WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Cross – Language Information Retrieval

[a]Dr.Zafar Ali Khan, [b]Tejaswini Madaiah Bhojan, [c]Srivika Srinivas Nunavathu

[a]Associate professor, program head-CSE, Department of Computer science & engineering, school of engineering, Presidency University, Itgalpur, Rajanakunte, Bengaluru, Karnataka. Email: zafaralikhan@presidencyuniversity.in, jafar.trichy@gmail.com Mob: +91-7338021015.
[b]Email: mbteju411@gmail.com, mobile: 9591082456
[c]Email: nsrivika@gmail.com, mobile: 8088102807

DOI: https://doi.org/10.55248/gengpi.2022.3.6.13

ABSTRACT:

In today's growing nation, local database storage and retrieval is crucial for developing countries. A huge amount of information on Internet is available in numerous language that can be access by anybody at any time. Data Retrieval systems are fundamentally linguistic: the content or context of documents must be described, and the inquirers needed for documents must be expressed. These descriptions and expressions are most frequentlyarticulated in free or controlled vocabularies that have some of the same characteristics as natural language. Therefore, the processes of document description or request formulation must be strongly related to the processes of description and inquiry in natural language. Natural language processing is used to (a) preprocess the documents to extract content-carrying terms, (b) discover inter-term dependencies and build a conceptual hierarchy specific to the database domain, and (c) process the user's natural language requests into effective search queries.

Since we assume that the information is primarily encoded as text, IR is also a natural language processing problem: to decide if a document is relevant to a given information need, one needs to be able to understand its content. CLIR deals with asking queries in one language and demand for retrieving documents in another language.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.

KEYWORDS : Cross language information retrieval, query translation, document translation, multiple–meaning words, compound words, dictionary-based translation, machine translation.

1.INTRODUCTION:

A classic IR system accepts the user information need in a form of query and gives back the documents that are relevant to the user need. With the explosion of knowledge on the web, it became necessary to break the language barriers for the monolingual IR systems. This may allow the users of IR systems to give query in one language and retrieve documents in different languages. IR system, with different source and target language is called CLIR system. CLIR enables the users to recover the set of documents different than the language of the query. It allows the userto enter their query in one language and regain the set of documents in the other languages. The main advantage ofCLIR is that the user can search the information without limited by the linguistic barriers. In Cross-languageinformation retrieval the language of the query is different from the language of the document. CLIR system is asystem in which a user is not restricted to only one language, it can formulate query in one language and then systemreturns the documents in the other language, Since in CLIR the language of query and the documents both can betranslated. CLIR requires a translation phase along with the usual monolingual retrieval phase. For this purpose, existing CLIR systems adopt various techniques explored in natural language processing (NLP) research. In brief, bilingual dictionaries, corpora, thesauri, and machine translation (MT) systems are used to translate queries or/and documents. CLIR same as the bilingual system simplifies the searching process for multi-lingual users and enablesthose peoples who know only single language to provide queries in their language and then get benefits fromtranslators for retrieving the documents of the other languages. In this paper, we propose a Hindi/English CLIR system for technical documents, focusing on translation of technical terms. As with MT systems, existing CLIR systems still find it difficult to translate technical terms and proper nouns, multiple–meaning wordswhich are often unlisted in general dictionaries.

2. SOME USE CASES:

Cases for the implementation of CLIR systems can be divided into two groups: (1) cases there the user does not speak the language of the documents, and (2) the circumstances there the user understands the documents but prefers to use a different language their question. These application conditions,

as well as other factors such as language correlation, are the expected number of queries and then influence the design choices that need to be considered in the development of the CLIR system.

a) In Web search, the natural emergence of one or more language (French/Greek/Hindi) causes differences when most information is in a small number of languages, but most users prefer to interact with the Web in another language. In such cases, the CLIR may provide access to information that would otherwise not be available.

b) Recall-Oriented CLIR application, which includes searches for medical records (147,75,89,88,93,11), patents (106) and search applications (30). End users of programs are often the most important professional for them to find all the documents on a defined topic. In such cases, a summary of the document in the language in question may be helpful.

c) The use of CLIR can be expected to be particularly important in regions where multilingualism is commonly used, such as countries with more than one official language. This may include English / French in Canada, Dutch / German in Belgium, 23 official languages in India. Users of CLIR programs in such settings may be able to understand multiple languages, but may be better able to, or more comfortable, use one language.

d) Questionnaire (QA) where instead of simply identifying the relevant documents to the user, the system seeks to use the information from the relevant document to provide a direct answer to the question

3. CORE TECHNOLOGY OF CLIR

The CLIR study came at a time when it was an important question how questions and texts presented in different languages can be done represented equally. Because that work is still shaping our vision for CLIR, we are starting to introduce how that is done. We organize our story into three main questions: (a) ‘what should be translated?’, (b) ‘what possible translations?’, And (c) ‘how should those translations be used?

a) What to Translate?

For CLIR, the question of what to translate exists at four levels. First we could ask what things should be translated — questions, scriptures, dual translation or none?

a.1) The Query or The Documents?

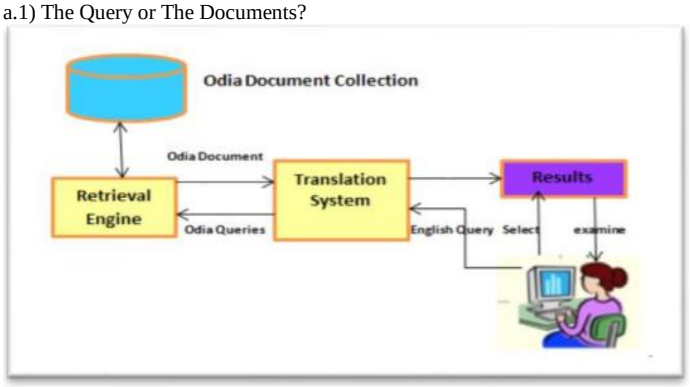


Figure 1. Query Translation Model

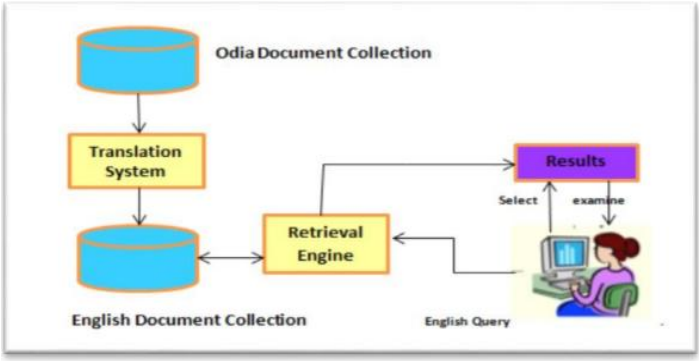


Figure 2. Document Translation Model

We divide the existing CLIR into four options regarding the implementation of the translation phase. The first method translates questions into the language of the document. The second method translates documents into the language of the query. The third method conveys both questions and documents in the representation of multilingualism. The fourth method does not convey questions or texts.

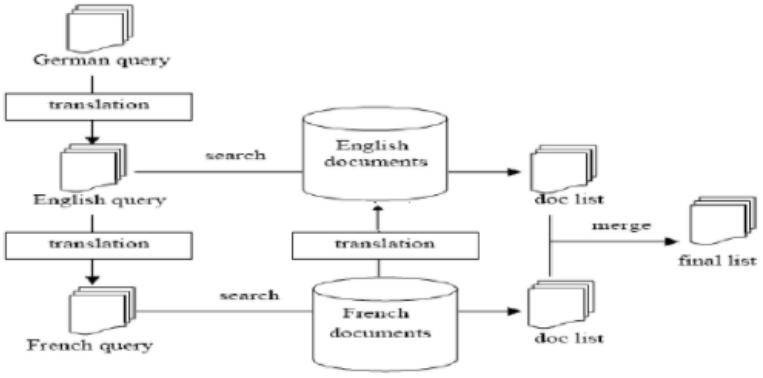
It seems rather obvious the query that should be translated, since queries are short, and documents are long. Indeed, query translation is widely used in CLIR experiments, specifically because it is efficient. But efficiency considerations may come out differently when the query workload is very high, as is the case in Web search. On the other hand, query translation has clear advantages when the goal is to cull through streaming content for documents that satisfy a standing query. The point here is that common sense in interpreting questions effectively may be true or false — much depending on the details of the job

Interpretation of a question is incorrect: In most cases the question does not provide enough context to automatically find the intended meaning of each

term in the question. Translation errors affect retrieval performance sensibly. When searching for a multilingual website, the query should be translated into each language

Document Translation can be the most desirable scenario in CLIR, if the purpose is to allow the users to search the documents different from their own language and receive results back in user's language. Document translation has another potential advantage beyond the (relatively rare) cases in which it is the more efficient choice. The words in a document occur in sequence,and we have good computational models for leveragingsuch sequences to make more accurate translations. Modern translation models are trained using paired sentences from paired documents, and pairings of documents with their translations are much more widely available than examples of queries paired with their translations. If people wrote queries like they wrote documents, this would not be a problem. But they don't. As a result, even when query translation would be more efficient, document translation might be more accurate. The document translation method has some advantage over the translation of the question. These include the following: (i) The longer text provides additional context for rendering translation, so that targeted language terms are selected more accurately. (ii) Translation errors should not hurt to retrieve too much, as they weigh on the entire document. (iii) The translation effort is made at the time of identification, thus obtaining immediate retrieval during operation. However, there are some problems with text translation, such as: (i) It takes a lot of calculation effort to identify clusters. (ii) Bad estimates made when there are more than two languages.

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



The third option is to translate both questions and documents. In this way - both questions and documents are translated into the same representation. This method requires additional storage space for translated texts but provides a measure when the same set of documents is required in multiple languages. A dual translation method also called a mixed translation method can be done in pivot language. Direct translation between two languages may not always be possible due to limitations of translation resources. To do this kind of translation, a third language tool or language is required between these languages, called pivot language. In this process, two types of methods are possible: a question or document is first translated into a pivot language, and then into the target language. For example, if we want to use Japanese queries to search Hindi content, we can translate both Japanese and Hindi into English, simply because there are additional Japanese-English and English language resources available. Hindi- English rather than Japanese-Hindi. Here in a sense, we have interpreted the question and the documents as invisible representations instead of representations associated with any one language.

We should note that non-translation at all is the fourth opportunity. If the character sets are similar, sometimes the appropriate words or loan names will match. While this is by no means a good idea in itself (except for the possible differences in vernacular languages such as British and American English), the existence of such "translations" can serve as important points for combining complex and efficient methods. they can be built.

a.2) Which Terms?

When we think of translation, we often think of translating words. But the meaning of the word is smooth. Is the poet a word? If so, are poets a different word, or is there something else (plural of a word)? When we think of words as intangibles in nature it means that poets and poets and even the poet are still many forms of the same word, and the poet is its root. But people who acquire knowledge have a much simpler view of the world. We do not pronounce what we use in our search by name, but we do pronounce words. And the name is whatever we choose to identify in the IR system. For example, terms may be capitalized or have lowercase letters, diacritic marks may be retained or removed; infectious characters can be kept as they are or made normal, and shortcuts can be removed or saved. Decisions made on each of these points are often different from those used where one-read-only translation is the goal because information retrieval activities (and thus retrieval measures) often place greater emphasis on memory (i.e., reducing falsehood) than performing a single machine translation function better. So, for CLIR, goals are not just what you point to; they are also what translates. However, that this introduces misunderstandings, as longer terms limit translation ambiguity, and shorter words provide more space to improve memory.

b) Which Translations are Possible?

Once we have identified which words, we need to be able to translate, our next questions are what possible versions of each term are possible, and (expand that) how likely is each translation to have the intended meaning?

b.1) We may look up the translation in another dictionary made by hand. In CLIR, this is called dictionary-based translation (regardless of whether the dictionary was developed from a human-readable dictionary). In most cases, this process is applied to questions, and is thus called Dictionary-Based Query Translation (DBQT). In fact, there are many types of handwritten dictionaries that can be mined; among the most common are bilingual grammar lists, bilingual dictionaries (containing not only word order but definitions, parts of speech, and examples of usage), multilingual thesauri (thus re-using technology from the first wave), multilingual ontologies (e.g., page articles in multiple languages dug into Wikipedia)
b.2) Lastly, the real source of all information is interpreting human intelligence, so if all else fails we can turn to the user for help. For example, the interactive system may provide limited evidence of possible translation to engage the searcher in a conversation about how best to translate the query name [134], or the systems may display search captions that include non-translated words in the hope that the searcher will be able to translate the word in context.

c) How to Use those Translations?

Most of the original work to translate a query-based query was to replace one word for the language of a document in each term of the question language. Since most words have the same name, 6 meaning they have more than one very different meaning, it naturally led to the question of which versions should be used. The usual method was to look up other words in the question to separate the intended meaning of each question term. However, one word may have both distant and closely related translations (e.g., the Hindi word jal can be translated into English as heat or water), so a better question (at least if more than one translation seems logical) would be which translations. should be used. That naturally leads to the question of how a few possible translations of a single word should be used, and that simply becomes a common question (as one well-known translation is simply a special case of that common question).
Initial attempts to use multiple translations, such as combining all the translations into one very long question, proved to be problematic. Yet even today it is not uncommon to see interpretations of unbalanced questions being misused as a basis for CLIR testing.

4. DIFFERENT TECHNIQUES FOR CLIR

A) Direct Translation -

i). Machine Translation - Machine translation, in simple terms, is a technique that makes use of software that translates text from one language to another language

a)1 Rule Based Machine Translation - Rule based MT uses linguistic information about source and target language. They propose a word-for-word translation system using subject verb agreement b) Statistical Machine Translation - Statistical machine translation generates translations using statistical methods based on bilingual text corpora.

ii). Dictionary Based CLIR - The most natural approach to cross-lingual IR is to replace each query term with most appropriate translations extracted automatically from Machine Readable Dictionaries (MRD). The translation using bilingual dictionaries is simple but Ballesteros and Croft (1996) and Hull & Grefenstette(1996) claim that it leads to a 40-60% loss in effectiveness as compared to monolingual retrieval.Loss can be due to factors as untranslatable search keys due to limitations in dictionaries, processing of derived or inflected word forms, phrase and compound translation and lexical ambiguity in source and target languages. To handle these problems, researchers have made use of domain specific dictionaries for the dictionary coverage problem, Stemming and morphological analysis to handle inflected words, POS tagging for phrase translation, corpus-based query expansion and query structuring for the ambiguity problem.

iii). Corpus Based Cross Lingual Information Retrieval - Corpus based CLIR methods use multilingual terminology derived from parallel or comparable corpora for query translation and expansion. There are two types of corpus:

a) Parallel Corpus - A parallel corpus is a collection where texts in one language are aligned with their translations in another language.

b) Comparable Corpus - Comparable corpus consist of texts that are not translations but share similar topics. They can be, e.g., newspaper collections written in the same time period in different countries.

B) Indirect Translation –

Indirect translation is the most common solution if any lack of resources that support direct translation. Indirect translation depends on the use of the presenter is placed between the source query and the target document collection. In the case of flexible translation, the question will be translated into content to work compared to the target text set.

5. CHALLENGES IN CLIR

Questions from users are usually very short, which creates more ambiguity in translating the questions, and reduces the accuracy of the results of retrieving different languages. As the problem of linguistic diversity in CLIR is more serious than monolingual IR, it is necessary to use strategies to improve multilingual recovery. In CLIR programs, users often present their query in their native language, and the program automatically searches texts written in other languages. Therefore, it is a challenge for CLIR to overcome the barrier between source language (SL) in question sentences and target language (TL) in documents to be searched. As discussed in the previous section, many CLIR systems use MT technology to solve this problem. As MT research itself has many challenges (such as accuracy), CLIR research also addresses critical issues and challenges that need to be addressed.

1) Ambiguity - The ambiguity occurs when words have multiple meanings also called homonymy or polysemy.Homonymy refers to a word

which has at least two entirely different meanings, for example the word "left" can either mean opposite of right or the past tense of leave. Polysemy refers to a word which can take on two distinct but related meanings such as the "head" of the family or the human's "head". So there becomes a problem when finding the most appropriate translation from various choices in the dictionary. Very frequently, translation of a word results in such a choice having to be made.

2) Using dictionary-based translation is a common practice in different IR systems but a significant decrease in performance is seen when the questions are contained.
words or phrases that do not appear in the dictionary. This is called the Out-of-Vocabulary. This is also expected of the best dictionaries. User input questions are usually shorter, and the extension of the query cannot help to replace missing words due to missing information.
3) Language structure - The challenge comes when language is complex. The English sentence includes subject, verb and object. In some languages, that is the order of the day. In Persian for example, a sentence is made up of a verb and an object before a verb. In some languages, such as Arabic, the pronoun is written as part of the verb, which determines the wording of the sentence and the gender of the subject.
4) Compound words are made up of two or three nouns and adjectives combined. There are three types of compound words:
Closed form (like keyboard, notebook, childlike, firefly)
Hyphenated form (like six-pack, mass-produced, over the counter)
Open form (like attorney general, full moon, real estate, post office)
Some words are ambiguous in expressing their meaning, such as under, afternoon, beach. However, there are compound words that mean only part of the word, such as bookworm and bellboy. There are also words that give a completely different meaning, such as in the case of a butterfly, which could mean butter or fly, or a term that has no relation to line or death.
5) Missing terms - There are languages where the names of certain things or actions are not present. In the United States, for example, an extra room in a house where guests are allowed to sleep is called a guest room. In Greek, such a room is called a ksnona, and in Italian, a guest room is called as camera per gliospiti.
6) Specialized Terminology and Proper Nouns - Special terms, such as scientific terms, are often difficult to translate and are often found in specialized dictionaries or term banks. Special words are usually less confusing than ordinary vocabulary although general vocabulary can have a special meaning when used in a particular subject

6. COMPARATIVE ANALYSIS OF CLIR AND OBSERVATION FOR INDIAN LANGUAGES

Recovery of multilingual knowledge of foreign languages such as English, French, Chinese etc. it has been an attractive place for researchers since time immemorial. But Indian languages have attracted attention only a decade ago. The work done by the researchers shows mixed results regarding the development of monolingual language retrieval from the Indian language perspective. Anurag Seetha and S. Das (2010) translated from the 2010 Fire 2010 test using Shabdanjali dictionary and extension queries via Hindi Wordnet. The method did not seem to work. This is because standard dictionaries have a lower writing problem. To eliminate this inefficiency Larkey and Connell (2003) used a probabilistic dictionary taken from the parallel corpus to translate English into Hindi and was obtained.successfully retrieving different languages. Similar or comparable companies have other useful CLIR resources. Parallel corpora are selected from CLIR because they provide more accurate translation information but due to their scarcity, similar companies are often used in CLIR.The above observation concludes that there is a wide range of research to develop existing algorithms or to develop new ones to improve the performance of the CLIR system.

7. APPROACH

In this section we suggest how to access multilingual information on the web and briefly discuss the components of the proposed design. The main components of the project are: Pre-processing, question translation, Word sense misunderstanding and retrieval of information.

A). Grammatical Complexities of English to Hindi - Translation

i) The basic word order in Hindi is Subject-Object-Verb (SOV) as against SVO word order in English. But in Hindi, the constituents of a sentence can be freely moved around in the sentence without affecting the core meaning

ii) Unlike English, vowel length and vowel nasalization are present meaningful in Hindi e.g.

(Kam) means "little" and (Kaam) means "work"

(Puuch) means "ask" and (puunch) means "tail"

iii) In English, extensions precede related words. In Hindi, such words are called postpositions because they follow the ruling words.

iv) There are no articles in Hindi. The definition of noun by is indicated by pronoun, context, or word order

v) All nouns in Hindi are masculine or feminine. This means that unreasonable sex is given to middle-class nouns in English e.g., "chair" is for a woman noun and "door" is a male noun in Hindi.

B). Preprocessing - The first step in any CLIR system is to process in advance the wording of the queries to speed up the translation process without compromising the quality of retrieval. This process is performed using a tokenization, stemming and stop word removal

i) Tokenization - Tokenization is a way of separating a piece of text into smaller units called as tokens. Here, tokens can be words, characters, or sub words. For example, consider the sentence: "Never give up".

ii) Stemming- It puts all the different types of word in one stem. For languages such as English with weak inflections, simple stemming algorithms can

be used. Such algorithms only remove plural conclusions. In languages with strong inflections, suffices are joined to the stem end to end. The advanced stemming algorithm can detect so many conclusions and delete them repeatedly. Porter Stemmer, Snowball stemmer etc. are well known for the advanced cutting algorithms.
iii) Stop Word Removal - stop words are those words which can be ignored in search query to increase the performance. example of short words are common functional short words like a, an, the etc.

C). Query Translation - In Query Translation, the query provided is translated from Source language into the target language and the query is found on the Web site to find the text in Target language. Interpretation Question often has a problem with translation ambiguity and this problem increases due to the limited number of contexts in the short questions. Interpretation of the question can be done using any one method which includes machine translation, dictionary-based or chorus-based method.

D). Ambiguity Removal in Translated Query–ambiguity is a common problem in all native languages i.e., there are many words in these languages that have more than one meaning. For example, an English noun 'plant' can mean a green plant or industry, or the word 'bank' means a financial institution or a river. The correct meaning of a word can be chosen based on the context in which it occurs. This function of automatically giving the most appropriate meaning to a polysemous word within a particular context is called word sense disambiguation.Word sense disambiguation can be solved by analyzing the co-occurrence of the word with other words in query that is needs to be translated. Disambiguation algorithms use a variety of resources and follow different strategies. Based on the use of resources and their processing methods, techniques can be classified as Knowledge-Based Methods (resources used Machine-Based Dictionaries, Thesaurus, Dictionaries), Supervised Reading.
In the sense of the word disambiguation, the meaning of the word is defined based on the company that holds it i.e., based on the words it originates. Similarly, the words in question provide an important idea for choosing the right translation / translation, albeit a small number. For example, in the question "nadi jal", here the translation of nadi refers to {river} and the translation of jal means {water, to burn}. Here, based on the context, we see that the choice of a second word translation is water as it is more likely to co-occur with river.

8. PROTOTYPE APPROACH

QUERY 1:WE MADE IT TOGETHER

A and B). Grammar Complexities -

i) Tokenization - 'We' + 'Made' + 'It' + 'Together'is the tokenized form by using whitespace between words.
ii) Stemming - Next using the Porter stemmer, the altered tokens are reduced to their base. But in this query stemming is not possible.
iii)Stop Word Removal – Removing the word 'it' using stop word list given by MIT. Query formed now is We Made Together.

C). Query Translation -

Hindi Translation of the query is हमनेएकसाथबनाया(hamane ek saath banaaya)

D). Ambiguity removal–
Hindi translation of the query isspecifiedi.e. it has only onesense.Therefore, the query is translated.

QUERY 2: Mother India

A and B). Grammar Complexities

i) Tokenization - 'Food' + 'India' is the tokenized form by using whitespace between words.
ii) Stemming - Next using the Porter stemmer, the altered tokens are reduced to their base. But in this query stemming is not possible.
iii)Stop Word Removal – No stop words available.

C). Query Translation -

Hindi Translation of the query isभारत माता(bhaarat maata)

D). Ambiguity removal-
The Hindi translation **"भारत"** is ambiguous i.e., it has multiple senses. It refers to country 'India' as well as a 'Name' (e.g., 'Bharat' is eating). The correct sense of a word can be identified based on the context of the query in which it appears using disambiguation algorithm. The correct translation of the word is related to the country India.

8. Conclusion:

Cross-lingual IR provides a new way of searching for texts using dozens of languages around the world and can serve as a basis for search not only between two languages but also for multiple languages. Today, most multilingual research includes a few popular languages such as English, Hindi, Spanish, Chinese, and French. Linguistic research has boosted the country's development. As the world is more technologically connected, a multilingual IR is needed for all languages. CLIR is a multidisciplinary institution that has been gaining more attention in the research community. Despite recent developments and new developments, there are still many things to consider. It is also made clear that respective work regarding Indian

languages has gained momentum over the last decade and there is more to come in this field. It is very clear from the observation that there is still a scope of improvement in the performance level of CLIR.

BIBLIOGRAPHY:

1]https://link.springer.com/content/pdf/bfm%3A978-94-017-2388-6%2F1.pdf
2]https://www.researchgate.net/publication/356159656_Cross-language_Information_Retrieval
3] https://www.sciencedirect.com/science/article/abs/pii/0306457394000558
4] https://www.cdac.in/index.aspx?id=mc_cli_cross_language_info
5] https://aclanthology.org/W99-0605.pdf
6] https://arxiv.org/abs/2112.13510
7] https://arxiv.org/abs/2111.05988
8] http://disi.unitn.it/~bernardi/Courses/DL/clir.pdf
9] https://en.wikipedia.org/wiki/Cross-language_information_retrieval
10] https://translate.google.co.in/
11]https://medium.com/lily-lab/a-brief-introduction-to-cross-lingual-information-retrieval-eba767fa9af6
12]https://www.google.com/imgres?imgurl=https%3A%2F%2Fmiro.medium.com%2Fmax%2F1400%2F1*tZwZAGNEue_zN_BRIPvuvA.png&imgrefurl=https%3A%2F%2Fmedium.com%2Flily-lab%2Fa-brief-introduction-to-cross-lingual-information retrievaleba767fa9af6&tbnid=3LHfz9Zf4xo3OM&vet=12ahUKEwispcXZiqv4AhXxkdgFHWC_BCsQMygCegUIARCkAQ..i&docid=B8M6BNq8LE7KNM&w=1400&h=987&q=cross%20language%20information%20retrieval&ved=2ahUKEwispcXZiqv4AhXxkdgFHWC_BCsQMygCegUIARCkAQ
13]https://uhn.ac.id/files/akademik_files/2004241144_2017_Journal%20of%20English%20Teaching%20as%20a%20Foreign%20Language%20Vol%203%20Issue%201%20June%202017_5.%20BERNIEKE%20DAMANIK%20%20Loan%20Words%20Used%20in%20the%20Students%E2%80%99%20Translation%20of%20English%20into%20Indonesian.pdf
14] https://www.geeksforgeeks.org/introduction-to-stemming/
15] https://www.cs.cmu.edu/~ref/mlim/chapter2.html
16] http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-90442016000200033
17] https://dl.acm.org/doi/pdf/10.1145/2379776.2379777
18] https://www.morganclaypool.com/doi/abs/10.2200/S00266ED1V01Y201005HLT008
19] https://digitalcommons.usu.edu/itls_facpub/18/
20] https://www.degruyter.com/document/doi/10.1515/9783110235005.399/html
21] https://surface.syr.edu/cgi/viewcontent.cgi?article=1018&context=istpub
22] https://www.microsoft.com/en-us/research/wp-content/uploads/2007/09/2007_CLEF_CLIRinIndLanguages.pdf
23] http://www.iro.umontreal.ca/~nie/IFT6255/Books/CLIR.pdf
24] https://paperswithcode.com/paper/query-translation-for-cross-language
25] https://www.igi-global.com/dictionary/cross-language-information-retrieval-web/6275
26] CROSS LANGUAGE INFORMATION RETRIEVAL: IN INDIAN LANGUAGE PERSPECTIVE Pratibha Bajpai1, Dr. Parul Verma2
27] Seetha Anurag, Das Sujoy, Kumar M., Evaluation of the English-Hindi Cross Language Information Retrieval System Based on Dictionary Based Query Translation Method. In: Proceedings of 10th International Conference on Information Technology (ICIT2007). Available at http://doi.ieeecomputersociety.org/10.1109/ICIT.2007.40.
28] English to Hindi A. Seetha, S. Das, J. Rana & M. Kumar (2010)
29] English to Hindi Larkey & Connell (2003)
30] Query Translation in Chinese-English Cross-Language Information Retrieval Zhang Yibo, Sun Le, Du Lin, Sun Yufang
31] Ballesteros, L. and Croft, W. B.(1996). Dictionary based methods for cross-lingual information retrieval. In Proceedings of the 7 '~ International DEXA Conference on Database and Expert Systems Applications, pp.791-801