Proposal for a Cooperation with Exchange of Researchers between the University of Bari and Fraunhofer IPSI on Personalization in Digital Libraries

Personalization of the content and service offer of a digital library is an important contribution to the mediator role of a digital library in order to meet individual information needs and preferences. In addition to the still fast growing amount of available heterogeneous content and metadata as well as of involved service and content providers, digital library R&D is currently strongly influenced by the upcoming new digital library architectural approaches with new federative and dynamic models for content and service provision.
These developments impose new challenges towards personalization support that have to be answered by improvements of the personalization approaches used in digital libraries and by the development of new ideas in this area. Promising approaches in this direction include a more comprehensive capturing of the characteristics of users, user groups, and their working context, the exploitation of ontologies in user modelling and personalization services, advanced approaches in learning user characteristics, and the development of standards for user modelling.
Both the University of Bari personalization team as well as the personalization team at IPSI work on advancing the methods for personalization support in digital libraries and in other information systems. The competences and research interests of the personalization team at IPSI and the personalization team at UNIBA complement each other in a synergetic way.

Fraunhofer-IPSI
The focus of the personalization team at Fraunhofer IPSI is on extended user modelling, intelligent community services based on extended user models and cross system personalization. Especially the following R&D work is carried out:

Personal Web Context
Traditional approaches to personalization in digital libraries manage only some dimensions (e.g., knowledge, or interest) of a user. However, the need arises for more multidimensional model of users. Users have differing cognitive patterns, are embedded in a community, engage in multiple tasks or goals and have competing simultaneous roles that are: interactive and related to other entities in a given domain.
For this reason IPSI is considering a context as means to representation the related ideas, situations, events or information that makes a more robust interpretation of the user possible. In the Personal Web Context (PWC), a person's context is defined in terms of their connectedness with domain entities and is defined and affected by the work they do, the things and people they know, or the activities they engage in. The Personal Web Context uses an ontology-based information infrastructure, referred to as a Resource Networks, describing the relevant entities in information collections and their typed relationships. The provision of valid and expressively annotated Resource Networks is the core of the Personal Web Context approach. Digital libraries managing a pre-selected and quality controlled information collection for a user community may provide such networks extracted from their information collection as value-added service to other parties, e.g. conference organizers, that want to implement the PWC approach as an additional service to their clients (e.g. conference participants).

Cross System Personalization
At present, personalization in digital library uses information collected about a user and his current context in a user profile which is traditionally restricted to a single system, although many user tasks require information handling activities that span several systems, like searching in different collections. Cross-systems personalization is a way to meet the challenge of delivering personalization to users whose interaction with an information system is part of a larger task that covers several interactions with different systems, thereby reusing the effort that has been invested in providing personalization information to one system.

University of BARI
The focus of the personalization team at University of Bari is on user modelling based on the analysis of transactions, analysis of the content and semantic-based personalization. Especially the following R&D work is carried out:

Learning user profiles from transactions
Among the issues the personalization community must deal with, the following are of special importance: how to provide personal recommendations based on a comprehensive knowledge of who users are, how they behave and how to extract this knowledge from the available data and store it in user profiles? To address these issues, we adopted an approach that uses information learned from transactional histories to construct individual profiles. The advantage of using this technique is that profiles generated from a huge number of transactions tend to be statistically reliable. We used *supervised learning techniques* to dynamically discover users' preferences from transactional data (e.g. recorded during past browsing activities of the users). The problem of learning user's preferences can be cast to the problem of inducing general concepts from examples labelled as members (or non-members) of the concepts. Classifications and the transactional data are gathered to form the user profile, which consists of two main frames: *factual* (personal and transactional data), and *behavioural* (user's

preferences). Profiles could be used to fit the search in any digital collection according to the user's interests. The profiles inferred by analysing transactional data are *coarse-grained*: they contain only the coarse categories of interest of users.

Learning user profiles from content

Recent advances at the intersection of Information Retrieval and Machine Learning offer some novel solutions supporting users in obtaining a good selection of information they are interested in. This results in the intensive development of methods for using machine learning techniques on text databases referred to as *text-learning*. In the content-based approach to text classification the system searches for the items similar to those user liked by comparing their content. Our research involves the analysis, design and implementation of algorithms that learn from observed data in order to build models able to make predictions about unseen data. We investigate learning from text collections for applications like building user profiles for locating relevant information on large document repositories. We are mainly concerned with the use of probabilistic techniques and relevance feedback methods for acquiring *fine-grained* profiles able to represent in a more detailed way the users' preferences.

Using ontologies for improving user profiles

Typical current personalization approaches model users' preferences, skills, knowledge and/or capabilities and store them in *user profiles* useful to deliver personalized content to users. Nowadays, the access to knowledge spaces needs to be user-tailored for achieving success. This means that not only the access to information is important, but also the relevance of quality itself matters. An effective profile-based access to *relevant* information is difficult if information and users are modelled according to their preferences, knowledge, etc., only at the *keyword* level. Innovative personalization services require extended *semantic-based user profiling* techniques (*semantic-based personalization*). The construction of ontology-based user profiles taking into account domain relationships goes beyond the state of the art of describing user profiles by term vectors and builds upon the idea of basing user profiles on concept ontologies (taxonomies and non-taxonomic relationships). This would bring doubtless advantages in terms of: profiles interoperability, formal reasoning applied to them and expressive power. Furthermore, profiles can be profitably represented by exploiting structures that are closer to, for instance, natural language ones, such as lexicons à la WordNet. This process will straightforwardly bring personalization into the mainstream of Semantic Web research, that aims at implementing the most interopeable hypermedia ever available for humans.

Bringing the efforts from the two groups together, the cooperation will focus on the following personalization-related topics:

- Extended ontology-based user models
- Cross-system personalization
- Hybrid personalization approaches based on extended user models
- Semantic-based personalization
- User profiles for user groups
- Standardization of user models

The two teams intend to strengthen their cooperation by exchanging researchers in order to enable close face-to-face discussion and cooperation of the aforementioned personalization topics. As a concrete step, a stay of three researchers of the UNIBA team at the IPSI institute in Darmstadt is planned for the period starting August 20, 2004 through September 21, 2004.

Short description of the researchers

Pasquale Lops is an assistant professor at the University of Bari since January 2004. His research is mainly concerned with the use of Machine Learning techniques for personalizing the access on digital collections (i.e. digital libraries, online catalogues). In more details it is mainly involved in the extraction of user profiles from the analysis of transactions. Moreover, he is carrying out some research on the integration of different paradigms for personalization, in order to build a new and more effective hybrid system. He addressed the area of the electronic commerce and recommender systems during COGITO (IST-1999-13347), FAIRSNET (IST-2001-34290) and FAIRWIS (IST-1999-12641) projects. Currently he is finalizing the PhD thesis in Computer Science and the main activities are related to the integration of different approaches for recommendation and personalization, in particular the use of content-based profiles for improving the process of neighbourhood formation in a classical collaborative filtering algorithm. Summarizing, his research interests include digital libraries, user profiling, personalization and hybrid recommender systems.

Marco Degemmis is a Ph.D. candidate in Computer Science at the University of Bari (Italy) since November 2001. His research is devoted to the application of supervised machine learning techniques for the induction of content-based profiles of users' preferences from text documents they liked in the past. He is also investigating on *content personalization*, by analyzing methods for tailoring information access to a specific user or group of users on the basis of the learnt profiles. The topic of his thesis concerns *semantic user profiles* obtained by integrating lexical resources (WordNet) and ontologies in the learning process in order to acquire the semantic meaning of the words by extracting concepts from unstructured text.

His current research, including *Text Categorization, Information Retrieval, Content-based Information Filtering,* is oriented towards developing digital library applications and advanced information retrieval systems based on user profiles.

Oriana Licchelli received her Laurea degree in computer science with honors from the University of Bari, Italy, in March 2001. Since November 2001 she is a Ph.D. Student, under the guidance of Prof. Giovanni Semeraro, at the Computer Science Dept. of the University of Bari. Her research interests focus on the analysis and application of Machine Learning techniques for personalizing the access on large Web repositories, such as digital libraries or other generic information sources. In particular, she is involved in the extraction of user profiles from the analysis of the interactions between the user and the Web sites. She addressed the area of the personalized digital collections during COGITO (IST-1999-13347), and FAIRSNET (IST-2001-34290) projects. Currently she is finalizing the PhD thesis in Computer Science and the main activities are related to the use of the personalized digital libraries within the e-Learning context, in more details the use of supervised learning technique to create user/student profile in order to improve searching in the digital library to provide personal recommendations for the student. Summarizing, her research interests include digital Libraries, user profiling and personalization in the e-learning domain.

Planned Activities for August/September 2004

- Internal Personalization Workshop at Darmstadt at August 26th, 2004
- Authoring of a white paper on Cross-System Hybrid Recommender
- Preparation of a common publication for at least one of the following conferences the based on the white paper and the cooperation results produced so far
    - CHI 2005 (main conference and/or workshops)
    - IUI 2005 (main conference and/or workshops)
    - UM 2005 (main conference)
    - ECDL 2005 (main conference)

Activities for the future cooperation

- Further common R&D work in the aforementioned areas
- Extension of the cooperation to include further DELOS partners like the University of Athens.