Association for Information Systems
AIS Electronic Library (AISeL)

BLED 2006 Proceedings | BLED Proceedings

2006

Personalisation of eSearch Services – Concepts, Techniques, and Market Overview

Kai Riemer
University of Münster, kai.riemer@sydney.edu.au

Fabian Brüggemann
University of Münster, fabianbrueggemann@web.de

Follow this and additional works at: http://aisel.aisnet.org/bled2006

Recommended Citation
Riemer, Kai and Brüggemann, Fabian, "Personalisation of eSearch Services – Concepts, Techniques, and Market Overview" (2006). *BLED 2006 Proceedings*. 3.
http://aisel.aisnet.org/bled2006/3

This material is brought to you by the BLED Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in BLED 2006 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

19th Bled eConference
eValues
Bled, Slovenia, June 5 - 7, 2006

Personalisation of eSearch Services – Concepts, Techniques, and Market Overview

Kai Riemer

Fabian Brüggemann

European Research Center for Information Systems (ERCIS)
University of Münster, Germany
kai.riemer@ercis.de
fabianbrueggemann@web.de

Abstract

The importance of information in today's society is still growing and information search has become an essential task in both the workplace and in private life. eSearch services provide access to the abundance of information available on the Internet by means of search engine technology. However, conventional search engines have certain limitations in dealing with the typical information overload problems. With the application of personalisation techniques search engine providers aim at moderating some of the problems by providing users with information access individualised to their needs. The aim of this paper is twofold. Firstly, techniques for personalisation of eSearch services are introduced. Secondly, the results of an empirical study of the market for eSearch services are presented. Typical examples illustrate eSearch personalisation in practice, and the diffusion of techniques and implications for further research in the domain are discussed.

1. Introduction

In today's knowledge society the importance of information is growing rapidly. The Internet is establishing itself as the medium of choice for the search of information in a variety of domains [4]. At first sight, the Internet seems to solve the typical search problem: It is widely available, it provides global information access, and it allows access to a wide variety of information including various multimedia formats. Furthermore, the quantity of available information is still growing [42]. The user, however, is confronted with a range of search problems [38]. Typically, the user's aim is to get the right information at the right time in a suitable quality with the lowest possible effort. However, on the Internet the user faces the problem of picking the right bits of information from a very large information pool [49,9]. In order to deal with the consequences of the resulting information overload, the relevant information has to be filtered in a convenient way [4].

But even contemporary search engines can hardly moderate the consequences of information overload presented by the Internet. The results list provided by a typical eSearch query is still very broad and often of poor quality; the sheer quantity of information is time consuming to overlook. One possibility to improve this rather unsatisfying situation is to personalise the search process to a user's individual needs. Personalisation of eSearch services is an increasingly popular topic with academics and search engine providers alike [28]. Many of the major providers are experimenting with personalisation techniques or have made announcements to do so [18,51,5]. The aim of this paper is to shed light on the concepts and techniques for personalisation of eSearch services and to provide an overview of the diffusion of such techniques in the marketplace. A brief outline of the typical eSearch process is given, before section 3 presents a variety of personalisation techniques. The results of the empirical investigation are discussed in the second part of the paper; this provides the basis for an outlook on trends and future developments.

2. The eSearch process

For information search on the Internet, users typically refer to eSearch services (search engines) [52]. Search engines are applications that provide users with the possibility to search for a set of keywords for which they expect the best outcome [33]. The result is a listing of web links, sorted by relevance, that should contain the desired information. To determine the relevance of results the eSearch provider typically uses a ranking algorithm. Most often, providers treat their algorithms as assets for strategic differentiation; thus algorithms are kept secret [4]. However, despite sophisticated ranking algorithms conventional eSearch results are far from optimal due to a variety of problems.

The selection of a suitable search engine itself presents a search problem given that there are approx. 70,000 services on offer world-wide, a lot of which focus on a specific topic or domain [19]. Within the eSearch process users often do not use the right amount or combination of search words for their specific problem [21,3]. Also, most users only launch one single search query [32]. And many users do not look past the first few hits presented by a search engine [35]. One of the most sever problems however is that traditional search engines do not take into account user intentions and do not allow to influence the ranking of search results. Consequently, traditional search engines confront the users with a set of results that is still quite broad and general, which requires a further manual search for relevant bits of information. Personalisation of eSearch services contributes moderating these problems.

3 Personalisation techniques

Personalisation is defined as the adjustment of information, products, offerings, or (parts of) web pages to the needs, preferences, and capabilities of an individual user [50]. In principle, almost any object of a website can be personalised [50]. Three dimensions can be differentiated referring to the personalisation of 1) content/information, 2) navigational systems, and 3) website layout [43,44,34]. As for search engines all three dimensions are of interest [50]. However, at the core of personalised eSearch is the content dimension, e.g. the personalisation of search results. However, personalisation of navigation and layout may also play a role in

helping the user to better handle the list of search results by changing its appearance [45].
The next section introduces user profiles as the basis for personalisation, while the following sections deal with different types of personalisation techniques: 1) User-oriented techniques use direct user inputs or profile data of a single user for personalisation. 2) Collaborative techniques are based on similarities between user profiles or take into account the membership of users in social networks. 3) Supporting techniques do not contribute to personalisation directly but support and complement the first two types of techniques.

3.1 User profiles as the basis for personalisation

Data about the user and his preferences are the basis for automated personalisation [36,39]. This user data is stored in profiles, which can be accessed, queried, and analysed by the personalisation system [10,16,24]. The creation and management of user profiles is a core component of most personalisation systems [8,20,31].

Explicit profiling

One way of creating a user profile is by explicitly gathering the data from the user [27]. The user can be given the possibility of expressing needs, wishes, and preferences by filling in a questionnaire or by evaluating a list of topic categories [27]. Explicit profiling has certain advantages in that users can indicate their preferences directly and they are given control of the personalisation process [2]. On the other hand, the shortcomings are that users are often not able to indicate their preferences correctly and that it takes additional effort to create the profile [24]. Also, preferences might become outdated over time [2].

Implicit profiling

Implicit profiling is carried out automatically by a software system with only a small amount of configuration required by the user. Essentially, it is based on the observation of user behaviour [40,9]. One way is to evaluate the user's search history in order to determine user preferences [24]. Studies show that this technique usually leads to good data quality [30]. User preferences can also be extracted from a list of recently visited web sites. The time spent on reading a web page, the number and frequency of visits, as well as the user adding a web page to his list of bookmarks can be interpreted as user interest [22,24]. Consequently, a list of preferred topic categories can be derived from the appearance of certain keywords within these web pages [30]. Automatic preference determination can also be carried out by observing user behaviour within the context of the computer desktop, e.g. by means of desktop search engines.
Changing user interests can be better taken into account by implicit profiling techniques [47]; the continuous examination of user behaviour supports building dynamic profiles [16]. However, implicit profiling is prone to error and thus comes at the expense of uncertainty. Conclusions drawn from user actions can be misleading, since not all actions necessarily indicate personal preferences, as might be the case when the user carries out search queries for friends and colleagues [17].

3.2 User-oriented personalisation techniques

User-oriented personalisation techniques are based on the actions and preferences of the individual user who launches the search query, while collaborative techniques (see later) take into account the profiles of other users as well.

Definition and restriction of the relevant search area

At the beginning of a search task, the user can be given the possibility to choose the area of the Internet relevant to the search. In doing so, the user can select topic categories for inclusion or exclusion in the search query [55]. Besides, meta-search engines might allow the user to choose the search engines or databases to be included in the search query [37], whereby the query can be restricted to specialised search engines that only cover a certain domain or area of interest [25,26]. The definition of the relevant search area can also be implemented as an iterative process that aims at improving an initial search query. Users might be allowed to explicitly discard topic areas from a list of search results [47]. By doing so, the search area can be further limited and thus the quantity of results is reduced, while at the same time the quality in terms of relevance is improved [54].

Keyword personalisation

Often, users select keywords that do not supply any relevant search results. To deal with this problem, search query personalisation aims at automatically amending search queries or supporting users in selecting suitable keywords:

- *Two-stage search with feedback:* The user formulates an initial search query and then evaluates the search results in terms of relevance ("relevance feedback"). To re-formulate the new search query the system automatically determines a new set of keywords from the list of results deemed relevant by the user. The keywords are then presented to the user for inclusion in a second search query that should then lead to better results [6].
- *Automated generation of keywords:* This technique also generates new keywords based on a relevance evaluation by the user. However, the keywords are automatically combined with the old keywords to launch a new search query without any necessary user interaction („top document feedback"). In addition, the system might save a set of individual keywords in the user's profile to complement future search queries ("past query feedback"). Research shows that these techniques can improve search results significantly [21].

Search query personalisation by logical deduction

This technique uses semantic relations in the data to personalise the search query [41]. It is based on the principle of logical deduction, which is known from expert systems. At the core of the technique is a hierarchy of logical statements and relations that is built into the system. The following example illustrates how it works. The set of statements might contain relations like "A Golden Retriever is a dog", "A dog is pet", "A cold indicates illness", "Veterinary surgeons can treat pets" and "Veterinary surgeons have an office in a certain area". When the user launches the search query "cold golden retriever", the system will react with the following steps. Firstly, the search query is changed to a semantic inquiry which is then assigned to one of the internal topic areas of the system. Secondly, the search query is amended with the help of the logical relations. The system might

change the search query to “veterinary surgeons, Bled, Slovenia”, whereby the location reference might be inferred from the user profile [47]. The disadvantage of this technique is that even for only a small area of interest a rather larger number of logical relations has to be built into the system for the system to be able to return sensible results.

Location-based personalisation

Today, work and private life are characterized by increasing mobility. Therefore, the localisation of information search becomes more and more important. Location information can be used to provide the user with regional or even local information. Objects of interest might be restaurants, banks, pharmacies, hotels, or petrol stations. Location-based personalisation can be based on data such as the user’s zip code, city, or region stored in the user profile. When a search query is launched from a mobile device it can be based on an automated device localisation [47]. In principle, this technique can be compared to restricting the search area to a certain location or combining keywords with location information (see above); however, location-based personalisation is treated as a separate technique due to its growing importance.

Personalisation by adjustment of page rank weights

While page rank algorithms usually remain hidden from the user and thus influencing its internal weighting criteria is not possible, profile-based personalisation can be built into the algorithm by the search engine provider. For doing so, user interests stored in the profiles can be used to adjust page rank weights in order to return more relevant results. Experiments indicate that the *Google* ranking algorithm can be amended for personalisation by aligning its internal weights to user data [29]. Generally, Google’s PageRank algorithm determines the relevance of a website recursively by the number of external links pointed to this site and the relevance (e.g. page ranks) of the linking websites [46]. In order to personalise the search process, those websites will be weighted higher that are linked to by websites of a certain topic area or by websites stored in the user profile (e.g. in the browsing history, search history, or bookmarks) [33]. By individually increasing the page ranks of these linking websites, the user’s preferences are taken into account, which leads to better results [11].

Personalisation by categorisation

Research shows that semantic webs and topic ontologies can contribute positively to personalisation [23]. For doing so, history data stored in the user profile is compared with a system of categories, i.e. a semantic web of topics. Subsequently, user data is scanned for categories and categories represented in the profile are selected and marked. For example, it might be determined that the word “Jaguar” found in the user history is linked to the semantic web category “nature/robbery cat/jaguar”. This category would then be marked in the user profile, while the category “cars/jaguar” would not. Once the user places a search query the identified categories can be used to restrict the list of results to only those websites that are within the scope of interest of the user [39]. By applying this technique, ambiguities in search query keywords can be detected and dealt with.

Suggestion of personalised search results

Initially, users often do not exactly know what information they are looking for specifically [7]. Here, a personalised eSearch system might support the user by actively suggesting relevant websites [1]. These personalised suggestions can be based on user preferences obtained from the user profile [44]. When the users already know websites typical for the relevant topic area, they can register these websites with the search engine. Data is then extracted from the websites to be stored in the user profile or to directly feed a search for websites with similar information [24].

Reorganisation of the results list

While some techniques personalise the search query (keywords) upfront or interfere with the internal algorithm, this technique comes into operation after a standard query has been processed. In order to only provide helpful and relevant results to the user, the list of results is filtered and rearranged according to user preferences [48]. Such a reorganisation can take place interactively via user inputs or be automated on the basis of data from the user profile. Those websites that lie within a specific topic area of the user profile might be weighted higher with regards to the rearrangement [24]. Besides, the system can contribute to the reduction of information overload by removing redundant and irrelevant information [24]. Rearrangement techniques are frequently combined with cluster techniques (see supporting techniques below).

3.3 Collaborative personalisation techniques

Collaborative relevance filtering

This technique uses the comparison of user profiles in such a way that similarities between profiles contribute to improving search results for one user. Collaborative filtering is based on the assumption that two users who show similarities of interest in some parts of their profiles also match in other areas of interest and as such might judge the same set of search results as relevant. Consequently, websites (and topics) that were deemed relevant by the matching users get a higher weight within the ranking algorithm [52]. This technique was derived from the collaborative filtering technique used by *Amazon.com* for its product recommendation system: "customers, who bought this product, also bought...."

Social network-based personalisation

With this technique it is the user not the eSearch system who chooses other users as reference points for similarity detection. To this end, social networks are created with search partners who show similar interests. Here again, users of the network might explicitly classify websites themselves or the system might put together user profiles. Being a member in a search network, the system can provide the user with websites that were classified being relevant by other users of the same network. Furthermore, the system might weight higher those websites that were classified by users in the social network or that are related to other users' interests [52]. Besides, users might have access to information such as the list of websites that all users of the network classified as relevant [47].

The following table provides a summary of all eSearch personalisation techniques discussed so far and indicates the usage of user profiles. The next parapgraph introduces supporting techniques that complement these personalisation techniques.

Personalisation technique	Main features	Usage of user profile
Definition and restriction of search area	Search engine selection, selection of topic areas, iteration possible	Preferred (topic) categories, search services and sources
Keyword personalisation	Two-stage search with feedback, derivation of suitable keywords for search query	Preferred (topic) categories
Logical deduction	Rewording of the search query by logical classification of the search query based on internal hierarchy of relations	Mainly access to internal hierarchy of relations rather than user profile
Location-based personalisation	Restriction of the search query to provide results with local context	Location information (e.g. zip code, residence)
Adjustment of page rank weights	Adjustment of the PageRank weights, based on similarities with user preferences	User interests, search history, web browsing behaviour, preferred websites (bookmarks)
Categorisation	Classification of the results on the basis of an ontology and matching with search words	Interests as hierarchy of topics (ontology)
Suggestion of search results	Selection of certain hits and suggestion as best result	Preferred (topic) categories
Reorganisation of results list	(Interactive) rearrangement of search results on the basis of user interests	Preferred (topic) categories
Collaborative relevance filtering	Relevance determination based on similarities between user profiles.	User interests, range of profiles of other users to search for similarities.
Social search networks	Relevance determination based on similarities with users in social networks.	Profiles of other network members.

Table 1: Overview of personalisation techniques

3.3 Supporting techniques

Adjustment of degree of personalisation

Search engines that use profiles to adjust page rank weights and to interfere with the internal search algorithm might permit users to adjust the degree to which user data affects the personalisation of the search results. With such a feature, the user is able to even better adapt the search results to his needs [54] in order to gain more control of the personalisation process, and to learn how profile data contrib-

utes to personalisation. By doing so, this technique might lead to a higher acceptance of personalised eSearch services.

Clustering of search results

One possible solution for the problem of ambiguity in search queries is the clustering of search results. Clustering is not a personalisation technique in itself, because clusters usually derive from a set of results that have not been personalised beforehand. However, clustering can be seen as an intermediate step to personalisation in that the technique provides a better separation and demarcation of topic areas within the results. Hence, it is easier for the user to find relevant websites, since the results are separated more clearly into topic areas. The next step towards personalisation can then be the rearrangement of clusters based on user profiles or direct user interaction (see above) [47,24].

Storage of search results

Another technique that supports the search process is the storage of results for future reference. Three different degrees of storage can be differentiated: 1) storage of keyword combinations in search queries, 2) storage of the results lists, and 3) storage of complete web pages. Storage of queries and results facilitates future information gathering and the formulation of new search queries.

Browser toolbar for direct access

In order to offer users direct access to eSearch functionality (e.g. the search query entry field) within their work context, search engines often provide a toolbar for integration with software products such as web browser programs.

4 Overview of the empirical study

Having differentiated and described several personalisation techniques in the section above, this chapter introduces the empirical investigation in regards to the diffusion of techniques in the marketplace. A selection of 15 search engines with personalisation features was chosen for evaluation. This sample represents the majority of the market and provides a good overview of personalisation diffusion in the marketplaces. A set of criteria for the classification of the eSearch services was derived from the literature study (see table 2). All eSearch services have then been classified based on these criteria by the two authors of this study independently. Independent evaluation was done in order to control for bias. While there was consesus in regards to most of the classifications, one deviation led to the redefinition of a category. In fact, the two techniques ‘definition of search area’ and ‘reorganisation of results’ were combined into one category before the empirical study and have only been plit up in two techniques as a result of the discussion initiated during the empirical work.

Profile building	Explicit: by customer configuration		Implicit: by analysis of user behaviour	
User-oriented techniques	Definition of search area	Keywords	Logical deduction	Location-based
	Page rank weights	Categorisation	Results suggestion	Reorganisation of results list
Collaborative techniques	Collaborative relevance filtering		Social search networks	
Supporting techniques	Adjustment of degree of personalisation	Clustering of search results	Storage of results	Browser toolbar
Additional dimensions	Navigation		Layout	

Table 2: Classification criteria to characterise eSearch services in the marketplace

The result of the classification of eSearch services is presented in table 3. In the table, an 'X' in a grey shaded cell indicates that a certain feature or technique is supported by the respective eSearch provider. The empirical population has been devided in browser-based and tool-based search services.

	Search service	Data gathering		User-oriented techniques								Collabo-rative		Supporting techniques				Add-on	
		explicit	implicit	Search area	Keywords	Log. deduction	Location-based	Page rank weights	Categorisation	Results suggestion	Reorganisation	Collab. filtering	Search networks	Degree of pers.	Clustering	Save results	Toolbar	Navigation	Layout
Browser-based	Amazon A9	X	X	X			X			X		X				X	X	X	X
	AskJeeves	X	X	X	X		X									X	X	X	
	Clusty	X		X							X				X		X	X	
	Eurekster	X	X	X				X		X			X				X		
	Filangy	X	X	X				X					X	X		X	X		
	Google		X	X			X	X						X		X	X		
	Grokker			X											X				X
	Iboogie	X		X							X				X				
	MSN	X	X	X			X			X	X			X			X		X
	Teoma			X	X														
	Yahoo!	X		X			X						X			X	X		
Tool-based	Bingoo	X		X						X	X					X		X	X
	Copernic		X	X							X					X	X		X
	Gurunet		X	X	X						X					X	X		
	Webferret		X	X	X						X					X			
	n=	9	9	15	4	0	5	3	0	4	7	1	3	3	3	9	10	4	5

Table 3: Application of eSearch personalisation in the marketplace

In the next paragraph, the adoption and diffusion of personalisation techniques in the marketplace are reflected upon, before implications of these results are discussed and an outlook on further trends and developments are given.

5 Adoption of personalisation techniques in practice

The results of the empirical study (table 3) show clearly that some techniques are already quite common in the marketplace while others have yet to find their way into the services of commercial search engine providers. Search area definition is the only personalisation technique implemented by all eSearch services in the sample and can be considered a standard feature. However, the personalisation effect of this technique is quite limited since the remaining search area usually is still quite broad. Beyond this, three groups of services in regards to the support of techniques can be differentiated.

5.1 Personalisation before and after the search query

This group of eSearch services offers a rather simple form of personalisation mainly by generating personalised search queries and/or reorganising the search results. Personalisation thus takes place before and after the actual search query by either feeding a personalised set of keywords to the query or filtering a list of otherwise unpersonalised search results.

AskJeeves, Teoma, Gurunet, and *Webferret* support keyword personalisation and thus suggest individual combinations of keywords to improve an initial search query. *AskJeeves* for example presents all search results in conjunction with their respective topic categories. The user is then able to use these categories to extend or restrict the search scope and launch a new search query. Essentially, this leads to a new combination of keywords that is determined based on the proximity of the initial keywords to internal topic categories.

A personalised rearrangement of search results is first and foremost a speciality of the tool-based eSearch services. However, these search tools are also quite limited in the extent of implemented personalisation, since all tool-based services are meta-search engines, so that personalisation of the algorithm or page rank weights is not possible. In browser-based services this technique is often combined with clustering of the results, as is the case with *Clusty* and *Iboogie*. Here, a personal view on the results list is provided to the user based on the clustering. Results lists in *Iboogie* for example are automatically clustered into groups of topics, which can then be selected and extended in order to navigate the results. By doing so, the user is provided with an individualised way of sorting through the results list. An advanced version of the results reorganisation is offered by *MSN*, where it is not based on clustering but on a set of criteria like “updated recently” or “visited frequently” that allows the user to generate a results list that is arranged according to own personal preferences. The degree to which these criteria are used by the system can also be determined by the user.

5.2 Personalisation on the algorithm level

While clustering and rearrangement of results provide the user with a personalised way of better dealing with the amount of (otherwise unpersonalised) search results, this second group of eSearch services goes beyond this by directly affecting the way the search results are identified. Based on preference data in the user profile, the idea is to directly limit the results to only those that are relevant for the user. Such a technique for example is applied by *Google*, where the relevance of results is derived at the algorithm level. Personalisation is based on the adjustment of internal weights of the page rank algorithm. This adjustment uses a comparison

of similarities between search results and the user profile data. The exact functioning of the *Google* algorithm however is kept secret. User profiles are constructed implicitly on the basis of data such as saved search results, browsed web pages, and news headlines the user clicked on. This data is evaluated in order to determine relevant topics of interest of the user.
While other eSearch services such as *Filangy* and *Eurekster* also personalise on the algorithm level they fall into the third category since they are based on collaborative techniques.

5.3 Collaborative personalisation

While personalisation at *Amazon A9* is mainly based on the collaborative relevance filtering algorithm adapted from Amazon's product suggestion mechanism [53], *Filangy*, *Eurekster*, and *Yahoo!* use social search networks created and maintained by the users. These services are based on the presumption that friends or acquaintances in social networks share similar taste and rate websites similarly. They are essentially based on the theory of small networks, which states that social networks can lead to strongly improved results in the search for information [12].
At *Eurekster* for example, users can join a search network (a 'search party') or invite friends to join their own network. In doing so, *Eurekster* works as filter for the information available in the social network. First of all, the user can restrict the search area to categories such as the web, blogs, topics, the own search history, and also the personal network. Once the user selects a search network as the active filter, *Eurekster* adjusts the relevance weights of its search algorithm and takes as the basis for relevance determination the websites that other network members have already evaluated. Websites that were classified as useful by a network user, e.g. by spending time reading a page, are then weighted higher for other users in the same search network [5]. Hence, the *Eurekster* algorithm continuously learns from the behaviour of network users and profile creation takes place in an implicit way on the basis of users' search histories.
Yahoo! as another example of collaborative techniques uses search networks in a different way in that users have the possibility to explicitly recommend and suggest web pages to other users. Here, personalisation is not based on a specialised search algorihtm.

6 Outlook and implications

The empirical investigation shows that most personalisation techniques identified from the literature have in some way been implemented in practice. However, the market for personalised eSearch services is still in its infancy. While there is no dominant design for eSearch personalisation at the moment, new personalisation techniques will emerge and be applied in the future and the domain will further develop [38]. Some trends can already be identified.

6.1 No dominant design yet

At the moment, no clear trend has emerged regarding a dominant design of eSearch personalisation; the market obviously is still in an early phase. At the moment, tool-based services present a more homogeneous picture than browser-based search engines. Furthermore, it is noticeable that collaborative and social

techniques are gaining in popularity, while two of the above mentioned techniques have not been applied in practice at all: logical deduction and categorisation, both of which originated from the academic domain. However, these new techniques that are based on concepts like ontologies and semantic web present considerable potential for future application. Other personalisation techniques like location-based search and automated results suggestion present themselves as add-on features that come in combination with one of the other techniques. Especially location-based search might gain importance in the future due to the ongoing trend towards mobility in both peoples' private and work lives. Furthermore new algorithms and ways of acquiring user data will emerge.

6.2 Trends and developments

New algorithms

Techniques like logical deduction and categorisation are not yet used in the marketplace. However, experimental implementations like the search engine KeyConcept developed by the University of Kansas demonstrate how new personalisation techniques like categorisation can be used. KeyConcept for example divides web pages automatically into categories based on keywords identified within the pages [23]. The outcome is an index, in which for each word and each category the associated web pages are stored. Users are then able to select a range of topic categories themselves that is subsequently stored as a profile. Once the user launches a search query the search words in combination with the topic categories from the profile are used to determine the relevant documents from the index [23]. While at the moment the user profile has to be configured explicitly, it is planned to determine relevant categories implicitly based on user behaviour.

New ways of data gathering

Besides new algorithms, a second area of research concentrates on new ways of data gathering. Currently, user profiles are either based on user activities captured in the user's search or browsing history, or on an explicitly compiled list of relevant topic categories. However, such information mainly refers to user interest. In the future, profiles could indicate user expectations in regards to search queries, for example whether search aims at finding information in regards to a product name, a manufacturer, a dealer, or a data sheet for a product. Also, the current search context of the user might be gathered from the documents found on the local computer [15,13,14,37]. Keywords might be automatically determined from the contents of desktop documents to indicate user interest. But not just new types of user information are of interest; new ways of gathering this data are also envisaged. An important aim of some recent projects is therefore the automation of the information acquisition process [19].

Context-embedded search

Finally, research targets new ways of embedding search with the user context. A personalised eSearch service might continuously run in the background within the user's working context (e.g. a third party software product) and constantly retrieve potentially needed information. Results can then be provided instantly. Hence, one trend that will grow in the future is the combination of web search and desk-

top search (see current efforts by companies such as Microsoft and Google). Data resulting from the desktop search might also be used to determine user preferences and thus further enhance the user profile. On the downside, such developments raise serious questions regarding data security and privacy.

Security, privacy, and user acceptance

The storage and usage of user data has both a legal and a trust dimension. On the one hand, privacy regulations and differing national law have to be considered. On the other hand, an open and trusting handling of the user data is necessary in order not to lose user confidence and acceptance [56]. For users it is important to keep control of their data as well as of the personalisation process [44,34]. If users are not sure how a certain system works (black box), they might refuse to use it, especially if the system produces results that are unexpected, inconsistent, and not intuitive [43].

7 Summary and conclusion

The aim of this study was to provide an overview of techniques for the personalisation of eSearch services and to examine the diffusion of such techniques in the marketplace. A number of personalisation techniques were identified and discussed based on an intensive literature review (see table 1). Using a classification scheme derived from the theoretical part of the study (table 2), fifteen eSearch services were examined regarding the support of personalisation techniques (table 3). From the discussion of the results, it becomes apparent that some interesting applications of personalisation techniques can be found in the market already, while the market itself is still in an early phase in regards to personalisation. Further research within this area will probably develop in two directions. Firstly, design-oriented research might aim at developing new techniques, while testing their application in user experiments before they will enter the commercial market. Secondly, more empirical studies might investigate the usage and acceptance of already existing techniques and thus gain new knowledge of the situational benefits and shortcomings of particular personalisation approaches. For both types of research this paper might serve as a conceptual basis.

References

1. Ansari, A.; Essegaier, S.; and Kohli, R. Internet Recommendation Systems. *Journal of Marketing Research*, 37 (2000), 363-375.
2. Ansari, A.; and Mela, C. F. E-Customization. *Journal of Marketing Research*, 15 (2003), 131-145.
3. Anthes, G. H. Search for tomorrow. *Computerworld*, 38, 14 (2004), 26.
4. Arasu, A.; Cho, J.; Garcia-Molina, H.; Papcke, A.; and Raghavan, S. Searching the Web. *ACM Transactions on Internet Technology*, 1, 1 (2001), 2-43.
5. Balas, J. L. When every search engine knows your name. *Computers in Libraries*, 25, 2 (2005), 33-35.
6. Belkin, N. J. Helping People Find What They Don´t Know. *Communications of the ACM*, 43, 8 (2000), 58-61.
7. Berreby, D. Getting to know you. Special Report: The Rise of E-Business. *IBM Systems Journal*, 1 (1999), 1-4.
8. Bharat, K.; Kamba, T.; and Albers, M. Personalized, Interactive News on the Web. *Multimedia Systems*, 6, 5 (1998), 349-358.

9. Birukov, A.; Blanzieri, E.; and Giorgini, P. (2005): Implicit: A Recommender System That Uses Implicit Knowledge To Produce Suggestions. University of Trento, Department of Information and Communication Technology: 1-8.
10. Blom, J. Personalization - a taxonomy. *Proceedings of the CHI 2000 Workshop on Designing Interactive Systems for 1-to-1 E-commerce*, 1-6 April 2000, 313-314.
11. Brin, S.; and Page, L. (1998): The Anatomy of a Large-Scale Hypertextual Web Search Engine. Proceedings of the 7th International World Wide Web Conference (WWW7). Brisbane, Australia: 107-117.
12. Buchanan, M. *Nexus - Small Worlds and the Groundbreaking Science of Networks*. New York: W. W. Norton and Company, 2002.
13. Budzik, J.; and Hammond, K. J. User interactions with everyday applications as context for just-in-time information access. *Proceedings of the 2000 International Conference on Intelligent User Interfaces*, New Orleans, Louisiana.
14. Budzik, J.; Hammond, K. J.; Birnbaum, L.; and Krema, M. Beyond similarity. *Proceedings of the 2000 Workshop on Artificial Intelligence and Web Search*.
15. Budzik, J.; Hammond, K. J.; Marlow, C.; and Scheinkman, A. Anticipating information needs: Everyday applications as interfaces to internet information servers. *Proceedings of the 1998 World Conference of the WWW, Internet and Intranet*, Orlando, Florida.
16. Datta, A.; Dutta, K.; D., V.; Ramamritham, K.; and Navathe, S. B. An Architecture to Support Scalable Online Personalization on the Web. *VLDB Journal*, 10, 1 (2001), 104-117.
17. Deng, L.; Chai, X.; Tan, Q.; Ng, W.; and Lee, D. L. Spying out Real User Preferences for Metasearch Engine Personalization. *Proceedings of the 6th WEBKDD workshop: Webmining and Web Usage Analysis (WEBKDD'04), in conjunction with the 10th ACM SIGKDD conference (KDD'04)*, Seattle, Washington, USA, August 22, 2004, 71-82.
18. Elgin, B. Web Search for Tomorrow. *Business Week*, 3883 (2004), 48.
19. Ellwein, C. *Suche im Internet für Industrie und Wissenschaft*. München: Oldenbourg Industrieverlag GmbH, 2002.
20. Field, A.; Hartel, P.; and Mooij, W. Personal DJ, an Architecture for Personalized Content Delivery. *Proceedings of the 10th International World Wide Web Conference (WWW10)*, Hong Kong, May, 1-5 2001, 1-8.
21. Fitzpatrick, L.; and Dent, M. Automatic Feedback Using Past Queries: Social Searching? *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, 306-313.
22. Gauch, S.; Chaffee, J.; and Pretschner, A. Ontology-Based Personalized Search and Browsing. *Web Intelligence and Agent Systems*, 1, 3-4 (2004), 219-234.
23. Gauch, S.; Madrid, J. M.; Induri, S.; Ravindran, D.; and Chadalavada, S. (2003): KeyConcept: A Conceptual Search Engine. University of Kansas, Information and Telecommunication Technology Center: 1-25.
24. Gauch, S.; and Speretta, M. Personalizing Search Based on User Search Histories. *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, Compiegne University of Technology, France, September 19-22, 622-628.
25. Gravano, L.; and Garcia-Molina, H. Generalizing GIOSS to vector-space databases and broker hierarchies. *International Conference on Very Large Databases (VLDB)* (1995), 78-89.
26. Gravano, L.; Garcia-Molina, H.; and Tomasic, A. GIOSS: Text-source discovery over the Internet. *ACM Transactions on Database Sytems*, 24, 2 (1999).
27. Greco, G.; Greco, S.; and Zumpano, E. Collaborative Filtering Supporting Web Site Navigation. *AI Communications*, 17 (2004), 155-166.
28. Hane, P. J. Partnerships, Personalization, and Personal Security Top the News. *Information Today*, 22, 6 (2005), 7-14.

29. Haveliwala, T. H. Topic-sensitive PageRank. *Proceedings of the 11th International World Wide Web Conference.*
30. Hirsh, H.; Basu, C.; and Davison, B. D. Learning to Personalize. *Communications of the ACM*, 43, 8 (2000), 102-106.
31. Hjelsvold, R.; Vdaygiri, S.; and Léauté, Y. Web-based Personalization and Management of Interactive Video. *Proceedings of the 10th International World Wide Web Conference (WWW10)*, Hong Kong, 129-139.
32. Jansen, B. J.; Spink, A.; and Saracevic, F. Real life, real user, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36 (2000), 207-227.
33. Jeh, G.; and Widom, J. Scaling Personalized Web Search. *Proceedings of the 12th International World Wide Web Conference*, 22nd of February 2002, 1-24.
34. Karat, C. M.; Brodie, C.; Karat, J.; Vergo, J.; and Alpert, S. R. Personalizing the user experience on ibm.com. *IBM Systems Journal*, 42, 4 (2003), 686-701.
35. Koutrika, G.; and Ioannidis, Y. A Unified User Profile Framework for Query Disambiguation and Personalization. *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA)*, Edinburgh, UK, 44-53.
36. Kramer, J.; Noronha, S.; and Vergo, J. A User-Centered Design Approach to Personalization. *Communications of the ACM*, 43, 8 (2000), 45-48.
37. Lawrence, S. Context in Web Search. *IEEE Data Engineering Bulletin*, 23, 3 (2000), 25-32.
38. Levy All Eyes on Google. *Newsweek*, 143, 13 (2004), 48-58.
39. Liu, F.; Yu, C.; and Meng, W. Personalized Web Search by Mapping User Queries to Categories. *Proceedings of the 11th International Conference on Information and Knowledge management*, McLean, Virginia, USA, 558-565.
40. Liu, F.; Yu, C.; and Meng, W. Personalized Web Search for Improving Retrieval Effectiveness. *IEEE Transactions on Knowledge and Data Engineering*, 16 (2004), 28-40.
41. Liu, H.; Lieberman, H.; and Selker, T. GOOSE: A Goal-Oriented Search Engine With Commonsense. *Proceedings of the 2002 International Conference on Adaptive Hypermedia and Adaptive Web Based Systems* (2002), 253-263.
42. Lyman, P.; and Varian, H. *How much information?* http://www.sims.berkeley.edu/research/projects/how-much-info/, accessed on 05.11.2005.
43. Manber, U.; Patel, A.; and Robison, J. Experience with Personalization on Yahoo! *Communications of the ACM*, 43, 8 (2000), 35-39.
44. Mesquita, C.; Barbosa, S.; and de Lucena, C. Towards the Identification of Concerns in Personalization Mechanisms Via Scenarios. *Proceedings of the 1st Conference on Aspect-Oriented Requirements Engineering and Architecture Design*, 25-31.
45. Notess, G. R. Customization Options for Web Searching. *Online*, 25, 1 (2001), 55-57.
46. Page, L.; Brin, S.; Motwani, R.; and Winograd, T. (1998): The Pagerank Citation Ranking: Bringing Order to the web. Stanford University.
47. Pitkow, J.; Schütze, H.; Cass, T.; Cooley, R.; Turnbull, D.; Edmonds, A.; Adar, E.; and Breuel, T. Personalized Search. *Communications of the ACM*, 45, 9 (2002), 50-55.
48. Prakash, K. S. S.; and Raghavan, S. V. Intelligent Search Engine: Simulation to Implementation. *Proceedings of the iiWAS 2004 - The sixth International Conference on Information Integration and Web-based Applications Services, 27-29 September*, Jakarta, Indonesien.
49. Ravindran, D.; and Gauch, S. Exploiting Hierarchical Relationships in Conceptual Search. *Proceedings of CIKM 2004* (2004), 238-239.
50. Riemer, K.; and Totz, C. The Many Faces of Personalization - An Integrative Overview of Mass Customization and Personalization. In Tseng, M. M.; Piller, F. T.

(eds.), *The Customer Centric Enterprise: Advances in Mass Customization and Personalization*. New York/Berlin: Springer, 2003, 35-50.
51. Rupley, S. Found What You´re Looking For? *PC Magazine*, 23, 9 (2004), 25.
52. Sugiyama, K.; Hatano, K.; and Yoshikawa, M. Adaptive Web Search Based on User Profile Constructed without Any Effort from Users. *Proceedings of the International WWW Conference*, New York, USA, 675-684.
53. Swartz, N. Amazon´s New Offering Manages Searches. *Information Management Journal*, 38, 6 (2004), 16.
54. Teevan, J.; Dumais, S. T.; and Horvitz, E. Beyond the commons: Investigating the value of personalizing Web search. *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA)*, Edinburgh, UK, 84-92.
55. Teevan, J.; Dumais, S. T.; and Horvitz, E. Personalizing search via automated analysis of interests and activities. *Proceedings of SIGIR '05* (2005), 449-456.
56. Volokh, E. Personalization and Privacy. *Communications of the ACM*, 43, 8 (2000), 84-88.