Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Query Expansion for Effective Retrieval Results of Hindi–English Cross-Lingual IR

Ganesh Chandra & Sanjay K. Dwivedi

To cite this article: Ganesh Chandra & Sanjay K. Dwivedi (2019) Query Expansion for Effective Retrieval Results of Hindi–English Cross-Lingual IR, Applied Artificial Intelligence, 33:7, 567-593, DOI: 10.1080/08839514.2019.1577018

To link to this article: https://doi.org/10.1080/08839514.2019.1577018

APPLIED ARTIFICIAL INTELLIGENCE
2019, VOL. 33, NO. 7, 567–593
https://doi.org/10.1080/08839514.2019.1577018

Taylor & Francis
Taylor & Francis Group

Query Expansion for Effective Retrieval Results of Hindi–English Cross-Lingual IR

Ganesh Chandra and Sanjay K. Dwivedi

Department of Computer Science, BBA (A Central) University, Lucknow, India

ABSTRACT

Information retrieval (IR) is the science of identifying documents or sub-documents from a collection of information or database. The collection of information does not necessarily be available in only one language as information does not depend on languages. Monolingual IR is the process of retrieving information in query language whereas cross-lingual information retrieval (CLIR) is the process of retrieving information in a language that differs from query language. In current scenario, there is a strong demand of CLIR system because it allows the user to expand the international scope of searching a relevant document. As compared to monolingual IR, one of the biggest problems of CLIR is poor retrieval performance that occurs due to query mismatching, multiple representations of query terms and untranslated query terms. Query expansion (QE) is the process or technique of adding related terms to the original query for query reformulation. Purpose of QE is to improve the performance and quality of retrieved information in CLIR system. In this paper, QE has been explored for a Hindi–English CLIR in which Hindi queries are used to search English documents. We used Okapi BM25 for documents ranking, and then by using term selection value, translated queries have been expanded. All experiments have been performed using FIRE 2012 dataset. Our result shows that the relevancy of Hindi–English CLIR can be improved by adding the lowest frequency term.

Introduction

Information access refers to the process of making information accessible and usable to user, which is available in various documents. Documents may have different formats, various sources and different languages. Traditional information retrieval (IR) systems are implemented mainly for monolingual documents. However, with the rapid development of Internet, the demand for searching information from multilingual documents is increasing, which results in the great challenge of how to match the user's query written in one language with the documents written in other languages.

CONTACT Ganesh Chandra ganesh.iiscgate@gmail.com Department of Computer Science, Babasaheb Bhimrao Ambedkar (A Central) University, Vidya Vihar, Lucknow 226025, India

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

Consequently, more sophisticated techniques are necessary to enhance the performance of retrieval system. Cross-lingual information retrieval (CLIR) (Gaillard et al. 2010) provides a convenient way that can solve the problems of language boundaries, where users can submit queries written in their own language and retrieve documents in another language (Pigur 1979).

In CLIR (Banchs and Costa-Jussà 2013), retrieval of information may be achieved by three types of translation: query translation, document translation, and both query and document translation (Sanchez-Martinez and Carrasco 2011). On the basis of resources, translation in CLIR can also be classified into three classes (Aljlayl and Frieder 2001): machine-readable dictionary (MRD)-based translation, machine translation (MT) and corpora (parallel or comparable corpora)-based translation.

Dictionary-based translation (Levow, Oard, and Resnik 2005) is a traditional approach of CLIR in which problems occur when queries contain words or phrases that appear in dictionary. Dictionary-based approach exploits MRDs and selection strategies such as random selection (Kwok 1997), select best and select all (Davis 1996).

The purpose of MT is to translate queries of one language into another language using a context. Many factors creates problem in MT of CLIR such as words with multiple meanings (polysemy) and sentences with multiple grammatical structures and grammar problems.

Corpus-based approach uses multilingual terms for query translation in CLIR. This approach can be classified into two types: parallel corpora-based and comparable corpora-based approach (Landauer and Littman 1990; Sheridan and Ballerini 1996). A parallel corpus contains a pair or set of documents that are identical but in different languages (i.e. original text and their translation). A parallel corpus is an expensive method that allows texts to be aligned and used in various applications such as computer-aided translator training and MT system. The comparable corpora are made up of similar documents in different languages, i.e. the pair documents are conceptually similar. A comparable corpus can be obtained from downloading electronic copies of newspapers and article, on the WWW for any specified domain.

With the development of social websites, every web user not only plays a single role of web information consumer but also an information creator. So CLIR becomes critical for web communication. Due to globalization, web users are more aware of the things like education, research and business, etc., and are interested to collect information from various languages of the world. Every user wants to retrieve information or documents in his/her native language to understand the retrieved documents more easily. Accessing information in user languages increases the demand for CLIR (Varshney and Bajpai 2013).

India is a multilingual country where languages or scripts are changed after few kilometers. Hindi is an official language of India and there is a need to provide local language support in web applications because a large amount of data or information of various domains, such as e-commerce, education, etc., require English language knowledge (Joshi, Bhatt, and Patel 2013). Internet environment increases the demand for Hindi–English CLIR (Ponte and Croft 1998). Query expansion (QE) is an effective technique to improve the performance of Hindi–English CLIR. QE adds related terms to query, overcomes word mismatch problem and improves the retrieval performance of CLIR.

The rest of the article is structured as follows: In the next section "Related Work" of CLIR is reported. In "Query Expansion" and "QE in CLIR" the importance of query expansion in searching of information in CLIR is described. In "Experimental Setup" the query translation, ranking of documents and term selection value are described. Then, in "Experimental Results" the different results are described. In "Discussion" the various results and their comparison are described. Finally, in "Conclusion" the most relevant conclusions derived from the experimental results are presented.

Related Work

The research on IR came into existence since the early 1970s whereas experiments for retrieving information across languages were first initiated by Salton in 1973 (Salton 1973). In 1986, Lesk analyzed lexical disambiguation using word overlap (Lesk 1986). However, most of the modern research on CLIR started in 1990s, and nowadays it has become one of the most important research topics in the area of IR. An overview of CLIR is given in (Ballesteros and Croft 1997; Oard 1998).

In 1998, Lisa et al. (Ballesteros and Croft 1998) developed an approach for resolving the ambiguity of query and phrasal translation using statistics co-occurrence analysis. In 2002, Kyung-Soon et al. (Lee, Kageura, and Choi 2002) developed a method to resolve the ambiguity of Korean-English CLIR using a clustering approach. In 2003, a surprise language exercise (Oard 2003) was conducted at ACM TALIP for the development of English to Hindi and Cebuano CLIR system. In this system, English language queries were used to retrieve documents of Hindi and Cebuano language.

In CLEF 2006, ad-hoc document retrieval task was reported by IIIT Hyderabad which involves Hindi and Telugu to English IR. In 2007, Seetha, Das and Kumar have performed some experiments for the evaluation of English–Hindi CLIR using dictionary-based query translation and the results are reported in (Seetha, Das, and Kumar 2007). Das et al. (2010) worked on the effect of QE in English–Hindi CLIR system using WordNet. This system uses Shabdanjali dictionary for query translation and expanded Hindi queries using Hindi WordNet. In 2011, S.M. Chaware et al. developed an approach to build an ontology for CLIR (Chaware and Rao 2011).

Table 1. Some prominent researches of Hindi–English CLIR.

Author	Title	Year
Shukla and Sinha (2015)	Categorizing sentence structures for phrase level morphological analyzer for English to Hindi RBMT	2015
Varshney and Bajpai (2013)	Improving retrieval performance of English-Hindi based cross-language information retrieval.	2013
Dwivedi (2012)	HSC based method for disambiguation of web queries in Hindi language	2012
Contractor et al. (2010)	Handling noisy queries in cross language FAQ retrieval.	2010
Gaillard et al. (2010)	Query expansion for cross language information retrieval improvement	2010
Seetha, Das, and Kumar (2009)	Improving performance of English-Hindi CLIR system using linguistic tools and techniques	2009
Mandal et al. (2008)	Bengali and Hindi to English CLIR Evaluation	2008
Chinnakotla et al. (2008)	Hindi to English and Marathi to English cross language information retrieval evaluation	2008
Bandyopadhyay et al. (2007)	Bengali, Hindi and Telugu to English ad-hoc bilingual task at CLEF 2007	2008
Pingali, Tune, and Varma (2008)	Improving recall for Hindi, Telugu, Oromo to English CLIR	2008
Mandal et al. (2007)	Hindi to English cross-language text retrieval under limited resources.	2007
Seetha, Das, and Kumar (2007)	Evaluation of the English-Hindi cross language information retrieval system based on dictionary based query translation	2007
Sekine and Grishman (2003)	Hindi-English cross-lingual question-answering system	2003

CLIR is a demanding research area in India because people are using different different languages for communication. The first major work on Hindi was done in TIDES surprise language exercise. Its purpose was to retrieve Hindi documents, provided by Linguistic Data Consortium (LDC), in response to English query. Recently, Indian government has initiated a project on "Development of Cross-Lingual Information Access System." In this project, user can apply a query and retrieve documents in any of the six Indian languages such as Bengali, Hindi, Telugu, Tamil, Marathi and Punjabi. Some of the prominent researches on QE in Hindi–English CLIR are described in Table 1.

Query Expansion

QE (Chandra and Dwivedi 2017; Daoud and Huang 2013) is the process of adding a new term to the original query to improve retrieval performance of IR, CLIR and MLIR. The purpose of QE is to improve the quantity, quality and relevancy of results retrieved by CLIR systems (Billerbeck 2005; Xu and Croft 2000). QE (Maxwell and Schafer, 2010) enhances the concept of query by adding related terms from top retrieved documents that are retrieved by original query.

Sometimes the queries which are entered by users are small or ambiguous. The word(s) of query are ambiguous if they have more than one meaning or senses. In that condition systems are not able to understand what the user wants to search. For example, if user enters a query "apple rate," then it may cause ambiguity (Chandra and Dwivedi 2014) for search engine because it is not clear either user wants to search rate of apple which is fruit or he wants to search the rate of Apple electronic product. Some important problems occurring in CLIR are word mismatch, vocabulary mismatch, query size, disambiguation and identifying relevant results.

One of the major problems of QE is "query drift." The term "query drift" refers to an alteration from the original content of the user. It exists in a system when the original meaning of the initial query is changed by QE. Let us consider the initial query "mughal" in terms of Akbar. A bad application of QE could transform the initial query into "mughal sarai," "mughal garden," "mughale azam song" and "mughal architecture," etc. Here the meaning of new queries and original queries has no relation. In CLIR, QE is performed by different modes: derivative term (semantically related terms to the initial query), synonyms, inflection (gender, number, tense, etc.), hyperonyms (sense of usual linguistic definition) and geographical expansion (Gaillard et al. 2010). The following three approaches are used in QE:

(i) Manual QE: In manual QE, user has freedom to choose the expansion terms.
(ii) Interactive QE: Interactive approach is based on feedback process where system suggests the user for QE.
(iii) Automatic QE: The automatic QE is performed without user intervention, so the whole process of QE is invisible to the user.

QE involves selection and searching of synonyms of words, finding of all various morphological forms of words and fixing spelling errors. The basic issues that deal with QE are as follows: source of term selection, methods of term selection, query term weightage, features generation and ranking of query terms, identification of relevant result, efficiency of QE, usability and parameter setting (Imran and Sharan 2009; Jothilakshmi, Shanthi, and Babisaraswathi 2013). QE is used to increase the performance of retrieved documents. Some of the application areas of QE are:

(i) Question answering system (QAS): The purpose of QAS is to provide a brief answer instead of full documents for certain natural language questions (Agichtein, Lawrence, and Gravano 2004). The problem of mismatch between question and answer vocabularies also occurs in QAS. Effectiveness of retrieval results in QAS can be increased by expanding the original question with terms that are expected to appear in documents containing answers (Riezler et al. 2007).
(ii) Information filtering (IF): IF is used to monitor and select documents that are relevant to user. Some common examples of IF are

e-commerce, electronic news, blogs and e-mail (Hanani, Shapira, and Shoval 2001). QE helps in selecting the best information resources (Zimmer, Tryfonopoulos, and Weikum 2008).

(iii) Multimedia IR System (MIRS): Nowadays, searching of multimedia documents, such as speech, image, video, etc., is an important research area. QE can be used in image retrieval, speech recognition, text retrieval and video retrieval by expanding the query terms for better results (Singhal and Pereira 1999).

QE in CLIR

In order to retrieve accurate information, the query which is given by user plays a very important role in CLIR. One of the most common factors behind the poor performance of CLIR as reported in the "Related Work" is lack of availability of resources and multiple representations of query words. To overcome the second issue, QE (Ballesteros and Croft 1997) is used to enhance the translated query with related terms extracted from the document collection. The basic approach of QE follows two steps: the identification of a presumed set of relevant documents; then, the selection of related terms used for query enrichment. This process of adding related terms to the translated query helps to improve the precision and quality of relevant documents.

Experimental Setup

In this work, QE has been performed for Hindi–English CLIR. Queries in Hindi (source language) are used to retrieve the relevant documents of English language (target language). In our previous work (Chandra and Dwivedi 2017), QE was performed for smaller setup with small number of queries. Extending the research we have now taken 50 queries of Forum for Information Retrieval Evaluation (FIRE) 2012 dataset. Google translator has been used for query translations from Hindi to English, and for all intermediate searches Google search engine has been utilized (help of linguistics has also been taken for certain queries for which Google translator has not given an accurate result). Table 2 shows the original 50 queries and their translation.

UAM Corpus Tool (http://www.wagsoft.com/CorpusTool/) developed at Autonomous University of Madrid by the computational linguist Mick O` Donnell and used to compute the frequencies of terms occurring in retrieved documents, and also the length of each retrieved documents has been computed using this tool as shown in Table 3. The Okapi BM25 (Billerbeck et al. 2003; Robertson et al. 1995; Sari and Adriani

Table 2. Queries and their translation.

Query	Hindi Query	English Query
1	वाई एस आर रेड्डी की मौत	YSR Reddy's death
2	संगीतकारों को भारत रत्न	Bharat Ratna musicians
3	नरेगा योजना	NREGA scheme
4	ऑस्ट्रेलियाई दूतावास बम वसि्फोट	Australian embassy bombing
5	यूरो अपनाने वाले देश	Countries adopting Euro
6	पहले 700 टेस्ट वकिेट लेने वाले क्रकिेटर	The first player to take 700 Test wickets
7	स्टीव इरवनि की मृत्यु	Steve Irwin's death
8	2008 गुवाहाटी बम वसि्फोट से क्षतति	Guwahati bombing damage in 2008
9	चामुंडा मंदरि भगदड़	Chamunda temple stampede
10	आदर्श हाउसगि सोसाइटी घोटाले इस्तीफा	Adarsh Housing Society scam resignation
11	ऑस्ट्रेलयिा में भारतीय छात्रों पर हमले	The attacks on Indian students in Australia
12	दलि्ली मेट्रो सेवा की शुरुआत	Beginning of Delhi Metro services
13	भारतीय नागरकि पाकसि्तानी जासूस	Indian Citizen Pakistani spy
14	शकि्षा अधकिार अधनियिम	Right to Education Act
15	बीजेपी से जसवंत सहि का बहषि्कार	Jaswant Singh Boycott from BJP
16	गोरखालैंड की मांग	Gorkhaland demand
17	श्रीलंकाई राष्ट्रीय क्रकिेट टीम पर हमला	Attack on Sri Lankan national cricket team
18	भारत की पहली महलिा स्पीकर	India's first woman Speaker
19	2001 साहति्य में नोबेल पुरस्कार वजिेता	2001 Nobel Prize Winner in Literature
20	2003 आशयिान कप वजिेता	2003 ASEAN Cup Winner
21	2001 भारतीय जनगणना	2001 Indian census
22	भुज भूकंप	Bhuj earthquake
23	धोनी कप्तान भारतीय टीम	Dhoni captain Indian team
24	पैगम्बर मोहम्मद कार्टून ववािद	Prophet Mohammad cartoon controversy
25	2002 नेटवेस्ट शृंखला का परणिाम	2002 NatWest Series results
26	इराक का प्रथम चुनाव	Iraq's First Election
27	प्रतषि्ठति व्यक्तयिों पर जूता फेकना	Dignitaries on the Shoe Throwing
28	भारत का पहला मानवरहति चन्द्रमा मशिन	India's First Unmanned Moon Mission
29	भारतीय संसद आतंकवादी हमला	Indian Parliament Attack
30	पोलयिो उन्मूलन अभयिान	Polio Eradication Campaign
31	अभयिुक्त अजमल कसाब	Accused Ajmal Kasab
32	सानयिा मर्जिा की शादी	Sania Mirza's Marriage
33	महेंद्र सहि धोनी राष्ट्रीय पुरस्कार	Mahendra Singh Dhoni National Award
34	वाममोर्चा ने कांग्रेस से समर्थन वापस लयिा	Left withdrew Support to the Congress
35	मगि दुर्घटना पश्चमि बंगाल	MIG Crash in West Bengal
36	वशि्व अहसिा दविस	World Non-Violence Day
37	फलि्म सेंसर बोर्ड महलिा अध्यक्ष	Film Censor Board Chairperson Woman
38	2010 ऑटो एक्सपो दलि्ली	Delhi Auto Expo 2010
39	हरभजन सहि ने श्रीसांत को थप्पड़ मारा	Harbhajan Singh Slapped Srisant
40	भारतीय एनीमेशन फलि्म उद्योग	Indian Animation Film Industry
41	ग्रामीण बैंक मुहम्मद यूनुस ववािद	Grameen Bank Muhammad Yunus Dispute
42	द वन्चिी कोड भारत रलीज़ ववािद	Da Vinci Code India Release Controversy
43	सरवाइकल कैंसर जागरूकता उपचार टीका	Cervical Cancer Awareness, Treatment Vaccine
44	पहला फार्मूला १ सीक्रटि भारत	India's first Formula 1 Circuit
45	स्टीव वॉ अंतर्राष्ट्रीय क्रकिेट सन्यास	Steve Waugh International Cricket Retirement
46	बलि और मेलडिा गेट्स फाउंडेशन परोपकारी क्रयिाकलाप भारत	Bill and Melinda Gates Foundation, the Philanthropic Activities in India
47	ग्रीस यूरो कप २००४ वजिय	Greece Won the Euro Cup 2004
48	इमरान खान कैंसर अस्पताल पाकसि्तान	Imran Khan's Cancer Hospital in Pakistan
49	आईफोन आईपैड डजिाइन लोकप्रयिता लॉन्च	IPhone iPad Design Popularity Launch
50	सैटेनकि वर्सेज ववािद	Satanic Verses Controversy

Table 3. Length of each document for 50 queries.

	Length of Documents										
Query	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10	Total
1	1083	1082	300	480	667	465	267	391	238	225	5198
2	3133	134	1156	651	854	532	758	384	2365	1696	11663
3	4650	361	1186	1114	1296	780	437	430	120	437	10811
4	1303	565	437	585	866	461	698	539	725	432	6611
5	737	370	6182	290	139	1443	330	432	269	259	10451
6	5872	10238	465	552	420	1339	370	498	602	1025	21381
7	4119	594	349	1138	1247	1019	574	288	478	751	10557
8	3049	639	746	135	726	564	89	241	10269	221	16679
9	360	569	1022	461	196	289	262	186	190	197	3732
10	2365	1660	352	131	728	1159	352	348	1004	448	8547
11	4031	338	661	2386	715	792	334	924	1257	774	12212
12	198	196	967	336	335	304	217	514	405	365	3837
13	420	493	333	929	970	1165	817	829	1603	1045	8604
14	2460	580	821	625	501	253	445	1005	587	1112	8389
15	743	888	427	314	384	295	398	201	691	247	4588
16	1855	1351	660	643	1414	598	3708	219	1638	1679	13765
17	3535	304	850	752	967	366	1173	321	343	379	8990
18	286	4769	271	962	457	389	246	272	324	395	8371
19	44	553	50	877	4471	3738	2800	287	15669	3465	31954
20	754	701	321	726	1769	1209	1595	480	416	153	8124
21	741	2987	247	741	276	235	864	493	491	88	7163
22	414	923	189	267	192	2176	1393	257	279	243	6333
23	545	2192	1057	695	4742	916	247	387	416	486	11683
24	9573	521	680	815	976	1413	631	251	1685	3931	20476
25	373	202	1120	1482	174	867	282	160	495	648	5803
26	980	2626	51	838	1228	956	744	483	1290	798	9994
27	640	362	4739	327	77	202	580	626	669	884	9106
28	4926	300	493	1229	2150	1002	365	669	418	877	12429
29	1288	4512	662	392	1547	316	743	464	838	127	10889
30	1160	78	144	93	5597	786	2614	3408	227	600	14707
31	1441	802	359	314	352	3683	1383	493	313	313	9453
32	423	6542	656	498	360	259	573	585	757	157	10810
33	6556	274	161	2884	671	1838	578	245	247	258	13712
34	2052	985	1786	996	757	595	297	1392	265	463	9588
35	190	126	115	102	116	189	90	124	51	115	1218
36	256	354	209	457	287	435	1514	276	400	891	5079
37	4369	72	733	331	334	688	268	1257	438	259	8749
38	1736	398	281	851	796	88	1723	741	590	323	7527
39	256	291	280	677	511	2324	12489	334	565	524	18251
40	1729	542	393	1722	1048	2045	618	1608	1069	1222	11996
41	918	1576	5069	669	433	2521	309	1132	609	533	13769
42	5590	4531	499	228	4589	8313	465	199	558	1732	26704
43	368	130	402	199	522	1022	1516	1591	240	3422	9412
44	970	931	229	1164	175	1728	431	761	662	549	7600
45	833	9651	413	975	740	584	581	534	692	790	15793
46	5763	522	394	1091	714	2111	487	2346	641	120	14189
47	4656	7774	432	1205	565	889	682	846	393	917	18359
48	896	4925	181	277	257	824	160	255	1220	426	9421
49	340	1427	1779	3660	2538	10355	437	917	545	918	22916
50	2190	539	3134	1628	7139	1144	833	819	754	2514	20694

2014) measure is an effective method for QE, and computation is described in Equation (1).

$$\mathrm{bm25(q,d)} = \sum_{t\in q} \log\left(\frac{N - f_t + 0.5}{f_t + 0.5}\right) \times \frac{(K_1+1) f_{d,t}}{k + f_{d,t}} \tag{1}$$

where q is a query containing terms t; d is a document; N is the number of documents in the collection; f_t is the number of documents containing term t and "$f_{d,t}$" is the number of occurrences of 't' in 'd'; and the computation of 'k' is described in Equation 2:

$$k = k_1((1-b) + b \times L_d/A_L) \tag{2}$$

where constants k_1 and b are set to 1.2 and 0.75, respectively; L_d and A_L are document length and average document length, respectively.

For each translated query (Table 2), document ranking for each retrieved document of each query (@ 10) has been obtained using Okapi BM25, as given in Table 4. For example, in case of Query No.1 (Table 2), i.e. "*YSR Reddy's Death,*" the length of first retrieved document, i.e. **length (L_d)** = 1083 obtained by using of UAM Corpus tool as described in Table 3. Now, the value of variable $k$can be computed as follows:

$$k = 1.2((1-0.75) + 0.75 \times 1083/519.8)) = 2.175$$

The Okapi BM25 value for the first document can be computed as follows:

$$\mathrm{bm25}(q,d) = \log\left(\frac{(10-9+0.5)}{(9+0.5)}\right) \times \frac{(1.2+1)*6}{2.172+6} + \log\left(\frac{(10-8+0.5)}{(8+0.5)}\right) \times \frac{(1.2+1)*13}{2.172+13} + \log\left(\frac{(10-7+0.5)}{(7+0.5)}\right) \times \frac{(1.2+1)*3}{2.172+3}$$

$$\mathrm{bm25}(q,d) = -2.71$$

Similarly, the Okapi BM25 value for other documents of this query and the rest of the documents of other queries can be computed. Table 4 shows the Okapi BM25 value and rank of each document for all 50 queries.

The expansion term(s) are to be obtained from narration and description of each query as available in FIRE and to obtain the most suitable expansion term for a query, the retrieved documents for that query serves as pool based on the computation of TSV (Rijsbergen 1979) which is computed in Equation 3:

$$\mathrm{TSV}_t = \left(\frac{f_t}{N}\right)^{r_t} \left(\frac{R}{r_t}\right) \tag{3}$$

Table 4. Okapi BM25 value and rank of each document.

Query	Okapi BM25 Value & Rank									
	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
1	−2.71	−2.708	−3.276	−2.564	−2.712	−2.169	−1.534	−2.268	−1.552	−2.39
	Rank9	Rank8	Rank10	Rank6	Rank7	Rank3	Rank1	Rank4	Rank2	Rank5
2	0.637	0.582	0.564	0.499	Query Terms	Query Terms	Query Terms	Query Terms	Query Terms	Query Terms
	Rank1	Rank2	Rank3	Rank4	Absent	Absent	Absent	Absent	Absent	Absent
3	−1.966	−1.718	−1.006	Query Terms	−1.939	−1.869	−1.965	−1.990	−1.834	−1.987
	Rank7	Rank3	Rank2	Absent	Rank5	Rank4	Rank6	Rank9	Rank1	Rank8
4	−7.701	−6.887	−7.14	−6.559	−7.041	−9.543	−6.025	−6.410	−6.546	−6.452
	Rank9	Rank6	Rank8	Rank5	Rank7	Rank10	Rank1	Rank2	Rank4	Rank3
5	−4.254	−4.362	−4.306	−4.481	−4.175	−4.179	−3.924	−4.293	−4.409	−2.633
	Rank6	Rank9	Rank8	Rank10	Rank4	Rank5	Rank2	Rank7	Rank3	Rank1
6	−3.331	−121.35	−2.826	−8.89	−2.654	−0.433	0.55	−2.617	−2.849	−2.666
	Rank8	Rank10	Rank6	Rank9	Rank4	Rank2	Rank1	Rank3	Rank7	Rank5
7	−2.93	−3.844	−4.053	−4.141	−3.055	−4.075	−2.665	−2.3	−4.05	−3.797
	Rank3	Rank6	Rank8	Rank10	Rank4	Rank9	Rank2	Rank1	Rank7	Rank5
8	−1.016	−1.566	−1.625	−1.006	−1.002	−1.542	−0.72	Query Term	−1.402	−0.728
	Rank5	Rank8	Rank9	Rank4	Rank3	Rank7	Rank1	Absent	Rank6	Rank2
9	−7.016	−6.941	−7.174	−6.047	−6.736	−6.426	−6.792	−5.777	−5.842	−7.196
	Rank8	Rank7	Rank9	Rank3	Rank5	Rank4	Rank6	Rank1	Rank2	Rank10
10	−9.091	−9.839	−9.18	−10.44	−11.215	−10.675	−10.633	−9.058	−9.249	−9.783
	Rank2	Rank6	Rank3	Rank7	Rank10	Rank9	Rank8	Rank1	Rank4	Rank5
11	−8.693	−9.541	−9.548	−8.652	−10.176	−8.248	−9.988	−10.223	−9.475	−10.14
	Rank3	Rank5	Rank6	Rank2	Rank9	Rank1	Rank7	Rank10	Rank4	Rank8
12	−3.86	−3.602	−4.266	−3.759	−4.484	−4.7	−4.725	−2.836	−4.203	−4.091
	Rank4	Rank2	Rank7	Rank3	Rank8	Rank9	Rank10	Rank1	Rank6	Rank5
13	−4.003	−2.958	−3.604	−4.468	−4.388	−4.1	−3.435	−2.762	−3.507	−1.090
	Rank7	Rank3	Rank6	Rank10	Rank9	Rank8	Rank4	Rank2	Rank5	Rank1
14	−4.643	−4.546	−6.449	−4.555	−4.629	−4.494	−3.494	−3.244	−2.709	−3.783
	Rank9	Rank6	Rank10	Rank7	Rank8	Rank5	Rank3	Rank2	Rank1	Rank4
15	−2.452	−3.582	−3.341	−4.879	−4.176	−5.855	−5.347	−5.334	−7.396	−4.313
	Rank1	Rank3	Rank2	Rank6	Rank4	Rank9	Rank8	Rank7	Rank10	Rank5

(*Continued*)

Table 4. (Continued).

Query	Okapi BM25 Value & Rank									
	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
16	−3.912 Rank10	−3.832 Rank9	−3.546 Rank3	−3.808 Rank8	−2.812 Rank2	−3.623 Rank4	−3.775 Rank6	−3.751 Rank5	−3.782 Rank7	−2.674 Rank1
17	−3.039 Rank6	−1.84 Rank3	−2.593 Rank4	−3.627 Rank9	−3.559 Rank8	−0.621 Rank1	1.141 Rank2	−3.153 Rank7	−3.786 Rank10	−2.526 Rank4
18	−3.205 Rank3	−5.076 Rank9	−4.963 Rank8	−3.362 Rank4	−4.03 Rank5	−5.901 Rank10	−1.746 Rank1	−4.734 Rank7	−4.251 Rank6	−3.092 Rank2
19	−3.901 Rank8	−3.609 Rank6	−2.231 Rank2	−2.829 Rank3	−3.95 Rank9	−4.026 Rank10	−3.06 Rank5	−3.813 Rank7	−1.761 Rank1	−2.869 Rank4
20	Query Term Absent	0.221 Rank1	0.642 Rank6	0.628 Rank5	0.292 Rank3	0.288 Rank2	Query Term Absent	Query Term Absent	Query Term Absent	0.328 Rank4
21	−3.657 Rank6	−3.728 Rank9	−2.685 Rank1	−3.659 Rank7	−9.038 Rank10	−3.718 Rank8	−3.346 Rank3	−3.032 Rank2	−3.527 Rank5	−3.449 Rank4
22	−3.208 Rank4	−3.323 Rank9	−0.567 Rank2	−3.233 Rank5	−3.272 Rank7	−3.275 Rank8	−0.619 Rank3	−3.413 Rank10	−3.246 Rank6	0.442 Rank1
23	Query Term Absent	−3.801 Rank4	−3.943 Rank5	−3.494 Rank3	−4.23 Rank9	−4.216 Rank8	−1.575 Rank1	−4.17 Rank7	−3.946 Rank6	−3.388 Rank2
24	−2.655 Rank7	−1.58 Rank3	−3.342 Rank10	−2.69 Rank8	−1.014 Rank1	−3.276 Rank9	−2.437 Rank6	−2.223 Rank5	−1.404 Rank2	−2.158 Rank4
25	−1.673 Rank7	Query Term Absent	−0.582 Rank1	−1.556 Rank5	−0.709 Rank2	−1.594 Rank6	−3.573 Rank9	−1.466 Rank4	−1.562 Rank8	−1.383 Rank3
26	−4.227 Rank3	−6.245 Rank8	−2.105 Rank1	−5–456 Rank7	−4.657 Rank5	−7.012 Rank10	−3.892 Rank2	−4.284 Rank4	−6.567 Rank9	−5.326 Rank6
27	−3.801 Rank4	−2.125 Rank1	−3.764 Rank3	−2.786 Rank2	−4.023 Rank5	−4.123 Rank6	−4.675 Rank10	−4.512 Rank8	−4.504 Rank7	−4.543 Rank9
28	−8.643 Rank9	−3.216 Rank7	−2.928 Rank4	−1.998 Rank1	−9.912 Rank10	−1.869 Rank3	−1.736 Rank2	−2.961 Rank5	−4.189 Rank8	−3.045 Rank6
29	−7.017 Rank4	0.546 Rank8	−0.461 Rank1	−1.698 Rank2	−3.505 Rank9	−3.597 Rank7	−4.019 Rank10	−0.631 Rank3	−6.412 Rank5	−1.019 Rank6

(*Continued*)

Table 4. (Continued).

	Okapi BM25 Value & Rank									
Query	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
30	−0.528	−3.144	0.326	−2.531	−3.689	−3.645	−4.463	−3.849	−3.374	−4.542
	Rank2	Rank4	Rank1	Rank3	Rank6	Rank5	Rank9	Rank8	Rank7	Rank10
31	−2.989	−2.961	−2.862	−1.116	−3.069	−2.896	−7.017	−2.985	−4.987	−7.601
	Rank6	Rank4	Rank2	Rank1	Rank7	Rank3	Rank9	Rank5	Rank8	Rank10
32	−6.229	−3.942	−2.172	−6.017	−1.613	−1.543	−2.912	−2.938	−4.254	−1.119
	Rank10	Rank7	Rank4	Rank9	Rank3	Rank2	Rank5	Rank6	Rank8	Rank1
33	−3.956	−3.569	0.582	−2.272	−9.011	−3.963	−6.491	−3.941	−3.922	−4.369
	Rank6	Rank3	Rank1	Rank2	Rank10	Rank7	Rank9	Rank5	Rank4	Rank8
34	−7.091	−4.625	−3.952	−3.632	−5.169	−3.985	−3.208	−4.253	−7.701	−7.261
	Rank8	Rank6	Rank3	Rank2	Rank7	Rank4	Rank1	Rank5	Rank10	Rank9
35	−1.365	0.581	−1.834	−1.345	0.326	−3.681	−2.829	0.582	−3.912	0.512
	Rank6	Rank3	Rank7	Rank5	Rank1	Rank9	Rank8	Rank4	Rank10	Rank2
36	−2.243	−3.106	−0.782	−2.165	−3.201	−3.205	−7.016	−2.633	−6.047	−6.449
	Rank3	Rank5	Rank1	Rank2	Rank6	Rank7	Rank10	Rank4	Rank8	Rank9
37	−4.256	−2.293	−0.648	−2.4	−4.526	Query Term	−1.354	−2.668	−6.551	−4.643
	Rank6	Rank3	Rank1	Rank4	Rank7	Absent	Rank2	Rank5	Rank9	Rank8
38	−3.053	−3.029	−0.728	−3.601	−8.124	−2.282	−5.484	−2.589	0.282	−8.102
	Rank6	Rank5	Rank2	Rank7	Rank10	Rank3	Rank8	Rank4	Rank1	Rank9
39	−4.643	−4.811	−4.963	−3.526	−5.415	−3.642	−1.490	−8.645	−5.717	−5.426
	Rank4	Rank5	Rank6	Rank2	Rank7	Rank3	Rank1	Rank10	Rank8	Rank9
40	−3.142	0.259	−5.179	1.421	−1.089	0.285	−9.095	−3.509	−1.548	−10.293
	Rank6	Rank2	Rank8	Rank1	Rank4	Rank3	Rank9	Rank7	Rank5	Rank10
41	0.493	0.486	−3.175	−1.158	−2.196	−8.562	−10.216	−3.015	−3.319	−5.194
	Rank1	Rank2	Rank6	Rank3	Rank4	Rank9	Rank10	Rank5	Rank7	Rank8
42	−3.418	0.296	−2.716	−0.172	−5.629	−3.164	−2.267	−7.516	−5.986	−12.352
	Rank6	Rank1	Rank4	Rank2	Rank7	Rank5	Rank3	Rank9	Rank8	Rank10
43	−5.428	−1.225	−2.618	Query Term	−2.367	−0.664	−3.618	−3.196	−7.196	−6.238
	Rank7	Rank2	Rank4	Absent	Rank3	Rank1	Rank6	Rank5	Rank9	Rank8
44	−4.463	−7.168	0.872	−1.187	0.298	−5.268	−8.162	−4.268	−7.248	−9.526
	Rank4	Rank7	Rank1	Rank3	Rank2	Rank6	Rank9	Rank5	Rank8	Rank10

(*Continued*)

Table 4. (Continued).

Query	Okapi BM25 Value & Rank									
	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10
45	−5.482	4.358	−5.165	−4.586	−2.684	−8.108	−5.824	−5.928	−7.128	−8.152
	Rank5	Rank1	Rank4	Rank3	Rank2	Rank9	Rank6	Rank7	Rank8	Rank10
46	−3.521	−2.616	−5.328	−1.982	−8.156	−8.164	−3.912	−8.562	−7.649	−8.268
	Rank3	Rank2	Rank5	Rank1	Rank7	Rank8	Rank4	Rank10	Rank6	Rank9
47	−1.128	−3.674	−2.196	−1.778	−5.565	−2.158	−4.626	−4.656	−10.179	−5.876
	Rank1	Rank5	Rank4	Rank2	Rank8	Rank3	Rank6	Rank7	Rank10	Rank9
48	−7.682	−2.695	Query Term	−4.921	−4.295	−7.989	−8.954	−3.532	−3.678	−9.002
	Rank6	Rank1	Absent	Rank5	Rank4	Rank7	Rank8	Rank2	Rank3	Rank9
49	−3.268	−1.198	−1.522	−2.816	−4.165	−12.165	−6.558	−3.527	−8.865	−8.267
	Rank4	Rank2	Rank1	Rank3	Ran6	Rank10	Rank7	Rank5	Rank9	Rank8
50	−4.179	−1.576	−4.476	−5.765	−2.169	−9.789	−10.268	−8.666	−6.854	−8.681
	Rank3	Rank1	Rank4	Rank5	Rank2	Rank9	Rank10	Rank7	Rank8	Rank6

where "R" is the number of top-ranked documents examined, and "r_t" is the number of documents that contain a particular term "t". Terms that have the lowest selection value and are not included in the original query are appended to form a new query. In order to identify the appropriate term for expanding the query using TSV, the three cases have been created and term(s) having minimum TSV has been added to the original query in all three cases .

Case 1: Selection of Keywords from all Documents (@10) of Query (Without Okapi BM25 Ranking)

In this case, our aim is to know the impact on QE before and after the ranking of retrieved documents through Okapi BM25. Therefore, the TSV has been computed without ranking by considering the originally retrieved documents.

For Query No.1, i.e. *"YSR Reddy's Death,"* for computation of TSV, six keywords (except query words) are taken from description and narration of query, and frequency of these keywords (in the document set) are obtained from UAM Corpus tool. TSV value for keyword "Andhra" having $f_t = 24$, $N = R = 10$ and $r_t = 7$ is computed as follows:

$$\mathrm{TSV_t} = \left(\frac{24}{10}\right)^7 \binom{10}{7} = 55037.65$$

TSV value of keyword, i.e. *"Andhra"* is 55037.65. Similarly, TSV value (shown in Table 5) for the rest of the keywords of this query has been obtained.

Two keywords *"Helicopter" & "Crash"* have minimum TSV (Table 5), so the two keywords will be added to the original query. Based on the matching patterns in FIRE narration and description of this query, the finally expanded query would become "*YSR Reddy's Death* ***Helicopter Crash***". Other QE have been performed using the same approach as shown in column 3 of Table 8.

Table 5. Term selection for query *"YSR Reddy's Death."*

Keywords	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7	Doc8	Doc9	Doc10	Total	TSV Value
Andhra	1	6	7	3	2			3	2		24	**55037.65**
Pradesh	5	5	6	3	2			3	1	1	26	**93972.17**
Chief	5	5	3	5	3	2	1	1	1		26	**54295.036**
Minister	4	4	2	6	2	2	1	1	1		23	**18011.52**
Helicopter				4	4	1		1	2		12	**627.056**
Crash				3	2	1		1	4	1	12	**627.056**

Case 2: Selection of Highest Frequency Keyword in Top 3 Ranked Documents

In this case, in order to compute TSV, we considered keyword with the highest frequency in top 3 ranked documents obtained by using Okapi BM25 (Table 3).

For Query No.1 (Table 2), i.e. "*YSR Reddy's Death,*" keywords having highest frequency in top 3 ranked documents (i.e. 6, 7, 9) are described in Table 6. Only three keywords having the highest frequency are considered (Table 6), out of six keywords as available in narration and description of the query (Table 6).

Two keywords "*Chief*" &"*Minister*" have minimum TSV (Table 6), so the two keywords will be added to the original query to form a new query. Based on the matching patterns in FIRE narration and description of this query, the finally expanded query would become "***Chief Minister*** *YSR Reddy's Death.*" Other QE have been performed using the same approach (column 4 of Table 8).

Case 3: Selection of Lowest Frequency Words in Top 3 Ranked Documents

In this case, in order to compute TSV, we considered keyword(s) with the lowest frequency in top 3 ranked documents obtained by using Okapi BM25 (Table 3).

For Query No.1 (Table 2) i.e. "*YSR Reddy's Death*," keywords having the lowest frequency in top 3 ranked documents (i.e. 6, 7, 9) are described in Table 7. Five keywords having the highest frequency are considered (Table 6), out of six keywords as available in narration and description of query as described in Table 7.

Two keywords "*Chief*" & "*Minister*" have minimum TSV, so these will be added to the original query to form a new query. Based on the matching patterns in FIRE narration and description of this query, the finally expanded query

Table 6. Term selection for query *"YSR Reddy's Death"* using the highest frequency words obtained from Okapi.

Keyword	Doc7	Doc9	Doc6	Total	TSV Value
Chief	1		2	3	**0.027**
Minister	1		2	3	**0.027**
Crash		4		4	**0.12**

Table 7. Term selection for query *"YSR Reddy's Death"* using the lowest frequency words obtained from Okapi.

Keyword	Doc7	Doc9	Doc6	Total	TSV Value
Chief	1	1		2	0.008
Minister	1	1		2	0.008
Pradesh		1		1	0.3
Helicopter			1	1	0.03
Crash			1	1	0.03

Table 8. Query expansion for case 1, case 2 and case 3.

Query	Translated English Query	Case 1 (Without Ranking)	Case 2 (Highest Frequency)	Case 3 (Lowest Frequency)
1	YSR Reddy's Death	YSR Reddy's Death ***Helicopter Crash***	***Chief Minister*** YSR Reddy's Death	***Chief Minister*** YSR Reddy's Death
2	Bharat Ratna Musicians	Bharat Ratna ***Awarded*** Musicians	Bharat Ratna Musicians ***(Vocalists)***	Bharat Ratna Musicians ***(Vocalists)***
3	NREGA Scheme	NREGA ***Main*** **S**cheme	NAREGA Scheme ***100 day's***	NAREGA Scheme ***Work***
4	Australian Embassy Bombing	***Front*** Australian Embassy Bombing	Australian Embassy Bombing ***Jakarta***	***Front*** Australian Embassy Bombing
5	Countries Adopting Euro	***Europe*** Countries Adopting Euro	Countries Adopting Euro ***Currency***	***Europe*** Countries Adopting EURO
6	First cricketer to take 700 test wickets	First Cricketer ***bowler*** to take 700 Test Wickets	***Shane Warne*** First cricketer to take 700 test wickets	***Shane Warne*** First cricketer to take 700 test wickets
7	Steve Irwin's death	***Hunter*** Steve Irwin's Death	***Crocodile hunter*** Steve Irwin's Death	***Crocodile hunter*** Steve Irwin's Death
8	Guwahati Bombing Damage in 2008	Guwahati Bombing Damage in ***October*** 2008	Guwahati 2008 Damage in Bomb ***blasting***	Guwahati Bombing Damage in ***October*** 2008
9	Chamunda Temple Stampede	Chamunda Temple Stampede ***Casualities***	***Jodhpur*** Chamunda Temple Stampede	Chamunda ***Devi*** Temple Stampede
10	Adarsh Housing Society scam Resignation	***Ashok Chavan*** Adarsh Housing Society Scam Resignation	***Maharashtra*** Adarsh Housing Society Scam Resignation in	***Chief Minister*** Adarsh Housing Society Scam Resignation
11	Attacks on Indian Students in Australia	The Attacks on Indian Students in Australia	Attacks on Indian Students in Australia	Attacks on Indian Students in Australia
12	Beginning of Delhi Metro Services	Beginning of Delhi Metro ***Rail*** Services	Beginning of Delhi Metro ***Rail*** Services	Beginning of Delhi Metro ***Rail*** Services
13	Indian citizen Pakistani spy	Indian citizen ***Diplomat*** Pakistani spy	***Arrested*** Indian citizen Pakistani spy	***Arrested*** Indian citizen Pakistani spy
14	Right to Education Act	Right to Education Act ***Lok Sabha***	Right to Education Act ***Rajya Sabha***	Right to Education Act ***Rajya Sabha***
15	Jaswant Singh Boycott from BJP	Jaswant Singh Boycott from BJP ***Controversial Book***	Jaswant Singh Boycott from BJP ***Controversial Book***	Jaswant Singh Boycott from BJP ***Controversial Book***
16	Gorkhaland Demand	Gorkhaland Demand ***Several Places***	Gorkhaland Demand ***Chief***	Gorkhaland Demand ***Chief***
17	Attack on Sri Lankan National Cricket Team	***Terrorist*** Attack on Sri Lankan National Cricket Team	Attack on Sri Lankan National Cricket Team ***Pakistan***	Attack on Sri Lankan National Cricket Team ***Pakistan***
18	India's First Female Speaker	***Meira Kumar*** India's First Female Speaker	***Meira Kumar*** India's First Female Speaker	India's First Female Speaker ***Lok Sabha***
19	2001 Nobel Prize Winner in Literature	***Naipaul*** 2001 Nobel Prize Winner in Literature	***Naipaul*** 2001 Nobel Prize Winner in Literature	***Naipaul*** 2001 Nobel Prize Winner in Literature
20	2003 ASEAN Cup Winner	***Team*** 2003 ASEAN Cup Winner	***Club*** 2003 ASEAN Cup Winner	***India*** 2003 ASEAN Cup Winner
21	2001 Indian Census	2001 Indian Census ***Conducted***	2001 Indian Census ***Conducted***	2001 Indian Census ***Conducted***
22	Bhuj Earthquake	Bhuj Earthquake ***Caused***	Bhuj Earthquake ***Gujarat***	***2001*** Bhuj Earthquake

(*Continued*)

Table 8. (Continued).

Query	Translated English Query	Case 1 (Without Ranking)	Case 2 (Highest Frequency)	Case 3 (Lowest Frequency)
23	Dhoni captain Indian team	***Appointment*** Dhoni Captain Indian Team	Dhoni captain Indian ***Cricket*** Team	Dhoni captain Indian ***Cricket*** Team
24	Prophet Mohammad cartoon controversy	***Depicting*** Prophet Mohammad Cartoon Controversy	***Depicting*** Prophet Mohammad cartoon controversy	***Muslims*** Prophet Mohammad cartoon controversy
25	2002 NatWest Series results	2002 NatWest Series Results ***Played***	***Victory*** 2002 NatWest Series results or	2002 NatWest Series results ***Played***
26	Iraq's First Election	Iraq's First Election, ***Conclusion***	Iraq's First Election, ***Peaceful***	Iraq's First ***General*** Election
27	Dignitaries on the Shoe Throwing	Dignitaries ***George W. Bush*** on the shoe throwing,	Dignitaries ***Incident*** on the Shoe Throwing	Dignitaries ***Incident*** on the Shoe Throwing
28	India's First Unmanned Moon Mission	***Launch*** India's First Unmanned Moon Mission	***Successful*** India's First Unmanned Moon Mission	***Chandrayaan-1*** India's First Unmanned Moon Mission
29	Indian Parliament Attack	Indian Parliament ***Terrorist*** Attack	Indian Parliament Attack ***2001***	***People's Reaction's*** *on* Indian Parliament Attack
30	Polio Eradication Campaign	***UNICEF*** Polio Eradication campaign	Polio Eradication Campaign in ***India***	Polio Eradication Campaign in ***India***
31	Accused Ajmal Kasab	***Allegations*** Accused Ajmal Kasab	Accused Ajmal Kasab ***Terrorist***	***Allegations*** Accused Ajmal Kasab
32	Sania Mirza's Marriage	***Tennis*** Sania Mirza's Marriage	Sania Mirza's Marriage ***Cricketer***	Sania Mirza's Marriage, ***Shoaib Malik***
33	Mahendra Singh Dhoni National Award	Mahendra Singh Dhoni ***Padma Sri*** National Award	Mahendra Singh Dhoni ***Padma Sri*** National Award	Mahendra Singh Dhoni ***Padma Sri*** National Award
34	Left withdrew Support to the Congress	Left ***Front*** Withdrew Support to the Congress	Left withdrew Support to the Congress ***Government***	Left ***Front's*** withdrew Support to the Congress
35	MIG Crash in West Bengal	MIG Crash in West Bengal, ***2001***	MIG ***Aircraft*** Crash in West Bengal	MIG ***Aircraft*** Crash in West Bengal
36	World Non-Violence Day	***Declaration*** of World Non-Violence Day	World Non-Violence Day, **UNO**	***2nd October*** World Non-Violence Day
37	Film Censor Board Chairperson Woman	Film Censor Board Chairperson ***Appointment*** woman	Film Censor Board Chairperson ***Appointment*** Woman	Film Censor Board Chairperson ***Appointment*** Woman
38	Delhi Auto Expo 2010	***Pragati Maidan***, Delhi Auto Expo 2010	***Pragati Maidan,*** Delhi Auto Expo 2010	***Pragati Maidan,*** Delhi Auto Expo 2010
39	Harbhajan Singh Slapped Srisant	Harbhajan Singh Slapped Srisant, ***IPL***	***Incident*** Harbhajan Singh Slapped Srisant	Harbhajan Singh Slapped Srisant, ***Action***
40	Indian Animation Film Industry	Indian Animation Film ***Created*** Industry	Indian ***Growing*** Animation Film Industry	Indian ***New*** Animation Film Industry
41	Grameen Bank Muhammad Yunus Dispute	Grameen Bank Muhammad Yunus dispute, ***Bagladeshi Government***	Grameen Bank ***Implemenation*** Muhammad Yunus Dispute	Grameen Bank ***Founder*** Muhammad Yunus Dispute
42	Da Vinci Code India Release Controversy	***Film*** "Da Vinci Code" India release Controversy	***Hurdles*** Da Vinci Code India Release Controversy	***Novel*** Da Vinci Code India Release Controversy

(*Continued*)

Table 8. (Continued).

Query	Translated English Query	Case 1 (Without Ranking)	Case 2 (Highest Frequency)	Case 3 (Lowest Frequency)
43	Cervical Cancer Awareness, Treatment Vaccine	Cervical Cancer Awareness, Treatment Vaccine	Cervical Cancer Awareness, Treatment Vaccine	Cervical Cancer Awareness, Treatment Vaccine
44	India's first Formula 1 Circuit	***Construction*** of India's First Formula 1 Circuit	***Planning*** India's first Formula 1 Circuit	***Organisers*** India's first Formula 1 Circuit
45	Steve Waugh International Cricket Retirement	***Batsman*** Steve Waugh International Cricket Retirement	***Captain*** Steve Waugh International Cricket Retirement	Steve Waugh ***Batsman*** International Cricket Retirement
46	Bill and Melinda Gates Foundation, the Philanthropic Activities in India	Bill and Melinda Gates Foundation, the Philanthropic Activities ***AIDS*** in India	Bill and Melinda Gates Foundation, the Philanthropic Activities in ***AIDS*** India	***Initiatives*** Bill and Melinda Gates Foundation, the Philanthropic Activities in India
47	Greece Won the Euro Cup 2004	Greece Won the Euro Cup ***Tournament*** 2004	Greece Won the Euro Cup ***Football*** 2004	Greece Won the Euro Cup ***Tournament*** 2004
48	Imran Khan's Cancer Hospital in Pakistan	***Cricketer*** Imran Khan's Cancer Hospital in Pakistan	Imran Khan's Cancer Hospital ***Plans*** in Pakistan	Imran Khan's Cancer Hospital ***Inauguration*** in Pakistan
49	IPhone iPad Design Popularity Launch	***New*** IPhone iPad Design Popularity Launch	***Apple's*** IPhone iPad Design Popularity Launch	IPhone iPad Design Popularity Launch ***Popularity***
50	Satanic Verses Controversy	Satanic Verses Controversy, ***1989***	***Novel*** Satanic Verses Controversy	Satanic Verses Controversy ***Issues***

would become "***Chief Minister*** *YSR Reddy's Death.*" Other QE have been performed using the same approach (column 5 of Table 8).

Experimental Results

In order to test queries of FIRE dataset, three measures are considered for evaluation: precision, average precision and mean average precision. For calculating these, the number of relevant documents is selected manually from the number of retrieved documents that are retrieved using Google search engine for each query. Precision is computed as per Equation (3), and results for each query are shown in Table 9 and Figure 1.

$$\text{Precision} = \frac{\text{RelevantRetrievedDocuments}}{\text{RetrievedDocuments}} \tag{4}$$

Average precision (Figure 2) is the average of the precision value obtained for the set of top K documents existing after each relevant document is retrieved, and this value is then averaged over information needs. Mean average precision, for a set of queries, is the mean of the average precision scores for each query, as shown in Table 10 and Figure 3.

Discussion

For our experiment, we used 50 Hindi queries from FIRE 2012 for retrieval of English documents. FIRE queries consist of three parts: a title of query, descriptions of query and narration of query that provides more details about what kind of documents should be considered relevant or irrelevant. In the QE – which is the focus of this paper – the additional query terms are extracted from ranked documents as well as without ranked documents using TSV (Blum and Langley 1997; Sari and Adriani 2014). Many researchers used Okapi BM25 for ranking of documents to achieve better effective results. Syandra et al. (Sari and Adriani 2014) achieved 88.51% relevant documents on Indonesian-English CLIR experiment using SVM (Support Vector Machine), Okapi BM25 and term selection, respectively. Bodo Billerbeck et al. (Billerbeck et al. 2003) also achieve better retrieval results in their paper Query Expansion using Associated Queries using Okapi BM25.

On comparing with our previous work (Chandra and Dwivedi 2017), wherein we performed all the experiments with a smaller setup, we found that the results are in line with the previous work.

Keywords are selected for QE in three different ways using TSV. In the first case, keywords were selected from narration and description of queries that are given in FIRE data set for top @10 documents without considering the rank of documents. For computing TSV values of each keyword, frequencies

Table 9. Precision before and after query expansion.

Query	Precision Value@10			
	Before Query Expansion	After Query Expansion		
		Case 1 (Without Ranking)	Case 2 (Highest Frequency)	Case 3 (Lowest Frequency)
1	0.2	0.4	0.5	0.5
2	0.2	0.3	0.6	0.6
3	0.7	0.8	0.7	0.9
4	0.7	0.8	0.7	0.8
5	0.1	0.6	0.5	0.6
6	0.1	0.3	0.4	0.4
7	0.2	0.4	0.5	0.5
8	0.2	0.7	0.5	0.7
9	0.6	0.7	0.8	0.7
10	0.7	0.9	0.8	0.9
11	0.5	0.5	0.5	0.5
12	0.2	0.5	0.5	0.5
13	0.3	0.5	0.7	0.7
14	0.5	0.6	0.7	0.7
15	0.2	0.4	0.4	0.4
16	0.3	0.5	0.6	0.6
17	0.3	0.6	0.7	0.7
18	0.4	0.5	0.5	0.6
19	0.6	0.7	0.7	0.7
20	0.4	0.6	0.6	0.8
21	0.5	0.6	0.6	0.6
22	0.5	0.7	0.7	0.6
23	0.2	0.5	0.6	0.6
24	0.3	0.6	0.6	0.4
25	0.5	0.7	0.6	0.7
26	0.6	0.6	0.7	0.8
27	0.6	0.7	0.8	0.8
28	0.5	0.5	0.6	0.8
29	0.6	0.7	0.6	0.8
30	0.3	0.5	0.7	0.7
31	0.5	0.6	0.7	0.6
32	0.4	0.5	0.5	0.7
33	0.5	0.8	0.8	0.8
34	0.5	0.5	0.6	0.5
35	0.4	0.6	0.7	0.7
36	0.4	0.7	0.7	0.9
37	0.4	0.6	0.6	0.6
38	0.5	0.7	0.7	0.7
39	0.4	0.8	0.8	0.9
40	0.6	0.7	0.8	0.9
41	0.7	0.8	0.8	0.9
42	0.4	0.7	0.7	0.8
43	0.5	0.5	0.5	0.5
44	0.4	0.6	0.7	0.8
45	0.4	0.5	0.6	0.5
46	0.5	0.7	0.7	0.9
47	0.6	0.7	0.6	0.7
48	0.5	0.7	0.7	0.8
49	0.5	0.6	0.7	0.8
50	0.5	0.7	0.6	0.8
Total	**21.6**	**30.4**	**31.9**	**34.4**

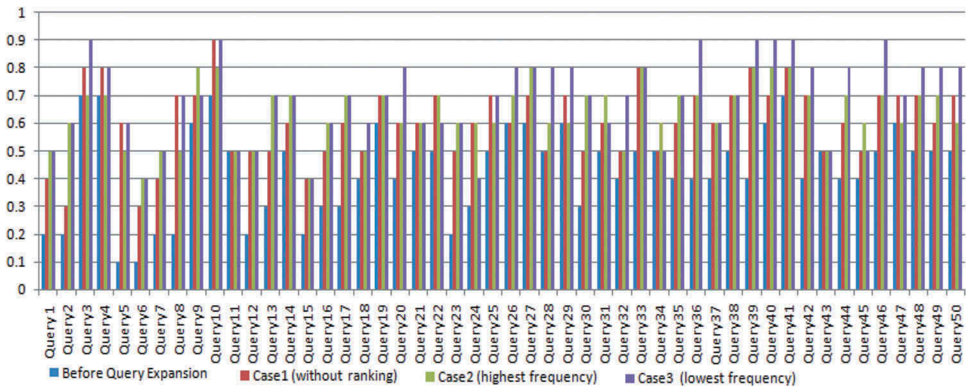

Figure 1. Precision values for all cases (i.e. case 1, case 2 and case 3).

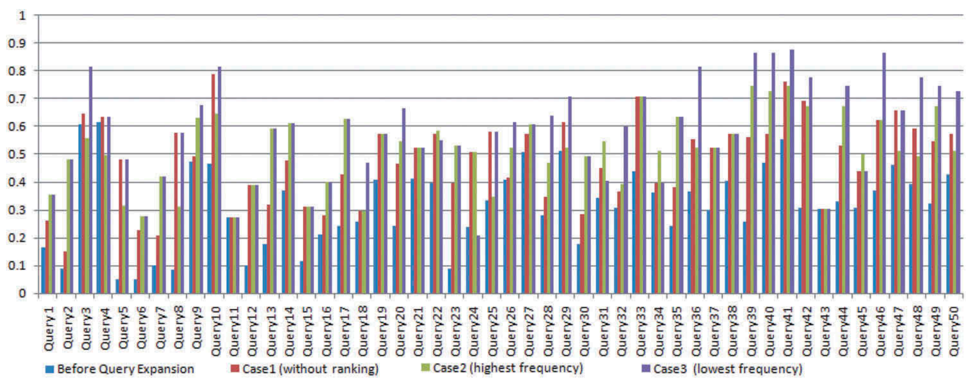

Figure 2. Average precision for all the cases (i.e. case 1, case 2 and case 3).

are calculated with the help of UAM corpus. Keywords which have minimum TSV value are added to the original query for QE.

In case 2 and case 3, keywords are selected from narration and description of FIRE queries only from top @3 documents that were ranked by Okapi BM25 method. In case 2, keywords that have the highest frequency in each document for single query were selected and after selection, we computed TSV values for all keywords. Keywords that have less TSV value were added to form a new query.

In case 3, keywords that have the lowest frequency in each document for single query were selected and after selection, we computed TSV values for all keywords. Keywords that have less TSV value were added to form a new query. In all of the above three cases, keywords that are not included in the original query are added to form a new query.

Computed mean average precision value for all the cases are as follows: mean average precision before QE is 0.3145 and after QE is 0.4781 (without

Table 10. Average precision and mean average precision value.

Query	Average Precision Value@10			
		After Query Expansion		
	Before Query Expansion	Case 1 (Without Ranking)	Case 2 (Highest Frequency)	Case 3 (Lowest Frequency)
1	0.1666	0.26	0.355	0.355
2	0.09	0.15	0.4795	0.4795
3	0.6083	0.6444	0.5574	0.8135
4	0.6164	0.6365	0.4972	0.6365
5	0.05	0.4796	0.315	0.4796
6	0.05	0.2266	0.2766	0.2766
7	0.1	0.2066	0.4175	0.4175
8	0.0833	0.5785	0.3134	0.5785
9	0.4724	0.4925	0.6294	0.677
10	0.4646	0.7903	0.647	0.8153
11	0.2726	0.2726	0.2726	0.2726
12	0.1	0.3891	0.3891	0.3891
13	0.1766	0.3207	0.5907	0.5907
14	0.3707	0.4775	0.6115	0.6115
15	0.1166	0.31	0.31	0.31
16	0.21	0.2796	0.4007	0.4007
17	0.2416	0.4281	0.627	0.627
18	0.2571	0.2962	0.2962	0.4707
19	0.408	0.5746	0.5746	0.5746
20	0.2428	0.4646	0.5464	0.6669
21	0.412	0.524	0.524	0.524
22	0.397	0.5749	0.583	0.549
23	0.09	0.3971	0.5314	0.5314
24	0.2375	0.5091	0.5091	0.2082
25	0.3341	0.5791	0.3479	0.5791
26	0.4073	0.4142	0.5246	0.617
27	0.5091	0.5749	0.6073	0.6073
28	0.2796	0.3462	0.468	0.6396
29	0.5133	0.616	0.524	0.7067
30	0.1766	0.2862	0.4906	0.4906
31	0.343	0.4507	0.5457	0.4057
32	0.3082	0.3646	0.393	0.6008
33	0.4383	0.7067	0.7067	0.7067
34	0.3606	0.3946	0.5133	0.3946
35	0.2432	0.3812	0.6365	0.6365
36	0.3666	0.5541	0.524	0.8153
37	0.2987	0.524	0.524	0.524
38	0.4049	0.5749	0.5749	0.5749
39	0.2582	0.5623	0.7453	0.8663
40	0.468	0.5749	0.7253	0.8663
41	0.5541	0.762	0.7453	0.8788
42	0.3082	0.6915	0.6732	0.7763
43	0.3057	0.3057	0.3057	0.3057
44	0.3321	0.5299	0.6732	0.7453
45	0.3082	0.438	0.4983	0.438
46	0.3707	0.6241	0.6241	0.8663
47	0.4612	0.6565	0.5133	0.6565
48	0.393	0.5907	0.4906	0.7763
49	0.3216	0.5464	0.6732	0.7453
50	0.4264	0.5746	0.5133	0.7253
Total	**15.725**	**23.9071**	**25.8156**	**29.2006**
Mean Average Precision	**0.3145**	**0.4781**	**0.5163**	**0.5840**

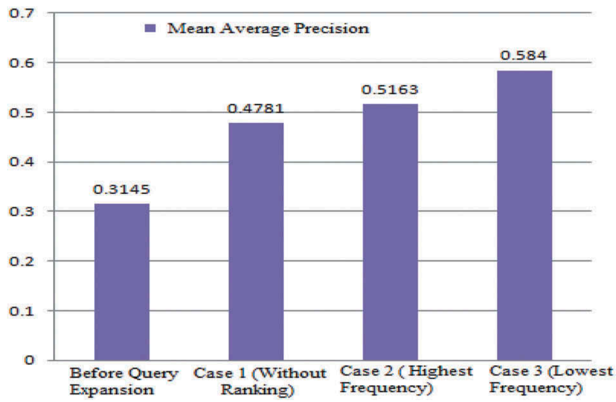

Figure 3. Mean average precision for all the cases (i.e. case 1, case 2 and case 3).

ranking, i.e. case 1), 0.5163 (with ranking, i.e. case 2) and 0.5840 (with ranking, i.e. case 3).

Retrieved results of relevant documents after QE (in all cases) are better than the retrieved results before QE. This shows that QE is one of the techniques to achieve better retrieval results in Hindi–English CLIR.

Retrieved results (i.e. mean average precision value) of ranked documents are higher in case 2 (0.5163) and case3 (i.e. 0.5840) as compared to without ranked documents in case 1 (i.e. 0.4781). This shows that term selection for QE by ranking using Okapi BM25 provides better retrieval results in comparison to retrieve results without considering the rank of documents in Hindi–English CLIR.

Mean average precision value of retrieved results, after QE in case 3, is higher than the remaining two cases (case 1, case 2). This shows that the selection of keyword having higher occurrence (i.e. high frequency) in a single document is less important than the occurrence of the keyword in multiple documents for QE.

Sometimes it is possible that the QE cannot be performed when suitable terms are not found in narration and description of queries for QE. Just like for Query No. 11, i.e. "*Attacks on Indian Students in Australia*," and Query No. 43, i.e. "*Cervical Cancer Awareness, Treatment Vaccine*," the QE is not performed in all the three cases.

Conclusion

The main problem of CLIR is poor performance that occurs due to query term mismatching, untranslated query word, multiple representations of query terms, etc. QE helps to resolve the problem of ambiguity in CLIR by adding suitable terms in a query to retrieve better results. QE method also improves the performance of a search engine. This paper focuses on term selection for QE. Term selection plays a vital role to expand the user's queries to increase the mean average precision of Hindi–English CLIR.

We have explored three different strategies, to select the most effective query terms for improvement of retrieved results in Hindi-English CLIR using QE. In case 1, for QE, appropriate keywords are taken from description and narration of queries of FIRE dataset without considering the rank of documents. In case 2 and case 3, for QE, appropriate keywords are taken from description and narration of queries of FIRE dataset by considering the rank of documents using Okapi BM25. In case 2 and case 3, keywords having the highest and lowest frequency are considered for QE, respectively.

Mean average precision value of retrieved results for case 1, case 2 and case 3 are 0.4781, 0.5163 and 0.5840, respectively. Among the three strategies explored, the best results are obtained in case 3 where QE is performed by adding lowest frequency words.

Results of case 2 and case 3 are better than case 1, which shows that without ranking (i.e. case 1), the quality of retrieved results of Hindi–English CLIR is not good as compared to results obtained by ranking using Okapi BM25 (i.e. case 2 and case 3). In the dynamic world of web, time factor and quality of retrieved documents are important issues of CLIR because the demand for accessing information in different languages is increasing with a fast speed.

References

Agichtein, E., S. Lawrence, and L. Gravano. 2004. Learning to find answers to questions on the web. *ACM Transactions on Internet Technology (TOIT)* 4 (2):129–62. doi:10.1145/990301.

Aljlayl, M., and O. Frieder. 2001, October. Effective Arabic-English cross-language information retrieval via machine-readable dictionaries and machine translation. Proceedings of the tenth International Conference on Information and knowledge management (pp. 295–302), Atlanta, GA, November 10, ACM.

Ballesteros, L., and W. B. Croft. 1997, December. Phrasal translation and query expansion techniques for cross-language information retrieval. ACM SIGIR Forum (Vol. 31, No. SI, pp. 84–91), Philadelphia, PA, ACM.

Ballesteros, L., and W. B. Croft. 1998, August. Resolving ambiguity for cross-language retrieval. In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval (pp. 64–71), Melbourne, Australia, ACM.

Banchs, R. E., and M. R. Costa-Jussà. 2013. Cross-language document retrieval by using nonlinear semantic mapping. *Applied Artificial Intelligence* 27 (9):781–802. doi:10.1080/08839514.2013.835232.

Bandyopadhyay, S., T. Mondal, S. K. Naskar, A. Ekbal, R. Haque, and S. R. Godhavarthy. 2007, September. Bengali, Hindi and Telugu to English ad-hoc bilingual task at CLEF 2007. In *Workshop of the Cross-Language Evaluation Forum for European Languages,* ed. C. Peters et al, 88–94. Heidelberg, Berlin: Springer.

Billerbeck, B. 2005, September. Efficient query expansion. PhD Thesis, RMIT University, Melbourne, Australia.

Billerbeck, B., F. Scholer, H. E. Williams, and J. Zobel. 2003, November. Query expansion using associated queries. Proceedings of the twelfth international conference on Information and knowledge management (pp. 2–9), New Orleans, LA, November 3–8, ACM.

Blum, A. L., and P. Langley. 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97 (1):245–71. doi:10.1016/S0004-3702(97)00063-5.

Chandra, G., and S. K. Dwivedi. 2014, December. A literature survey on various approaches of word sense disambiguation. Computational and Business Intelligence (ISCBI), 2014 2nd International Symposium on (pp. 106–09), New Delhi, India, IEEE.

Chandra, G., and S. K. Dwivedi. 2017. Query expansion based on term selection for Hindi–English cross lingual IR. *Journal of King Saud University-Computer and Information Sciences*. doi:10.1016/j.jksuci.2017.09.002.

Chaware, S. M., and S. Rao. 2011. Ontology approach for cross-language information retrieval. *Published in International Journal of Computer Technology and Application* 2:379–84.

Chinnakotla, M. K., S. Ranadive, O. P. Damani, and P. Bhattacharyya. 2008. Hindi to English and Marathi to English cross language information retrieval evaluation. In *Advances in multilingual and multimodal information retrieval*, ed. V. Jijkoun, Th. Mandl, H. Müller, D. W. Oard, V. Petras and D. Santos, 111–18. Berlin: Springer. ISBN:978-3-540-85759-4. doi:10.1007/978-3-540-85760-0_14.

Contractor, D., G. Kothari, T. A. Faruquie, L. V. Subramaniam, and S. Negi. 2010, October. Handling noisy queries in cross language faq retrieval. Proceedings of the 2010 conference on empirical methods in natural language processing (pp. 87–96), MIT, MA, October 9–11, Association for Computational Linguistics.

Daoud, M., and J. X. Huang. 2013. Mining query-driven contexts for geographic and temporal search. *International Journal of Geographical Information Science* 27 (8):1530–49. doi:10.1080/13658816.2012.756883.

Das, S., A. Seetha, M. Kumar, and J. L. Rana. 2010. Post Translation Query Expansion using Hindi Word-Net for English-Hindi CLIR System. In Published in Forum of Information Retrieval Evaluation, FIRE (pp. 53–41), Gandhinagar, India, DAIICT.

Davis, M. W. 1996, November. New experiments in cross-language text retrieval at NMSU's computing research lab. In TREC.

Dwivedi, S. K. 2012. A highest sense count based method for disambiguation of web queries for Hindi language web information retrieval. *International Journal of Information Retrieval Research (IJIRR), IGI Global, USA, 2012 (DBLP)* 2 (4):1–11.

Gaillard, B., J. L. Bouraoui, E. G. De Neef, and M. Boualem. 2010, May. Query expansion for cross language information retrieval improvement. Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on (pp. 337–42), Nice, France, IEEE.

Hanani, U., B. Shapira, and P. Shoval. 2001. Information filtering: Overview of issues, research and systems. *User Modeling and User-Adapted Interaction* 11 (3):203–59. doi:10.1023/A:1011196000674.

Imran, H., and A. Sharan. 2009. Thesaurus and query expansion. *International Journal of Computer Science & Information Technology (IJCSIT)* 1 (2):89–97.

Joshi, H., A. Bhatt, and H. Patel. 2013. *Transliterated search using syllabification approach*. Delhi, India: Forum for Information Retrieval Evaluation.

Jothilakshmi, R., N. Shanthi, and R. Babisaraswathi. 2013. A survey on semantic query expansion. *Journal of Theoretical & Applied Information Technology* 57 (1): 128–138.

Kwok, K. L. 1997, March. Evaluation of an English-Chinese cross-lingual retrieval experiment. Working Notes of AAAI-97 Spring Symposiums on Cross-Language Text and Speech Retrieval (pp. 110–14), Palo Alto, CA.

Landauer, T. K., and M. L. Littman. 1990. Fully automatic cross-language document retrieval using latent semantic indexing. In Proc. of the 6th Conference of UW Center for New OED and Text Research, 1990 (pp. 31–38). doi:10.1099/00221287-136-2-327.

Lee, K. S., K. Kageura, and K. S. Choi. 2002, August. Implicit ambiguity resolution using incremental custering in Korean-to-English cross-language information retrieval. Proceedings of the 19th International Conference on Computational Linguistics-Volume 1 (pp. 1–7), Taipei, Taiwan, Association for Computational Linguistics.

Lesk, M. 1986, June. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In Proceedings of the 5th annual international conference on Systems documentation (pp. 24–26), Toronto, Canada, ACM.

Levow, G. A., D. W. Oard, and P. Resnik. 2005. Dictionary-based techniques for cross-language information retrieval. *Information Processing & Management* 41 (3):523–47. doi:10.1016/j.ipm.2004.06.012.

Mandal, D., M. Gupta, S. Dandapat, P. Banerjee, and S. Sarkar. 2008. Bengali and Hindi to English CLIR evaluation. In *Advances in multilingual and multimodal information retrieval*, ed. V. Jijkoun, Th. Mandl, H. Müller, D. W. Oard, V. Petras and D. Santos, 95–102. Berlin: Springer. ISBN:978-3-540-85759-4. doi:10.1007/978-3-540-85760-0_14.

Mandal, D., S. Dandapat, M. Gupta, P. Banerjee, and S. Sarkar. 2007, September. Bengali and Hindi to English cross-language text retrieval under limited resources. In Working Notes for the CLEF 2007 Workshop. doi:10.1094/PDIS-91-4-0467B.

Maxwell, T., and B. Schafer. 2010. Natural language processing and query expansion in legal information retrieval: Challenges and a response. International review of law. *Computers & Technology* 24 (1):63–72.

Oard, D. W. 1998. A comparative study of query and document translation for cross-language information retrieval. In Proceedings of the Third Conference of the Association for Machine Translation and the Information Soup (pp. 472–83), Berlin, Springer.

Oard, D. W. 2003. The Surprise Language Exercises. *ACM Transactions on Asian Language Information Processing (TALIP)* 2 (2):79–84. doi:10.1145/974740.

O'Donnell, M. 2008, January. http://www.wagsoft.com/CorpusTool/

Pigur, V. A. 1979. Multilanguage information-retrieval systems: Integration levels and language support. *Automatic Documentation and Mathematical Linguistics* 13 (1):36–46.

Pingali, P., K. K. Tune, and V. Varma. 2008. Improving recall for Hindi, Telugu, Oromo to English CLIR. In *Advances in multilingual and multimodal information retrieval*, ed. V. Jijkoun, Th. Mandl, H. Müller, D. W. Oard, V. Petras and D. Santos, 103–10. Berlin: Springer. ISBN: 978-3-540-85759-4. doi:10.1007/978-3-540-85760-0_14.

Ponte, J. M., and W. B. Croft. 1998, August. A language modeling approach to information retrieval. Proceedings of the 21st annual international ACM SIGIR conference on research and development in information retrieval (pp. 275–81), Melbourne, Australia, ACM.

Riezler, S., A. Vasserman, I. Tsochantaridis, V. Mittal, and Y. Liu. 2007, June. Statistical machine translation for query expansion in answer retrieval. *Proceedings of the 45th Annual Meeting-Association for Computational Linguistics* 45 (1):464–71.

Rijsbergen, C. J. V. 1979. *Information retrieval*. 2nd ed. Glasgow: Dept. Of Computer Science, University.

Robertson, S. E., S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford. 1995. Okapi at TREC-3. *Nist Special Publication Sp* 109:109.

Salton, G. 1973. Experiments in multi-lingual information retrieval. *Information Processing Letters* 2 (1):6–11. doi:10.1016/0020-0190(73)90017-3.

Sanchez-Martinez, F., and R. C. Carrasco. 2011. Document translation retrieval based on statistical machine translation techniques. *Applied Artificial Intelligence* 25 (5):329–40. doi:10.1080/08839514.2011.559906.

Sari, S., and M. Adriani. 2014, October. Learning to rank for determining relevant document in Indonesian-English cross language information retrieval using BM25. Advanced Computer Science and Information Systems (ICACSIS), 2014 International Conference on (pp. 309–14), Jakarta, Indonesia, IEEE.

Seetha, A., S. Das, and M. Kumar. 2007, December. Evaluation of the English-Hindi cross language information retrieval system based on dictionary based query translation method. Information Technology, (ICIT 2007). 10th International Conference on (pp. 56–61), Orissa, India, IEEE.

Seetha, A., S. Das, and M. Kumar. 2009, January. Improving performance of English-Hindi CLIR system using linguistic tools and techniques. In Proceedings of the First International Conference on Intelligent Human Computer Interaction (pp. 261–71), India, Springer.

Sekine, S., and R. Grishman. 2003. Hindi-English cross-lingual question-answering system. *ACM Transactions on Asian Language Information Processing (TALIP)* 2 (3):181–92. doi:10.1145/979872.

Sheridan, P., and J. P. Ballerini. 1996, August. Experiments in multilingual information retrieval using the SPIDER system. Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 58–65), Zurich, Switzerland, ACM.

Shukla, S., and U. Sinha. 2015, March. Categorizing sentence structures for phrase level morphological analyzer for English to Hindi RBMT. In 2015 International Conference on Cognitive Computing and Information Processing (CCIP) (pp. 1–6). IEEE.

Singhal, A., and F. Pereira. 1999, August. document expansion for speech retrieval. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 34–41), Berkeley, CA, ACM.

Varshney, S., and J. Bajpai. 2013, December. Improving retrieval performance of English-Hindi based cross-language information retrieval. MOOC Innovation and Technology in Education (MITE), 2013 IEEE International Conference in (pp. 300–05), Jaipur, India, December 20 – 22, IEEE.

Xu, J., and W. B. Croft. 2000. Improving the effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems (TOIS)* 18 (1):79–112. doi:10.1145/333135.333138.

Zimmer, C., C. Tryfonopoulos, and G. Weikum. 2008, July. Exploiting correlated keywords to improve approximate information filtering. Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval (pp. 323–30), Singapore, ACM.