International Journal of Information Systems and Computer Technologies (IJISCT) Vol. 2, No 1.
ISSN: 2791-3635(Print), 2791-3643(Online). DOI: https://doi.org/10.58325/ijisct.002.01.0043

Personalized Query Expansion- A Scoping Review

Afsa Hameed[*1]

[1]American school of creative studies, Sharjah, United Arab Emirates.

Corresponding Author: Afsa Hameed (afsah@ascs.sch.ae)

ABSTRACT
The rapid expansion of data on the web has largely increased the problems associated with efficient searching and retrieval of the required data. To resolve the ineffectiveness of information retrieval systems query expansion technology is employed. This research details different storage techniques along with the personalization techniques and the data filtering techniques. Finally the study concludes by presenting a comparison of different recommendation techniques used for query expansion. Moreover, the main focus is laid on the expansion of queries by using each user's profile and search history. This study will enable to retrieve the search results in accordance with the interests of people. Furthermore, it will enhance the efficiency of the search process and the search engine quality will be complemented to a great extent.

Keywords: Query expansion, personalized, user profile

1. INTRODUCTION

Mostly the users pose short queries that seem ambiguous and remain unable to retrieve efficient results. The users make use of short keywords that enhance the ambiguity of query being posed by the user. Executing such queries does not provide does not retrieve the required information and thus the user feels uncomfortable. The queries are also recommended to the users and a proper solution of this problem is the analysis of the behavior of the user and mimic the pattern for query transition that are adopted by users who succeed in finding the needed information [1].

Query expansion is a process in which query is expanded by using some additional keywords to make the query more representative and clearer. The addition of extra keywords has shown to improve the search results [2-4]. There are many researchers that firmly agree on a point that the primary cause of the lack of effectiveness of information retrieval system is inaccuracy of the query formed by the just few keywords that lack the information need of the user [5]. The main motivation behind query expansion is the removal of ambiguity of and the clarity of information need by adding few more words in the query [6].

1.1 Personalized query expansion

In the past few years query expansion has gained importance in literature [6-9] . The information need of the users is satisfied by providing query term suggestion by taking inti consideration the interest of user. The personalized search systems not only retrieve the documents that are relevant to the query terms but also that are relevant to the user interest.

The main feature of this system is the provision of personalized suggestions thus the logs of the user interest are maintained for suggestion of query terms. A vital issue in personalized query expansion is the recording and storing of personalized data .User profile is maintained by some systems (or user model) [10] , whereas aggregate view of the data used is maintained by some other systems [11, 12].

1.2 Techniques for storage of user's data

Several techniques and data structures that are used to represent the information about usage and user are grouped in to two types. A technique in which feature vector is made, it is vector made up of term and associated [13].The second way to express the user information id semantic network [6]. In this technique the words and their semantically associated terms are named as nodes and associated nodes and the weights are assigned to them and the links between them.

Result adaption or query adaption or both could be used for the personalization of the information retrieval systems. In result adaption the results are ranked according to the interest of user [13, 14]. On the other hand in query adaption the query of the users is expanded to make it clearer and to retrieve relevant results [15]. Various techniques are used for query expansion including co-occurrence based technique [6, 15, 16], and thesaurus based techniques [8]. In personalized query expansion additional terms are used from individual user profiles to make query results better.

2. RELATED WORK

Personalized web has been studied extensively. There are many approaches that make use of query log and the click-through analysis [7, 9]. There are numerous systems that explore data on desktop and as well as external resources [8, 15]. Various techniques also exist that focus on user task and activity log [14, 16]. This technique is being implied in commercial search engines [11]. Personalization has two approaches in general. Another approach was developed to learn about the user's interests from tags and their bookmarks, then ranking of the results is done according to user's interest [13]. [8] had investigated about personalized search result soon the basis of user's social relations, then ranking is done according to the social relations of the user. [14] investigated data gathering from numerous online social systems for making personalized adaptable. An interest profile of each user was created by combining different social information streams and then ranking of results is done on the basis of different compound scores. Although the work of this group is attractive, the results tend to be unsatisfactory if relevant results are not retrieved instead of complex methodology.

3. PERSONALIZATION TECHNIQUES

Different refining strategies are used for personalization like ontology model, computational intelligence and data extracting methods.

1. Data Mining Techniques: A method which extract interesting patterns of information from a bundle of data in order to find relevant data is known as data mining. clustering, association rule mining and classification are some methods of personalization [17-19]

2. Clustering: A method which divided data into clusters or groups, similarities in between clusters are decreased whereas within each cluster are increased that will help to identify people with common interests depending on the user's profile.

3. The creation of association rules, a common data mining technique used in the design of recommender systems, defines item-to-item correlations by connecting database items that have a certain relationship [18]. When a database is being evaluated, it poses a challenge for association rule-based systems to produce suggestions since larger item sets are still unable to meet the minimum requirements of constraints. The method of dimensionality reduction could be used on the dataset in order to minimize this problem [18]. Clustering can analyze the user patterns and can form peer groups, while the association rules can be used to discover patterns between these items, Classification is used to incorporate all this information for making a recommendation [18]. Classification depends on set of "training data". Each category based on browsing history, demographics of consumer is given a label, and this is used to train the model. Once each label is optimized through the training steps, real user data is input, and the appropriate recommendations are given [20].

4. Web Mining Techniques: it define as retrieval of intriguing patterns from online information sources is known as web mining [21]. The several methods utilized for web mining.

5. Useful data, information, and knowledge can be gleaned from the content of online pages via web content mining, has two category that are named as web search results mining and web page mining. Web page mining discovers the patterns directly from the web page's content [22, 23]. Whereas Web structure mining shows the structure of the web sites and show the way they are connected. The links of web structure mining are named further as internal and external link [24].

6. Using an Ontology Based Personalization method
   Ontology is formal specification and description of knowledge (a set of concepts and the relations between them in their domain) [25]. It gives common understanding of topics that are to be communicated between systems and users [26]. Personalized ontology [27] is conceptual model that specifically describes the background knowledge of user. Web users may have varied expectations for the same query search [28]. A general personalization framework based on the conceptual modeling of user's navigational behavior was also presented [24].

7. Filtering strategies are methods for removing information from massive amount of information by making use of automated techniques prior to recommendation of data to end user.
   A. Content Based Filtering Technique: Content-based filtering systems are entirely based on individual preferences of users [23]. The system tracks the behavior of each user and then recommend that they are like user interests of past. Vector similarities such as cosine similarity could be used, and prediction could be done based on interest of user for particular item [29].
   B. Rule Based Filtering Technique: Decision tree is drawn to show various rules. In this a set of questions are asked from users, questions are derived from decision tree. This approach fails to make recommendations for those patterns that do not exist in the association rules.
   C. Collaborative Filtering Technique: The opinion of the user is predicted, and recommendations are offered on the basis of opinions of users and opinions of other similar minded users.

   A summary of all the recommendation techniques is given and their advantages and disadvantages are listed.

TABLE.1. Comparison of different recommendation techniques.

S.No	Example Recommender System	Technique used	Advantages	Disadvantages
1.	DBSCAN [106],OPTICS [80],BIRCH [81].	Clustering Algorithms	• Faster Recommendation. • Better performance	• Not accurate
2.	MovieLens [119].	Association rules	• Provide accurate Prediction about the user interest. • Fast to implement. • Not much storage space required. • Fast to execute.	• Difficulty in producing the recommendations when the data base is sparse. • Not suitable if preferences change rapidly. • Rules can be used only when enough data validates them.
3.	Levis [92], Netflix [95].	Memory Based Recommender Systems.	• Easy implementation • New data can be added easily and incrementally • Need not consider the content of the items being recommended. • Scale well with correlated items	• Dependent on human ratings • Performance decrease when data are sparse. • Cannot recommend for new users and items • Limited scalability
4.	GroupLens [93].	Model based Collaborative Filtering.	• Better at addressing the sparsity and scalability • Improve prediction performance	• Expensive model building. • Lose useful information for dimensionality reduction.
5.	TAN-ELR, NB-ELR [121].	Classification Technique	• Accurate. • Deals Sparsity problem effectively	• Scalability. • Time Consuming in model training.
6.	Personality Diagnosis [115].	Hybrid Recommenders	• Improved prediction performance. • Overcomes sparsity problem.	• Increased complexity. • Expensive. • Need external information that usually not available.

7.	Newsweeder [118],LIBRA	Content-Based Filtering	• Efficiently address Cold start problem and Sparsity Problem	• Requires content that can be encoded as meaningful features. • User interest must be represented as a learnable function. • Overspecialization problem.
9.	ATHENA [106].	Ontology	• No cold start issue. • More accurate	• Knowledge engineering is required. • Expert opinion may not match with user preferences.
10.	PEN recsys [120].	Web usage mining	• More accurate	• Much storage space is required.

4. CONCLUSION

The systems for retrieving information are now being used in all departments and custom-made is used to make the information more relevant. An overview of recommendation systems and personalization strategies was provided in this work. Many suggestion systems have been developed based on content-based refined, collaborative refined and an algorithms of hybrid recommendation. Additionally, their admiring benefits and drawbacks have been discussed. In research, the system which suggest user benefits and interest, is growing day by day. The usage of specialist big data platforms with appropriate mining techniques for personalization is required due to the necessity for improved accuracy and processing massive volumes of different information.

REFERENCES

[1] S. Sadoughi and S. J. I. S. Zarifzadeh, "An efficient hybrid query recommendation using shingling and hashing techniques," vol. 104, p. 101928, 2022.

[2] B. Billerbeck, F. Scholer, H. E. Williams, and J. Zobel, "Query expansion using associated queries," in *Proceedings of the twelfth international conference on Information and knowledge management*, 2003, pp. 2-9.

[3] D. Cai, C. Van Rijsbergen, and J. M. Jose, "Automatic query expansion based on divergence," in *Proceedings of the tenth international conference on Information and knowledge management*, 2001, pp. 419-426.

[4] L. A. F. Park and K. Ramamohanarao, "Hybrid pre-query term expansion using latent semantic analysis," in *Fourth IEEE International Conference on Data Mining (ICDM'04)*, 2004, pp. 178-185: IEEE.

[5] C. Carpineto and G. J. A. C. S. Romano, "A survey of automatic query expansion in information retrieval," vol. 44, no. 1, pp. 1-50, 2012.

[6] P.-A. Chirita, C. S. Firan, and W. Nejdl, "Personalized query expansion for the web," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 7-14.

[7] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li, "Towards context-aware search by learning a very large variable length hidden markov model from search logs," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 191-200.

[8] F. Liu, C. Yu, W. J. I. T. o. k. Meng, and d. engineering, "Personalized web search for improving retrieval effectiveness," vol. 16, no. 1, pp. 28-40, 2004.

[9] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *Acm Sigir Forum*, 2017, vol. 51, no. 1, pp. 4-11: Acm New York, NY, USA.

[10] M. Speretta and S. Gauch, "Personalized search based on user search histories," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, 2005, pp. 622-628: IEEE.

[11] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior information," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 2006, pp. 19-26.

[12] B. Smyth and E. J. I. r. Balfe, "Anonymous personalization in collaborative web search," vol. 9, no. 2, pp. 165-190, 2006.
[13] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 155-162.
[14] Q. Wang and H. Jin, "Exploring online social activities for adaptive search personalization," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 999-1008.
[15] M. Bertier, R. Guerraoui, V. Leroy, and A.-M. Kermarrec, "Toward personalized query expansion," in *Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, 2009, pp. 7-12.
[16] C. Biancalana and A. Micarelli, "Social tagging in query expansion: A new way for personalized web search," in *2009 International Conference on Computational Science and Engineering*, 2009, vol. 4, pp. 1060-1065: IEEE.
[17] Y. Zhang and J. Koren, "Efficient bayesian hierarchical user modeling for recommendation system," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 47-54.
[18] M.-S. Chen, J. Han, P. S. J. I. T. o. K. Yu, and d. Engineering, "Data mining: an overview from a database perspective," vol. 8, no. 6, pp. 866-883, 1996.
[19] P. S. Yu, "Data mining and personalization technologies," in *Proceedings. 6th International Conference on Advanced Systems for Advanced Applications*, 1999, pp. 6-13: IEEE.
[20] D. W. J. E. o. L. Oard and Information Sciences, "Multilingual information access," pp. 3682-3687, 2010.
[21] T. Wang, Y. J. W. T. o. I. S. Ren, and applications, "Research on personalized recommendation based on web usage mining using collaborative filtering technique," vol. 6, no. 1, pp. 62-72, 2009.
[22] Y. H. Cho, J. K. Kim, and S. H. J. E. s. w. A. Kim, "A personalized recommender system based on web usage mining and decision tree induction," vol. 23, no. 3, pp. 329-342, 2002.
[23] B. Mobasher, R. Cooley, and J. J. C. o. t. A. Srivastava, "Automatic personalization based on web usage mining," vol. 43, no. 8, pp. 142-151, 2000.
[24] S. Acharyya and J. Ghosh, "Context-Sensitive Modeling of Web Surfing Behaviour Using Concept Trees," in *Proc. of the 5th WEBKDD Workshop, Washington*, 2003.
[25] M. Eirinaki and M. J. A. T. o. I. T. Vazirgiannis, "Web mining for web personalization," vol. 3, no. 1, pp. 1-27, 2003.
[26] T. R. J. I. j. o. h.-c. s. Gruber, "Toward principles for the design of ontologies used for knowledge sharing?," vol. 43, no. 5-6, pp. 907-928, 1995.
[27] K. Curran, C. Murphy, and S. Annesley, "Web Intelligence in Information Retrieval," in *Web Intelligence*, 2003, pp. 409-412.
[28] P. Castells, M. Fernandez, D. J. I. t. o. k. Vallet, and d. engineering, "An adaptation of the vector-space model for ontology-based information retrieval," vol. 19, no. 2, pp. 261-272, 2006.
[29] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," in *Proceedings of the 2nd ACM Conference on Electronic Commerce*, 2000, pp. 158-167.