CROSS-LANGAUGE SEARCH ENGINE

Test Strategy

TEST STRATEGY

CROSS LANGAUGE SEARCH ENGINE

Westend • Aundh
Pune, Maharastra,India 411007

Revision History

Date	Version	Author	Page affected	Reason for change
18 July 2014	1.0	Anand Kumar Sharma Anand Kulkarni	Draft	

Table of Contents

Chapter 1

1. Introduction

1.1 About Search Engine

Search engines are programs that search documents for specified keywords or search query and retrieve a list of the documents where the keywords or search query were found. A search engine is really a general class of programs; however, the term is often used to specifically describe systems like Google, Bing and Yahoo! Search that enable users to search for documents on the World Wide Web. Search results are generally presented in a line of results often referred to as search engine results pages. The information may be a specialist in web pages, images, information and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler

1.2 Cross language Search Engine

In a Cross Language search, a user submits a query in one language; the system conducts several steps to find out the relevant Web pages written in a different language

At last, the system translates the results into a language of the user's query (using MT). For example, an English–speaking user who needs information on alternative medicine written in Hindi can submit search terms in English to a system. The system will translate the query into Hindi, and return the search results in English back to the user. In this example, translation between the two involved languages happens on query level (translate the English query into Hindi)

It is search engine where an user will be able to give a query in one Indian language and S/he will be able to access documents available in the language of the query, and another user desired language.

Results will be presented to the user in the language of the query. The results can also be presented in the language in which the information originally resided

1.2 Purpose

The purpose of this document is to propose to evaluate the Cross Language Search Engine, which can be referred by any testing agency for determining the overall quality of Cross Search engine.

1.3 Scope

The scope of this document is to define Cross language Search Engine evaluation strategy which mainly includes importance of test data, various testing techniques and subjective evaluation metrics.

1.4 Abbreviations & Acronyms Used In This Document

CLIA	Cross Lingual Information Access
SRS	Software Requirements Specifications
DeitY	Department Of Information Technology
GIST	Graphics And Intelligence Based script Technology
NER	Named Entity Recognition
MWE	Multi Word Expression
URL	Uniform Resource Locator
IE	Information Extraction
ZWJ	Zero Width Joiner
ZWNJ	Zero Width Non joiner
GUI	Graphical User Interface
IME	Input Method Editor
DCG	Discounted Cumulative Gain
nDCG	Normalized Discounted Cumulative Gain
ACR	Acronyms

2. World Wide Web

The World Wide Web can be divided into two major areas, the surface web and the deep web

2.1 Surface Web

The surface web are a massive interlinked collection of digital information. It is often visualized as a large graph similar to a spider's web. Hyperlinks give the web its highly inter-connected web structure. They are pointers to other documents and collections on the web. Each HTML web page typically has outgoing hyperlinks to other HTML pages, and incoming hyperlinks from other HTML pages, creating a spider-web like system of interconnected data.

2.2 Deep web

It is assumed that a large search engine like Google covers most of the information available on the internet. However with a popular search engine reveals that most of the collection contents are not in the Google results. There are many collections on the internet that contain rich sources of information that are not available to traditional search engines. These collections are collectively called the deep web. Deep web collections are phone directories, subject directories, patent collections, news articles, and holiday booking interfaces. Why is this rich and authoritative information not available to search engines? The simple answer is that there are generally no hyperlinks directly pointing to the information contained in these collections. The only practical way to access and evaluate the information contained in these collections is to enter a query into a search box on the collection's HTML interface, and then to browse the results.

2.3 Crawling Web

Web search engines are comprised of an spider, an indexer and an retrieval engine.
The spider is a program that is used to gather information from the world wide web. The spider follows hyperlinks across the web collecting information from HTML web pages. Search engines such as Google start from a small set of pages on the World Wide Web and recursively follow all the hyperlinks to other pages. This makes it possible to reach most of the surface web in a relatively short time. Some of the major search engines attempt to index the entire surface web into a centralized index. For Each major search engine has a different implementation of a spider, each with a different search strategy. For example, many spiders refuse to collect information from pages generated by database enabled scripting languages such as PHP and ASP, leaving massive amounts of high quality data unindexed. Some spiders limit the depth to which they crawl a website, sometimes only collecting the index page of the site, other times only

3. Need of Evaluation of Search Engine

Due to the massive growth of the World Wide Web, applications called “search engines” have emerged as a fast and efficient method of finding electronic information. Information on almost any subject imaginable can be rapidly growing on web. Information and knowledge continues to grow, search engines will become increasingly important to our day to day work. Most large search engines are based on an “inverted index” model. This is similar to a concordance, where all information is broken into terms and stored in a sorted index of terms called the “index”. This enables searches to be quickly and efficiently performed.

For each term in each document, statistical information is stored about it, such as the number of times it occurred in the document and where it was placed in relation to other terms.

Information lookup is performed by taking a query, breaking it into terms, and then doing a search of the index for the query terms. Once the terms have been found in the index, the search engine then runs a series of algorithms to select the most relevant document based on the query terms.Words which occur frequently, stop words, are discarded. A stopword is any word which has no semantic content. Common stop words are prepositions and articles, as well as high frequency words that do not help retrieval. These words can be removed from the internal model of the query, document, or collection without causing loss of precision and recall.

3.1 Type of Query

Classification of User Intent: Action, Information and Navigation – "Do-Know-Go"

Sometimes it is helpful to classify user intent for a query in one or more of these three categories:

• **Action intent** – Users want to accomplish a goal or engage in an activity, such as download software, play a game online, send flowers, find entertaining videos, etc. These are "do" queries: users want to do something.

• **Information intent** – Users want to find information. These are "know" queries: users want to know something.

• **Navigation intent** – Users want to navigate to a website or webpage. These are "go" queries: users want to go to a specific page.An easy way to remember this is Do-Know-Go. Classifying queries this way can help you figure out how to rate a webpage. Please note that many queries fit into more than one type of user intent.

Action intent Queries – "Do"

The intent of an action query is to accomplish a goal or engage in an activity on the Web. The goal or activity may be to download, to buy, to obtain, to be entertained by, or to interact with a resource that is available on the Web. Users want to do something. Here are some examples of goals and activities:

• Purchase a product.
• Download software for free or for money.
• Pay a bill online.
• Play a game online.
• Print a calendar.
• Send flowers.
• Organize photos or order prints online.
• Watch a video clip.
• Copy an image or piece of clipart.

• Take an online survey.

• View entertaining webpages, such as pictures, gossip, videos, etc.

Helpful pages for an action query are pages that allow users to do the activity or accomplish the goal.

Query Likely User Intent Description of The Landing Page

[geography quiz], Take an online geography quiz. Page with a working online geography quiz.

[Beatles poster], Find an image of a Beatles poster or perhaps purchase a Beatles poster.

Page on which to view or purchase a Beatles poster.

[download adobe reader], Download software. Official free download page on the Adobe website.

[fairy tale coloring pages], Print coloring pages. Trustworthy page with printable coloring pages.

[online personality test], Take an online personality test. Page with a working online personality test.

[Hindi English dictionary],Translate Hindi words into English or English words into Hindi.

Page that translates words from English to Hindi and Hindi to English.

Information Queries – "Know"

An information query seeks information on a topic. Users want to know something; the goal is to find information. Helpful pages have high quality, authoritative, and comprehensive information about the query.

Query Likely User Intent Description of the Landing Page

[Switzerland],Find travel and tourism information for planning a vacation or holiday, or find information about the Swiss geography, languages, economy, etc. Page about Switzerland on a well-known travel guide. Informative CIA World Factbook webpage on Switzerland.

[cryptology use in WWII]

Find information about how cryptology was used in World War II.

United States Air Force Museum article about

cryptology use during WWII.

[how to remove candle wax from carpet], Find information on how to remove candle wax from carpet.Page

Navigation Queries – "Go"

The intent of a navigation query is to locate a specific webpage. Users have a single webpage or website in mind.This single webpage is called the target of the query. Users want to go to the target page.The most helpful page for a navigation query is the navigational target page.

Query Likely User Intent URL of the Target Page Description of the Target Page

[ibm]

Go to the IBM homepage. http://www.ibm.com/ Official homepage of the IBM Corporation.

[youtube]

Go to the YouTube homepage. http://www.youtube.com/ Official homepage of YouTube.

[ebay]

Go to the eBay homepage. http://www.ebay.com/ Official homepage of eBay.

[harvard college admissions],

Go to the Harvard College admissions page on the Harvard University website.

http://admissions.college.harvard.edu/index.html
Harvard College Office of Admissions page on the official
Harvard University website.
[best buy store locator], English (US)
Go to the store locator page on the Best Buy website.
http://www.bestbuy.com/site/olspage.jsp?id=cat12090&type=page
Store Locator page on the official Best Buy website.

4. Information retrieval testing

In field of Multi-language information retrieval, it can be classified into four major areas

Viz

Multilingual Retrieval

Bilingual Retrieval

Monolingual Retrieval

Domain Specific Retrieval

Various Evaluation Method:

4.1 Single Best Target (For Monolingual Search)

In this method of testing, evaluator knows which the best target page for a given query is. The best results should appear at the top of the list. Evaluator has a set of queries and for each query user also has single best target matching for that query.

4.1.1 Selecting test queries

a) Only those phrases are selected as queries where it focus user's intended target and information corresponding to that query is available in corpus. Only Navigational query should be used. A navigational query is one that usually has only one satisfactory result.

b) Skip searches for which there is more than one best target OR You're not sure what the user is trying to find. e.g. a search for "झांसी" could either refer to the "झांसी की रानी" OR "झांसी का किला".

c) Best target should be in crawl corpus of CLIA search engine.

Keep only the phrases where you feel very confident about the user's intended meaning the well-phrased queries. If there's any doubt about what the user wanted, skip it.

Example:

Query	Expected Result
झांसी	http://yatrasalah.com/touristPlaces.aspx?id=18
ग्वालियर का किला	http://bharat.gov.in/knowindia/culture_heritage.php?id=39
अतुल्य भारत	http://hi.incredibleindia.org/
पर्यटन मंत्रालय भारत सरकार	http://kol.indusnettechnologies.com/gov/mot/

Table D

Testing the results returned by search server

Each query is fed to the search server, search server returns matching results, User have to analyze results and pinpoint where the best target actually falls in the list, and count how many spaces it is from the first position. Currently we are testing only top 5 results returned by search server for each query.

Assign score to best matched result according to its position.

Once all queries have returned search results, Evaluator should analyze results and locate where the expected result is in top 5 search result.

Assign score to best matched result according to its position.

Position	1	2	3	4	5	If expected result is not in top 5
Score	10	8	6	4	2	0

Table E

Single Best Target Statistics

After carrying out all tests following statistics will be generated for analysis.

- Total number of queries according to their expected result position in result set
- Percentage of queries according to their expected result position in result set.
- Score of each query according to position of expected result in result set.
- Mean score of all queries.

Example: For the queries mentioned in Table D statistics mentioned as in below table (Table F) will be calculated.

Statistics of queries according to expected result position in result set		
Position	**Total queries**	**Percentage of queries**
1	1	25.00%
2	2	50.00%
3	0	0.00%
4	1	25.00%
5	0	0.00%
Not Found in top 5	0	0.00%
Total Queries = 4		
Average score of all queries = (1*10+2*8+0*6+1*4+0*2)/4 = 7.5 out of 10		

Table F

4.2 Discounted cumulative gain (For Cross Language)

Discounted cumulative gain (**DCG**) is a measure of effectiveness of a web search engine algorithm often used in information retrieval. Using a graded relevance scale of documents in a search engine result set, DCG measures the usefulness, or *gain*, of a document based on its position in the result list. The gain is accumulated from the top of the result list to the bottom with the gain of each result discounted at lower ranks.

For DCG calculation 10 keyword queries and 10 phrase queries (total 20 queries) for each language will be evaluated. Ranking will be done on the top 5 documents retrieved by CLIA for every search query. Total of 3 evaluator will required for each language to rate the relevancy.

Steps to be followed:

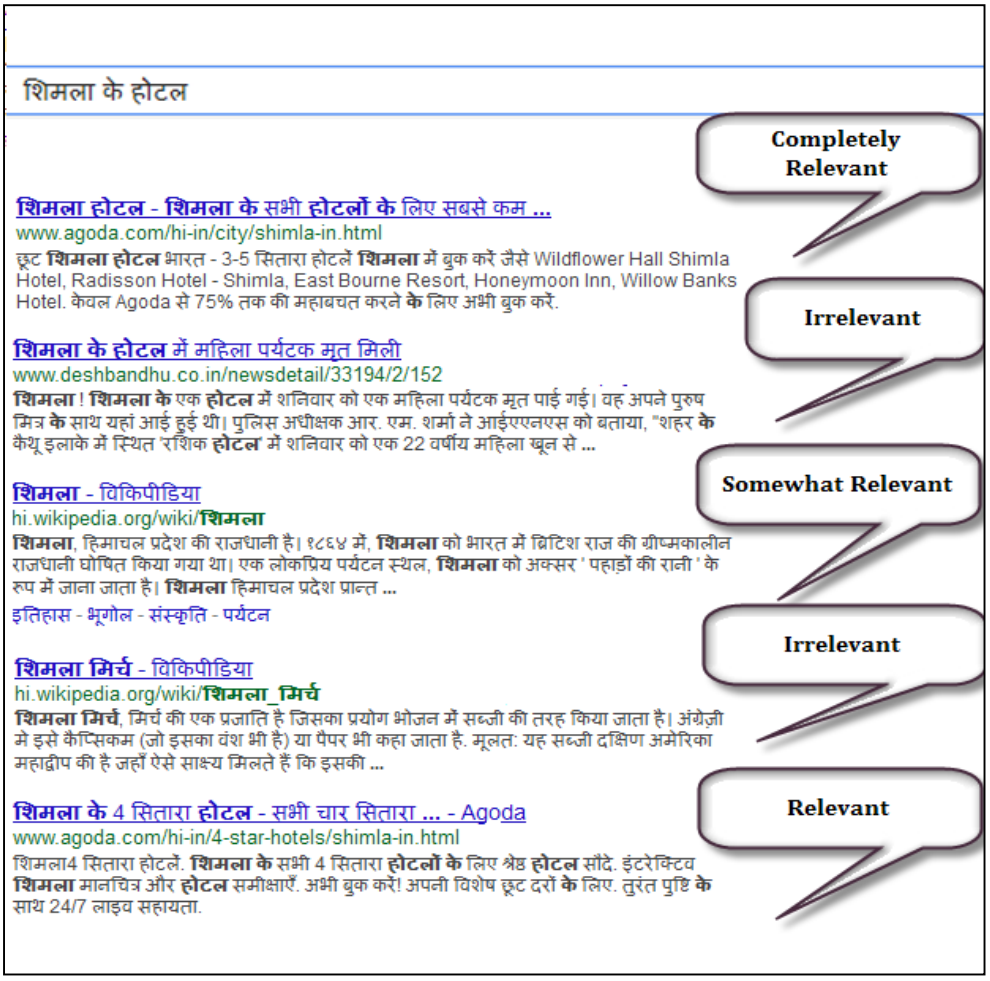
1. Search result for each query that is to be fired on CLIA system is captured in three languages.
2. Top 5 search result in three languages is to be graded on following basis.

Completely Relevant (Based on user query, search result provides correct information that user need.**)**	3
Relevant (Based on user query the result is correct with context.**)**	2
Somewhat Relevant(Based on used query search engine has provided information that can further be used to get more information**)**	1
Irrelevant (The result has failed to provide any correct information for search query**)**	0

Table G

To explain the evaluation scale we have taken a query in Hindi that is "**शिमला के होटल**".Let's assume that this query produce five result. One the basis above scale these top 5 query is to be evaluated

In this query user intention is to find hotel in shimla. If we have to judge from user prospective then second and fourth result are irrelevant. First result explain about Hotel and and its cost so it can be assumed as relevant from user prospective. Similarly third result is from Wikipedia which given some information about Hotel in Shimla. Last result is about list of 4 star hotels which again is Relevant from user prospective.



Discounted cumulative gain calculations:

The premise of DCG is that highly relevant documents appearing lower in a search result list should be penalized as the graded relevance value is reduced logarithmically proportional to the position of the result. The discounted CG accumulated at a particular rank position is defined as:

$$DCG_p = rel_1 + \sum_{i=2}^{p} rel_i / log_2(i)$$

Formula 1

rel_i is the graded relevance of the result at position

Normalized DCG

Search result lists vary in length depending on the query. Comparing a search engine's performance from one query to the next cannot be consistently achieved using DCG alone, so the cumulative gain at each position for a chosen value of should be normalized across queries. This is done by sorting documents of a result list by relevance, producing the maximum possible DCG till position, also called Ideal DCG (IDCG) till that position. For a particular query, the *normalized discounted cumulative gain*, or nDCG, is computed as:

$$nDCGp = \frac{DCGp}{IDCGp}$$

Formula 2

The nDCG values for all queries can be averaged to obtain a measure of the average performance of a search engine's ranking algorithm. Note that in a perfect ranking algorithm, the DCG_p will be the same as the $IDCG_p$ producing an nDCG of 1.0. All nDCG calculations are then relative values on the interval 0.0 to 1.0 .

Example:

To explain the process of calculation of the nDCG an evaluator will be asked to judge the relevance of each document (search result) to the query. Each search result will be judge on the scale mention in "Table A" abo ve. The most relevant result will be marked as 3 and irrelevant document will be marked as 0. Let's assume that for a query we have 10 search results. We will consider top 5 results to evaluate

in 3 languages. On the basis of judgment of evaluator let's assume evaluator has provided following scores for top 5 search result.

Result at Position 1	3
Result at Position 2	2
Result at Position 3	3
Result at Position 4	0
Result at Position 5	1

Table H

On the basis on Table B following calculation is done

Position of Result (i)	rel_i	$Log_2 i$	$rel_i/log_2 i$
1	3	0	NA
2	2	1	2
3	3	1.585	1.892
4	0	2.0	0
5	1	2.322	0.431

Table I

DCG_5 of this ranking is calculated using Formula 1

DCG_5= 3+ (2+1.892+0+0.431) =7.323

In order to compare the DCG value must be normalized. To normalize DCG values, an ideal ordering for the given query is needed. For this example, that ordering would be the monotonically decreasing sort of the relevance judgments provided by the experiment participant, which is as shown below

Result at Position 1	3
Result at Position 3	3
Result at Position 2	2
Result at Position 5	1
Result at Position 4	0

Table J

Position of Result (i)	rel_i	$Log_2 i$	$rel_i/log_2 i$
1	3	0	NA
2	3	1	3.0
3	2	1.585	1.261
4	1	2	0.5
5	0	2.322	0

Table K

The DCG for this ideal ordering is:

$IDCG_5$= 3+ (3+1.261+0.5+0) = 7.761

Using Formula 2 for $nDCG_5$

$$\mathbf{nDCG_5=DCG_5/IDCG_5=7.323/7.761=0.943}$$

This calculation for nDCG is to be done for language in which query is fired and for Cross language search for English and Hindi. This means that total 20(number of query)*5(no. of search result for a query)*3(cross language) =300 search result will be evaluated for one language. Similarly it has to be done for other language

Average of the nDCG is calculated for 20 queries in different language.

Average of 20 queries for nDCG is calculated for all 3 evaluator for respective language. Average of all the three evaluator will be consider as final nDCG for respective language

$$nDCG_{language} = \frac{\sum_{i=1}^{3} nDCG}{3}$$

4.3 Precision and Recall

What is Precision?

Precision is defined as the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search.

$$\text{Precision} = \frac{|\{\text{relevant search result}\} \cap \{\text{retrieved search result}\}|}{|\{\text{retrieved search result}\}|}$$

What is Recall?

Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents.

While Recall and precision have been standard method for information retrieval. Recall is always a difficult method to calculate as it requires the knowledge of the total number of relevant items in the collection.

$$\text{Recall} = \frac{|\{\text{relevant search result}\} \cap \{\text{retrieved search result}\}|}{|\{\text{relevant search result}\}|}$$

F- Measure

It is harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

F = 2 (Precision *Recall)/ Precision +Recall

Or it can be represented as F = 2/(1/ Precision)+(1/Recall)

4.4 Snippet Generation:

To evaluate the snippet generation a set of queries per language for each language for will be fired to the portal.

The snippet output will be evaluated by the evaluator who will read the snippets and decide if they are relevant to the query topic or not. The top 5 snippet will be considered for every query .Total number of snippet to evaluate is

Evaluator will have to grade the snippet on grading scale as mention below.

Completely Relevant(The Snippet is consistent with the page and is couched in correct language)	3
Relevant (The Snippet provides a clear picture of the page but its grammatical accuracy is low)	2
Somewhat Relevant(The Snippet allows the user to understand the page but is not still very clear)	1
Irrelevant(Snippet judged totally divergent from the page)	0

Table L

Example:

To explain the process of evaluation let's have evaluator will rate the snippet of particular language as mention below.

Snippet at Position 1	3
Snippet at Position 2	2
Snippet at Position 3	3
Snippet at Position 4	0
Snippet at Position 5	1

Table M

On the basis on Table L following calculation is done.

Position of Result (i)	rel_i	$Log_2 i$	$rel_i/log_2 i$
1	3	0	NA
2	2	1	2
3	3	1.585	1.892
4	0	2.0	0
5	1	2.322	0.431

Table N

DCG_5 of this ranking is calculated using Formula 1

DCG_5= 3+ (2+1.892+0+0.431) =7.323

In order to compare the DCG value must be normalized. To normalize DCG values, an ideal ordering for the given query is needed. For this example, that ordering would be the monotonically decreasing sort of the relevance judgments provided by the experiment participant, which is as shown below.

Snippet at Position 1	3
Snippet at Position 3	3
Snippet at Position 2	2
Snippet at Position 5	1
Snippet at Position 4	0

Table O

Position of Result (i)	rel_i	$Log_2 i$	$rel_i/log_2 i$
1	3	0	NA
2	3	1	3.0
3	2	1.585	1.261
4	1	2	0.5
5	0	2.322	0

Table P

The DCG for this ideal ordering is:

$IDCG_5$= 3+ (3+1.261+0.5+0) = 7.761

Using Formula 2 for $nDCG_5$

$$\mathbf{nDCG_5=DCG_5/IDCG_5=7.323/7.761=0.943}$$

This calculation for nDCG is to be done for language in which query is fired and for Cross language search for English and Hindi. This means that total 10(number of query)*5(no. of search result for a query)*3(cross language) =150 search result will be evaluated for one language. Similarly it has to be done for other language.Average of the nDCG is calculated for 10 queries in different language

Average of 10 queries for nDCG is calculated for all 3 evaluator for respective language. Average of all the 3 evaluator will be consider as final nDCG for respective language.

$$nDCG_{language} = \frac{\sum_{i=1}^{3} nDCG}{3}$$

4.5 Snippet translation:

To evaluate the snippet translation a set of queries for each language for will be fired to the portal. Top 5 snippets will be considered for every query for snippet translation testing.

The following language pairs will have to be tested for translation.3 evaluators for each language pair will be used all of which are common users.

- Each evaluator will be provided two sheets. One sheet for Blind Testing and one for Open Testing.
- In the case of Blind Testing, the evaluator will not have access to Source snippet in English and Hindi and in Open Testing s/he will be provided with the source snippet output along with the translated snippets.
- Then the evaluator will grade translated snippet outputs on 0 to 4 scales.
- **0 – 4 grading scale** would be provided to the evaluator for use while grading the translated snippets.

0	The translated output is not at all comprehensible to you.
1	Keywords comprehensible and rest of the snippet makes little sense.
2	Comprehensible after accessing the source (English/Hindi) text.
3	Comprehensible with difficulty.
4	Snippet is comprehensible.

Table R

- The graded out put through both blind and open testing, provided by each evaluator for a given language will be processed as per formula provided below.

For each language pair:

$$\text{Evaluation mean of grading for each set of translations} = \frac{\sum_{T=1}^{3} Grades\ given\ by\ evaluator\ T}{3}$$

Final evaluation of grading for all translations generate

$$= \frac{\sum_{L=1}^{3} Evaluation\ mean\ for\ all\ snippets\ translated\ for\ langauge\ set\ L}{3}$$

$$\text{Total average grading} = \sum_{q=1}^{10} Evaluation\ mean\ for\ of\ each\ set\ of\ translations$$

4.6 Summary generation:

To evaluate the Summary Generation a set of Queries will be fired to the portal.

The top 5 snippets will be considered for every query for Summary generation. The following language pairs will have to be tested for translation. 3 evaluators for each language pair will be used .As in the case of evaluation of snippet translation, this assessment will also be done in two ways- blind testing and open testing. Blind testing will assess whether the summary generated is comprehensible enough and open testing evaluates whether the contents of the web page has been summarized accurately in the summary generated. A Grading scale of 0-4 will be set-up:

0	The summary generated is not at all comprehensible to you.
1	Summary generated makes little sense.
2	Comprehensible after accessing the source web page.
3	The summarization has been done to an extent but not completely.
4	Summary has summarized web page correctly.

Table S

Each evaluator will be provided two sheets. One sheet for Blind Testing and one for Open Testing. In the case of Blind Testing, the evaluator will not have access to original contents of web page in source language (language of query, Hindi and English) and in Open Testing s/he will be provided with the contents of the web page along with the generated summary.

Percentile will be calculated based on below formulae.

$$\text{Evaluation mean of grading for each set of summary} = \frac{\sum_{T=1}^{3} Grades\ given\ by\ evaluator\ T}{3}$$

-

$$\text{Final evaluation of grading for all summary generated} = \frac{\sum_{L=1}^{3} Evaluation\ mean\ for\ all\ summary\ generated\ for\ langauge\ set\ L}{3}$$

$$\text{Total average grading} = \sum_{q=1}^{10} Evaluation\ mean\ for\ all\ summaries\ for\ language\ set\ L$$

4.7 Named Entities:

Identification of named entities in every language and their accurate translation/transliteration is one of the most crucial parameters of CLIA performance. Named entities if incorrectly translated will result in an ineffective information retrieval.

A test bed of sets of named entities for each language has been collected for this testing. Testing will be done to validate if the named entities input are translated or transliterated in Hindi and English for effective search and retrieval. Named entities viz. **हवा महल** (in Hindi), **মশাকরণ** (in Bangla) should be translated/Transliterated to "Hawa Mahal" in English (and not Palace of Winds) and Writer's Building in English respectively.

Evaluation will be carried out using grading scale of 0-1

0	Properly identify the NER
1	Does not identify the NER.

Table T

4.8 Acronyms

The system will be tested as to its capability to handle the acronyms in a given language. Certain acronyms such as **सौ** in Marathi are ambiguous and hence will need special contextual disambiguation.

The system will be tested as to its capability to handle all acronyms in the source language(s) and across the target languages. A test bed of sets of Acronyms for each language has been collected for this testing

Evaluation will be carried out using grading scale of 0-1

0	Properly identify the acronyms
1	Does not identify the acronyms.

Table U

Chapter 5

5. Test Data Preparation

In order to evaluate CLIA system for it is advised to collect the data domain wise as well as different data type. This will help to give much more precise result. For preparation of Test data, Linguist assistance can be taken.

For CLIA system for Indian language following category of data set should be considered

Keywords Queries

Phrase Queries

NER

Acronyms

Normalization

Spelling variations

Singular/Plural

Lemmatizer

Synonyms

Spell Checker

1) Keyword Queries

विशाखापट्नम

ऋषिकेश

हवामहल

भारतीय पुरातत्व सर्वेक्षण

स्वर्ण मंदिर

पर्यटन भूगोल

रमणीय-कमनीय उदयपुर

पिछोला झील

घी-त्यार एक लोक उत्सव

Phrase Queries

धनोल्टी से मसूरी का सफ़र

गांधी स्थल की यात्रा

कर्नाटक की ऐतिहासिक इमारतें

हवा महल का इतिहास

राजस्थान का स्वर्ग

वाटिका इन डिलक्स उदयपुर

स्वर्ण मंदिर का शहर अमृतसर

अमृतसर के हॉस्टल

ममता कहानी देवी नागरानी

कब्र का मुनाफा द्वितीय संस्करण का लोकार्पण

NER

राजस्थान

जलाल उद्दीन मोहम्मद अकबर

पर्यटन मंत्रालय

सारनाथ

मसूरी-नैनिताल

नालंदा युनिवर्सिटी

काशी

जिम कॉर्बेट नॅशनल पार्क

डलहौज़ी

काशिनाथ

Acronyms

बीबीसी

नाटो

सार्क

एड्स

नाबार्ड

बिस्कोमान (बिहार राज्य सहयोग क्रय-विक्रय संघ समित)

भाजपा

याहू

Normalization Queries

साफ़-साफ

मौका -मौक़ा

जरुरत-ज़रूरत

फोटो -फ़ोटो

Spelling variations

इण्टरनेट- इंटरनेट

हिन्दी - हिंदी

फिल्म -फ़िल्म

किराय-किराए

रविन्द्र -रविंद्र

स्थाई -स्थायी

अक्सर -अकसर

स्कूल, स्कुल

Singular /Purual

डिबिया – डिबियां

माला - मालाएं

नदी -नदियाँ

टैक्सी-टैक्सियाँ

दुकान-दुकानें

झाड़ी-झाड़ियाँ

Lemmatizer

साफ़-सफाई

पवित्र-पवित्रता

संगठन-संगठित

Synonyms

नाव - नौका

भारतवर्ष -हिंदुस्तान- इंडिया

होटल, होटेल, सराय

आहार, खाद्य, खाना, भोजन

भगवन ,प्रभु ,इश्वर

Spell Checker

राजस्था

अमृतस

पुरातत्

मंदि

भगोल

नग्रह

6. Conclusion

In this above document we have tried to introduce the evaluation of Cross language search engine. We evaluated the methodology use to map the theory with actual implementation of CLIA system.

The Test data is manually prepared for specific domain for different language. We have used Top-n-Precision for evaluating the CLIA system. Apart from this Discounted cumulative Gain method is also implemented to calculate the accuracy of search engine.

Further to test accuracy of monolingual search result SINGLE BEST TARGET is used

For this subjective evaluation, the evaluators needs to be chosen with some idea of search engine and a training session must be organized for the evaluator especially to make them understand the parameters of evaluation.

Chapter 7

7. References

http://en.wikipedia.org/wiki/Discounted_cumulative_gain

http://www.stanford.edu/class/cs276/handouts/EvaluationNew-handout-1-per.pdf

http://maroo.cs.umass.edu/pdf/IR-557.pdf

http://searchenginewatch.com/article/2066485/The-Search-Engine-Perfect-Page-Test

http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1042&context=cs_faculty_pubs

http://alistapart.com/article/testing-search-for-relevancy-and-precision

http://static.googleusercontent.com/media/www.google.com/en/us/insidesearch/howsearchworks/assets/searchqualityevaluatorguidelines.pdf

Evaluating a Cross-language Semantically Enrich Search Engine Leyla Zhuhada,member ,IEEE,and Olfa Nacreous ,*2010 Seventh international Conference of information Technology*

New Measurement for Search Engine evaluation proposed and tested *Liwen Vaughan Faculty of information and Media Studies, University of Western Ontario, London, Information Processing management 40 (2004) 677-691*

Evaluating Web Search Engine using Click through Data *Ben Carterette , Rosie Jones*