

Third Int'l Workshop on "Personalized Access, Profile Management, and Context Awareness in Databases" (PersDB 2009)

Sihem Amer-Yahia
Yahoo! Research New York
111 W 40th Street, New York
NY, USA
sihem@yahoo-inc.com

Georgia Koutrika
Computer Science Dept., Stanford University
353 Serra Mall, Stanford
CA, USA
koutrika@stanford.edu

1. INTRODUCTION

Proliferation of database-driven web sites has brought upon a plethora of information access and dissemination applications. Monitoring and trading stock portfolios, news notification, weather tracking, and even simple search are just a few examples. In addition, emerging applications such as Web-based communities, wikis, social networks, mashups and folksonomies enhance creativity, information sharing, and collaboration among users providing richer interaction possibilities. Now, users can not only access content but they can also generate, share and modify content (both theirs and others'), compose their applications, enhance their interface, etc.

In all these applications, different notions of user information, such as preferences, community memberships and social interactions, and context information, such as a user's social network, location, time, and other features of a user's environment, are of paramount importance in order to improve and personalize user experience. In this context, new challenges emerge for user-centric, context-aware database systems for storing and managing different aspects of user and context information, for data management and computing taking into account personal, social and contextual information about users and for customizability of their behavior.

There are several research efforts that are trying to address research problems that arise when developing such systems. For example, user preferences may be associated with the search process and interpreted as preferences over the desired system answers (e.g., [3, 4]) or they may be associated with aspects of the data management process specifying user requirements for response times, data freshness, quality of service etc, and used in mechanisms, such as caching, query admission control, resource allocation, and scheduling (e.g., [2, 5]). Likewise, a user's context can be the social network or community where a user belongs. A user may belong to more than one network and community. Identification of a user's community is important as well as building personalized recommendations, user collaboration and sharing schemes that address communities [1, 6].

The purpose of the PersDB Workshop ("Personalized Access, Profile Management, and Context Awareness in Databases") is to bring together researchers and practitioners both from the academia and the industry and provide a forum for presentation of the latest research results, new technology developments, and new applications in the areas of personalized access, profile management, and context awareness in database systems.

The 3rd Int'l Workshop on "Personalized Access, Profile Management, and Context Awareness in Databases" (PersDB 2009) was held in conjunction with VLDB in Lyon, France. We have received 18 papers spanning different topics including context-driven

databases, recommendations, search and social networks. We accepted 6 papers, having a healthy acceptance rate of about 33%.

2. WORKSHOP OUTLINE

2.1 Invited Talk

Recommender systems have been extremely successful in reaching relevant content to users and they are very popular on sites, such as Amazon, Netflix and Google News. These systems incorporate endorsements of items by other users and/or ratings provided by the same user and they recommend items to the user that are likely to be of interest. *Laks V.S. Lakshmanan* (University of British Columbia, Canada), who gave the keynote talk titled "*Recommender Systems Revisited: from Items to Transactions*", described a new idea: developing recommendation strategies and systems not just for recommending items but for users performing transactions.

The invited speaker described the notion of an *exchange market* (e.g., *ReadItSwapIt.co.uk*), a social network where users register items (e.g., toaster, lawn mower) they are willing to give away to other users in exchange for items in their wish list which they have registered with the system. Users either swap items or more generally exchange items in cycles. He described approximations as well as heuristic algorithms for recommending exchange transactions to users assuming relatively simple transaction cycles.

Recommending transactions where items are swapped is a new idea which raises many challenging problems that go beyond item recommendations and opens several interesting directions for future work. For instance, one challenge is to consider more complicated transactions involving several interesting parties that need to come together in order to achieve a transaction. Another interesting direction is to consider exchange markets that change over time and model different possible objectives that may characterize transactions, such as fairness and average waiting time.

2.2 Panel: "How Far Should We Personalize?"

Given the proliferation of data and applications, different possibilities as well as different requirements for personalizing user experience emerge at various levels (e.g., content, UI, services, etc). For instance, people may like different services on their cell phone or in Facebook. They may be interested in different content depending on their location, task, preferences or group they belong to. Different presentation features serve different people better. For example, some users like lengthy explanations, others may want to see reviews. Endless personalization possibilities seem to exist in different applications, from personalized search and ads to personal mashups. At the same time, there are several arguments against (over-) personalization.

For example, if we do not have correct information about a user, personalization may hurt accuracy. In addition to gathering and maintaining a profile, on-the-fly personalization can be expensive. Over-personalization may lead to over-specialization. Making a recommendation of something a user would definitely like is valuable but what about serendipity and diversity? In advertising, there is also a delicate balance between the accuracy of ads and their total irrelevance. From the publishers point of view, they would rather serve relevant content than not, from the users' perspective, more relevant ads may be annoying and more intrusive because they may exploit sensitive, personal, information (e.g., a user announces the birth of a child in an email to friends, ads start suggesting where to buy baby stuff).

With all these in mind, the workshop panel's subject was "How Far Should We Personalize?" and its objective was to discuss when, what, how, and to what extent we should or should not personalize. Five panelists coming both from academia and industry shed light on the subject from different perspectives.

Sihem Amer-Yahia (Y! Labs, USA) pointed out that personalization is helpful if we have enough information about the user and the topic. Unfortunately, most users are tail users and there is little information about all possible topics. Indicatively, she mentioned that in delicious, only 3.5% of the URLs are tagged by more than 5 users. Given the fact that we may not have enough information for all topics and all users, personalization needs to be performed in a granular fashion: For head users/topics, *personalization* is suitable. For lesser known users/topics, *groupalization* at various levels is more suitable. Finally, for unknown users/topics, we need to resort back to traditional methods with no personalization.

On the other hand, *Yannis Ioannidis* (University of Athens, Greece) argued that in many scenarios we have various different sources of information for a user, including explicit statements, such as ratings, implicit feedback, such as selecting a product, and social activities, such as blogging. In addition, a user connects to other people in many ways and these connections can possibly lead to learning more about a particular user. Hence, in these cases, the problem is more complex. First, we need to understand the meaning and contribution of each possible piece of information that may be explicitly or implicitly related to a user, and then we need appropriate *methods to extract, detect, propagate, derive and synthesize user information*.

Christian Jensen (Aalborg University, Denmark) described the importance of *geo-context awareness* for the mobile Internet and for service customization. Mobile phones are now providing GPS functionality. An emerging standard is also implemented in Firefox to enable sites to ask the browser for the user's location, which includes current location, destinations visited, routes followed, and so forth. At the same time, there is the "dark side" of personalization: For example, knowing someone's location may be misused in a number of ways. There is a need to develop privacy-preserving techniques that offer (useful) tradeoffs among the degree of privacy, the cost or performance and the quality of service. In addition, it is important to offer transparency, i.e., enable users to understand the privacy proposition and opt-in in a flexible and fine-grained way. In brief, we need to enable *privacy-aware personalization*. *Elisa Quintarelli* (Politecnico di Milano, Italy) also discussed the notion of context in personalization, where the context can also include other information apart from location, such as user roles, interests, and temporal information.

Finally, *Evi Pitoura* (University of Ioannina, Greece) pointed out that personalization is not just about optimizing relevance for a particular user. Other factors must be considered, such as *diversity* and building *trust*. In addition, it is easier to judge positively a search

result/recommendation/query ranking when there is some explanation of its derivation. On the other hand, *over-personalization* may be annoying: too many personal options, too many personal ads may just overwhelm and drive away the user.

2.3 Paper Presentations

2.3.1 Collaborative Efforts

Vagelis Hristidis and *Eduardo Ruiz* in their paper "*CADS: a Collaborative Adaptive Data Sharing Platform*" describe the challenges and initial design ideas for building a collaborative adaptive data sharing platform, which facilitates data annotation at insertion-time and leverages these annotations at query-time. CADS learns with time the information demand (query workload), which is then used to create adaptive insertion and query forms. The goal of CADS is to allow the effortless sharing of documents, while at the same time serving semi-structured queries.

Giorgos Giannopoulos, *Theodore Dalamagas* and *Timos Sellis* in their work "*Collaborative Ranking Function Training for Web Search Personalization*" present a framework for improving the ranking function training and the re-ranking process for web search personalization. They utilize clickthrough data from several users in order to create multiple ranking functions, each one corresponding to different topic areas. Those ranking functions are combined each time a user poses a new query in order to produce a new ranking, taking into account the similarity of the query with each of the topic areas.

2.3.2 Personalization

Alessandro Campi, *Mirjana Mazuran* and *Stefania Ronchi* in their paper titled "*Domain Level Personalization Technique*" learn user preferences by tracking user choices on search results that are presented in clusters. They apply user preferences to (a) rank the clusters of a search result list, showing in the first positions the clusters containing more interesting contents from the user viewpoint, (b) rank the documents contained in each cluster accordingly, and (c) recommend a set of terms taken from the user's profile for query expansion.

Huiming Qu, *Jie Xu* and *Alexandros Labrinidis* in their paper titled "*Guiding Personal Choices in a Quality Contracts Driven Query Economy*" deal with user preferences regarding quality-of-service (QoS) and quality-of-data (QoD). Given an environment where user preferences over different quality dimensions are expressed using quality contracts that are attached to queries, they look at how these contracts can be adapted over time. They propose an adaptive strategy, which monitors a user's queries and the server's responses and automatically adapts the quality contracts of subsequent user-submitted queries.

2.3.3 Recommendations

Sofiane Abbar, *Mokrane Bouzeghoub* and *Lopes Stephane* in their paper "*Context-Aware Recommender Systems: a Service-Oriented Approach*" propose a context-aware recommender system. The proposed system consists of several personalization services that help introduce the notion of context in recommendations.

Finally, *Kostas Stefanidis*, *Marina Drosou* and *Evi Pitoura* in their paper "*You May Also Like Results in Relational Databases*" consider recommending to the users of a database not only tuples in the results of their queries but additional tuples that may be of potential interest. They propose three approaches to compute such results that exploit: (a) the content and schema of the current query result and database instance, or (b) the history of previously submitted queries to the database system, e.g. by using query logs,

or (c) resources external to the database, such as related published results and reports.

3. WORKSHOP CONCLUSIONS

Given the proliferation of data and applications, different possibilities for personalizing user experience as well as for adapting aspects of the data management process, such as response times, data freshness and quality of service, emerge. The wide use of advanced mobile devices opens up the way to new opportunities for personalized services that take into account not just user preferences but also the user context, such as location and situation.

Obviously, we need to see personalization at various levels and from different aspects. We need to offer personalization in a granular level by dynamically taking into account how much we know about a user or a topic to determine how far we should go. One emerging trend is leveraging the collaborative efforts and the wisdom of the crowds to better serve the needs of different user communities. We can also take advantage of the users' social activities, such as resource sharing and connecting, to derive information about individual users or groups of users and offer targeted services and content. How we can achieve all these is an open challenge.

Finally, in personalization, we need to consider several factors not just improving relevance. For example, we need to offer diverse results and we need to build trusted services. Privacy is a highly overlooked issue. Users need to understand and be able to control how personalization interferes with their privacy. To this end, we need to build privacy-aware personalization.

4. REFERENCES

- [1] S. Amer-Yahia, S. Basu Roy, A. Chawla, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. In *VLDB Conference*, 2009.
- [2] M. Cherniack, E. Galvez, M. Franklin, and S. Zdonik. Profile-driven cache management. In *ICDE Conference*, 2003.
- [3] W. Kießling and W. Kostler. Foundations of preferences in database systems. In *VLDB Conference*, 2002.
- [4] G. Koutrika and Y. Ioannidis. Personalized queries under a generalized preference model. In *ICDE Conference*, 2005.
- [5] H. Qu and A. Labrinidis. Preference-aware query and update scheduling in web-databases. In *ICDE Conference*, 2007.
- [6] Y. Zhou, G. Cong, B. Cui, C. Jensen, and J. Yao. Routing questions to the right users in online communities. In *ICDE Conference*, 2007.