Ontology-based Personalization and Recommender System in Digital Libraries

Daniela Kjurchievska[1]

Abstract – **The widespread use of the internet has resulted in Digital Libraries (DL) that are increasingly used by diverse communities of users for diverse purposes, and in which sharing and collaboration have become important social elements. In that sense, Academic Digital Libraries (ADL) have emerged as a result of current technology in learning and researching environment, which offers myriad of advantages especially to students and academicians from one side, and advances in computing and information system technologies from other side; thus had been introduced in universities and to the public. This is due to dramatic change in learning environment through the use of Digital Library System (DLS) which impacts on these societies' way of performing their study/research. A simple search function increasingly leads to user dissatisfaction as user's needs become more complex and as the volume of managed information increases. Proactive DL, where the library evolves from being passive and untailored, are seen to offer a great potential to overcome those issues and include techniques such as personalization and recommender systems. Personalization is viewed as an application of data meaning and machine learning techniques to build models of user behavior that can be applied to the task of predicting user needs and adapting future interaction with the ultimate goal of improved user satisfaction.**

Keywords – **Personalization and Recommendation Systems, Digital Library, Ontology, Ontology-Based Personalization System.**

I. INTRODUCTION

The emerging generation of DL is more heterogeneous along several dimensions. The collections themselves are become more heterogeneous in terms of their creators, content, media, and communities served. The range of library types is expanding to include long-term personal digital libraries, as well as DLs that serve specific organizations, educational needs, and cultural heritage and they vary in their reliability, authority, and quality. The user communities have become heterogeneous in term of their interests, background, skill levels, ranging from novices to experts in a specific subject area. The growing diversity of DLs, the communities accessing them, and how the information is used requires that the next generation of DLs be more effective at providing information that is tailored to a person's background knowledge, skills, tasks, and intended use of the information.

Information retrieval technologies have matured in the last decade and search engines do a good job of indexing contents available on the Internet and making it available to users, even the user knows exactly what he is looking for but often, search engines themselves can return more information than the user could possibly process. Also, most widely used search engines use only the content of DL documents and their link structures to access the relevance of the document to the user's query.

Hence, no matter who the user of a search engine is, if the same query is provided as input to the search engine, the results returned will be exactly the same.

The need to provide users with information tailored to their needs led to the development of various information filtering techniques that build profiles of users and attempt to filter large data stream, presenting the user only those items that it believes to be of interest.

The goal of personalization is to provide users with what they want or need without requiring them to ask for it explicitly. This doesn't in any way imply a fully automated process, instead it encompasses scenarios where the user is not able to fully express exactly what they are looking for, but the interaction with an intelligent system can lead them to items of interest.

Intelligent Techniques for Personalization is about leveling all available information about users of the DL to deliver a personal experience. The "intelligence" of these techniques is at various levels ranging from the generation of useful, actionable knowledge through to the inferences made using this knowledge and available domain knowledge at the time of generating the personalized experience for the user. As such, this process of personalization can be viewed as an application of data mining and hence requiring support for all the phases of a typical data mining cycle including data collection, pre-processing, pattern discovery and evaluation, in off-line mode, and finally the deployment of the knowledge in real-time to mediate between the user and the DL.

II. THE PERSONALIZATION PROCESS

Personalization can be defined as the way in which information and services can be tailored in a specific way to match the unique and specific needs of the individual user or community of users. This is achieved by adapting the presentation and/or the services presented to the user by taking into account the user's tasks, background, history, device, information needs, location, etc., essentially the user's context. Personalization can be user-driven which involves a user directly invoking and supporting the personalization process by providing explicit input, i.e., the user explicitly initiates actions and provides example information in order to control the personalization. On the other hand, personalization

[1]Daniela Kjurchievska, Technical Faculty at University "St. Kliment Ohridski" - Bitola, Republic of Macedonia, E-mail: daniela.curcievska@uklo.edu.mk

can be completely automatic, where the system observes some user activity and identifies the input used to tailor some aspects of the system in personalized way. These two examples of user-driven and automatic personalization are the extreme ends of the spectrum and many personalization tools will have elements of both approaches.

III. Classification of Approaches to Personalization

In this section we discuss various dimensions along which personalization systems can be classified based on the data they utilize, the learning paradigm used, the location of the personalization and the process that the interaction takes with the user.

A. Individual Vs Collaborative

The term personalization impresses upon the individuality of users and the need for systems to adopt their interfaces to the needs of the user. This requires data collected on interactions of users within the system to be modeled in user-centric fashion.

A personalization system may choose to build an individual user model, which is a data structure that represents user interests, goals and behaviors. The more information a user model has, the better content and presentation will be tailored for each individual user. The user model is created through a user modeling process in which unobservable information about a user is inferred from observable information from that user, for example, using interactions with system. User model can be created using a user-guided approach, in which the models are directly created using the information provided by each user, or an automatic approach, in which the process of creating a user model is hidden from the user. This approach commonly requires content descriptions of items to be available and is often referred to as content-based filtering systems.

An alternative approach to recommendation is not only the profile for active user but also other users with similar preferences, referred to as the active user's neighborhood, when recommending items. In that sense, the DLs may be viewed as common working place where users may become aware of each other, open communication channels, and exchange information and knowledge with each other or with experts. This means that it is quite possible that users may have overlapping interests if the information in a DL matches their expectations, backgrounds, or motivations. Such users might well profit from each other's knowledge by sharing opinions or experiences or offering advice. This approach is referred to as social or collaborative filtering.

A major disadvantages of approaches based on an individual profile include the lack of serendipity as recommendation are focused on the users previous interests. Also, the system depends on the availability of content descriptions of the items being recommended. On the other hand the advantage of this approach is that it can be implemented on the client side, resulting in reduced worries for the user regarding privacy and improved data collection for implicit user preference elicitation.

The collaborative approach also suffers from a number of disadvantages, not least the reliance on the availability of rating for any item prior to it being recommendable, often referred to as the new item rating problem. Also, a new user needs to rate a number of item before he can start to obtain useful recommendations from the system, referred to as the new user problem.

B. Reactive Vs Proactive

Reactive approaches view personalization as a conversational process that requires explicit interaction with the users either in the form of queries or feedback that is incorporated into the recommendation process, refining the search for the item of interest to the user.

Proactive approaches, on the other hand, learn user preferences and provide recommendations based on the learned information, not necessarily requiring the user to provide explicit feedback to the system to drive the current recommendation process. Proactive systems provide users with recommendations, which the user may choose to select or ignore.

C. User Vs Item Information

Personalization systems vary in the information they use to generate recommendations. Typically, the information utilized by these systems includes:

- Item Related Information: This includes content descriptions of the items being recommended and a product/domain ontology
- User Related Information: This includes past preference rating and behavior of user, and user demographics

Most systems that use user related information, tend to be based on past user behavior such as the items they have bought or rated in the past.

In addition to the system that depend solely on item related or user related information, a number of hybrid systems have been developed that use both types of information.

D. Memory Vs Model Based

As mentioned before the process of personalization consists of an off-line and online stage. In the off-line stage the key tasks are the collection and processing of data pertaining to user interests and the learning of a user profile from the data collected. Learning from data can be classified into memory based learning (also known as lazy learning) and model based learning (or eager learning) based on whether it generalizes beyond the training data when presented with a query instance (online) or prior to that (off-line).

Traditional Collaborative filtering and content-based filtering based systems that use lazy learning algorithms, are examples of memory-based approach to personalization, while

item-based and others collaborative filtering approach that learn models prior to deployment are examples of model-based personalization systems.

E. Client Vs Server Side

Approaches to personalization can be classified based on whether they have been developed to run on a client side or on the server side. The key distinction between these personalization approaches is the breadth of data which is available to the personalization system. On the client side, data is only available about the individual user and hence the only approach possible on the client side is Individual.

On the server side, the system has the ability to collect data on all its visitors and hence both Individual and Collaborative approaches can be applied.

IV. Personalization Techniques

The traditional systems for personalization of DL, based on the previously described approaches are:

- Content-Based Filtering
- Traditional Collaborative Filtering
- Model-Based Technique
  - Item-Based Collaborative Filtering
  - Clustering-Based Approach

Beside these traditional personalization systems, a number of hybrid approaches to personalization have been proposed. These hybrid recommenders have been motivated by the observation that each of the recommendation technologies in the past has certain deficiencies that are difficult to overcome within the confines of a single recommendation approach.

One form of hybrid recommender that has recently been gaining a lot of attention is that which is based on the use of ontologies to describe the relationship between all the elements which take part in a DL scenario of use.

V. Ontology-Based Personalization System for DL

Every day a huge amount of newly created information is electronically published in DL, whose aim is to satisfy users' information needs. Both, the collectors and the user communities become more heterogeneous and this growing diversity has changed the initial focus of providing access to digital content and transforming the traditional services into digital ones to a new handicap where the next generation of libraries should be more proactive offering personalized information to their users taking in consideration each person individually.

In order to build such personalization system, several multidisciplinary aspects must be addressed: first, there are cognitive and behavioral aspects that determine the way users perform search and examine the obtained results. Second, personalization issues must be addressed from a user-centered point of view, under the approach of human computer interaction and third, there are technological and knowledge engineering aspects related to the way all this information is structured for both updating and querying purposes. In this point, Ferran et al. purposed the set of desired functionalities and requirement of an ideal scenario for DL which include personalization capabilities by means of ontologies. The use of ontologies for describing the possible scenarios of use in a DL brings the possibility of predicting user requirement in advance and to offer personalized services ahead of expressed need. They suggest building ontologies by using other sub-ontologies which describe the basic element of the personalization system: users, digital resources, action, navigational profiles, etc.

In this system, collaborative filtering approaches are used for guidance and providing recommendation to the user. That means that the system automatically collects information about the user's action and determines the relative importance of each content by weighting all the collected information among the large amount of users. In this DL scenario of use both navigation techniques are also valid, simple searches starting from a single search term or advanced search using multiple criteria. The basic idea of this approach is the efforts for finding a useful piece of information in DL carried out by an individual, and which can be stored in structured way and then shared for future users with similar necessities.

Two elements determine the functionalities of the desired PS, the user's profile which include navigational history and user preferences, and the information collected from navigational behavior of the DL users. User profile should include all the information relevant to the user: personal information, which can be publicly made available by each user in order to facilitate the discovery of similar interest and navigational history and behavior records, which will be used altogether with the personal information by PS to build the set of recommendations that will help each user in browsing and searching the DL.

Depending to the users' navigation, two different behaviors can be identified, exploratory and goal-oriented navigation. The exploratory navigation can be mainly oriented to obtain a general vision of the available resources in the library. Depending on user profile, this navigation would have different implicit intentions. In the case of goal-oriented navigation, it is usually considered that the user is looking for a resource. Both searches can be classified in different use cases. For example, in case of searching for an author, if the user is a student, the recommendations associated to search results should be oriented to the area of the course subject, taking into account the navigation of other students and also the teacher's recommendations. If the user is a researcher, recommendation should be oriented by different criteria dependent on the searches that have been carried out by other investigators, or to magazines, books and conferences where searched author had published. Recommendation are generated by using the knowledge extracted from the searching and browsing profiles of users with similar interests and knowledge integrated in the ontology, or by following citation of similar documents.

The use of ontology could be also interesting when it comes to incorporating new functionalities into existing DL, by describing the relationships between elements. For instance, if

a teacher defines one or more books as recommended bibliography for given subject, students enrolled in such subject should be aware of those books when performing searches related to the subject.

It is important to clarify that ontologies are not built for describing the contents of a DL, but for describing the way users browse and search contents, with the aim of building a PS based on accurate recommendations. The ontology itself is composed of sub-ontologies which describe all the interesting relationships between the elements of the small micro-scenarios that emerged in using of DL system. Into the DL, the creation of ontology will help library managers to construct tailored libraries for each subject. Every library is built on the explicit recommendation from a teacher. With ontology, those specialized libraries could be built from the use that previous student gave to the resources and new information could be added from use of the library by experts.

It is remarkable that the use of ontologies can be also extended to implement and transfer the concept of user profiles and user navigational behavior to other DLs and databases, so when a user leaves one service to connect into another one, the user profiles can be transferred from one database to another through the appropriate semantic web services.

VI. CONCLUSION

The need to provide users with information tailored to their needs led to the development of various information filtering techniques that build profiles of users and attempt to filter large data stream, presenting the user only those items that it believes to be of interest.

Beside the traditional approaches for personalization systems, ontologies are powerful tool for describing complex scenarios of use such as in DLS. The use of ontologies promotes the integration of new scenarios into existing ones. New system functionality and requirement can be added by including the appropriate description into the ontology framework that defines the DL scenario of use.

REFERENCES

[1] B. Mobasher and S.S. Anand , “Intelligent Techniques for WEB Personazation”, pp. 1–36, Springer-Verlag Berlin Heidelberg 2005

[2] N. Ferran, E. Mor, J. Minguillón, "Towards personalization in digital libraries through ontologies", Library Management, Vol. 26 Iss: 4/5, pp.206 – 217, 2005

[3] M. Lytras, M. Sicilia, J. Davies, V. Kashyap, "Digital libraries in the knowledge era: Knowledge management and Semantic Web technologies", Library Management, Vol. 26 Iss: 4/5, pp.170 – 175, 2005

[4] M.E. Renda, U. Straccia , “A personalized collaborative Digital Library environment: a model and an application “, Original Research Article Information Processing &Manag., Volume 41, Issue 1, Pages 5-21, January 2005

[5] S. Grigoriadou, A. Kipourou, E. Mouratidis , M. Theodoridou , “Digital Academic Libraries: an important tool in engineering education”, 7th Baltic Region Seminar on Engineering Education, St Petersburg, Russia, 4-6 September, 2003

[6] A. F. Smeaton, J. Callan, ”Personalization and recommender systems in digital libraries”, International Journal on Digital Libraries, Volume 5, Issue 4, pp 299-308, 2005

[7] E. Frias-Martinez, G. Magoulas, S. Chen, R. Macredie, “Automated user modeling for personalized digital libraries”, International Journal of Information Management, Volume 26, Issue 3, Pages 234–248, 2006

[8] M. Kumar,“Academic Libraries in Electronic Environment: Paradigm Shift”, International Conference on Academic Libraries (ICAL), New Delhi, 2009

[9] M. Nisheva-Pavlova, P. Pavlov, “Search Engine in a Class of Academic Digital Libraries” In: T. Hedlund, Y. Tonta (Eds.), “Publishing in the Networked World: Transforming the Nature of Communication. 14th International Conference on Electronic Publishing, 16-18 June 2010, Helsinki, Finland”, Edita Prima Ltd, Helsinki, 2010, ISBN 978-952-232-085-8, pp. 45-56. (2010)

[10] M. Nisheva-Pavlova, “Providing and Maintaining Inter-operability in Digital Library Systems’, Proceedings of the Fourth International Conference on Information Systems and Grid Technologies (Sofia, May 28-29, 2010), St. Kliment Ohridski Uiversity Press, 2010, ISBN 978-954-07-3168-1, pp. 200-208. (2010)

[11] R.A. Razilan, W. Dollah, F.A.Saaid, S. Diljit, “Academic Digital Library’s Evaluation Criteria: User-Centered Approach”, International Journal of Social and Human Sciences, 2009