Evaluating Bias in Retrieval Systems for Recall Oriented Documents Retrieval

Sanam Noor[1] and Shariq Bashir[2]
[1]Department of Computer Science, University of Peshawar, Pakistan
[2]Centre of Science and Engineering, New York University Abu Dhabi, UAE

Abstract: *The evaluation of a retrieval system has always been the focus of research. Most of the retrieval systems seem to be more efficient for precision oriented documents than recall oriented documents since there is a difference between both the recall and precision oriented documents. Therefore, a system that is efficient for the retrieval of precision oriented documents does not need to be good for recall oriented documents as well. Evaluation of retrieval system is very necessary in order to determine whether these methods are suitable for recall oriented documents retrieval or not. We evaluate different retrieval systems for recall oriented documents retrieval. Our main focus is on finding the bias in retrieval systems. We use different retrieval systems for evaluation; in which four are query expansion techniques while the other three retrieve documents without using query expansion techniques. Patent documents are used for analyzing the effectiveness of retrieval systems. Accessibility of documents is measured by retrievability measurement. Lorenz curve and Gini coefficient are used for measuring bias in systems. Our experiments results show that Term Frequency Inverse Document Frequency (TFIDF) is less biased. While exact method show high retrievability inequality. In query expansion techniques language modelling shows less inequality.*

Keywords: *Retrieval systems evaluation, search systems bias analysis, retrievability measurement, patent retrieval.*

Received February 28, 2013; accepted April 4, 2013; published online April 17, 2014

1. Introduction

Information Retrieval (IR) is the automatic identification of those documents in a large document collection that are relevant to an explicitly stated information need [5]. Precision and recall are two famous methods for measuring the performance of IR systems. Precision measures how precise the search is the higher precision; the less irrelevant document will be retrieved. When every retrieved document is relevant the value of precision is 1 and when every retrieved document is irrelevant precision is 0. Recall measures how complete the search is the higher recall, the less missing documents. When all relevant documents are retrieved the value of recall is 1. There is an inverse relation between recall and precision. High recall is achieved at the expense of precision. High recall means not to miss any relevant document, which requires searching all documents in a collection that decreases precision. Search engine can increase its recall by retrieving more documents at the cost of increasing irrelevant documents. Recall is thus an expression of how exhaustive a search for documents is. Precision oriented documents are those documents where the retrieval of all documents is not mandatory such as news. While in recall-oriented documents the retrieval of all documents is mandatory such as patent documents, legal documents or literature. In case of recall oriented documents exhaustive search is required to get the required information. The information needed by the user is spread over multiple documents; user has to look at multiple documents to get the required information. In case of patent documents equal retrievability of all documents is very necessary.

Different retrieval systems need to be evaluated for recall oriented documents to find out whether these techniques are suitable for recall oriented documents or not. Retrieval systems are used to retrieve documents relevant to the user's queries. In our experiments, we use seven retrieval systems Term Frequency Inverse Document Frequency (TFIDF), BM25 [23], exact match, Kullback Leibler Divergence (KLD) [16, 26], Term Selection Value (TVS), Language Model (LM) [16] and clustering approach [18, 19]. TFIDF, BM25 and exact method retrieve document without using query expansion. KLD, TVS, LM and clustering are query expansion techniques. TFIDF is a ranking function, which determines the weight of a particular term in a document. It measures how important a term is to a document in a collection. BM25 measures how relevant a document is to a query based on TFIDF, document length and other statistics [23]. Exact match is based on boolean operation. Exact match retrieves documents according to criteria specified by the user in query. Query expansion is a technique used to reformulate and enhance the user query to improve search results. Query expansion uses different techniques to select expansion terms. TVS, KLD [16], LM and Pseudo Relevance Feedback (PRF) documents selecting using clustering [18, 19] are different techniques of query expansion.

Retrievability is the ease at which a document can be retrieved through a system [1, 2, 3, 4]. If the search tool limits what can be accessed within the collection then there is an increased likelihood that documents which are biased against cannot be easily retrieved [22, 24]. Missing one patent document can lead to copyright infringement and can cause million-dollar lawsuit [7, 8, 9, 11, 13, 14].

Retrieval system is the mean by which we access documents. Retrieval systems are playing vital role in providing access to documents in collection. In recall-oriented documents such as patent documents search accessibility of each and every relevant document is very vital. Since, the role of retrieval systems is very important in accessibility of documents. This provides motivation for analyzing the influence of retrieval systems on accessibility of documents.

The remaining paper is organized as follows: Section 2 describes the related research work. In section 3 we have described retrieval systems and the proposed methodology. Section 4 presents the results of experiments. Finally section 5 concludes this paper.

2. Related Work

Due to novel and recently proposed domain, there is no extensive research done on retrievability measure. However, in the past there exist a number of studies on the web coverage of search engines and these are somewhat related to this domain. In the following section, we provide an overview of the major works of both domains:

a) Bias Analysis on The Basis of Web Coverage.
b) Bias Analysis on The Basis of Documents Retrievability.

2.1. Bias Analysis on the basis of Web Coverage

Lawrence and Giles [17] performed a study to analyze the coverage bias of web search engines. For this purpose they used 6 search engines and a large query log from a scientific organization. These queries should return the same set of pages for all 6 engines, as they thought that these engines have similar coverage since they are indexing the same set of documents. To express the coverage of the engines with respect to the size of the web, they used 128 million pages from northern light search engine at the time of their experiments as an absolute value. Their experiments revealed that no single search engine covers more than 57.5% of the estimated full web. They also showed that some large search engines only cover less than 5% of the web. Finally, the authors concluded that the solution to the problem of search engines not indexing the whole web is to use Meta search engines or to define Goal-Driven search engines that have a specific focus e.g., sports or scientific literature.

Vaughan and Thelwall [25] performed a study on the coverage of web pages from 42 countries to discover the index bias of three major search engines. For this purpose, they used their own research crawler and crawled domains from 42 countries. A large number of queries were submitted to three search engines and their developed research crawler. The bias quantification was on the basis of site coverage ratio, and it was computed on the number of pages covered by the search engines divided by the number of pages covered by their research crawler. The main limitation of their study was that it did not consider the constantly changing nature of the web, as their developed crawler could remain behind the indexes of search engines since they did not have similar number of resources available as major search engines have.

Mowshowitz and Kawaguchi [20] undertook a study to discover bias in fifteen major commercial search engines. In order to generate queries, they used the ACM computing classification system as queries and the top 30 results of each search engine were recorded. Their large experiments results confirmed that there was some bias in all search engines. Their proposed bias measurement uses the number of unique domains as a ranked array based on the combination of all web search results returned by the queries. However, this measurement could itself introduce bias into the experiments, as it is not based on all possible results of the web but only on the combinations of the web pages returned from the search engines. Secondly, their measurement cannot show if there is a bias against particular results if all of the included search engines are biased against similar results.

Lauw *et al.* [15] found that deviation (controversy) in the evaluation scores of objects in the reviewer-object models can also be used for discovering bias. They observed that bias and controversy of reviewers to objects are mutually dependent to each other. This dependency indicates that there will be more biased if there is high deviation towards less controversial object. To identify this controversy and bias they proposed a reinforcement model. Their approach of discovering bias can also be applied in the web search setting. In this case, the reviewers can be regarded as web search engines and the objects that they are reviewing (ranking) are web pages. On the basis of this approach, search engines will be more biased if they give high ranks to low ranked web pages of other search engines.

Owens [21] conducted a recent study on the bias analysis of search engines. One major concern of their study was to discover whether the search engines unfairly lead users to particular sites over other sites. For this purpose they discovered the relative news bias of 3 search engines. They reported this relative bias amongst search engines in the forms of political bias and predilection for specific sites. They performed the experiments over 9 weeks, and posed a large number of realistic and currently topical queries to the news sections of 3 search engines. On the basis of their experiments results they showed that there are significant biases towards predilections for a certain news sources in all search engines.

All these studies revealed a range of possible biases, for example, if one site has more coverage than the other. These studies are usually motivated by the view that the search engines may be providing biased content and these measures are aimed at being regulatory in nature, whether the sites in a particular geographical location are favoured, or whether the search engines are biased given a particular topic. As opposed to web coverage our work focuses on individual documents retrievability and this can be also used to detect such biases.

2.2. Bias Analysis on the Basis of Document Retrievability

Azzopardi and Vinay [2] proposed the concept of accessibility (findability, retrievability) of documents. They adopted this concept from transportation planning and used it in the context of IR. The IR system is like being at a bus stop where every possible bus route is available, (i.e., the universe of all possible queries) and we can select any route desired, at anytime [1]. Their work focused on measuring the accessibility of documents in the collection given the IR system used to access these documents. The influence of the IR system is examined to restrict or promote access to the information within the collection as opposed to other restrictions. They proposed one of most important function called document accessibility function or retrievability measure [1, 2]. Azzopardi and Vinay [2] proposed a methodology to evaluate IR model. They evaluate four IR model TFIDF, BM25, BM25i and LM. Retrievability measure is used for measuring retrievability of documents. The experiments results show that BM25i favours documents with more incoming links and TFIDF and LM1000 tends to favour documents with less outgoing links. They analyzed the impact of IR system on collection so that, the collection IR model can be improved for better access.

Bashir *et al.* [4] used the retrievability function proposed by [2] to analyze the effectiveness of IR model for patent documents. They used four IR models i.e., BM25, BM25F, TFIDF and Exact method in their experiments. They extracted terms from claim sections of patent documents and expanded these terms into two, three and four terms by using query expansion. They identify relevant and irrelevant queries before using it for retrieval. Rather than using just one measurement they analyze documents accessibility using four different measurements. They showed how documents retrievability is affected by using relevant, irrelevant and set of all queries. Their proposed approach is suitable for “invalidity search” and “Patentability” retrievability measurements.

Bashir *et al.* [3] proposed query expansion technique based on documents clustering for PRF. Their experiments show that clustering approach for PRF is an effective approach for increasing the findability of individual documents and decreasing the bias of a retrieval system. Their proposed system show less bias than other system.

3. Experiments

The main purpose of our research is to evaluate different retrieval systems for recall oriented documents. Our focus is on evaluating bias in retrieval systems. We compare the performance of different retrieval systems for patents documents. We evaluate retrieval systems to find out whether these models equally retrieve all documents or there exist any kind of bias. Seven state-of-the art retrieval models including standard query expansion methods are used for evaluating the retrievability inequality. These are:

1. TFIDF.
2. OKAPI Retrieval Model (BM25) [23].
3. Exact Match.
4. LM with Term Smoothing [26].
5. KLD for Query Expansion [16].
6. Term Selection Value for Query Expansion (QE-TS) [16].
7. PRF Selection using Clustering, Lee *et al.* [18].

TF-IDF is a ranking function which determines the weight of a particular term in a document. It measures how important a term is to a document in a collection. Term Frequency (TF) is the number of time a term appears in a document. Term is more important to a document if it appears more often in a document. IDF is used to measure the importance of term in a collection. TF-IDF will have high weight if TF is high and document frequency is low. BM25 is a ranking function used by search engines to rank matching documents according to their relevance to a given search query. BM25 ranks a set of documents based on the query terms appearing in each document, regardless of the inter-relationship between the query terms within a document. Exact match retrieves documents according to criteria specified by the user in query. Only those documents that exactly match the query criteria are considered relevant and are retrieved by the system. TVS is a Query expansion technique. Query expansion is a technique used to reformulate and enhance the user query to improve search results. TVS is used to choose expansion terms for the query, and these terms are then added to the user query. KLD is also called Relative Entropy, it is used to measure distance or make comparison between two documents. KLD is one of the methods used for ranking candidates terms for query expansion. LM is one of approach used in IR to rank documents. Beside documents retrieval language models can be applied to relevance feedback and query expansion. A document can be a good match for query if it is likely to generate the query. Each document has its own language model. When query *q* is submitted, documents are ranked based on probability that document will generate query. Clustering is used to make clusters of similar documents and isolates irrelevant documents. Clustering can also be used to remove query ambiguity. One of the methods proposed by Lee *et al.* [18] is to select dominant documents and using that dominant document for expansion. First documents are retrieved by using query likelihood model with Dirichlet smoothing. Next for top n documents clusters are created using K-NN method to find dominants documents.

The main steps for analyzing retrieval systems are as follow:

- *Step 1*: Create queries by using two approaches Query Generation Frequent Terms (QG-FT) and Query Generation with Document Relatedness (QG-DR). QG-FT extracts terms from claim section that have support greater than three. QG-DR creates queries by using K-NN algorithm.
- *Step 2*: Run the queries and find retrievability of patents. On basis of queries patents will be retrieved and ranked by retrieval system
- *Step 3*: Sort retrievability *r(d)* of patents in ascending order. Find cumulative *r(d)* of patents documents.
- *Step 4*: Represent the retrievability *r(d)* of patents by Lorenz curve. Draw Lorenz curves for cumulative retrievability *r(d)*.
- *Step 5*: Use Gini coefficient to summarize bias in Lorenz curve.

3.1. QG-DR

Steps of QG-DR are as follows [12]:

1. Create set of related documents by using K-nearest algorithm.
2. Define LM for source documents and collection.
3. Sort the terms in the vocabulary based on their contribution to relative entropy. Use relative entropy to compare source document set to the collection.

$$Score(t) = PR(t)log\,\frac{PR(t)}{PC(t)} \tag{1}$$

4. Identify the terms that contribute most to the entropy. In CQG queries are generated based on its contribution to relative entropy. The most discriminating term is used to create initial query. Two term query is then created by combining the first term with second most discriminating terms. In same way queries are created.
5. This process is repeated until no terms are left in vocabulary.

Following the steps of QG-DR we create queries. After queries generation we read those queries and find the retrievability score of retrieval systems.

3.2. Retrievability Measure

Retrievability measure how likely a document *d* can be retrieved in top *c* ranked results for all queries in *Q*. Given a collection *D*, an IR system accepts a user query *q* and returns a ranking of documents, which are deemed to be relevant to user query from collection *D* by IR system [2].

Retrievability measure of a document *d* is calculated as follows:

$$r(d) = \sum_{q \in Q} f(K_{dq}, c) \tag{2}$$

$f(K_{dq}, c)$ is a generalized utility/ cost function where K_{dq} is the rank of *d* in the result for query *q*, and *c* denotes the maximum rank that a user is willing to proceed down the ranked list. The function returns a value of 1 if $K_{dq} \leq c$ and *0* otherwise.

For measuring retrievability score of documents we read the queries and found the retrievability score of each document for different retrieval systems. After running the queries we obtain the retrievability of documents. For measuring inequality we use Gini coefficient and Lorenz curve. Gini coefficient and Lorenz curve are two interlinked methods of measuring inequality. Gini coefficient compares Lorenz curve with the line of perfect equality and is calculated as follows:

$$G = \frac{\sum_{i=1}^{N} (2 * i - N - 1) * r(di)}{N * \sum_{i=1}^{N} r(di)} \tag{3}$$

We draw Lorenz curve on the basis of retrievability scores [10]. It shows the retrievabality inequality of different retrieval systems. It shows us how bias is the system. Gini coefficient is use to summarize bias in Lorenz curve. We calculate Gini coefficient for rank cut off values of 30 and 90.

4. Results

We use a collection of US Patents Classification (USPC) in our experiments. We use USPC class 433 (Dentistry) and class 424 (Drug, bio affecting and body treating compositions). These patents are available on ww.uspto.gov. Total numbers of documents in Classes (433, 424) are 43, 225 and total numbers of unique terms are 325, 921. We use two approaches for queries generation. In QG-FT we extract those terms from the claim section that have support greater than three. We combine these terms to form two, three and four terms queries.

In Query generation with document relatedness QG-DR approach cluster of the related document is created by using K-NN algorithm [6, 12]. Cluster for each document is created by using 35 neighbors. Language modelling is applied on clusters to extract relevant terms.

$$Score(t) = P_R(t)log\,\frac{P_R(t)}{P_C(t)} \tag{4}$$

Where, $P_R(t)$ is the probability of term *t* in cluster (related set of document) and $P_C(t)$ is the probability of term *t* in whole collection. Those terms (top seventy terms) that contributed most to relative entropy were extracted from clusters and were used as queries [19].

To find the retrievability scores of patents queries are read by the retrieval system. On basis of queries relevant patents are retrieved by the retrieval system. When retrieval system completes its processing we obtain retrievability scores of patents. Retrievability score of patents show the number of times it has been retrieved by retrieval system in top ranked documents.

First, we sort retrievability scores in ascending order then we calculate cumulative retrievability of patents. To present visually bias in retrieval system we draw Lorenz Curves for cumulative retrievability. Finally Gini coefficient is used to present bias in retrieval system as a single value.

Figures 1, 2 and 3 show retrievability inequality of different retrieval system with rank cut off factor 30. As we can see from above figures that TFIDF consistently show less bias as compare to all other retrieval systems (Including Query expansion based retrieval system), while Exact method shows greatest bias for all length of queries. Exact method shows worse performance. For two terms queries BM25, KLD and Lee *et al.* [18] show similar results and show less bias. KLD shows less bias for two terms queries but its performance degrades as the number of terms in query is increased. KLD does not perform well for three and four terms queries. Lee *et al.* [18] and QE-TS do not perform well. In Query Expansion based techniques LM shows better performance. Figures 4, 5 and 6 show Lorenz Curves of retrieval systems. Queries are generated by QG-DR approach and rank cut off factor is 30. Figures 4, 5 and 6 show retrievability score of different retrieval system with rank cut off factor 30 and queries are generated by QG-DR approach.

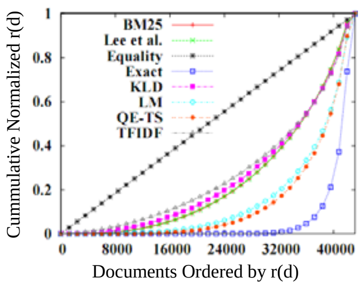

Figure 1. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Two terms queries are used by generating queries with QG-FT.

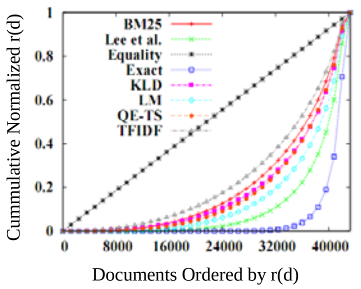

Figure 2. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Three terms queries are used by generating queries with QG-FT.

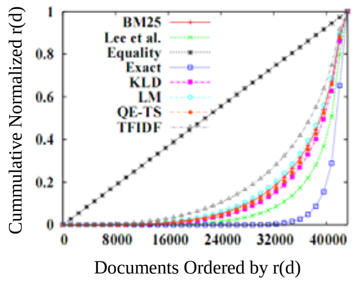

Figure 3. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Four terms queries are used by generating queries with QG-FT.

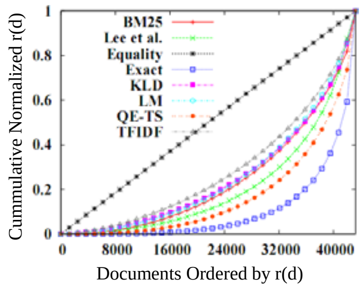

Figure 4. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Two terms queries are used by generating queries with QG-DR.

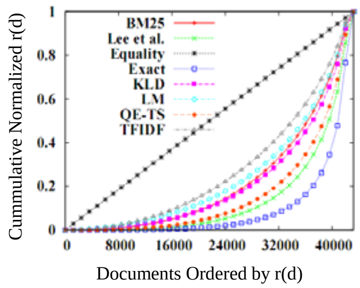

Figure 5. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Three terms queries are used by generating queries with QG-DR.

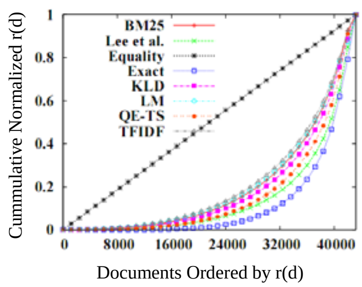

Figure 6. Lorenz curve showing retrievability inequality between documents for USPC (433, 424) collection. Four terms queries are used by generating queries with QG-DR.

As we can see from above figures that overall TFIDF show less bias and perform better than all other retrieval systems. While exact method shows worse bias and shows worse performance as compare to all other retrieval systems. For two terms queries KLD, LM and BM25 show almost similar results, they show less bias. For three terms queries BM25 shows less inequality. In Query Expansion based techniques LM shows better performance. QE-TS and Lee *et al.* [18] show high inequality.

Table 1 shows Gini coefficient values of different retrieval system for two, three and four terms queries with rank cut off value of 30. As we can see from Table 1 that performance of retrieval system is degraded as we increase the number of terms in queries. For longer length queries the Gini coefficient values are high as compare to shorter length queries. Retrieval systems show high bias for longer queries. We can see from Tables 1 and 2 that performance of retrieval systems are slightly better for QG-DR approach as compare to QG-FT. QG-DR uses clustering for queries generation. QG-DR makes the group of relevant documents and helps in removing irrelevant documents. Queries are then created from the cluster of relevant documents and remove irrelevant queries. Even through the retrieval system perform slightly good for queries generated by QG-DR as compare to QG-FT but the difference is so small that it's negligible.

Table 2 shows Gini coefficient values of different retrieval system with rank cut off value of 90. We can see from Table 2 as the value of rank cut off c is increased to 90 the bias in system decreased. The reason is that as the user proceeds down the list more relevant documents are retrieved and result in lower bias in the system.

Table 1. Gini-coefficient scores with rank cut-off factor (c=30) and different retrieval models for USPC (433, 424) collection.

Retrieval Model	Query Generation Mechanism	Two Term Queries	Three Term Queries	Four Term Queries
BM25	QG-FT	0.50	0.58	0.67
	QG-DR	0.51	0.51	0.65
TFIDF	QG-FT	0.48	0.53	0.63
	QG-DR	0.47	0.51	0.58
Exact	QG-FT	0.76	0.8	0.8
	QG-DR	0.70	0.85	0.75
LM	QG-FT	0.53	0.59	0.67
	QG-DR	0.50	0.55	0.64
KLD	QG-FT	0.50	0.63	0.71
	QG-DR	0.51	0.61	0.68
QE-TS	QG-FT	0.67	0.74	0.7
	QG-DR	0.63	0.71	0.72
Lee *et al.* [18]	QG-FT	0.50	0.74	0.77
	QG-DR	0.61	0.76	0.74

Table 2. Gini-coefficient scores with rank cut-off factor (c=90) and different retrieval models for USPC (433, 424) collection.

Retrieval Model	Query Generation Mechanism	Two Term Queries	Three Term Queries	Four Term Queries
BM25	QG-FT	0.42	0.54	0.67
	QG-DR	0.46	0.57	0.63
TFIDF	QG-FT	0.35	0.55	0.65
	QG-DR	0.43	0.47	0.59
Exact	QG-FT	0.61	0.67	0.73
	QG-DR	0.56	0.65	0.65
LM	QG-FT	0.38	0.52	0.65
	QG-DR	0.42	0.52	0.6
KLD	QG-FT	0.41	0.54	0.67
	QG-DR	0.44	0.54	0.62
QE-TS	QG-FT	0.54	0.65	0.68
	QG-DR	0.53	0.61	0.67
Lee *et al.* [18]	QG-FT	0.42	0.65	0.68
	QG-DR	0.49	0.62	0.65

5. Conclusions

We perform experiments on patent data set and analyze different retrieval system. We use two approaches for generating queries i.e., QG-FT and QG- DR. We didn't see any significant difference in the performance of both approaches.

Our experiments results show that TFIDF is less biased. While exact method show greatest retrievability inequality. Exact method consistently shows worse results for different length of queries. The performance of BM25 and LM is good. They show less retrievability bias and almost show similar results. KLD performance is good as compare to Clustering and QE-TS. Clustering method and QE-TS do not perform well for patent documents. Overall TFIDF shows better performance as compare to all other retrieval system. In query expansion techniques LM gives good results. After analyzing the results, we find that due to bias in retrieval system some documents are hard to find in a collection. It's recommended to use those retrieval systems for retrieval of patent documents that are less biased or give equal priority to all documents. Keeping in view the bias and limitation of retrieval systems in future new retrieval system can be designed that provide equal access to documents and are more suitable for patent documents retrieval.

References

[1] Azzopardi L. and Vinay V., "Accessibility in Information Retrieval," *in Proceedings of the 30th European Conference on IR Research*, Glasgow, UK, pp. 482-489, 2008.

[2] Azzopardi L. and Vinay V., "Retrievability: An Evaluation Measure for Higher Order Information Access Tasks," *in Proceedings of the 17th ACM conference on Information and knowledge management*, New York, NY, USA pp. 561-570, 2008.

[3] Bashir S., "Improving Retrievability with Improved Cluster-based Pseudo Relevance Feedback Selection," *Expert Systems with Application Journal*, vol. 39, no. 8, pp. 7495-7502, 2012.

[4] Bashir S. and Rauber A., "Improving Retrievability of Patents in Prior-art Search," *in proceedings of the 32nd European Conference on Information Retrieval Research*, Milton Keynes, UK, pp. 457-470, 2010.

[5] Batri K., Murugesh V., and Gopalan N., "Effect of Weight Assignment in Data Fusion Based Information Retrieval," *the International Arab Journal of Information Technology*, vol. 8, no. 3, pp. 244-250, 2011.

[6] Custis T. and Al-Kofahi K., "A New Approach for Evaluating Query Expansion: Query-Document Term Mismatch," *in Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development on*

Information Retrieval Location, Amsterdam, Netherlands, pp. 575-582, 2007.

[7] Doi H., Seki Y., and Aono M., "A Patent Retrieval Method using a Hierarchy of Clusters at TUT. NTCIR '05," *in Proceedings of NTCIR-5 Workshop Meeting*, Tokyo, Japan, 2005.

[8] Fall C., Torcsvari A., Benzineb K., and Karetka G., "Automated Categorization in the International Patent Classification," *ACM SIGIR Forum*, vol. 37, no. 1, pp. 10-25, 2003.

[9] Fujii A., Iwayama M., and Kando N., "Introduction to the Special Issue on Patent Processing," *Information Processing and Management: an International Journal*, vol. 43, no. 5, pp. 1149-1153, 2007.

[10] Gastwirth J., "The Estimation of the Lorenz Curve and Gini Index," *The Review of Economics and Statistics journal*, vol. 54, no. 3, pp. 306-316, 1972.

[11] Itoh H., Mano H., and Ogawa Y., "Term Distillation in Patent Retrieval," *in proceedings of the ACL-2003 workshop on Patent corpus processing*, Sapporo, Japan, pp. 41-45, 2003.

[12] Jordan C., Wattters C., Gao Q., "Using Controlled Query Generation to Evaluate Blind Relevance Feedback Algorithms," *in Proceedings of Joint Conference on Digital Library*, NC, USA, pp. 286-295, 2006.

[13] Konishi K., "Query Terms Extraction from Patent Document for Invalidity Search," *in Proceedings of NTCIR '05: NTCIR-5 Workshop Meeting*, Tokyo, Japan, pp. 1-6, 2005.

[14] Konishi K., Kitauchi A., and Takaki T., "Invalidity Patent Search System at NTT Data," *in Proceedings of NTCIR-4 Workshop Meeting*, Tokyo, Japan, pp. 1-7, 2004.

[15] Lauw H., Lim E., and Wang K., "Bias and Controversy: Beyond the Statistical Deviation," *in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, USA, pp. 625-630, 2006.

[16] Lavrenko V. and Croft W., "Relevance Based Language Models," *in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Louisiana, USA, pp. 120-127, 2001.

[17] Lawrence S. and Giles C., "Accessibility of Information on the Web," available at: http:// nike.psu.edu/ classes/ ist597/ 2003-fall/ papers /

[18] Web-access.pdf, last visited 1999.

[19] Lee K., Croft W., and Allan J., "A Cluster-Based Resampling Method for Pseudo-Relevance Feedback," *in Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 235-242, New York, NY, USA, 2008.

[20] Liu X. and Croft W., "Cluster-Based Retrieval using Language Models," *in Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, UK, pp. 186-193, 2004.

[21] Mowshowitz A. and Kawaguchi A., "Assessing Bias in Search Engines," *Information Processing Management*, vol. 38, no. 1, pp. 141-156, 2002.

[22] Owens C., "A study of the relative bias of web search engines toward news media providers," *Master Thesis*, University of Glasgow, 2009.

[23] Page L., Brin S., Motwani R. and Winograd T., "The Page Rank Citation Ranking: Bringing Order to the Web," *Technical Report*, Stanford Digital Libraries, 1998.

[24] Robertson S. and Walker S., "Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval,: available at: http://www.informatik.uni-trier.de/~ley/db/ conf/ sigir/ RobertsonW94. html, last visited 1994.

[25] Singhal A., Buckley C., and Mitra M., "Pivoted Document Length Normalization," available at: http://nlp.stanford.edu/IR-book/html/htmledition/ pivoted-normalized-document-length-1.html, last visited 1996.

[26] Vaughan L. and Thelwall M., "Search Engine Coverage Bias: Evidence and Possible Causes," *in Information Processing and Management Journal*, vol. 40, no. 4, pp. 693-707, 2004.

[27] Zhai C. and Lafferty J., "A Study of Smoothing Methods for Language Models Applied to Information Retrieval," *ACM Transaction on Information Systems*, vol. 22, no. 2, pp. 179-214, 2004.

Sanam Noor is currently pursuing her PhD in computer science from Department of Computer Science, University of Peshawar. She is also working as a Lecturer of computer science at FG Degree College for Women, Peshawar. She has MS (CS) from IBMS/CS Agricultural University. Her research interests include information retrieval, computer networks and data mining.

Shariq Bashir is PostDoc fellow at center of science and engineering, New York University, Abu Dhabi. He is also an assistant professor of Computer Science at (NUCES-FAST), Islamabad. He received his PhD in computer science from Vienna University of Technology, Austria. His research interests include information retrieval, unsupervised IR systems evaluation, retrievability measurement, query expansion, learning to rank and opinion-based entity ranking. He has published more than 30 research papers in leading international conferences and journals of information retrieval and data mining.