Fusion Based Linked Object Web Ranking for Social Networks

Pushpa R. Suri[1] and Harmunish Taneja[2]

[1] Department of Computer Science and Applications, Kurukshetra University
Kurukshetra, Haryana, India, Pin - 136119
[2] Department of Information Technology, M.M. Engineering College,
M.M. University, Mullana, Haryana, India, Pin - 133203
pushpa.suri@yahoo.com, harmunish.taneja@gmail.com

Abstract. Search is no doubt the most critical aspect of information computing over web. Ranking is pivotal for efficient web search as the search performance mainly depends upon the ranking results. The trade-off between the systems costs of extracting a single right web object and a set of potential results continues to be major design issue. Social networks have ensured the expanding disproportion between the face of World Wide Web stored traditionally in search engine repositories and the actual ever changing face of Web. As definition of search is changing, socially-enhanced interactive search methodologies are the need of the hour. In this paper new integrated ranking model based on page rank paradigm for sorting web objects from multiple social forums is proposed. This model identifies relationships between web objects in separate social networks particularly in the light of music search. It can be extended to video and other multimedia forms.

Keywords: Web objects, Ranking, Fusion Based Rank (FBR), Linked Object Web Ranking**,** Social Networks, Music search.

1 Introduction

Searching from billions of objects, makes it challenging to harness single right result. Also, ranking needs to be redefined beyond the boundaries of query driven search in today's scenario. Traditionally the entities in each search results are similar in sense they are web pages. With the proliferation of social sites there is a big bang explosion of the web users that are not just passive. Enhanced ranking needs to be prepared for selecting and merging heterogeneous results such as videos, images, news and local business listings and not just text features. Web expansion has transformed the "just enter in Google" search approach to the "likes" button of facebook that allows to submit personalized opinion on absolutely any object of interest on web. The "likes" play significant role in increasing or decreasing the popularity of the web object contained in social network. Social networks have paced the wave for involving users at all level of search and creating a raised level of interest in manual indexing [1]. Consequently its impact on the ranking method is imperative. All search engines use some

N. Meghanathan et al. (Eds.): CCSIT 2012, Part III, LNICST 86, pp. 76–86, 2012.
© Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2012

kind of web object link-based ranking in their ranking algorithms. This has been the result of the success of Google, and its PageRank link algorithm [2]. Traditionally from small amount of labeled data to more recently, machine learning technology are exploited to create ranking function. But existing technologies are inadequate for search applications offered by various modern social networks. A new learning framework that could rank related web objects that may be heterogeneous from multiple distinct and or similar social forums is desired. Vertical search engines refer to the search services restricted by specific information, such as music, image, video and academic search [3]. Popular vertical search engines like Academic Search [4], Google Scholar[5], Froogle [6], Movie Search [7], Image Search [8][9], Video Search [7], News Search [10] and many more have become increasingly effective in serving users with specific needs. But social networks reflect real world data sets as heterogeneous, semi-structured and multi-relational. Ranking function is most critical part of the search engine with two key factors viz. relevance and quality [11].

The social network users are the volunteers to actually rate the music which is a development to be noted for enhancing the traditional ranking methodology. If search results for query based on a song are coming from different social networks, then it is not an easy task to rank them because ranking of song in one social network may be higher and in the second it may be less. Motivated by such observation, the proposed framework i.e. FBR (Fusion based Rank) for Social networks fuse the ranking scores of searched web objects (music) from heterogeneous social networks by identifying the relation link between web objects from different social forums. We propose an integrated ranking method in heterogeneous social network by recognizing relatedness among different web objects that could be reflected in the rankings particularly a more famous a singer is, the more likely the songs will receive majority likes. Consequently more famous songs lead to more popularity of singers than those that are less cited. The paper is organised as follows. Section 2 outlines the challenges and choices faced in ranking web objects from multiple social networks and presents the attribute based comparison of popular social networking sites. Section 3 elaborates the framework for fusion based web object ranking in social networks. Section 4 describes the experimental study and results on the effectiveness of the FBR framework on information computing in a simulated music social network and finally in section 5 the paper is concluded.

2 Background

Traditionally search engines have only had access to three major types of data describing web pages, i.e., web page content, link structure and query or click through log data. A not so new entry to this family of document types is the user generated content describing the pages directly is less explored [12]. Networked web pages and their ranking has been noticed at length like PageRank [13], HITS [14]. The theory of PageRank is very simple, i.e, “good pages reference good pages” and are labelled with have higher PageRank. But ranking social network web objects, on the other hand, is employed for exploring links between similar web object from heterogeneous origins. Object-level ranking has been well studied for vertical search engines. But algorithms evolved from PageRank [13], PopRank [3] and LinkFusion [15] proposed to rank objects coming from multiple communities suggests that the ranking is restricted by the unavailability of well-defined graphs of heterogeneous data in some kinds of vertical domains in which objects. The challenge of integrating ranks of web objects from multiple heterogeneous social networks lies in their distinct ranking systems. The ranking interval or scale including the minimal and maximal ratings for each web object is varied from 5 points to 10 points rating scale. Another important issue is the rating criteria of different social forums. Also there may be ambiguity in the quality of the same score [16]. Future of WWW is where individuals are linked together in addition to web pages. The rise of social networks and its hold over common man interest from academician to a naive web surfer indicates the inefficiency of traditional search theories in general. Table 1 presents the comparative elaboration of the top fifteen most popular social networking sites with attributes like monthly visitors and various ranks (as on July 13, 2011) *eBizMBA Rank* [17], *Alexa* Global Traffic Rank [18] and U.S. Traffic Rank from both *Compete* [19] and *Quantcast* [20] that are constantly updated at each website's.

Table 1. Comparisons of the Top 15 Most Popular Social Networking Sites

Name	Logo	Estimated Unique Monthly Visitors	eBizMBA Rank	Compete Rank	Quantcast Rank	Alexa Rank
Facebook	facebook	700000000	2	3	2	2
Twitter	twitter	200000000	15	30	5	9
Linkedin	Linkedin	100000000	33	57	26	17
mySpace	my	80500000	50	26	44	79
Ning	Ning	60000000	143	180	120	128
Tagged	TAGGED	25000000	255	382	151	141
Orkut	orkut	15500000	401	570	540	93
hi5	hi5	11500000	479	983	392	62
myyearbook	myyearbook YOU'VE GOT FRIENDS!	7450000	617	522	293	1036
Meetup	Meetup	7200000	635	644	732	528
Badoo	badoo	7100000	653	1346	489	125
bebo	b	7000000	655	944	434	588
mylife	my life	5400000	865	118	688	1789
friendster	friendster	4900000	955	1920	643	301
Multiply	MULTIPLY	4300000	1136	2446	677	285

3 FBR (Fusion Based Rank): Framework

Manually the ranking criterion can be standardized by score normalization and transformation if the search involves same song or singer [21]. This section presents an integrated ranking framework FBR to analyse inter- and intra-type links and to sort web objects in different social networks at the same time. The conceptual scheme of proposed integrated ranking framework is shown in figure 1.

Two kinds of links among web objects are defined: intra-type links: which represent the relationship among web objects within a social network, and inter-type links: which represent the relationship among web objects between different social networks.

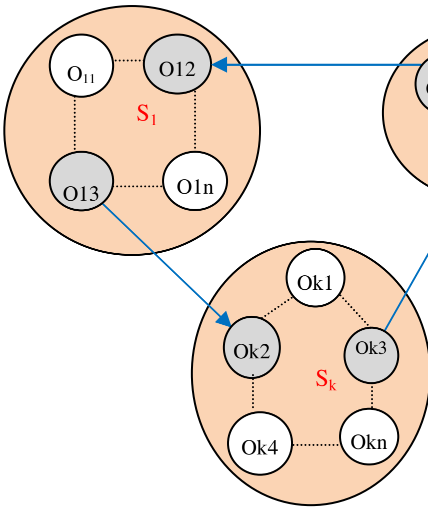

Fig. 1. Integrated social networks for enhanced Fusion based ranks: FBR

FBR exploits the two layer hierarchical approach as shown in Fig. 1 for identifying the inheritance and aggregation relationships between classes by creating super classes that may encapsulate the services or features common to other classes.

There are k graphs (S_1, S_2,......,S_k) where each $S_i = (O_{in}, E_i)$ is the un-weighted undirected graph (social network) of web objects O_{in} with links as E_i. The classes may belong to different social networks S_1, S_2,...S_k. Social Network S_i has a collection of related web objects (S_i: O_{i1}, O_{i2}, O_{i3}.....O_{in}) where i varies from 1 to k and k is the social networks integrated. If the web objects of different social networks share such a relationship as “has” or “contains” between them, an aggregation link is identified which is further implemented for rank computation.

The rank computed by each social forum is labelled as the partial rank and the framework add a popularity factor based on this partial rank to each link of the web object relationship graph pointing to the queried object. Different types of relationship are given different popularity factor. The proposed model calculates the rank for web objects with the activity flow of the FBR method as shown in Fig 2. The main components of the proposed framework are as below:

- Query Handler: The Query Handler receives the query terms based user requests and extracts one or more web objects from the entered query.
- Web Object Matching: All the related web objects contribute to the final rank of the search results. This module filters the related objects in various social networks under study.
- Relationship score: Relationships among objects contribute to their popularity factor and this module computes the score of searched web objects in terms of its relationships over other social networks.
- Fused Global Rank: This module computes the final Global rank indicating the actual level of relevancy of the search results.

Algorithm for FBR method (Fig. 2) is as below:

Algorithm: FBR	
Step 1	Rank web objects independently according to the PageRank paradigm [11].
Step 2	Web objects Partial ranks are computed by performing link analysis equipped with user's feedback as manual indexes.
Step 3	Furnish the Partial ranks for each social network S_i by exploiting Intra-type links.
Step 4	The popularity of web object is propagated to other social networks by exploiting inter-type links connected through popularity factor.
Step 5	The popularity factor is combined with the partial ranks to estimate a global rank for the searched web object.

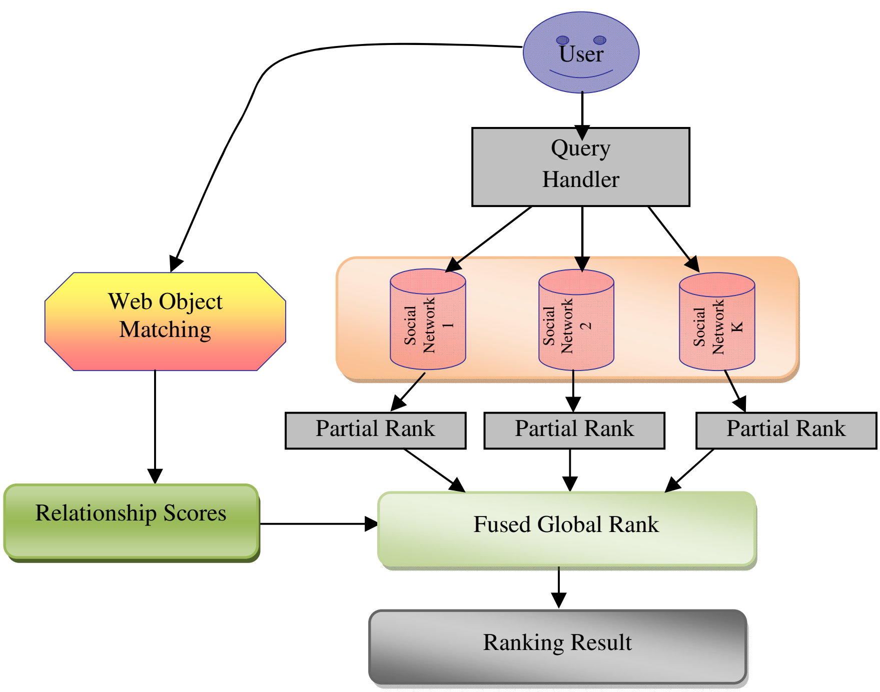

Fig. 2. Activity flow of FBR Method

4 Experimental Study and Results

There exist a number of redundant popular songs between any two social forums, because music web site may submit song to many social forums in order to obtain the feedback or comments. The proposed ranking method FBR emphasizes that although the ranks of duplicate songs in different social forums should be almost equal but practically this is not the case. The rank variation is ruled by the interest of the people connected to the social network. In the proposed work, Object orientation concepts help identifies the relationships of the searched web object within and across social networks. Each social network is viewed as a sub graph where each web object and its relationship with other web objects is reflected in the form of multiple level inheritance.

Fig. 3. Music Object relationship graph

The attribute value of web objects is obtained by iteratively calculating the periodic user's feedback. It is clear that the more popular the web object is, the more likely the user is interested in it. For ranking the most unique feature of the proposed approach is the web object graph with heterogeneous links e.g. a music object song may be searched by a set of other music objects that may be sung by a set of singers object and included in other web object album as shown in Fig. 3.

4.1 Datasets

Our study requires two sets of data: web objects and popular social networks. We explored various options for datasets, and eventually decided not to use data from popular social network services, such as Orkut, Facebook as the datasets in these sites are rich on social networks, but poor on web documents. We used data from YouTube which offers both rich web document data and extensive social network information. We first started with ten internet users randomly chosen with high degree and different interests, and obtained their friends and music interests. Then we used these friends as the new players and fetched the friends and music metadata from these. This process was iterative and stopped until no more options were available or the number of retrieved web users exceeded a pre-defined threshold.

4.1.1 Web Object Data

In the YouTube music dataset, each web object contains rich metadata such as album title, genre, tags, popular artists, description, uploader, rating, etc. These metadata were downloaded with our crawler and stored locally. The music is labelled with 10 main categories (Pop, Rap, Rock, Jazz, Today's hit, Reggae, Country, International, Comedy, and Latin etc.) as provided by YouTube. In this experiment, we defined a user's interest with one of these music categories. If the majority of a user's videos fall into one category, we regard that web object with higher relatedness score. With the metadata of music web objects, the indexes for music documents are built. The values in the fields: title, tags, genre and description were parsed into terms to create an inverted index using, in which each term in a specific field points to a collection of music documents. These indexes can be easily used to generate term document relatedness scores.

4.1.2 Social Network

In YouTube, each user has attributes such as name, gender, location, date of birth, friends, music uploaded, and about-me. Social networks can then be constructed based on the information obtained. An inter-type link was created between two web objects if they exist in similar context. In this experiment, we used two fully connected social networks based on the downloaded music data:

- S_1: Social Network 1: a larger social network that consists of 18,576 different registered users and 39,271 music uploaded by users;
- S_2: Social Network 2: a smaller network that has 2,464 users and 7,978 music uploaded by users.

4.2 Evaluation

To examine the effectiveness of the proposed Fusion based ranking algorithm, we compared following two algorithms:

Baseline: tf-idf [11] was used as the base line method.

Fusion based Rank (FBR): *FBR*(Q,O,Rel) = *f*(S_i Rank(Q, Web object), Rel (Rel Score))

where Q is the query, Rel score is the similarity value between web objects in social networks. tf-idf values are normalized.

In each category, we choose ten queries that are related to the music category but at the same time could also refer to things that do not belong to the category. Queries in the music category were about album titles, singer etc. The top twenty returned music of each query were mixed and presented to two web surfers for independent evaluation. They evaluated each returned result based on to what extent the returned music was relevant to a query as well as to what extent the music was related to the user's interest. Table 2 shows the scores used to rate the relevance of a return. The highest score of a return is 5 and the lowest score is 0. After a pilot experiment, the inter-rater reliability among two raters was 70%, indicating that they reached a reasonable agreement about the relevance criteria [22].

Table 2. Evaluating Content Relevance

	High	Medium	Low
High	5	3	0
Medium	3	2	0
Low	0	0	0

We used *Normalized Discounted Cumulative Gain (NDCG)* [23] metric to evaluate the ranking algorithms. This method measures the usefulness of the ranking result based on the relationship between the relevance scale of web objects and their position in the ranking. The premise of DCG is that highly relevant documents are more useful when they have higher ranks in the result list. The NDCG at position k is given by:

$$NDCG_k = \frac{DCG^R_K}{DCG^T_K}$$

$$DCG^X_K = Rel_1(X) + \frac{\Sigma\, Rel_i(X)}{\log i}$$

where i *goes from 2 to k*

$Rel_i(X)$ shows the level of relevancy for the result at position i in rank X.

DCG^T_K is the value for the optimal rank at position k.

4.3 Results

We evaluated the performance of two different ranking approaches. The NDCG values were averaged over all queries without considering the differences in the size of social networks and the degrees. As shown, FBR method performed better than the baseline algorithm. Fig. 4 shows that, for the first 20 search results, the NDCG values of FBR are higher than that of the baseline algorithm.

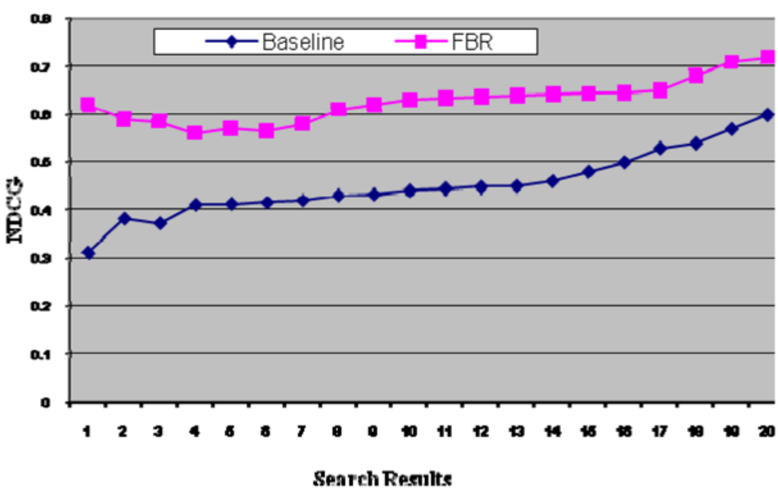

Fig. 4. NDCG: Fusion Based Rank Vs Baseline

The performance of these two algorithms is compared in two social networks with different size. To make the results more comparable, we only used searchers for the music category in both networks. The NDCG results from two social networks are shown in Fig. 5.

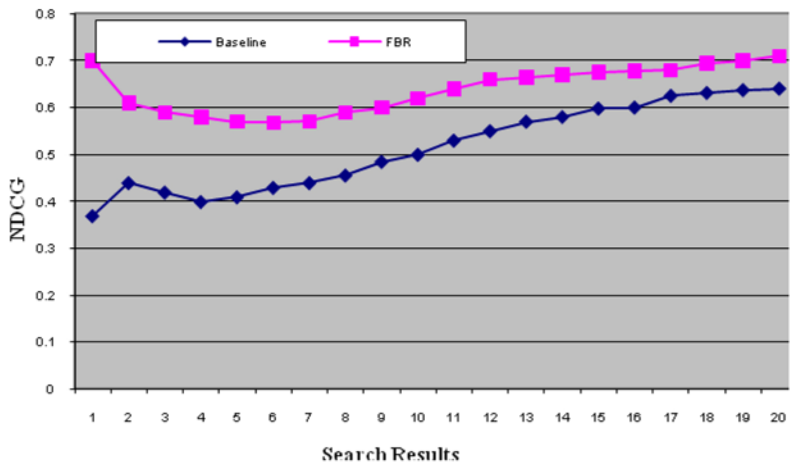

Fig. 5. NDCG: Fusion Based Rank Vs Baseline for Larger Social Network 1

The NDCG scores are averaged over all queries from searchers of music for the two networks. As shown, for the larger social network 1, the FBR method performs better than the baseline algorithm. But the difference in the smaller social network 2 is minimal as shown in Fig.6.

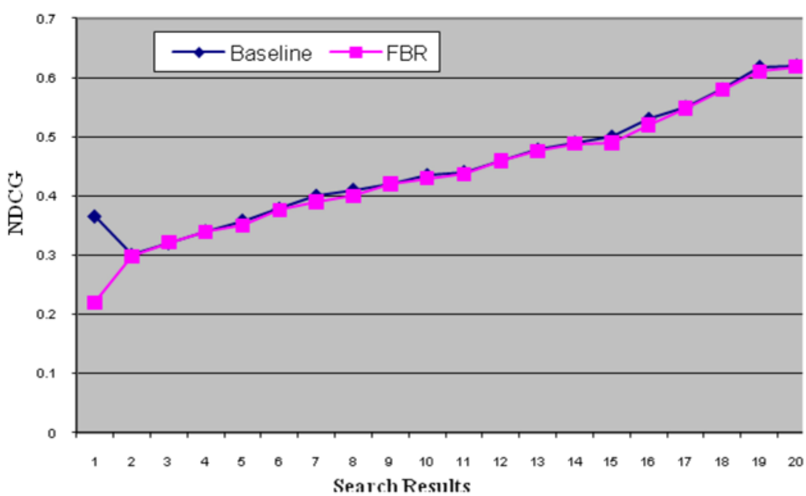

Fig. 6. NDCG: Fusion Based Rank Vs Baseline for Small Social Network 2

5 Conclusion

Social networks facilitate the formation of user generated content in diverse formats. Such networks smooth not only the classification user oriented web objects but also social indexing. Users send comments and feedbacks, likes or dislikes resources, upload and share multimedia contents, communicate online with social contacts, pour in wiki-style knowledge bases, maintain personal bookmarks.

In this paper, the web object ranking problem in case of social networks where traditional ranking algorithms are no longer valid is explored. The fusion based rank (FBR) method integrates many social networks and identifies the aggregation relationship graph highlighting the links among the related or duplicate web objects. The results of our evaluation studies indicate that overall, the FBR framework can return better search results than the traditional tf-idf ranking algorithm in terms of relevance, the ranking effectiveness of returned search results as it considers the global information of a social network. But its performance is appreciable only as the network size reaches a certain magnitude. FBR method is more effective than the baseline approach as it considers the global information of social networks. FBR method benefits more from large social networks than from small networks. The results reflect that the proposed fusion methods are practical and efficient solution to web object ranking in social networks. This framework can be applied with modifications to other multimedia applications.

References

1. Voss, J.: Tagging Folksonomy and Co-Renaissance of Manual Indexing. In: Proceedings of the International Symposium of Information Science, pp. 234–254 (2007)
2. Ding, C., He, X., Husbands, P., Zha, H., Simon, H.: Pagerank, Hits and a Unified framework for link analysis. In: NERSC Division, Lawrence Berkeley National Laboratory, 94720, pp. 1–12. University of California, Berkeley (2001-2002)
3. Nie, Z., Zhang, Y., Wen, J.-R., Ma, W.Y.: Object-level ranking: Bringing order to web objects. In: Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, pp. 567–574 (2005)
4. Web Site: Microsoft academic search, `http://academic.research.microsoft.com/`
5. Web Site: Google scholar paper search, `http://Scholar.google.com`
6. Froogle. Google product search, `http://froogle.google.com`
7. GoogleVideo. Google video search, `http://video.google.com`
8. Google. Google image search, `http://images.google.com`
9. Yahoo. Yahoo image search, `http://images.yahoo.com`
10. GoogleNews. Google news search, `http://news.google.com`
11. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Libraries, pp. 1–17 (1998)
12. Paul, H., Georgia, K., Garcia-Molina, H.: Can social bookmarking improve web search? In: First ACM International Conference on Web Search and Data Mining, WSDM, Stanford, pp. 195–206 (2008)
13. Brin, S., Page, L.: The anatomy of a large-scale hyper textual Web search engine. In: 7th WWW Conference, Brisbane, Australia, pp. 107–117 (April 1998)

14. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. Journal of the ACM 46(5), 604–632 (1999)
15. Xi, W., Zhang, B., Chen, Z., Lu, Y., Yan, S., Ma, W.-Y., Fox, E.A.: Link fusion: A unified link analysis framework for multi-type interrelated data objects. In: Proceedings of the 13th International Conference on World Wide Web, pp. 319–327 (2004)
16. Bao, S., Xue, G.R., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing web search using social annotations. In: WWW, pp. 501–510 (2007)
17. http://www.ebizmba.com/articles/social-networking-websites
18. Web Site: http://www.alexa.com/
19. Web Site: http://www.compete.com/
20. Web Site: http://www.quantcast.com/
21. Eck, D., Lamere, P., Bertin-Mahieux, T., Green, S.: Automatic Generation of Social Tags for Music Recommendation. In: Proceedings of Neural Information Processing Systems (NIPS), pp. 1–8 (2007)
22. Landis, R., Koch, G.: The measurement of observer agreement for categorical data. Biometrics 33, 159–174 (1977)
23. Jarvelin, K., Kekalainen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of the ACM Conference on Information Retrieval (SIGIR 2000), pp. 41–48 (2000)