.

4.1 User-system interaction

One of the most important issues related to the development of personalized systems, is the advent of digital devices used for access to content at any time. All of these devices, from handheld Personal Digital Assistants (PDAs) to the Personal Video Recorders have gained in processing power and storage space. In this sense they are becoming very similar to computers and can run Java applications, connect to the Internet, download, store and display content items of any type [13]. This means that the observation of user preferences and a user’s content selections is no longer bound to a single device but is possible in almost any situation at any time: At work, at home on a trip, etc. Consequently, personalized systems will be able to get much more information about the user. This approach will of course require more sophisticated user modeling techniques. An especially difficult issue is the acquisition of user preferences from different devices at different times and in different contexts, and their synthesis into a single user model.

Related to this issue is the question of user-system interaction. New types of user interfaces will have to be developed, which will enable seamless communication and interaction. Some authors report that users like to communicate with computers as if they were human; therefore new user interfaces using humanoid avatars and voice processing are being developed [10]. These enhancements may not directly improve the quality of personalized selections, but will nevertheless improve the overall functionality of such systems, as the users will benefit from easier and more ’human-like’ interaction.

It is well known that users prefer communication based on interaction that requires as little as possible of explicit feedback. The first step in this direction is the possibility to analyze content usage history in home digital devices, from which some conclusions about user preferences can be made [1]. This approach basically uses the information about a user's actions related to particular content items like movies, videos, etc. Using the information about how many times a particular movie of a certain genre, with a certain cast crew, etc., was watched, or maybe deleted, conclusions about user's preferences can be made (see Section 4). Some nonintrusive methods of getting user feedback are presented in [26], [27]. However, nonintrusive ratings (such as the time spent reading an article) are often inaccurate and cannot fully replace explicit ratings provided by the user.

More straightforward solutions use very simple explicit feedback techniques like "thumbs-up" "thumbs-down" used in the TiVo system, where a (positive or negative) rating of a particular content item requires only a single push of a button.

At the same time much more intriguing and controversial approaches are being developed. In the most mature stage of development seem to be the systems that are able to track a user's focus on the screen of any device [10]. By combining this information with the information about the contents on the screen they can get very reliable information about the user's current interests. This technology seems especially useful for the analysis of interest in textual documents and Web pages. Another step forward from the analysis of a user's (visual) focus seems to be in the capturing of additional information about the user's state of mind, his/her current mood, fatigue, etc. It is well known that the appropriateness of content selection is dependant on these parameters, therefore, the ongoing research projects are investigating the possibilities to identify people's mood based on their biometrical signals (heart rate, temperature, sweating, amount of exhaled gases, etc.) [11]. These systems will be able to identify a user's current mood and will combine this information with the information about his/her general preferences in order to recommend suitable music or movie selection for example. Complementary initiative comes from the authors of the TV Anytime standard [2], which have in the content description schemes foreseen a field describing the type of mood for which some TV programme is suitable.

While being enthusiastic about the progress of the ongoing research, we should be aware of the accompanying issues that these systems might bring. The main issue is undoubtedly the security and privacy of a user's data, which not only includes the preferences but also the information about a user's health condition or reactions to particular content types, captured by the future recommender systems. Even though these issues are beyond the scope of this paper, we should bear in mind that inadequate protection of such information may bring a lot of potential trouble to future users and also affect the wide acceptance of these products.

4.2 Semantic search and understanding of content

The problem of contextual understanding of keywords is mostly present in case of textual documents, especially web pages. The current situation in the web is such that data is generally "hard coded" in HTML files. The concept terms used are semantically ambiguous, so there is no way of telling which of the possible meanings is the right one. For example: a search query "north pole" may provide thousands of result pages with content about the famous geographical location, a company with the same name or even a pub. This ambiguity is transferred into the user profile when information about "preferred" keywords is extracted from web pages and stored. These words are handled using statistical tools, which are mostly based on advanced 'counting' of word occurrences. However, taking into account the wider context of the documents would give much more information about the documents and consequently user preferences. In order to resolve this issue, an initiative called the Semantic web has been started [12]. To put it simply, the idea is to describe specific term meanings, their relationships and context information with the help of schemas and ontologies. Once such information is attached to documents they become much more useful and easily processable, e.g. for categorizing. The descriptions are made using RDF (Resource Description Format). With the development and use of inference logic, this approach may become a very powerful tool for information processing and retrieval.

The situation is similar with multimedia content types. A solution to this problem is being offered by the MPEG-7 standard [1]. Namely, the MPEG-7 standard aims to describe all types of MM content (audio and speech, moving video, still pictures, graphics and 3D models) including information on how objects are combined in scenes and their semantic meanings and relationships (temporal, spatial, etc.). The standard has powerful mechanisms for content description, but is on the other hand very complex, therefore widespread use is still questionable.

The problem of semantic content descriptions unfortunately does not end when content is described using structured, semantically sound metadata standards. Even if every content creator would take enough time to describe each of the created content items according to an appropriate metadata standard, there would still be many similar content items described with different metadata standards. Agreements in metadata specifications are necessary to achieve semantic interoperability among communities. However, more than one metadata specification exists and it is impossible to achieve a general consensus about the use of one, and only one specification. To minimize the time spent in maintaining content metadata and to maximize the use among a broad range of users, the compatible parts of one metadata specification should be available in other related specifications [7]. Therefore, metadata crosswalks are necessary to facilitate interoperability among heterogeneous, but at the same time related specifications. A crosswalk is an explicit mapping of one metadata specification to another. Creating a crosswalk is a difficult and error-prone task, which requires in-depth knowledge and experience of the metadata specifications being mapped [7]. Developing a crosswalk between two metadata specifications requires steps like harmonization, semantic mapping and additional rules. An example of metadata mapping is presented in Figure 2. In this example, the semantically equivalent metadata elements such as abstract and keywords are both present in two different metadata specifications (on the left and right side of the Figure 2) and are labeled with the same colors.

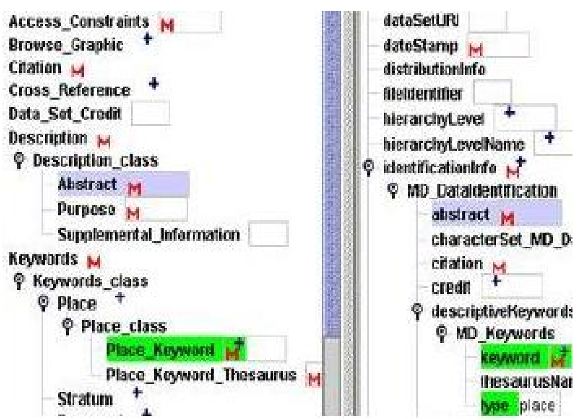

FIGURE 2: AN EXAMPLE OF SEMANTIC METADATA MAPPING BETWEEN TWO METADATA STANDARDS [7].

Even though the identification of semantically matching concepts may seem simple for a human, it is not so for computers, especially when considering the complex structures that the metadata standards may have.

4.3 User modeling standards (MPEG-7 user profile)

Despite the development in the field of personalized content selection, there are surprisingly few standards regarding the descriptions of user preferences. Apart from the widespread use of word vectors, there is actually only one standard in this field: the MPEG-7 user profile. The idea was to standardize the description of user preferences and the means of usage of MM content [1].

In order to enable the exchange of information about user preferences with 3rd party services, they also decided to standardize the information about content usage history. This part of the standard is specifically oriented towards user interaction with the personal digital video recorder or similar devices, but can also be used elsewhere. For this purpose, two description schemes (DS) were designed: The UsageHistoryDS and the UserPreferencesDS.

The UsageHistoryDS (see Figure 3) enables annotation of user's actions (play, record, delete etc.) with respect to a particular content item. These annotations form a detailed record of the interaction history between the user and the content, which represents a basis for meaningful representation of user's preferences. The latter are 'stored' in the UserPreferencesDS.

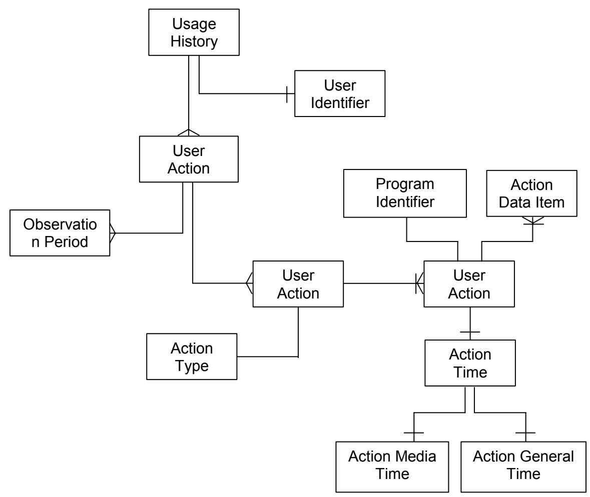

FIGURE 3: THE USAGEHISTORY DESCRIPTION SCHEME, USED FOR DESCRIBING THE INFORMATION ABOUT THE INTERACTION BETWEEN THE USER AND CONTENT.

UserPreferencesDS (see Figure 4) enables annotations of user's preferences regarding content creation (favorite titles, actors, directors, locations of content creation etc.), content classification (favorite genres, subjects, languages, etc.), source preferences (favorite media formats, dissemination mediums etc.) and some others. Although the standard mandates the structure of both description schemes, it does not mandate the algorithms used for mapping of usage history to user preferences. These are left open for the developers.

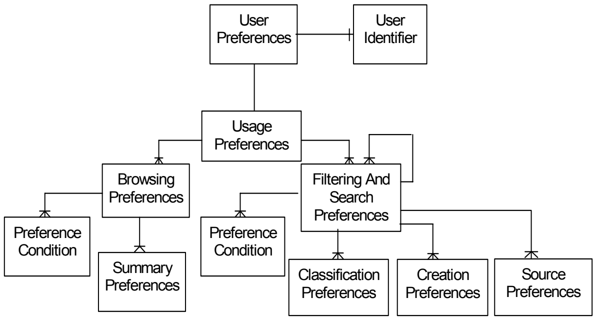

FIGURE 4: THE USERPREFERENCES DESCRIPTION SCHEME, USED FOR DESCRIBING THE INFORMATION ABOUT USER'S PREFERENCES.

A typical data and content flow of a personalized retrieval system, using user preference description and usage history description is presented in Figure 5.

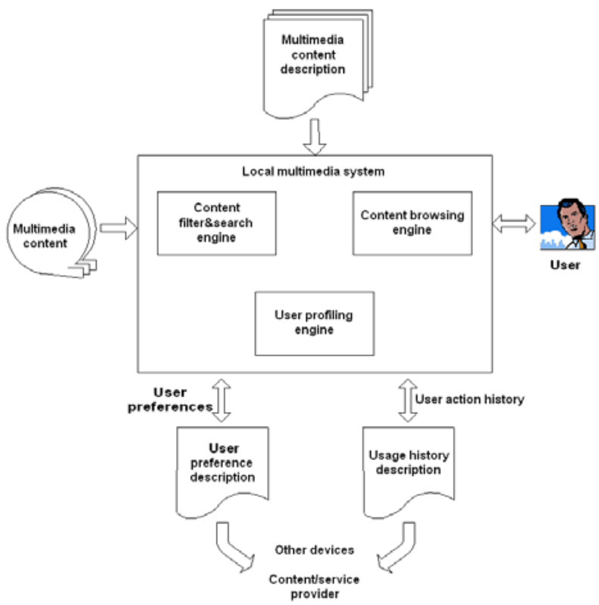

FIGURE 5: MM SYSTEM AND USER INTERACTION [1].

5. CONCLUSION

Personalized content retrieval is becoming a widespread technology, penetrating in different fields of usage. A number of personalization systems exist today, selecting and recommending content to a vast number of users. In the future, users can expect improvements in the field of user interfaces, which will enhance the interaction with systems, contextual understanding of terms and topics, enable exchange of user related information and consequently personalize experience on almost any device. This scenario will come true only if supporting technologies are provided, like unobtrusive biometrical sensors, further advancement of digital devices in terms of processing power and available storage, improved personalization algorithms etc. These approaches may seem very controversial, as they provide an insight into the user's most personal world by analyzing his/her mood and even knowing his/her biometrical data. Therefore, the user should have the final word when deciding which details (if any at all) of his/her user model should be shared or used for other purposes. We should bear in mind that data, gathered by personalized systems are very interesting to many commercial companies, as well as to individuals. We should make sure that this data is not misused and that user privacy is not compromised. In the opposite case, such systems and products will never be accepted by a wide audience.

ACKNOWLEDGMENT

This work would not have been possible without the grant of the Slovenian Ministry of Education, Science and Sport, No. S-64303 and our work on the myTV project (IST-1999-11702).

REFERENCES

[1] B.S Manjunath, P. Salembier and T. Sikora, "Introduction to MPEG-7, Multimedia content description interface," John Wiley & sons Ltd., England, 2002.
[2] TVAnytime: Specification series: S-3 metadata, SP0003 v1.3, part, A, 2002. ftp://tva:tva@ftp.bbc.co.uk/pub/Specifications/SP003v13.zip (URL).
[3] Dublin Core Metadata Initiative, http://dublincore.org/ (URL)
[4] D.J Hand, H. Manilla and P. Smyth, "Principles of data mining", The MIT Press, Cambridge, Massachusetts, USA, 2001.
[5] V.V. Raghavan, W.I. Grosky and R. Kasanagottu, "Information Retrieval on the World Wide Web," *IEEE Internet Computing, Vol. 1, No. 5,* 1997.
[6] A. Kobsa, "Generic user modeling systems," *User Modeling and user adapted interaction, Vol.11, pp. 49-63*, 2001.D. O' Sullivan, B. Smyth, D.C. Wilson, K.
[7] E.L. Bermudez, "ONTOMET: Ontology metadata framework", Ph.D. thesis, Drexel University, Philadelphia, USA, 2004.
[8] J. Masthoff, "Group modeling: Selecting a sequence of television items to suit a group of viewers," *User Modeling and user adapted interaction, Vol.14, pp. 37-85,* 2004.
[9] R. Burke, "Hybrid Recommender Systems Survey and Experiments," *User Modeling and User Adapted Interaction, Vol.12, pp. 331--370*, 2002.
[10] M. Witkowski, Y. Arafa and O. de Bruijn, "Evaluating User Reaction to Character Agent Mediated Displays using Eye-tracking Technology," *Proceedings of Workshop on Information Agents in E-commerce; Agents and Cognition*, AISB Convention,York, UK, 2001.
[11] P. Goulev and E. Mamdani, "Investigation of human affect during electronic commerce activities," *IEE Eurowearable'03,* Birmingham, UK, 2003.
[12] S. Palmer, The semantic web: An Introduction, http://infomesh.net/2001/swintro (URL), 2001.
[13] DVB - Multimedia Home Platform Specification 1.1 draft, tm2485, tam668r12, European Broadcasting Union,

http://www.etsi.org (URL), May 2001.

[14] M. Berthold and M. J. Hand (ed.), "Intelligent Data Analysis," Springer Verlag Berlin, 2002.

[15] S. Belongie, C. Carson, H. Greenspan and J. Malik, "Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval," Proceedings of the Sixth International Conference on Computer Vision, 1998.

[16] N. Vasconcelos and A. Lippman, "Embedded Mixture Modeling for efficient probabilistic content-based indexing and retrieval", *In SPIE Multimedia Storage and Archiving Systems III*, 1998.

[17] X. Zhu, H. Zhang and L. Wu, "New query refinement and semantics integrated image retrieval system with semiautomatic annotation scheme," *Journal of electronic imaging 10(4), pp. 850-860*, 2001.

[18] G. Salton, Automatic Text processing, Addison-Wesley, 1989.

[19] Y. Zhang, J. Callan and T. MInka, "Novelty and redundancy detection in adaptive filtering," Proc. 25th Ann. International ACM SIGIR Conf., pp.81-88, 2002.

[20] B. Bezerra, F. Carvalho, G. Ramalho and J.D. Zucker, "Speeding up Recommender Systems with Meta-prototypes," *Lecture Notes in Computer Science, Advances in Artiffcial Intelligence, 16th Brazilian Symposium on Artiffcial Intelligence - SBIA 2002,* pp. 227--236, 2002.

[21] J. Zimmerman, K. Kurapati, A. Buczak, D. Schaffer, S. Gutta and J. Martino, "TV Personalization system. Design of a TV Show Recommender Engine and Interface," *Personalized Digital Television. Targeting Programs to Individual Viewers. Kluwer Academic Publishers, Human-Computer Interaction Series, Vol. 6*, 2004.

[22] D. Billsus and M.J. Pazzani, "A hybrid user model for news story classiffcation, *"In Proceedings of the Seventh International Conference on User Modeling*," Banff, Canada, pp. 99-108, 1999.

[23] D. Billsus and M.J. Pazzani, "Learning collaborative information filters," Proc. Int'l conf. Machine learning, 1998

[24] K. Yu, X. Xu, J. Tao, M. Ester, and H.-P. Kriegel, "Probabilistic Memory-Based Collaborative Filtering," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 1, pp.56-69, 2004.

[25] B. Sarvar, G. Kaypis, J. Konstn, and J. Riedl, "Item based collaborative filtering recommendation algorithms," Proc. 10th Int'l WWW Conf., 2001.

[26] D. W. Oard and J. Kim, "Implicit feedback for recommender systems," Proc. Recommender systems. Papers from 1998 workshop, Technical report WS-98-08, 1998.

[27] A.I. Schein, A. Popescul, L.H. Ungar, and D.M. Pennock, "Methods and Metrics for Cold-Start Recommendations", 25th *Ann. Int'l ACM SIGIR Conf.,* 2002.

[28] McDonald, and A. Smeaton, "Improving the quality of the personalized electronic program guide", *User Modeling and user adapted interaction, Vol.14, No. 1, pp. 5-36,* 2004.