

Item-Based Filtering and Semantic Networks for Personalized Web Content Adaptation in E-Commerce

Panayiotis Koutsabasis and John Darzentas

University of the Aegean,
Department of Product and Systems Design Engineering,
Hermoupolis, Syros, Greece, GR-84100
{kgrp, idarz}@aegean.gr

Abstract. Personalised web content adaptation systems are critical constituents of successful e-commerce applications. These systems aim at the automatic identification, composition and presentation of content to users based on a model about their preferences and the context of interaction. The paper critically reviews related work in the field and presents an integrated approach for the design of personalization and web content adaptation in e-commerce that places emphasis on item-based collaborative filtering and on short-term, dynamic user models represented as semantic networks. The proposed approach for personalised web content adaptation can provide different types of interesting recommendations taking into account the current user interaction context in a computationally inexpensive way. It is also respectful of user personal information and unobtrusive with respect to user feedback.

Keywords: personalisation, web content adaptation, item-based collaborative filtering, user model, semantic networks, e-commerce.

1 Introduction

Over the last decade it has been realised that personalised web content adaptation can significantly enhance user experience in e-commerce. According to Bunt et al. [4] personalised content adaptation “*involves identifying the content most relevant to a given user and context (jointly referred to as the interaction context), as well as how this content should be organized*”. The large number of approaches, techniques and systems available makes it difficult for designers to identify and develop personalization and web content adaptation components that address specific e-commerce application requirements.

The recent book entitled *The Adaptive Web* [3] reviews several approaches for modelling technologies (including user profiles, data mining and document modelling), adaptation technologies (including content-based, collaborative and hybrid information filtering), applications (including e-commerce, e-learning, etc.) and challenges (including group collaboration, privacy and usability engineering). Furthermore, Adomavicius and Tuzhilin [1] present a thorough review of algorithms and techniques for content-based, collaborative and hybrid filtering approaches and identify extensions in a number of directions including improvement of understanding of

users and items, incorporation of contextual information, support for multi-criteria ratings, and provision of more flexible and less intrusive types of recommendations. In addition, Montaner et al [19] analyse 37 different systems from the perspective of user profile generation, maintenance and exploitation approaches and techniques and conclude that “*the combination of modelling particular user preferences, building content models and modelling social patterns in intelligent agents seems to be an ideal solution*”.

Personalised web content adaptation has grown into a large research field that attracts scientists from different communities such as: hypertext, user modelling, machine learning, natural language generation, information retrieval, intelligent tutoring systems and cognitive science [3]. The aims of the paper are to (a) review related work in the field and (b) to present the architectural design of personalised web content adaptation system for an e-commerce application that hosts virtual product exhibitions in the area of traditional folk art. The architecture is generic and may be applied to any e-commerce system. In addition, the proposed approach is novel regarding the combinatory employment of item-based filtering and semantic networks for representing short-term, dynamic user models that are generated without user intervention.

2 Related Work

Related work in the field is reviewed according to the following important design considerations that are relevant to any type of personalised web content adaptation system: types of data sources, goals for data mining, user profile generation and representation and recommendation approaches. The review concludes with outlining important research challenges.

2.1 Types of Data Sources

Data sources for personalised web content adaptation can be divided into two main groupings: structured, such as databases or metadata-encoded information and unstructured, such as plain text. The principle data source of an e-commerce application concerns items (or products), which are usually stored in a structured database that can provide automatic responses to user queries. Other types of data sources include: retailer information, user demographic information (which may be part of the user profile), usage statistics (which is typically constructed off-line via data mining), and so on. All these types of data sources are typically structured databases.

In the case of unstructured data sources, there is a need to provide structure to enable search; this usually happens in a vector representation [19] (or “bag of words”), with each word being assigned a weight using methods such as $tf \cdot idf$ (term frequency times inverse document frequency)¹ [20]. To improve recall, techniques like removing stop words and stemming (keeping only the root of the word) are used. The similarity with other documents or profiles is usually measured with the cosine metric (of the

¹ Terms frequently appearing in a document (tf), but rarely in other documents (idf), are more likely to be relevant to the topic of the document.

angle between the two vectors). There are many other approaches to represent documents and calculate similarity; a useful review is given in [15].

2.2 Goals for Data Mining

Data mining, and particularly web usage mining has emerged as an important component of contemporary personalisation and content adaptation systems. Web usage mining can be defined as “*the automatic discovery and analysis of patterns in click-stream and associated data collected or generated as a result of user interactions with web resources on one or more web sites*” [18]. According to Mobasher [17] the data mining cycle includes the three phases of: data collection and pre-processing, pattern discovery and evaluation, and applying the discovered knowledge in real-time.

The goals and implementation of data mining differ depending on the specific application. The primary data sources used in web usage mining are the server log files, which typically include web server access logs and application server logs. Of particular interest to web personalisation and adaptation is the data collection and pre-processing phase of data mining which includes data fusion (from multiple logs), data cleaning, page view identification, sessionization (distinguishing user actions or episodes of a single interaction session), episode identification and finally data transformation that includes the aggregation and integration to e-commerce application databases [17].

2.3 User Profile Generation and Representation

Most personalisation and content adaptation systems make use of user profiles. User profiles may include demographic information and may also represent the interests or preferences of either a group of users or a single person [10]. User profiles may be created explicitly when users directly specify their preferences or implicitly by user agents that monitor user activity. User profiles can be dynamic or static, depending on whether their content changes over time. Furthermore, user profiles may be long-term or short-term depending on whether they are maintained for a long period of time in the system and they reflect preferences that do not change frequently over time or vice-versa.

According to Gauch et al. [10] user profiles are generally represented as sets of weighted keywords, semantic networks and concept-based profiles. User profiles represented as weighted keywords are updated by web documents that are interesting to users. When user profiles are represented as semantic networks, the nodes correspond to keywords or bags of words and the arcs to weights: the bigger the weight, the more relevant the connected nodes. In concept-based profiles, the nodes represent abstract topics considered interesting to the user, rather than specific words or sets of related words as in semantic networks. In general there are many variations in the implementations of user profiles based on the above approaches. Recently, ontologies have been employed to represent user profiles on the basis of domain classifications, however with interest in search engines in general [9] or in particular engines for research papers [16].

2.4 Recommendation Approaches

Recommendation approaches (also referred as information filtering and recommender systems) are usually classified in the following categories [2]:

- Content-based: recommendations are made on the basis of estimating the similarity of new items with those that the user has preferred (purchased, positively reviewed, etc.) in the past. Pure content-based approaches suffer from that they cannot produce novel recommendations to users because the user profile is the only source of information about user preferences and does not change frequently. Another drawback is that in many cases users are required to explicitly provide their preferences.
- Collaborative: recommendations are made on the basis of estimating the similarity of the current user's profile with other users; items that other users prefer will be recommended to the current user. Pure collaborative filtering algorithms are computationally expensive and will not work until other users have provided items with ratings.
- Hybrid: recommendations are made combining both aforementioned approaches in various ways [5].

Other approaches, that are usually combined with content-based and collaborative recommenders include:

- Cluster models [24] that maintain profiles for clusters of users that have similar preferences. Recommended items are pulled from other users belonging in the same cluster.
- Demographic filtering approaches [20] that suggest items taking into account demographic information like age, occupation, education, nationality, place of living, etc.
- Knowledge-based: that recommend items based on domain-dependent inferences about user needs and preferences [5].
- Item-based filtering [13][14], which matches each of the user's preferred items to other similar items and then combines those similar items into a recommendation list. According to [21], item-based techniques first analyze the user-item matrix to identify relationships between different items, and then use these relationships to indirectly compute recommendations for users.

2.5 Challenges for Personalised Web Content Adaptation

Offline vs. Real-Time Computation. Contemporary personalisation and web content adaptation systems typically use a two-stage process for generating recommendations. The first stage is carried out offline, where web usage data from previous interactions is collected and mined and a model for making recommendations is created [17]. The second stage is carried out in real-time as users start interacting with the web site. During interaction, the model for making recommendations for the particular user is retrieved and data from the current user session is embedded into the model. Web content adaptation systems must perform most time-consuming calculations offline in order to perform fast and scale up when large amounts of data are made available.

Explicit vs. Implicit User Information Collection. Many recommender systems are intrusive in the sense that they require explicit feedback from the user and often at a significant level of user involvement [1]. Explicit user information collection typically happens through HTML forms (e.g. to directly specify preferences, to provide demographic information, etc.), as well as when users are asked to provide ratings, reviews and opinions for items. Obviously, these methods assume that users will be willing to provide this information in advance. However, related studies have shown that this assumption is unrealistic: users are reluctant to provide personal information to computer applications in general [12] mainly due to privacy concerns; they are even laconic when they provide queries [8]! On the other hand, Gauch et al. [10] identify a number of implicit user feedback techniques: browser cache, proxy servers, browser agents (e.g. the Google toolbar), desktop agents (e.g. Google desktop), web logs and search logs (based on cookies). Each technique has advantages and disadvantages with the most prominent that most still require some user intervention.

The ‘Cold-Start Problem’ in Generating Recommendations. Nearly all approaches for item recommendation suffer from the ‘cold-start’ problem, which refers to the fact that there may not be (interesting) recommendations to make to users, when there are few user profiles or ratings available. For example, in a content-based recommendation approach there may be no (interesting) recommendations for new users, unless they explicitly provide feedback about their preferences. There are many workarounds to this problem, like introducing a knowledge-based component that does not rely on historical information, to boost the process [7].

Contextual Web Content Adaptation. The large majority of recommender systems make their recommendations based only on the user and item information and do not take into consideration additional contextual information that may be crucial in some applications [1]. For example, a user that likes alternative rock music may log in to their favourite music recommendation system in order to buy a gift to a friend that likes classical music. The system should take into account this particular context of use and recommend accordingly, however this requires that it can scale up well during a particular user session.

Multi-Criteria Ratings. Most of the current recommender systems deal with single criterion ratings. However in some applications it seems quite useful to incorporate multi-criteria ratings into recommendation methods. For example, a scientific book may be reviewed on novelty, soundness, use of language, etc.; a movie on plot, direction, photography, and so on. According to Adomavicius and Tuzhilin [1] multi-criteria ratings have not yet been examined in the recommender systems literature.

Evaluation. There are not established approaches for evaluation of personalisation and web content adaptation [11]. The goals and measures of evaluations differ significantly across studies ranging from accuracy and relevance to item coverage and usability; thus the selection (or rejection) of some evaluation measures is an issue. In addition, some recommendation techniques and algorithms have been designed with specific computational requirements in mind (e.g. many collaborative filtering algorithms have been designed specifically for data sets with many more users than items); which results to satisfactory results in only some cases of data sets [14].

3 The Design of Personalised Web Content Adaptation for Virtual Product Exhibitions in E-Commerce

This section presents the architectural design of personalised web content adaptation system for an e-commerce application that hosts virtual product exhibitions in the area of traditional folk art.

Traditional folk art includes a huge variety of products that ranges from traditional clothing and decoration to artefacts of everyday life such as kitchen appliances and jewellery. These products often have a high commercial value and address a large market comprising clients who want to decorate public and private spaces (such as hotels, offices, restaurants, etc.), tourists, collectors, and interested industries such as the packaging and film industry. Of particular interest to this project are ceramics that have a multitude of uses ranging in packaging, decoration and household artefacts.

The traditional folk art sector is fragmented consisting of many SMEs that design and manufacture artefacts on the basis of self-learning and family tradition. The virtual exhibition of traditional folk art products will provide new opportunities for retailers in the sector to promote their products through the web. The traditional folk art sector typically addresses 'foreigners', i.e. customers that are not accustomed to the tradition of others, and in this respect the global market.

The aims of the personalised web content adaptation system are to complement the dynamic page segments of the e-commerce application² in order to:

- Provide interesting recommendations about items that are relevant to the current user interaction context, thus enabling contextual content adaptation;
- Provide a number of different types of recommendations based on different information filters;
- Integrate complementary approaches for recommendations thus constituting a hybrid recommendation system, in an attempt to address the “cold-start” problem in recommender systems;
- Ensure that the recommendation algorithms are computationally inexpensive;
- Ensure that customers interact with the e-commerce system unobtrusively of requests related to providing their preferences or other personal information;

The design of the proposed web content adaptation system is presented in the next section in terms of the conceptual architecture and its basic components.

3.1 Conceptual Architecture of Personalised Web Content Adaptation

The conceptual architecture of the proposed personalised web content adaptation system is depicted in Fig. 1. The main ideas of the proposed system are to combine item-based recommendations which are computationally inexpensive and sensitive to user interaction context with user profiling/modelling represented as a short-term, semantic network to enable fast memory-based operations. In addition, the user is not burdened by requests to provide explicit personal information, which is collected and

² Except for the item recommendations function, the e-commerce application is deterministic [22], i.e. it returns the same set of items to all customers with the same query at a certain time.

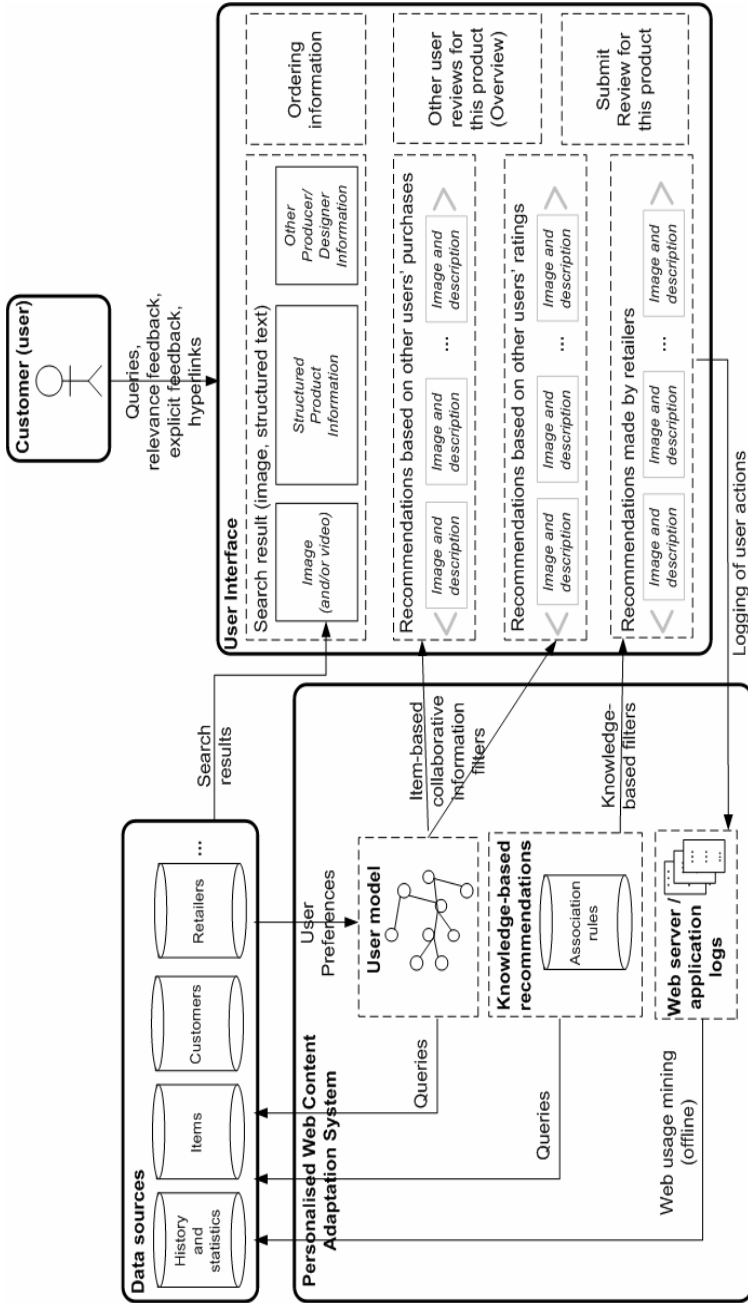


Fig. 1. Conceptual architecture of the personalised web content adaptation system for virtual product exhibitions in e-commerce

analysed by a data mining component offline. Finally, a knowledge-based component is proposed that uses association rules related to item properties to aid the item-based recommendation in order to get around the cold-start problem.

User Interface. The web user interface is dynamically generated after a user has selected an item from the pool of search results (a draft wireframe of the user interface is depicted as part of Fig. 1). The user interface is composed of a set of fragments. The ‘search result’ fragment presents detailed information about the item selected, which can consist of related image(s), video(s), and structured text description. There are fragments from which users can provide ordering information (e.g. a shopping basket, a wish-list etc.) and see ratings and reviews for this item provided by other users. Furthermore, users can submit their own ratings and reviews for the item available – to enable relevance feedback for producing recommendations. All the aforementioned fragments are retrieved directly from the e-commerce application databases.

In addition, there are recommendation fragments, which provide certain types of recommendations to customers based on different information filters. These fragments are dynamically generated from by the user model in the case of item-based collaborative information filters and from the knowledge-based component in the case of knowledge-based recommendations. These fragments present recommendations of other items to the user that can be selected to continue the navigation of the e-commerce application.

Item-Based Collaborative Filtering. Item-based collaborative filters recommend items that have been purchased or rated by other users in corresponding user interface fragments. If there is not (reliable) knowledge about recommendations, these page fragments are simply not added to the user interface (e.g. for a new user).

Item-based collaborative filtering happens in two-stages. The first stage calculates offline the similarity of each item in the e-commerce database with all other items on the basis of users’ preferences (e.g. purchases and ratings, obviously with a different level of importance). More specifically, in our design the algorithm first identifies all items that have been purchased or rated by a particular customer. Then, the algorithm records these preferences for all customers. After recording all relationships between items (jointly purchased and/or rated with other items by customers), the similarity of items can be computed in various ways: for example, Sarwar et al. [21] propose cosine-based similarity, correlation-based similarity and adjusted cosine similarity for item-based collaborative filtering in particular. We are to initially implement weight and similarity metrics according the vector space model (tf*idf weighting scheme and cosine metric similarity), which is the most widely tested and used. The algorithm is similar to [14], with the difference that we also keep track of rated items for user recommendations. The complexity of the algorithm is $O(I^2 \cdot U)$ in the worst case (i.e. if an item has been purchased or rated by nearly all customers!), where I is the number of items and U the number of users, however in practice it is much closer to $O(I \cdot U)$ [14].

User Model as a Semantic Network. The second stage of item-based collaborative filtering happens in real-time by creating the user model and employing this during interaction. In our design of item-based filtering, the user model is a semantic network that is retrieved in memory each time a user selects an item from the pool of search results. According to [15] “a semantic network is formed by a directed graph, whose nodes are organized in hierarchic structures, while the arcs connecting them represent the binary relations between them, such as relations ‘is-a’ and ‘part-of’”.

In our design, the semantic network consists of a central node which is the selected item (the ‘search result’ in Fig. 1.) and it is connected with nodes that are related items to this item (and have been identified by the item-based collaborative filtering algorithm offline, as described above). We define two types of relations: “purchased together” and “rated” (by other users). Each relation also holds a degree of similarity that can be used to prioritise the presentation to the user. An example of a simple semantic network of this type that shows related nodes at one level of depth is depicted in Fig. 2. Obviously the semantic network can be retrieved and built in memory in further depth, if required.

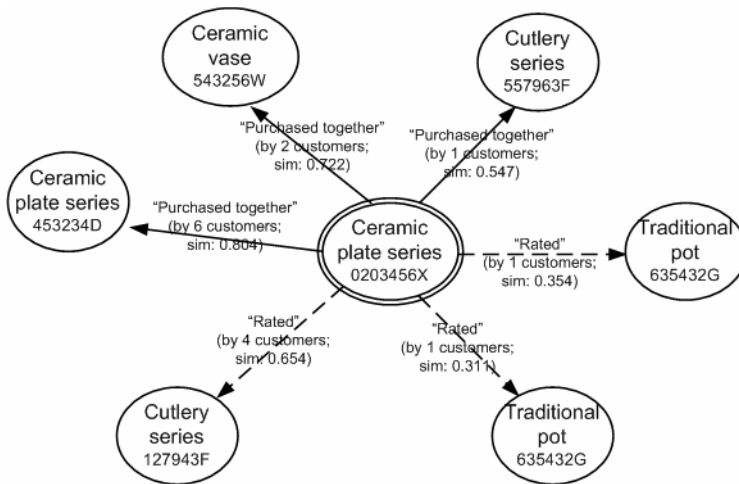


Fig. 2. An example of a simple semantic network whose nodes are items related with “purchased together” and “rated” relationships

Whenever the user selects an item from the pool of search results, the semantic network for this item is retrieved from the data sources and constructed in memory. Further to this, summary information about related items is retrieved and displayed to the user according to the type of relationship of the selected item. Each type of relationship corresponds to an information filter. The degree of similarity of each neighbouring node-item to the selected item is taken into account to prioritise the order of presentation of the recommended items. The complexity of the algorithm is then $O(I \cdot \log I)$, i.e. the computational cost to retrieve an item I from the sorted index.

Knowledge-based Recommendations. Knowledge-based recommendations are employed to address to some extent the cold start problem in recommender systems. This type of recommendation is designed so that it relies on association rules that are inserted by knowledge engineers at the retailer site, when adding new items (products) to the data sources. For the purposes of this project, creation of digital content will take place and this content will then be inserted to the e-commerce system. During this phase of the project, the e-commerce system backend will require from retailers to associate related products.

Web Usage Mining. The proposed approach also incorporates web usage mining, mainly with regard to data collection and pre-processing that result to updating structured databases with user behavioural data such as user queries made and items browsed. Although this data is not particularly related to the current design of the personalised web content adaptation system, it will be required for further development, for example to let the user see previously browsed items or create related information filters.

3.2 Other Related Systems

The proposed personalisation and web content adaptation system is currently under development, and thus it cannot be evaluated by a user-centred approach. Therefore, a number of related systems are presented that have similar design ideas.

The most widely known system that uses item-based collaborative filtering is Amazon.com to generate several types of recommendations. In [14] it is reported that “*our algorithm’s online computation scales independently of the number of customers and number of items in the product catalog, produces recommendations in realtime, scales to massive data sets, and generates high quality recommendations*”. It needs to be noted that several aspects of the Amazon.com approach to recommendation are not known, such as the user profile–item matching technique and techniques used to find similar users [19].

Another system that has implemented item-based algorithms for evaluation purposes is MovieLens, which is a movie recommender. The experimental study of item-based algorithms in MovieLens suggests that “*item-based algorithms provide dramatically better performance than user-based algorithms, while at the same time providing better quality than the best available user-based algorithms*” [14].

With respect to the employment of semantic networks to the representation of user profiles, Gauch et al. [10] review a number of systems that build semantic networks either with keywords as nodes and similarity weights as arcs, or bags of words that correspond to concept-nodes that are related to other concept-nodes again via weighted arcs. In our approach we extend those ideas by incorporating items in the nodes (that can be seen at the level of abstraction of either a bag of words or structured information) and by connecting those items with both similarity weights and types of relations that correspond to information filters. We also anticipate that our approach for representing the semantic network can be further extended to include more types of relations, e.g. such as “items browsed” by other users.

4 Summary and Future Work

The paper first presented a critical review of related work in the area of personalised web content adaptation. Important research challenges in the field include those of: offline vs. real-time computation, explicit vs. implicit user information collection, the cold-start problem in producing recommendations, contextual web content adaptation, multi-criteria ratings and evaluation of the personalisation and web content adaptation process.

The main result of the paper is however the architectural design of an integrated approach to personalised web content adaptation that addresses the first four of the aforementioned challenges. The conceptual architecture of the proposed system was discussed mainly in terms its two distinguishing features of item-based collaborative filtering and the dynamic creation of short-term user models that are represented as semantic networks.

The proposed system is under development and will be integrated to an open source e-commerce platform for hosting virtual product exhibitions of traditional folk art products.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering* 17(6) (2005)
2. Balabanovic, M., Shoham, Y.: Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM* 40(3), 66–72 (1997)
3. Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): *Adaptive Web 2007*. LNCS, vol. 4321, pp. 409–432. Springer, Heidelberg (2007)
4. Bunt, A., Carenini, G., Conati, C.: Adaptive Content Presentation for the Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 409–432. Springer, Heidelberg (2007)
5. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. *UMUAI* 12(4), 331–370 (2002)
6. Burke, R.: Knowledge-based Recommender Systems. In: Kent, A. (ed.) *Encyclopedia of Library and Information Systems*, vol. 69 (sup. 32) (2000)
7. Burke, R.: Hybrid Web Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
8. Carroll, J.M., Rosson, M.-B.: The paradox of the active user. In: Carroll, J.M. (ed.) *Interfacing thought*, pp. 80–111. MIT Press, Cambridge (1987)
9. Gauch, S., Speretta, M., Pretschner, A.: Ontology-based User Profiles for Personalised Search. In: Sharman, R., Kishore, R., Ramesh, R. (eds.) *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*. Springer, Heidelberg (2007)
10. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User Profiles for Personalized Information Access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 54–89. Springer, Heidelberg (2007)
11. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems (TOIS)* 22(1), 5–53 (2004)

12. Jansen, B.J., Spink, A., Saracevic, T.: Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management* 36(2), 207–227 (2000)
13. Karypis, G.: Evaluation of Item-Based Top-N Recommendation Algorithms. Technical Report, CS-TR-00-46, Computer Science Dept., University of Minnesota (2000)
14. Linden, G., Smith, B., York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing*, 76–80 (January-February 2003)
15. Micarelli, A., Sciarrone, F., Marinilli, M.: Web Document Modelling. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 155–192. Springer, Heidelberg (2007)
16. Middleton, S.E., Shadbolt, N.R., De Roure, D.C.: Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems* 22(1), 54–88 (2004)
17. Mobasher, B.: Data Mining for Web Personalization. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 90–135. Springer, Heidelberg (2007)
18. Mobasher, B.: Web usage mining and personalization. In: Singh, M.P. (ed.) *Practical Handbook of Internet Computing*. CRC Press, Boca Raton (2005)
19. Montaner, M., Lopez, B., De La Rosa, J.: A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review* 19, 285–330 (2003)
20. Pazzani, M.: A Framework for Collaborative, Content-Based, and Demographic Filtering. *Artificial Intelligence Rev.*, 393–408 (December 1999)
21. Salton, G., McGill, M.: *Introduction to Modern Information Retrieval*. McGraw-Hill, New York (1983)
22. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms, Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th International World Wide Web Conference (WWW10)*, Hong Kong, May 1-5 (2001)
23. Tanudjaja, F., Mui, L.: Persona: A Contextualized and Personalized Web Search. In: *Proceedings of the 35th Hawaii International Conference on System Sciences – 2002*. IEEE, Los Alamitos (2002)
24. Ungar, L., Foster, D.: Clustering Methods for Collaborative Filtering. In: *Proc. Workshop on Recommendation Systems*. AAAI Press, Menlo Park (1998)