International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-4)

Research Article | **April 2015**

Refinement of Search Results of the Google using Cross Lingual Reference Technique and GPS

Monika Sharma, Dr. Sudha Morwal
Department of Computer Science,
Banasthali University Jaipur,
Rajasthan, India

Abstract—

We present a system which improves the search result of the Google using the Cross language information retrieval (CLIR) and Location based web search. Cross-language information retrieval Cross-language information retrieval (CLIR) is type of information retrieval which deals with the retrieval of the information written in a language other than the language of the user's query. Location based web search is searching of web content or the information available on the web according to the geographical location. Google is a widely used search engine that uses the keyword based matching to find the documents that match with the user's query. It will retrieve the set documents as a search result that contains query terms. Google is a keyword based search engine. It performs the Keyword matching by checking the keywords in the Google's database and retrieves the documents containing the query term in the same language as the language of the query, but it does not retrieve the equivalent words of the query terms in other native languages. For e.g. if the location of the user is "Maharashtra "then it will give the search result in "Marathi language". Search words have multiple meanings or appear in multiple contexts or in multiple languages.

Keywords— Information Retrieval, World Wide Web, CLIR, GPS

I. INTRODUCTION

"Refinement of Search results of the Google using the Cross Lingual Reference and GPS Technique" improves the search results of the Google by providing the location based web search. Based on the current location of the user or the hosting device, the native language of the user will be determined.

Current location of the user will be obtained with the help of Google's GPS service i.e. Google Geolocation API. Now native language information retrieval will be implemented with the help of CLIR. This System aims to perform the location based search that uses the equivalent word of the query terms in the native languages based on the current location of the user to provide the highly relevant search results of the Google. We include different Indian native languages such as Hindi, Bengali, Telugu, Marathi, Tamil, Gujrati, Kannada, Punjabi and Urdu. This system allows the users to retrieve documents in their own native languages rather than the original query language.

A. Cross Lingual Information Retrieval

CLIR enables the users to retrieve the set of documents other than the language of the query. It allows the user to enter their query in one language and retrieve the set of documents in the other languages. CLIR is beneficial for the user in searching the information without limited by the linguistic barriers [1].

Cross-language information retrieval is a type of information retrieval in which the language of the query l is different from the language of the document. In CLIR system a user is not restricted to only one language, he can formulate query in one language and then system returns the documents in the other language. Since the language of query and the documents to be translated in CLIR. CLIR system simplifies the search process for multilingual users and enables those peoples who know only one language to provide queries in their language and then get help from translators for retrieving the documents of the other languages [2].

B. Location based Web Search

Location based web search is searching of web content or the information available on the web according to the geographical location. Location based web search take your geographic location into consideration and it will give the search result of the Google according to the native language of the current location [3].

The relation of web information to a geographical location has gained much attention. Location based web search will provide the search result based on the current location of the user. Search result will depends upon the present physical location of the user. In geographic or location-based web search, the search results must not only be related to the term in a query but also must be geographically related to a current location which is associated with a query. Location-based search engines associate locations to the information available on the web in order to answer queries [3].

C. How Google Search works

Google is the most popular search engine on the World Wide Web handling more than three billion searches every day. Google uses the keyword matching to find the documents that are relevant to user's query. Google checks each of keyword in its database and then retrieve the list of documents that contains keyword. Google runs on a distributed network of thousands of computers and can perform fast parallel processing. Parallel processing is a method of computation in which many calculations can be performed simultaneously, significantly speeding up data processing [6]
The Keyword based search does not provide the relevant information to the User. There is a requirement of "Cross Language Information Retrieval" that enables the user to search the query Keyword in Multiple languages. Search result of the Google can be refined by using CLIR.

D. Limitations of Google Search

Google is keyword based search engine and it uses the keyword matching to retrieve the documents that contain the keywords in query. Google will not provide the search result of the query in terms of the geographical location. The searching of the query on Google does not depend upon the physical location of the user or the hosting device. It will give the result of search only in English language but not in the native language of the user based on current location. This will give irrelevant search results to the native language users. So it will provide the location based information retrieval according to the native language of the user. Search result of the Google has been refined by using CLIR and location based web search.

E. Refinement of Google Search Result

Google will perform the keyword matching of the query to provide the search results. It will provide the search result of the query term in only English language .Google will not provide the search result of the query in the native language of the user. But if the search result will be given in native Indian languages based on the location of the user then it will be more beneficial to the user .Users will get the required information in their own native languages so it becomes easier for the user to understand the web content in native languages. By using the Location based web search means providing the search result on the basis of the geographical locations will improve the searching of Google.
CLIR is helpful for providing the search result in the native language of the user. CLIR does not limit the search result to keyword matching but it will also provide the different language documents having equivalent word to the search query as a search result. The location based web search is implemented with the help of CLIR approach. The World Wide Web made it possible to access the information resources available without any geographical limits and boundaries. Many of the documents on the www are available in multiple languages which are designed for local native language users. The location based search provides more user friendly approach which uses geographical location and provides the location based information. Firstly the location of user is detected and then provides the information in the native language according to location of user. The English query will be converted into native language query with the help of CLIR.

II. GENERAL APPROACHES

For the refinement of the search result of the Google, we have used Cross Language Information Retrieval and Location based web search to provide search result on the basis of the physical location of user.

A. Approach used for CLIR

CLIR provide the set of documents as a search result in the language different from the language of the query. This can be achieved by using the Query translation approach. Query translation became the widely used technique for the CLIR to access documents in a different language than the query [4]. For query translation, we can use online translation service i.e. Google Translate, train a Statistical Machine Translation system , Machine Readable Dictionaries to translate query terms or make use of large scale multilingual knowledge sources like Wikipedia . CLIR based on Google Translate achieve 90% of the accuracy of monolingual search.
In this project we use Google Translate query translation approach. Online translation can be applied to the query entered by the user. Online query translation can be achieved by using one of the Google Translate API which will convert the query into the native language based on the current location of the user [5].

B. Approach used for Location based web search

To achieve the location based web search we make use of the Geolocation API. Geolocation API will be used for obtaining the current address of the user. The latitude and longitude of the user and hosting device will be obtained by the use of Geolocation API.Then using reverse geocoding the latitude and longitude will be converted back into address [9].

III. PROPOSED SYSTEM

In this project a System has been designed to improve search result of Google using Cross Lingual Cross Language Information Retrieval. We have proposed a methodology for retrieving the search results of the Google using gps based or location based. In this the user will get the result of the Goggle search result in their own native languages automatically. Using this methodology we detect a user's location and provide customized search based on that location.
So location based search will be helpful for users for searching in their own language. Block diagram of proposed system is shown in fig 1.

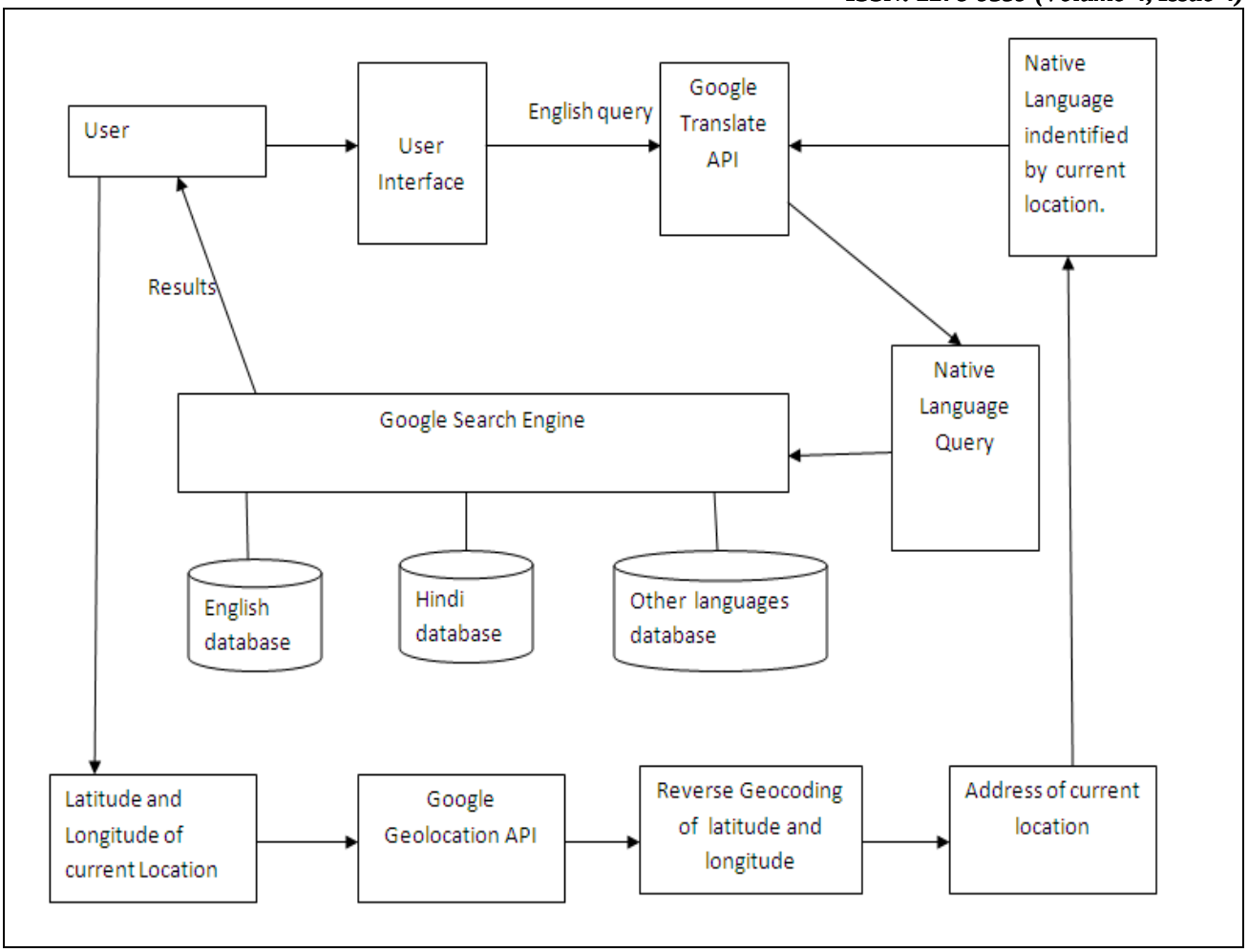

Fig.1. Block diagram of the proposed system

The two modules of the project are GPS Tracking Module and Search Module.

A. GPS Tracking Module

The GPS Tracking module provide the current location of the user. This module will firstly calculate the latitude and longitude of the current location of the hosting device .Then based on the values of latitude and longitude, the address will be calculated. The GPS Tracking module will use the geolocation API for providing the address of the user.GPS tracking module will show full address of the user .then based on the location of the user ,native language of that area will be retrieved .The GPS module will be helpful for performing the Google Search according to locations. Query submitted by the user will then converted into that language and pass to Google search engine then Google will give the search result in that language .So using this project it is possible to obtain the search results according to location so more beneficial to native peoples.

1). *Geographic coordinates Calculation*

Firstly we calculate the Geographic coordinates i.e. latitude and longitude of the current location the geographical coordinates will be obtained with the help of Geolocation API.

getCurrentPosition ( ) method retrieves the current geographic location of the user and the hosting device. The location is expressed in terms of geographic coordinates i.e. latitude and longitude.

For E.g. The retrieved value of the geographic coordinates of the current location are given below

Latitude= 26.913564899999997

Longitude= 75.80108489999999

2). *Reverse Geocoding*

Reverse geocoding is the process of reverse coding of a location having geographical coordinates i.e. latitude and longitude to a readable address. This allows the identification of nearby street address, places, and subdivisions such as neighbourhood, county, state, or country [8].

Reverse Geocoding is performed using the Geolocation API.Using the Reverse Geocoding we convert the value of latitude and longitude into address. Reverse geocoding will convert the set of geographic coordinates latitude and longitude (26.913564899999997, 75.80108489999999) into the address given below.

L-1 To L-4, Panch Batti, C Scheme, Ashok Nagar, Jaipur, Rajasthan 302001, India

3). *Address*

Save the whole address into a text file **Location.tx**t so that it can be used in search module for providing the search result in the native language of the current location.

4). *Language Identification*

Based on the address of the current location, the native language of that area will determined so that the result of search as shown in TABLE 1. The TABLE 1 shows different geographical locations along with their corresponding native languages. User can also select the language according to his requirement in which language he wants to retrieve the search result.

For E.g. the current location is Jaipur Rajasthan so the native language will be Hindi so search module will give the search result in Hindi language.

Current location Address- L-1 To L-4, Panch Batti, C Scheme, Ashok Nagar, Jaipur, Rajasthan 302001, India

Language-Hindi

TABLE 1: Different Geographical Locations along with their native languages.

S.No.	**Location**	**Language**
1.	Bihar, Chhattisgarh, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Mizoram, Rajasthan, Uttar Pradesh, Uttarakhand , Delhi	Hindi
2.	West Bengal	Bengoli
3.	Andhra Pradesh	Telugu
4.	Maharashtra	Marathi
5.	Tamil Nadu, Puducherry	Tamil
6.	Gujarat	Gujrati
7.	Karnataka	Kannada
8.	Punjab	Punjabi
9.	Jammu and Kashmir	Urdu
10.	Arunachal Pradesh, Nagaland, Sikkim, Lakshadweep and others	English

B. Search Module

The Search module enables the user to enter the query in one language and then provide the information about the query terms in their own native languages as a search result of Google. This module will provide the GUI screen where user enters their query in English language and then the system will give result in local language of the current location. as shown in Fig.1.

1).*Query Generation*

User enters the query through the user Interface in English Language.

For E.g. Users search a Query "Pranav Mukharjee".

2). *Query Translation*

This System involves translation of query written in English language to other Indian languages so that the semantically similar pages in the native languages will be retrieved as search result on the basis of current location. Based on the current location and native language of the user obtained by the GPS module, the query translation can be applied to the query. The query entered by the user in English language is translated into the native language of the user so that the user will get result in their native language..

For e.g. the current location of the user is Rajasthan, then this system will convert the English query into Hindi query.

Query-Pranav Mukharjee

Translated Query- प्रणव मुखर्जी

3). *Refined Search Result*

Google will provide the search result of the query terms in English language. Search result of the Google will provide the set of documents containing the "Pranav Mukharjee" terms as shown in fig 2.

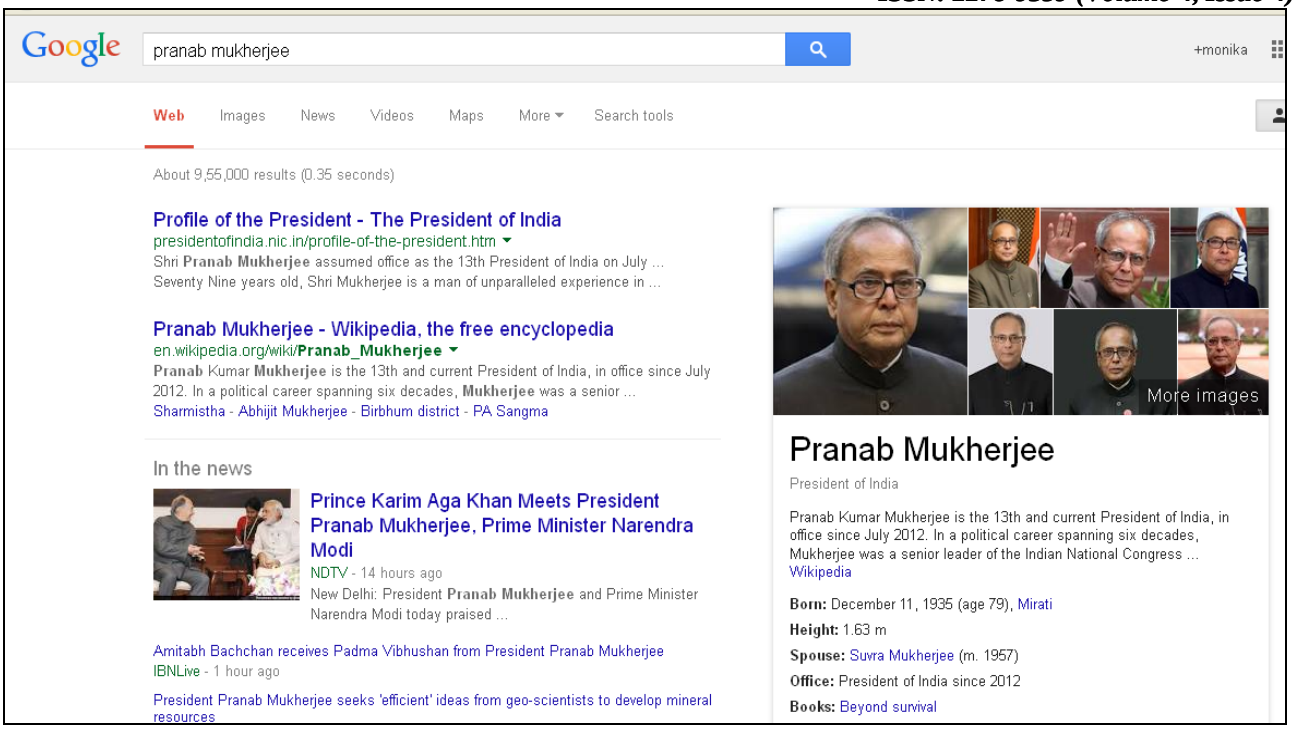

Fig. 2. Search Result of the Google

But the Refined System will provide the list of the documents containing the Hindi equivalent of the query "प्रणव मुखर्जी "of the "Pranav Mukharjee" as a search result. So the user will get the search result of the Google in his native language i.e. Hindi based on current Location i.e. Rajasthan as shown in fig 3.

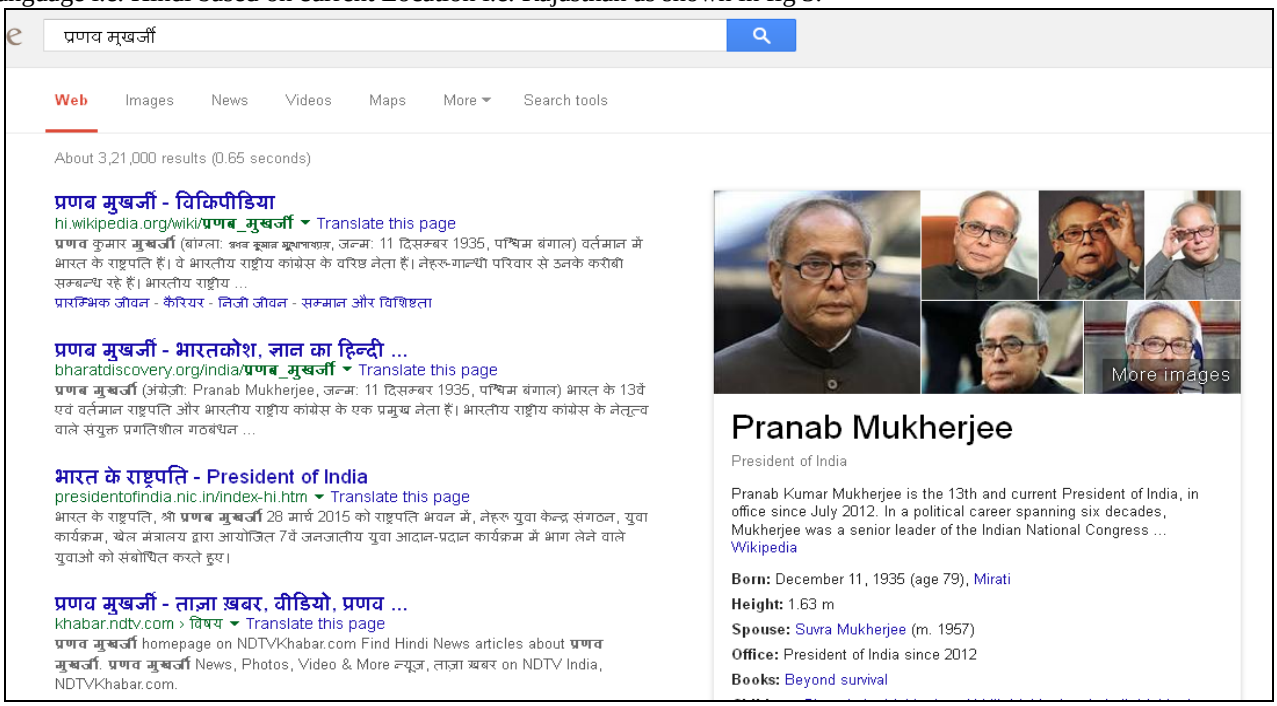

Fig. 3. Refined Search Result of the Google on the basis of Native language

IV. APPLICATIONS

1.This System can be helpful for immigration department. For eg. Immegration department interact with thousands of the Indian native Language speakers which are not able to understand English Languages ,So this project will be beneficial for all those peoples by providing the search result in their native languages.

2.This System can be used for the multilingual population regions so that the people can retrieve the search result in their native languages according to their choice.

3.This system can be used for Tours and Travelling companies. The website of the companies provides the information about the tourism sites to the peoples in their own native languages so that it will be more interactive to peoples accessing their website.

V. BENEFITS

1. This type of the System is useful for the native languages users.
2. User can also search in the other languages according to their choice. If the native language of the user is Hindi and he wants to retrieve the documents of the Tamil languages then this project is also beneficial.

VI. LIMITATIONS

This System provides the search result of the Google to the users in their own native languages according to the current location of the user or user can also search in any language based on his requirement. Some of the challenges are given below.

- This project uses Google Translate API for language translation hence it only support the languages provided by the Google API. If we want to add a new language that is not supported by Google then it is not possible.
- This project supports only the locations or regions which are supported by the Geolocation API.

VII. CONCLUSION

Location based search becomes popular for search engines and their users since many of the web documents are available in native languages. Location based search provides the search result of the Google in the native languages of the user. . Current location information can be used to provide local or native language information using the GPS module. CLIR refines the Google search result by retrieving the semantically similar pages of multiple languages. This project allows the search engine to not only retrieve the documents having query terms in English language but it retrieves the documents having semantic equivalent of the query in other native languages based on the location of the user. This project allows the search engine to provide search result more relevant and impressive to the User query.

REFERENCES

[1] J. Cardeñosa, C Gallardo, Adriana Toni," Multilingual Cross Language Information Retrieval A new approach".
[2] H. Alizadeh, R. Fattahi "Applying Natural Language Processing Techniques for Effective Persian- English Cross-Language Information Retrieval", International Journal of Information Science and Management
[3] Saied Asadi ,Xiao fang Zhou ,Hamid R.Jamali ,Hossein Vakil Mofrad, "Location-Based Search Engines Tasks and Capabilities: A Comparative Study" , Volume 1, Number 2, December, 2004
[4] B.Herbert,G. Szarvas,I Gurevych "Combining Query Translation Techniques To Improve Cross-Language Information Retrieval",2011
[5] P Buitelaar, K Netter, Feiyu Xu," Integrating Different Strategies for Cross-Language Information Retrieval in the MIETTA Project".
[6] http://en.wikipedia.org/wiki/Google_Search
[7] http://www.googleguide.com/google_works.html
[8] http://en.wikipedia.org/wiki/reverse_geocoding
[9] https://developers.google.com/maps/articles/geolocation

