A comprehensive survey on cross-language information retrieval system

Gouranga Charan Jena, Siddharth Swarup Rautaray
School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, Odisha, India

Article Info	ABSTRACT
	Cross language information retrieval (CLIR) is a retrieval process in which the user fires queries in one language to retrieve information from another (different) language. The amount of information generated in different languages and disseminated via the World Wide Web and social media platforms is growing exponentially. It is important to ensure that online information resources and services are efficiently and equitably accessible to all users, regardless of their linguistic and cultural backgrounds. The need for exploring and building more specialized information systems that enable speakers of an Odia language to discover valuable information beyond linguistic and cultural barriers. This study is aimed to build an experimental CLIR system between one of the under-resourced languages (i.e. Odia), and one of the most commonly used online languages, (i.e. English). However, we have not yet considered any prototype approach to develop English-Odia CLIR system. In this research we studied on the various IR variants and techniques used in CLIR system. Further, based on available literature, a number of challenges and issues in CLIR have been identified and discussed. Firstly, we have focused on building basic linguistic resources and IR tools required for designing an English-Odia CLIR using Apache Lucene engine. Secondly, some of the major linguistic resources that we developed during the course of this study include a Unicode compatible Odia Keyboard, normalised Odia corpus, a generic computational model of Odia morphology, a bilingual dictionary and an Odia stemmer from scratch. While our English-Odia dictionary has been adopted and used for query translation.

Corresponding Author:

Gouranga Charan Jena,
School of Computer Engineering,
KIIT Deemed to be University,
Bhubaneswar, Odisha, India.
Email: jenagouranga2000@gmail.com

1. INTRODUCTION

A huge amount of information on internet is available in different language that can be access by anybody at any time. The ability to search and retrieve information in multiple languages is becoming increasingly important and challenging in today. Consequently, Cross-lingual (language) information retrieval search engines have received more research attention and are increasingly being used to retrieve information on the internet. The area of information access has evolved to perform many sophisticated tasks such as the information retrieval, question answering tasks, summarization, multimedia information retrieval, text mining and clustering and Web information retrieval. Cross-language information retrieval is a retrieval process in which the user fires queries in one language to retrieve information from another language. Information retrieval tries to identify relevant documents for an information need, expressed as a query. The problems that an Information retrieval system should deal with include document indexing (which tries to extract important indexes from a document and weigh them), query analysis (similar to document indexing)

and query evaluation (i.e. matching the query with the documents). Each of these problems has been the subject of many studies in IR. The meaning of the term IR can be very broad. It might be defined as finding material (usually text) that satisfies an information need from within large collections (usually stored on computers").In modern parlance, the word 'search' has tended to replace "information retrieval". The term 'search' is quite ambiguous, but in context we use the two synonymously. There are different types of retrieval techniques used to design an Information retrieval system i.e. a) Boolean retrieval b) Dictionaries and tolerant retrieval c)XML retrieval d) Probabilistic IR e) Language models for IR f) Text classification and Naive Bayes g) Vector space classification h) Supper vector machines and machine learning on documents i) Flat clustering j) Hierarchical clustering k) Matrix decompositions and latent semantic indexing l) Web search basics m) Web drawing and indexing etc. However, we have not discussing the details of the each retrieval Model here. But these are the backbone of the any information retrieval Process. There are five basic steps involved in traditional Information retrieval system i.e. *1st Step:* Indexing, *2nd Step:* Index Compression*, 3rd Step:* Computing scores in a complete search system, *4th Step*: Evaluation in Information retrieval, *5th Step:* Relevance feedback and query expansion. Nowadays the types of Information Retrieval system extended in to many areas such as: Image and multimedia IR, Music retrieval, Parallel and peer-to-peer IR and many more. Information retrieval can be monolingual, bilingual or multilingual. Broadly the Information retrieval (IR) be classified in to four categories as:

1.1 Monolingual information retrieval

The monolingual retrieval is the most basic and most common type, which like its name reveals, uses only one language for both the query and information retrieval. Monolingual information retrieval is a retrieval process in which the user presents queries in one language to retrieve information in the same language.

1.2 Bi-lingual information retrieval

Bi-lingual information retrieval facilitates the uses of queries in one language to access documents in other languages. If two languages are involved, for example if the document is written in Hindi and the query is made in English, information retrieval is said to be bilingual.

1.3 Cross-lingual information retrieval

Cross-Language Information Retrieval is quickly becoming a mature area in the information retrieval world. The goal is to allow a user to issue a query in language L and have that query retrieve documents in language L'. The idea is that the user wants to issue a single query against a document collection that contains documents in a myriad of languages. An implicit assumption is that the user understands results obtained in multiple languages. If this is not the case, it is necessary for the retrieval system to translate the selected foreign language documents into a language that the user can understand. Surveys of cross-language information retrieval techniques and multilingual processing include [1], [2] various approaches on CLIR adopted by researchers few decades ago and their limitations. Cross-language information retrieval (CLIR) is a retrieval process in which the user presents queries in one language to retrieve information in another language. CLIR approaches are decomposed into two major research fields: - the first is **dictionary based approach** and **machine translation (MT)** and the second is **Concept Driven Approach** (CDA).

1.4 Multi-Lingual Information Retrieval

MLIR facilitates the uses of queries in one language to access documents in different languages (more than two languages involved).

2. STATE OF THE ART

Cross-Language Information Retrieval is a vivid research field due to existence of numerous languages used by people across the world. In this section we try to provide a bird eye view of various works done in the field of Cross-Language Information Retrieval by analysing some of the comprehensive surveys and articles published in the field during the time span of 2001 to 2016. The following Table 1 depicts a detailed of work done by authors along with respective years as well as key findings.

Through the Table 1 we intended to provide a crisp detail of various surveys done in the field of cross-language information retrieval in general.

Table 1. Analysis of some comprehensive surveys and articles

Ref./ Author	Year	Scope of Analysis	Key Findings
Ari Pirkola et.al [3]	2001	Reviewed literature on dictionary-based cross-language information retrieval (CLIR) and presents CLIR research done at the University of Tampere (UTA). The structured query model and report finding for four different language pairs concerning the effectiveness of query structuring is presented.	This paper considers the main problems associated with dictionary-based CLIR, which are 1) Untranslatable search keys due to the limitations of general dictionaries, 2) the processing of Inflected words, 3) phrase identification and translation and 4) lexical ambiguity in source and target language. This paper proposed a standard method in dictionary-based CLIR is to replace each source language key by all of its target language equivalents given by a dictionary to overcome above said problems.
Mustafa Abusalah et.al [4]	2007	Reviewed a literature survey based on CLIR system in short. They have described previous work in CLIR, current problems in CLIR, and made recommendations for future work.	Described previous work in CLIR, current problems in CLIR, and made recommendations for future work. It describes CLIR systems that have combined query translation to parallel-corpora could show better results as parallel corpora has rich context to cover the weak context of the query. The semantic web can play an important role in CLIR through the use of ontologies. Ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary. Ontology is an explicit specification of a conceptualization. Ontologies can be implemented in translation systems to extract conceptual relations for monolingual and cross language IR.
Manoj Kumar Chinnakotla et.al [5]	2008	Developed a Hindi to English and Marathi to English CLIR systems and describes a query translation based approach using bi-lingual dictionaries. Query words not found in the dictionary re transliterated using a simple rule based transliteration approach. The resultant transliteration is then compared with the unique words of the corpus to return the 'k' words most similar to the transliterated word. The resulting multiple translation/transliteration choices for each query word are disambiguated using an iterative page-rank style algorithm which, based on term-term co-occurrence statistics, produces the final translated query. Using the above approach, for Hindi, a Mean Average Precision (MAP) of 0.2366 using title and a MAP of 0.2952 using title and description are achieved. For Marathi, a MAP of 0.2163 using title is achieved.	This paper discussed a query translation based approach using bi-lingual dictionaries. Query words not found in the dictionary are transliterated using a simple rule based transliteration approach. Developed a Hindi to English and Marathi to English CLIR systems developed for the CLEF 2007 Ad-Hoc Bilingual Task. Our approach is based on query translation using bi-lingual dictionaries. Transliteration of words which are not found in the dictionary is done using a simple rule based approach. It makes use of the corpus to return the 'k' closest possible English transliterations of a given Hindi/Marathi word. Disambiguating the various translations/transliterations is performed using an iterative page-rank style algorithm which is based on term-term co-occurrence statistics.
Sujoy Das et.al [6]	2009	Made an observation that the dictionary based query translation approach has been widely used by researchers of CLIR. The translation ambiguity and target polysemy are the two major problems of dictionary based CLIR.	This paper discussed that the dictionary based query translation approach has been widely used by researchers of CLIR. The translation ambiguity and target polysemy are the two major problems of dictionary based CLIR. In this paper, the researchers have investigated part of speech and co-occurrence based disambiguation techniques for English-Hindi CLIR system.
Mustafa abusalah et al. [7]	2009	Reports an experiment to evaluate a CLIR system that uses ontology to improve query translation in the travel domain. The ontology-based approach significantly outperformed the Machine Readable Dictionary translation baseline using Mean Average Precision as a metric in a user centred experiment.	This paper suggests the ontology-based approach significantly outperformed the Machine Readable Dictionary translation baseline using Mean Average Precision as a metric in a user centred experiment.
Nurjannaton Hidayah Rais et.al [8]	2010	Discusses research on query translation events in Malayalam-English Cross-Language Information Retrieval (CLIR) system.	This paper research work on query translation events in Malayalam-English Cross-Language Information Retrieval (CLIR) system. The researchers have assumed that by improving query translation accuracy, they can improve the information retrieval performance.
Saurabh Varshney et.al [9]	2013	This paper studied the effect of target polysemy and translation ambiguity in dictionary based query translation approach for English- Hindi CLIR system.	Have studied the effect of target polysemy and translation ambiguity in dictionary based query translation approach for English- Hindi CLIR system. CLIR provides new paradigms in searching documents through myriad varieties of languages across the world and it can be the baseline for searching not only among

Ref./ Author	Year	Scope of Analysis	Key Findings
Rekha Warrier et.al [10]	2015	This paper describes a detailed literature survey based on various CLIR techniques, advantages and limitations in short.	two languages but also in multiple. Creating accurate metadata in different languages in documents or good translation of key information in documents can help improve the quality of the index and retrieval. After the evaluation of both the pure dictionary and the ontology systems, the ontology based system scored higher in terms of precision. In future development ontology will be enhanced and extended by using annotation tools to align new concepts to the ontology and then test it again with the dictionary system. Other areas for investigation include ease of use, the use of relevance feedback, the effect of more extensive use of concept relations and possibly experiments with larger data sets. This paper discussed different types of CLIR techniques and advantages and disadvantages of each technique.
Sanjay K. Dwivedi et.al [11]	2016	This paper describes a detail survey on CLIR system and number of challenges, issues present in the CLIR system. This paper takes an overview of the new application areas of CLIR and reviews the approaches used in the process of CLIR research for query and document translation. It also describes the diversity of information and language barriers are the serious issues for communication and cultural exchange across the world. To solve such barriers, cross language information retrieval (CLIR) system, are nowadays in strong demand.	This paper describes a detail survey on number of challenges and issues in CLIR have been identified and discussed. Cross languages information retrieval systems would play a very important role in allowing the people to go through the documents and literatures of other languages thus breaking the language barrier. This paper gives a broad overview of the speedy demanding work in the field of CLIR by exploring its aspiration, difficulties, basic tools, major works and future research goals.

3. CLASSIFICATION OF CLIR TECHNIQUES

Broadly, the CLIR techniques are divided in to two categories i.e. Bilingual MRD and MT and Concept based field. Then these two broad categories subdivided in to many sub categories depicted in diagram Figure 1.
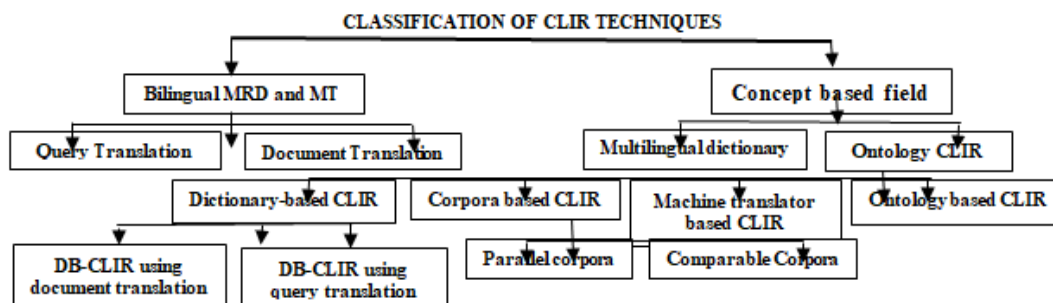

Figure 1. CLIR techniques

3.1. Based on "Bilingual MRD and MT", CLIR system is divided into
a) Query Translation

Query translation approaches use dictionaries, corpora based , machine translation, language specific stemmers, thesauri, and automatically generated bilingual term lists to implement the necessary translation of the user query in language L to the target query language L'. An excellent discussion of dictionary-based query translation techniques is given in [12]. In this work, a study of various techniques is presented and effectiveness of specific techniques is measured. Query translation involves translating the query to the target language. Table 2 describes, ambiguity occurs when words have multiple meaning which also referred to as polysemy. In dictionary based approach of query translation, ambiguity is high due to not picking the right meaning according to the user context in the query as compare to corpora based translation. Corpora based translation approach of query translation picks right meaning of the word according to the context. However, the cost of development of corpus is more expensive than dictionary based translation approach. The availability of Unicode corpus for a under resource language is rare. In machine translation

approach of query translation ambiguity is less as the user have an option to choose the meaning of a word from list of meaning of a word. Dictionary based approach of query translation and Corpora based translation approach of query translation system can be test based on an analysis of the internal structure of the component.

Table 2. Comparison between Techniques of Query Translation

Parameters	Dictionary Based Translation Approach	Corpora Based Translation Approach	Machine Translation Based Approach
Ambiguity	High	Low	Low
Offline Translation	Possible	Possible	Not Possible
Working Architecture	Visible as like white box testing.	Visible as like white box testing	Works similar to black box testing
Development expenses	Less expensive	More expensive than DBT	More expensive
Translation Availability	Highly available in many languages	Available only in few languages	Available only in few languages

An user fires an English query from the cross language IR interface and the query is being tranlated to Odia language. Then it will search the Odia document collection and retrives the desired Odia document as its search result. The above Figure 1 describe the working architechure of the Query Translation approach.
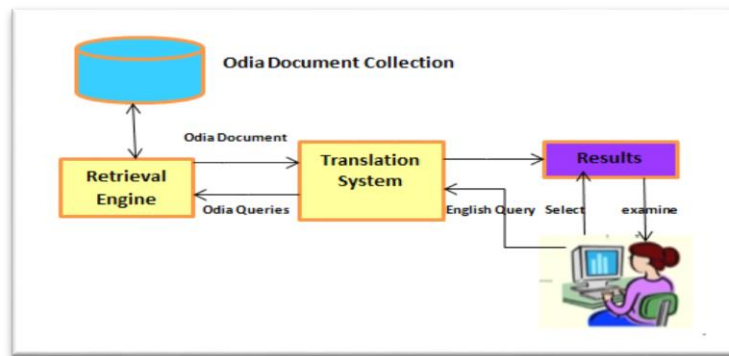

Figure 1. Query Translation Model

b.   Document Translation

A simple way to "translate" documents is to use any of the query translation approaches that we have already discussed. Since the documents contain natural language text, it is also possible to run machine translation algorithms to translate from language. L to language L'. Although machine translation algorithms are not perfect, they have been used as a foundation for CLIR. An advantage of these algorithms is that the use of the full natural language text provides more evidence when selecting a potential translation. Table 3 describes, document translation translates the document into the source language (i.e. the language used for the query). For document translation approach we have used bilingual corpus in the backend. The size of the bilingual/parallel corpus should large to get the better search result. The preparation of bilingual corpus is high cost and overhead. The linguist must have mastery over the two languages to prepare the bilingual corpus.

Table 3. Difference between Query and Document Translation

Parameter	Query Translation	Document Translation
Size	Small	Large
Language	Prior knowledge of translation language is not required	Prior knowledge of translation language is required
Overhead	Low	High
Recovery	When these are translated inappropriately, the IR engine has no chance to recover	Chance to recover
Ambiguity	Maximum chances of occurring ambiguity	Minimum chances of occurring
Cost	Low cost	High Cost

An user fires an English query from the cross language IR interface then it will search the English document collection and retrives the desired English document as its search result.Then it will translate the English search result to the Odia Document and gives the user in Odia search results. Figure 2 describes the working architechure of the Document Translation approach.
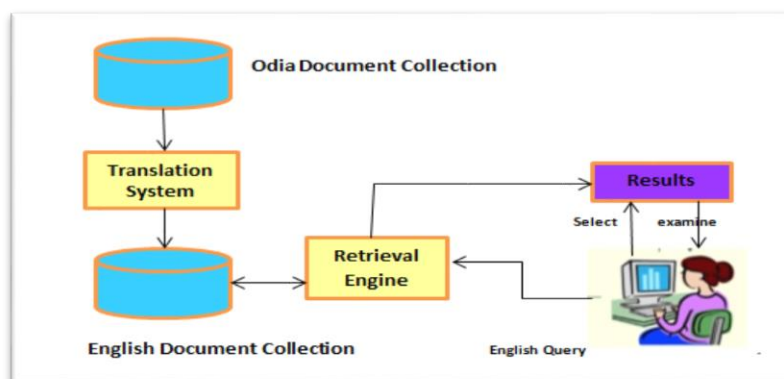

Figure 2. Document Translation Model

3.2. Based on second research field which is "Concept based field", CLIR system has two approaches
  a. Multilingual dictionary (Dictionary based).
  b. Ontology CLIR techniques can be classified into different categories based on translation resources:
   i) Dictionary-based CLIR technique (DB-CLIR)
   ii) Corpora based CLIR technique (Parallel corpora and Comparable Corpora)
   iii) Machine translator based CLIR technique
   iv) Ontology based CLIR technique (Concept driven field)

4. RESEARCH GAP
Queries from users are often too short, which produce more ambiguity in query translation process. And it reduces the accuracy of the cross language retrieval results. Since the problem of language mismatch in CLIR are more serious than in monolingual IR. It is necessary to exploit techniques for improving the multilingual retrieval performance. In CLIR systems, users often present their query in their native language, and then the system automatically searches documents written in other languages. Therefore, it is a challenge for CLIR to conquer the barrier between the source language (SL) in query sentences and the target language (TL) in documents to be searched. Most of the CLIR systems utilize Machine Translation (MT) technology to resolve this problem. As MT research itself has a number of issues (such as accuracy), the research in CLIR also faces critical issues and challenges that must be addressed. Due to this translation causes a reduction in the retrieval performance of CLIR as compared to monolingual IR system. The main reason for this reduced performance are missing specified vocabulary, missing general terms and wrong translation due to ambiguity.

4.1. Ambiguity
Ambiguity [13] occurs when words have multiple meaning which also referred to as homonymy or polysemy. Ambiguity in IR is semantic and syntactic in nature, whereas ambiguity in CLIR is semantic and lexical. So the probability of occurrence of ambiguity in CLIR is higher than normal IR, due to the availability of different languages [14].

4.2. Complexity in New Applications
Question/Answering is relatively a new stream of Information Retrieval. In Question/Answering end-users throw a question in a form of query and retrieve answers for that. However, challenge is to retrieve answers of English questions in different language.

4.3. Specialized Terminology and Proper Nouns

Specialized terminology, such as scientific names, is often difficult to translate and is often found in specialized dictionaries or term banks. Specialized terminology tends to be less ambiguous than regular vocabulary although regular vocabulary can have a specialized meaning when used in a certain subject area.

5. PROPOSED SOLUTION

Cross lingual information retrieval for foreign languages like English, French, Chinese etc. has been an appealing area for researchers from long time. But Indian languages have grabbed attention only a decade back. The work done by researchers show mixed results in terms of improvement over monolingual retrieval in Indian language perspective. Parallel or comparable corpora are yet other useful resources for designing a CLIR System. Parallel corpora are preferred in CLIR because they provide more accurate translation knowledge but due to their scarcity, comparable corpora are often used in CLIR system. The above observation concludes that there is a wide scope of research to improve existing algorithms or developing new one to improve the performance level of CLIR system. However, we have not yet decided any concrete prototype approach for developing English-Odia CLIR System. But now we are focusing on collecting resources like parallel corpus, bilingual dictionary and probalistic dictionary etc. and building components like stemmer, transliteration and SMT etc. system required for building an effective CLIR system. We presume that our proposed prototype system will prove to be competent with other existing systems in future if we successfully collect the resources, build the component and modified the existing algorithm.

6. RESULT AND ANALYSIS

Anurag Seetha & S. Das [15] performed translation on Fire 2010 Hindi test collection using Shabdanjali dictionary & query expansion by Hindi WordNet. The method proved to be ineffective. It is because general dictionaries have low coverage problem. To remove this inefficiency Larkey and Connell [16] used probabilistic dictionary derived from parallel corpus for English to Hindi translation and achieved effective cross lingual retrieval. Other researchers have made use of machine translation for cross lingual retrieval. D.Thenmozhi & C. Aravindan [17] used MT on agricultural domain and retrieved pages with MAP of 95%. MT systems produce high quality translations only in limited domains and are very expensive too. It involves the cost of creating bilingual dictionary, parallel corpora and the construction and evaluation of MT system. R. Udupa & J. Jagarlamudi [18] used Probabilistic translation lexicon produced by Statistical Machine Learning while A.Imam & S. Chowdhury [19] used SMT using parallel corpus for English to Bangla translation. Table 4 describes that the authors in past had used different translation approaches, size of the data set for evaluating their cross language information retrieval system. We here just showing their results, whatever translation approaches and their test dataset they used to evaluate the cross language information retrieval system.

Table 4. Result Analysis of CLIR Techniques for Indian Languages

Languages	Translation	Size of test data/ Performance	Specific Features
English to Hindi Larkey & Connell [16]	Probabilistic dictionary derived from parallel corpus	41697 Hindi news articles/ method contributes to effective Hindi retrieval	It combines the ranked lists from the Inquery search and the Language Modeling search to obtain the final ranking of retrieved documents.
Hindi to telugu to English P.Pingali & V.Verma [20]	Bilingual Dictionary	English news corpus of LA Times 1995 containing 113005 documents & 56472 documents from Glasgow Herald of 1995.The system is robust	Simple techniques such as dictionary lookup with minimal lemmatization such as suffix removal is not sufficient for Indian Languages CLIR
Bengali & Hindi to English D. Mandal & P. Banerjee [21]	Machine Translation using Bilingual dictionary	English news corpus of LA Times 2002 containing 135153 documents/ Map for Bengali-English queries is 7.26 & for Hindi-English queries is 4.77	Queries with named entities provided better results as compared to the queries without named entities implying the importance of a very good bilingual lexicon and transliteration tool in CLIR for Indian languages.
English to Hindi & Hindi to English S. Sethuramalingam & V. Varma [22]	Bilingual Dictionary	English corpus consisted of 125,638 news articles from the Telegraph, Calcutta edition while Hindi corpus consisted of 95215 news articles published in Jagran. English-Hindi CLIR performance is 58% while reverse is 5%.	Disjunctive query formulation using weighted keywords give an overall better performance in both CLIR and Multi Lingual scenario

7. CONCLUSION AND FUTURE WORK

Cross-lingual Information Retrieval provides new paradigms in searching documents through multitude varieties of languages across the world. It can be the baseline for searching not only among two languages but also in multiple languages. Today, most of the cross-lingual research involved only few famous languages like English, Hindi, Spanish, China and French. As the world becomes more connected by technology, CLIR in every language is needed. Creating accurate metadata in different languages in documents or good translation of key information in documents can help improve the quality of the index and retrieval. CLIR is a multidisciplinary area that has been increasingly gaining more attention from the research community. Despite recent advances and new developments, there are still many aspects to be explored. Other areas for investigation include ease of use, the use of relevance feedback, the effect of more extensive use of concept relations and possibly experiments with larger data sets. In Indian context, this is one of the hotspots of linguistic diversity (350 languages) in the globe. The fact that a dominant language of one region may be a language of a linguistic minority in other region. CLIR systems would play a very important role in allowing the people to go through the documents and literatures of other languages thus breaking the language barrier. We work out here to give a broad overview in the field of CLIR by exploring its aspiration, difficulties, basic tools, major works and future research goals. In reviewing this information, it becomes possible to gain a larger picture of the CLIR field and building an experimental English-Odia CLIR system.

REFERENCES

[1] Oard D, and Diekema A. CLIR Annual Review of Information Science; Vol. 33, Williams, M. (ed.), Information Today Inc., Medford, NJ; 1998.

[2] Oard D. and Diekema A. Cross-Language Information Retrieval, In: Williams, M. (ed.), ACM, New York, NY (1999).

[3] Ari Pirkola et.al, Dictionary Based CLIR: Problems, Methods, and Research Finding. Sep. 2001; Volume 4: pp 209-230.

[4] Mustafa Abusalah, John Tait & Micheal Oakes. *Literature review of Cross language information retrieval;* 2007.

[5] Chinnakotla, M.K. et al. *Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation.* Dept. of CSE, IIT Bombay, India; 2008.

[6] Sujoy Das et.al, *Disambiguation Strategies for English-Hindi Cross Language Information Retrieval System*; 2009.

[7] Abusalah, M. et.al, *CLIR using Multilingual Ontology as Translation and Query Expansion Base*; September 2009.

[8] Nurjannaton Hidayah Rais, Muhamad Taufik Abdullah, Rabiah Abdul Kadir. *Query Translation Architecture for Malay-English Cross-Language Information Retrieval System*; 2010.

[9] Varshney S. & Bajpai J. *Improving performance of English-Hindi CLIR using transliteration of query terms;* 2013.

[10] Warrier R, Sharvari S. Govilkar. *A Survey on Various Clir Techniques. International Journal of Technical Research and Applications* e-ISSN: 2320-8163:Volume 3, Issue 2,Mar-Apr 2015; PP. 152-156

[11] Sanjay K. Dwivedi and Ganesh Chandra. A Survey On Cross Language Information Retrieval, *International Journal on Cybernetics & Informatics (IJCI):* Vol. 5, No. 1, February 2016.

[12] Levow, G., Oard, D., and Resnik, P. *Dictionary-based techniques for cross-language information retrieval, Information Processing and Management; 2004.*

[13] Chandra, G.; Dwivedi, S.K., "*A Literature Survey on Various Approaches of Word Sense Disambiguation,*" in *Computational and Business Intelligence,* 2014 2nd International Symposium on, vol.: no., pp.106-109, 7-8 Dec. 2014.

[14] Diekema, Anne R.,"Translation events in cross-language information retrieval: lexical ambiguity, lexical holes, vocabulary mismatch, and correct translations",2003.

[15] Seetha A, Das S, & Kumar M. *Evaluation of the English-Hindi CLIR System Based on Dictionary Based Query Translation Method.* In Information Technology, ICIT 2007; 10th International Conference on IEEE, pp. 56-61.

[16] Larkey L.S, And Connell M.E. Structured queries, Language modeling, and relevance modeling in cross-language information retrieval; 2003.

[17] Thenmozhi D, and Aravindan C. Tamil-English Cross Lingual Information Retrieval System for Agriculture Society. *International Forum for Information Technology in Tamil Conference*; October 2009.

[18] R. Udupa, J Jagarlamudi and K Saravanan. *Hindi-English Cross-Language Information Retrieval. Microsoft Research India at Fire;* 2008.

[19] Imam A, S. Chowdhury. English to Bangla CLIR using parallel corpus and SMT. *International Conference on Computer and Information Technology (ICCIT)*; pages 566 – 571, Bangladesh 2011.

[20] Pingali P, Varma V. *Hindi and Telugu to English Cross Language Information Retrieval at CLEF 2006;* In Working Notes of Cross Language Evaluation Forum.

[21] Mandal D. et al., *Bengali and Hindi to English Cross language Text Retrieval under Limited Resources*; 2008.

[22] Sethuramalingam S, & Varma V. *IIIT Hyderabad's CLIR experiments for FIRE-2008; In The working notes of First Workshop of Forum for Information Retrieval Evaluation (FIRE)*, Kolkata; 2008.