An Extensive Literature Review on CLIR and MT activities in India

Kumar Sourabh

Abstract:
This paper addresses the various developments in Cross Language IR and machine transliteration system in India, First part of this paper discusses the CLIR systems for Indian languages and second part discusses the machine translation systems for Indian languages. Three main approaches in CLIR are machine translation, a parallel corpus, or a bilingual dictionary. Machine translation-based (MT-based) approach uses existing machine translation techniques to provide automatic translation of queries. The information can be retrieved and utilized by the end users by integrating the MT system with other text processing services such as text summarization, information retrieval, and web access. It enables the web user to perform cross language information retrieval from the internet. Thus CLIR is naturally associated with MT (Machine Translation). This Survey paper covers the major ongoing developments in CLIR and MT with respect to following: Existing projects, Current projects, Status of the projects, Participants, Government efforts, Funding and financial aids, Eleventh Five Year Plan (2007-2012) activities and Twelfth Five Year Plan (2012-2017) Projections.

Keywords: Machine Translation, Cross Language Information Retrieval, NLP

1. INTRODUCTION:

Information retrieval (IR) system intends to retrieve relevant documents to a user query where the query is a set of keywords. Monolingual Information Retrieval - refers to the Information Retrieval system that can identify the relevant documents in the same language as the query was expressed whereas, Cross Lingual Information Retrieval System (CLIR) is a sub field of Information Retrieval dealing with retrieving information written in a language different from the language of the user's query. The ability to search and retrieve information in multiple languages is becoming increasingly important and challenging in today’s environment. Consequently cross-lingual (language) information retrieval has received more research attention and is increasingly being used to retrieve information on the Internet. While there are numerous search engines that are currently in existence, few support truly cross-language retrieval. Many search engines are monolingual but have the added functionality to carry out translation of the retrieved pages from one language to another, for example, Google, yahoo and AltaVista.

Query and the documents are needed to be translated in case of CLIR. But this translation causes a reduction in the retrieval performance of CLIR. Most approaches translate queries into the document language, and then perform monolingual retrieval. There are three main approaches in CLIR are machine translation, a parallel corpus, or a bilingual dictionary. Machine translation-based (MT-based) approach uses existing machine translation techniques to provide automatic translation of queries. The information can be retrieved and utilized by the end users by integrating the MT system with other text processing services such as text summarization, information retrieval, and web access. It enables the web user to perform cross language information retrieval from the internet. Thus CLIR is naturally associated with MT (Machine Translation)

Work in the area of Machine Translation in India has been going on for several decades. During the early 90s, advanced research in the field of Artificial Intelligence and Computational Linguistics made a promising development of translation technology. This helped in the development of usable Machine Translation Systems in certain well-defined domains. The work on Indian Machine Translation is being performed at various locations like IIT Kanpur, Computer and Information Science department of Hyderabad, NCST Mumbai, CDAC Pune, department of IT, Ministry of Communication and IT Government of India. In the mid 90's and late 90's some more machine translation projects also started at IIT Bombay, IIT Hyderabad, department of computer science and Engineering Jadavpur University, Kolkata, JNU New Delhi etc. Research on MT systems between Indian and foreign languages and also between Indian languages are going on in these institutions. The Department of Information Technology under Ministry of Communication and Information Technology is also putting the efforts for proliferation of Language Technology in India, And other Indian government ministries, departments and agencies such as the Ministry of Human Resource, DRDO (Defense Research and Development Organization), Department of Atomic Energy, All India Council of Technical Education, UGC (Union Grants Commission) are also involved directly and indirectly in research and development of Language Technology by providing funds and financial aids for major projects being carried out in the field of MT and CLIR.

2. CLIR State of Art: Indian Language Perspective

Bengali and Hindi to English CLIR

Debasis Mandal, Mayank Gupta, Sandipan Dandapat,Pratyush Banerjee, and Sudeshna Sarkar Department of Computer Science and Engineering IIT Kharagpur, India presented a cross-language retrieval system for the retrieval of English documents in response to queries in Bengali and Hindi, as part of their participation in CLEF1 2007 Ad-hoc bilingual track. They followed the dictionary-based Machine Translation approach to generate the equivalent English query out of Indian language topics. Their main challenge was to work with a limited coverage dictionary (of coverage _ 20%) that was available for Hindi-English, and virtually non-existent dictionary for Bengali-English. So they depended mostly on a phonetic transliteration system to overcome this. The CLEF results point to the need for a rich bilingual lexicon, a translation disambiguator, Named Entity Recognizer and a better transliterator for CLIR involving Indian languages [1].

Hindi and Marathi to English Cross Language Information Retrieval

Manoj Kumar Chinnakotla, Sagar Ranadive, Pushpak Bhattacharyya and Om P. Damani Department of CSE IIT Bombay presented Hindi to English and Marathi to English CLIR systems developed as part of their participation in the CLEF 2007 Ad-Hoc Bilingual task. They took a query translation based approach using bi-lingual dictionaries. Query words not found in the dictionary are transliterated using a simple rule based approach which utilizes the corpus to return the 'k' closest English transliterations of the given

- *Kumar Sourabh is pursuing research (PhD) in Department of CS and IT University of Jammu, PH-+919469163570. E-mail: kumar9211.sourabh@gmail.com*

Hindi/Marathi 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, they achieve a Mean Average Precision (MAP) of 0.2366 in title which is 61.36% of monolingual performance and a MAP of 0.2952 in title and description which is 67.06% of monolingual performance. For Marathi, they achieve a MAP of 0.2163 in title which is 56.09% of monolingual performance [2].

Hindi and Telugu to English Cross Language Information Retrieval

Prasad Pingali and Vasudeva Varma Language Technologies Research Centre IIIT, Hyderabad presented the experiments of Language Technologies Research Centre (LTRC) as part of their participation in CLEF 2006 ad-hoc document retrieval task. They focused on Afaan Oromo, Hindi and Telugu as query languages for retrieval from English document collection and contributed to Hindi and Telugu to English CLIR system with the experiments at CLEF [3]

FIRE-2008 at Maryland: English-Hindi CLIR

Tan Xu and Douglas W.Oard College of Information Studies and CLIP Lab, Institute for Advanced Computer Studies, University of Maryland participated in the Ad-hoc task cross-language document retrieval task, with English queries and Hindi documents. Their experiments focused on evaluating the effectiveness of a "meaning matching" approach based on translation probabilities. The FIRE Hindi test collection provides the first opportunity to carefully assess some of the resources and techniques developed for the Translingual Information Detection, Extraction And Summarization (TIDES) program's "Surprise Language" exercise in 2003, in which a broad range of language engineering tools were constructed for Hindi in a comparatively short period. The results reported appear to confirm that some of the language resources developed for the Surprise Language exercise are indeed reusable, and that meaning matching yields reasonably good results with less carefully constructed language resources than had previously been demonstrated [4].

English to Kannada / Telugu Name Transliteration in CLIR

Mallamma v reddy, Hanumanthappa Department of Computer Science and Applications, Bangalore University, They present a method for automatically learning a transliteration model from a sample of name pairs in two languages. Transliteration is mapping of pronunciation and articulation of words written in one script into another script. However, they are faced with the problem of translating Names and Technical Terms from English to Kannada/Telugu. [5].

Kannada and Telugu Native Languages to English Cross Language Information Retrieval

Mallamma v reddy, Hanumanthappa Department of Computer Science and Applications, Bangalore University conducted experiments on translated queries. One of the crucial challenges in cross lingual information retrieval is the retrieval of relevant information for a query expressed in as native language. While retrieval of relevant documents is slightly easier, analyzing the relevance of the retrieved documents and the presentation of the results to the users are non-trivial tasks. To accomplish the above task, they present their Kannada English and Telugu English CLIR systems as part of Ad-Hoc Bilingual task by translation based approach using bi-lingual dictionaries. When a query words not found in the dictionary then the words are transliterated using a simple rule based approach which utilizes the corpus to return the 'k' closest English transliterations of the given Kannada/Telugu 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. Finally they conduct experiments on these translated queries using a Kannada/Telugu document collection and a set of English queries to report the improvements, performance achieved for each task [6].

Bilingual Information Retrieval System for English and Tamil

Dr.S.Saraswathi, Asma Siddhiqaa.M, Kalaimagal.K, Kalaiyarasi.M addresses the design and implementation of BiLingual Information Retrieval system on the domain, Festivals. A generic platform is built for BiLingual Information retrieval which can be extended to any foreign or Indian language working with the same efficiency. Search for the solution of the query is not done in a specific predefined set of standard languages but is chosen dynamically on processing the user's query. Their research deals with Indian language Tamil apart from English. The task is to retrieve the solution for the user given query in the same language as that of the query. In this process, an Ontological tree is built for the domain in such a way that there are entries in the above listed two languages in every node of the tree. A Part-Of-Speech (POS) Tagger is used to determine the keywords from the given query. Based on the context, the keywords are translated to appropriate languages using the Ontological tree. A search is performed and documents are retrieved based on the keywords. With the use of the Ontological tree, Information Extraction is done. Finally, the solution for the query is translated back to the query language (if necessary) and produced to the user [7].

Recall Oriented Approaches for improved Indian Language Information Access

Pingali V.V. Prasad Rao Language Technologies Research Centre International Institute of Information Technology Hyderabad: investigated into Indian language information access. The investigation shows that Indian language information access technologies face severe recall problem when using conventional IR techniques (used for English-like languages). During this investigation they crawled the web extensively for Indian languages, characterized the Indian language web and in the process came up with some solutions for the low recall problem. They focused their investigation on the loss of recall in monolingual and cross-lingual based IR and text summarization. The following are some of their major contributions.

- They showed that Indian language information access technologies that use state-of-the-art technologies used by English like languages, face low recall. They observed the recall loss to be relatively higher when the target language corpus is English.
- They came up with a unified information access framework which can address the problems of monolingual and Cross-lingual Information Retrieval and Text Summarization.
- They showed that, word spelling normalization is an essential component of Indian language information access systems and proposed a linguistically motivated rule based approach and showed that this approach works better than the various approximate string matching algorithms.
- They modeled the problem of Dictionary based query translation as an IR problem [8].

A high recall error identification tool for Hindi Treebank Validation

Bharat Ram Ambati, Mridul Gupta, Samar Husain, Dipti Misra Sharma. Language Technologies Research Centre, International Institute of Information Technology Hyderabad, proposed tool that has been used for validating the dependency representation of a multi-layered and multi representational tree bank for Hindi. The tool identifies errors in the Hindi annotated data at POS, chunk and dependency levels. They proposed a new tool which uses both rule-based and hybrid systems to detect errors during the process of treebank annotation. They tested it on Hindi dependency treebank and were able to detect 75%, 62.5% and 40.33% of errors in POS, chunk and dependency annotation respectively. For detecting POS and chunk errors, they used the rule-based system. For dependency errors, they used the combination of both rule-based and hybrid systems. The proposed approach works reasonably well for relatively smaller annotated datasets [9].

English Bengali Ad-hoc Monolingual Information Retrieval Task Result at FIRE 2008

Sivaji Bandhyopadhyay, Amitava Das, Pinaki Bhaskar Department of Computer Science and Engineering Jadavpur University, Kolkata. Their experiments suggest that simple TFIDF based ranking algorithms with positional information may not result in effective ad-hoc mono-lingual IR systems for Indian language queries. Any additional information added from corpora either resulting in query expansion could help. Application of certain machine learning approaches for query expansion through theme detection or event tracking may increase performance. Document-level scoring entailment technique also could be a new direction to be explored. Application of word sense disambiguation methods on the query words as well as corpus would have a positive effect on the result. A robust stemmer is required for the highly inflective Indian languages [10].

Using Morphology to Improve Marathi Monolingual Information Retrieval

Ashish Almeida, Pushpak Bhattacharyya IIT Bombay. They study the effects of lexical analysis on Marathi monolingual search over the news domain corpus (obtained through FIRE-2008) and observe the effect of processes such as lemmatization, inclusion of suffixes in indexing and stop-words elimination on the retrieval performance. Their results show that lemmatization significantly improves the retrieval performance of language like Marathi which is agglutinative in nature. Also, it is observed that indexing of suffix terms, which show spacio-temporal properties, further improve the precision. Along with these, the effects of elimination of stop-words are also observed. With all three methods combined they are able to get mean average precision (MAP) of 0.4433 for 25 queries [11].

A Query Answering System for E-Learning Hindi Documents

Praveen Kumar, Shrikant Kashyap, Ankush Mittal Indian Institute of Technology, Roorkee, India developed a Question Answering (QA) System for Hindi documents that would be relevant for masses using Hindi as primary language of education. The user should be able to access information from E-learning documents in a user friendly way, that is by questioning the system in their native language Hindi and the system will return the intended answer (also in Hindi) by searching in context from the repository of Hindi documents. The language constructs, query structure, common words, etc. are completely different in Hindi as compared to English. A novel strategy, in addition to conventional search and NLP techniques, was used to construct the Hindi QA system. The focus is on context based retrieval of information. For this purpose they implemented a Hindi search engine that works on locality-based similarity heuristics to retrieve relevant passages from the collection. It also incorporates language analysis modules like stemmer and morphological analyzer as well as self constructed lexical database of synonyms. The experimental results over corpus of two important domains of agriculture and science show effectiveness of their approach. [12]

Om: One tool for many (Indian) languages

Ganpathiraju Madhavi, Balakraishnan Mini, Balakrishnan N., Reddy Raj (Language Technologies Institute, Carnegie Mellon University, Pittsburgh) (Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560 012, India) describe the development of a transliteration scheme Om which exploits this phonetic nature of the alphabet. Om uses ASCII characters to represent Indian language alphabets, and thus can be read directly in English, by a large number of users who cannot read script in other Indian languages than their mother tongue. It is also useful in computer applications where local language tools such as email and chat are not yet available. Another significant contribution presented in their research is the development of a text editor for Indian languages that integrates the Om input for many Indian languages into a word processor such as Microsoft WinWord. The text editor is also developed on Java platform that can run on UNIX machines as well. They propose this transliteration scheme as a possible standard for Indian language transliteration and keyboard entry [13].

A multimodal Indian language interface to the computer

Hema A Murthy, C Chandra Sekhar Dept. of Computer Science & Engineering IIT Madras, Chennai developed a multimodal interface to the computer that is relevant for India. Although India's average literacy level is about 65%, less than 5% of India's population can use English for communication. And even though the world-wide web and computer communication has given us access to information at the click of a mouse, 95% of our population is excluded from this revolution due to dominance of English. To overcome this problem they propose to set up an Indian Language Systems Laboratory at IIT Madras. Their initial goal will be to develop a multimodal interface to the computer that is relevant for India, i.e., one that enables Indic computing [14]. The components of this Indian language interface will be:

1. Keyboard and display interface
2. Speech interface
3. Handwriting interface

Part of Speech Taggers for Morphologically Rich Indian Languages

Dinesh Kumar Gurpreet Singh Josan Department of Information Technology DAV Institute of Engineering & Technology Jalandhar, Punjab, INDIA

Their research, reports about the Part of Speech (POS) taggers proposed for various Indian Languages like Hindi, Punjabi, Malayalam, Bengali and Telugu. Various part of speech tagging approaches like Hidden Markov Model (HMM), Support Vector Model (SVM), and Rule based approaches, Maximum Entropy (ME) and Conditional Random Field (CRF) have been used for POS tagging. Accuracy is the prime factor in evaluating any POS tagger so the accuracy of every proposed tagger is also discussed in this paper [15].

Post Translation Query Expansion using Hindi Word-Net for English-Hindi CLIR System

Sujoy Das, Anurag Seetha, M. Kumar, J.L. Rana have investigated impact of query expansion using Hindi WordNet in the context of English-Hindi CLIR system. The WordNet is a lexical database, machine readable thesaurus Hindi language. They have translated English query using Shabdanjali dictionary. The translated queries have been expanded using Hindi WordNet and nine query expansion strategies have been formulated. In these runs title field of topic was used for query formulation and expansion and in one run title + description field was used for query formulation and expansion. The queries are translated, then expanded and are submitted to the retrieval system to retrieve documents from the Fire Hindi Test collection. Their observations suggest that simple query expansion using Hindi WordNet is not effective for English-Hindi CLIR system [16].

Cross-Lingual Information Retrieval System for Indian Languages

Jagadeesh Jagarlamudi and A Kumaran Multilingual Systems Research Microsoft Research India Bangalore, INDIA attempted in building a CLIR system with the help of a word alignment table learned, from a parallel corpora, primarily for statistical machine translation. Presented their results of participation in the Indian language sub-task of the Adhoc monolingual and bilingual track of CLEF 2007. In post submission experiments they found that, on CLEF data set, a Hindi to English cross lingual information retrieval system using a simple word by word translation of the query with the help of a word alignment table, was able to achieve » 73% of the performance of the monolingual system. Empirically they found that considering 4 most probable word translations with no threshold on the translation probability gave the best results. On CLEF 2007 data set, their official cross-lingual performance was 54.4% of the monolingual performance and in the post submission experiments they found that it can be significantly improved up to 73.4%. [17]

Indian Languages IR using Latent Semantic Indexing

A.P.SivaKumar, Dr.P.Premchand, Dr.A.Govardhan focused on improving a Hindi-English cross language information retrieval using latent semantic indexing. For that they collected parallel corpus which contains both Hindi and English documents which are semantically equal and performed singular value decomposition to get a CLIR system. Their tests depicted that the latent semantic indexing improves the results three times to that of direct matching method. [18]

Approximate String Matching Techniques for Effective CLIR among Indian Languages

Ranbeer Makin, Nikita Pandey, Prasad Pingali and Vasudeva Varma International Institute of Information Technology, Hyderabad, India presented an approach to identify cognates and make use of them for improving dictionary based CLIR when the query and documents both belong to two different Indian languages. The effectiveness of their retrieval system was compared on various models. The results show that using cognates with the existing dictionary approach leads to a significant increase in the performance of the system. Experiments have also led to the finding that Indian Language CLIR system based only on the cognates approach performs better, on an average, than the dictionary approach alone. [19]

Using Fuzzy String Search Based on Surface Similarity

Sethuramalingam S, Anil Kumar Singh, Pradeep Dasigi and Vasudeva Varma Language Technologies Research Centre International Institute of Information Technology Hyderabad, India conducted CLIR experiments between three languages which use writing systems (scripts) of Brahmi-origin, namely Hindi, Bengali and Marathi and found significant improvements for all the six language pairs using a method for fuzzy text search based on Surface similarity (general term for orthographic and phonetic similarity between any two words of a language). For evaluation, they used the CLIR data released at the FIRE workshop, 2008. Their paired T-test results indicate that retrieval scores are statistically significant. [20]

CLIA

The CLIA (Cross Lingual Information Access) Project is a mission mode project being executed by a consortium of academic and research institutions and industry partners. Cross-language information retrieval enables users to enter queries in languages they are fluent in, and uses language translation methods to retrieve documents originally written in other languages. Cross-Language Information Access is an extension of the Cross-Language Information Retrieval paradigm. The objective of Cross-Language Information Access is to introduce additional post retrieval processing to enable users make sense of these retrieved documents. This additional processing takes the form of machine translation of snippets, summarization and subsequent translation of summaries and/or information extraction. [21]

3. MT Activities: Indian Language Perspective

Machine translation system is software designed that essentially takes a text in one language (called the source language) and translates it into another language (called the target language). This section will summarize briefly the machine translation systems for Indian languages. An overview of existing machine translation systems:

Year	Systems	Organization / Developers	Translation
1991	ANGLABHARTI	IIT Kanpur Prof. R.M.K. Sinha	English to Indian languages
1995	Anusaaraka	IIT Kanpur Prof. Rajeev Sangal	Indian language to another IL
1999	The Mantra (Machine assisted Translation tool)	C-DAC, Mumbai	English text into Hindi specified domains
2002	English – Hindi translation system	Lata Gore	Eng to Hindi weather narration domain
2002	VAASAANUBAADA	Kommaluri Vijayanand	Bilingual Bengali-Assamese
2003	UNL Based English-Hindi Machine Translation	IIT-B Pushpak Bhattacharya	English to Hindi
2004	ANGLABHARTI-II	IIT Kanpur	English to IL
2004	The Matra system	(KBCS) (NCST), (CDAC) Mumbai	English to Hindi
2004	ANUBHARTI	IIT Kanpur Prof. R.M.K. Sinha	Hindi to IL
2004	Shiva and Shakti	IIIT Hyderabad IIS-B and Carnegie Mellon University	English to Hindi
2004	ANUBAAD	Jadavpur University, Kolkata Dr. Sivaji Bandyopadhyay	English to Bengali
2004	Hinglish	R. Mahesh K. Sinha and Anil Thakur	Hindi to English

2006	IBM-English-Hindi	IBM India Research Lab	English to Hindi
2007	Punjabi to Hindi	Gurpreet Singh Josan Punjabi University Patiala	word-to-word Punjabi-Hindi
2009	Sampark	Consortium of institutions: IITs IIITs Ana University IIS-B and CDAC	IL to IL
2009	Hindi to Punjabi	Vishal Goyal Punjabi University Patiala	word-to-word Hindi-Punjabi

Table 1 Overview

3.1 MT in India (Participation and Funds / financial aids)

ANGLABHARTI 1991 ANGLABHARTI II 2004

- Ministry of Home Affairs, Government of India, have played instrumental roles by funding these projects. The project is primarily based at IIT Kanpur, in collaboration with ER&DCI, Noida, and has been funded by TDIL. The first prototype was built for English to Tamil in 1991 and later a more comprehensive system was built for English to Hindi translation. Translation System has been made available for following languages pair as technology demonstrator:

  i) English to Bangla
  ii) English to Punjabi
  iii) English to Malayalam
  iv) English Urdu

List of Organizations participating in AnglaBharti Mission:

- IIT Mumbai working on Marathi & Konkani and will be developing Angla Marathi & Angla Konkani.
- IIT Gwahauti working on Asamiya & Manipuri and will be developing AnglaAsamiya & AnglaManipuri.
- CDAC Kolkata working on Bangla and will be developing AnglaBangala.
- CDAC(GIST group) Pune working on Urdu, Sindhi & Kashmiri and will develop AnglaUrdu, AnglaSindhi & AnglaKashmiri.
- CDAC Thiruananthpuram working on Malyalam and will be developing AnglaMalayalam.
- TIET Patiala working on Punjabi and will be developing AnglaPunjabi.
- JNU New Delhi working on Sanskrit and will be developing AnglaSanskrit.
- Utkal University Bhuvaneshwar working on Oriya and will be developing AnglaOriya. [22][23]

ANUSAARAKA 1995

- Started at IIT Kanpur and now shifted to IIIT Hyderabad. Got Funding from Ministry of Information Technology, under their program for (TDIL) and financial support from Satyam Computers Private Limited

The MANTRA 1999

- Mantra system was started with the translation of administrative document such as appointment letters, notification, and circular issued in Central government from English to Hindi. The system is ready for use in its domains. Sponsored by the Department of Official Language (DOL), Ministry of Home Affairs, Government of India, Mantra-Rajbhasha has been developed as personal computer, Intranet and Internet versions. The project has been funded by TDIL, and later by the Department of Official Languages.

Sampark (2009)

- Developed by the Consortium of institutions. Consortium of institutions include IIIT Hyderabad, University of Hyderabad, CDAC(Noida, Pune), Anna University, KBC, Chennai, IIT Kharagpur, IIT Kanpur, IISc Bangalore, IIIT Alahabad, Tamil University, Jadavpur University. Currently experimental systems have been released namely {Punjabi, Urdu, Tamil, Marathi} to Hindi and Tamil-Hindi Machine Translation systems.

- Funded by TDIL program of Department of Electronics and Information Technology (DeitY), Govt. of India

Shiva and Shakti (2004

- Carneige Mellon University USA, international institute of information technology, Hyderabad and Indian institute of science, Bangalore, India. Shiva is an Example-based system. It provides the feedback facility to the user. Therefore if the user is not satisfied with the system generated translated sentence, then the user can provide the feedback of new words, phrases and sentences to the system and can obtain the newly interpretive translated sentence.

- Shakti is a statistical approach based rule-based system. It is used for the translation of English to Indian languages (Hindi, Marathi and Telugu). Shakti system combines rule-based approach with statistical approach whereas Shiva is example based machine translation system

UNL-based English-Hindi MT System

- The UNL (Universal Networking Language) is an international project of the United Nations University, with an aim to create an Interlingua for all major human languages. IIT Bombay is the Indian participant in UNL. They are also working on MT systems from English to Marathi and Bengali using the UNL formalism. [24] [25] [26]

Anuvadaksh (English to Indian Language Machine Translation System)

- Department of Electronics and Information Technology (DeitY) came up with the mission of consortia for Machine Translation (MT) Systems. English to Indian Language Machine Translation (EILMT) consortium has been formed as a part of this mission.
  Anuvadaksh is a state-of-the-art solution that allows translating the text from English to six other Indian languages i.e.
  1.Hindi
  2.Urdu
  3.Oriya
  4.Bangla
  5.Marathi
  6.Tamil
  This is a collaborative effort of the consortium institutes which have brought forward the integration of four Machine Translation Technologies:
  1.Tree-Adjoining-Grammar (TAG) based MT
  2.Statistical based MT (SMT)
  3.Analyze & Generate rules (AnalGen) based MT
  4.Example Based MT (EBMT)

- The technical modules such as Named Entity Recognizer [NER], Word Sense Disambiguation [WSD], Morph synthesizer, Collation & Ranking and

Evaluation modules have been developed by different consortium institutes. The Language Vertical tasks have been carried out by various consortia members. Anuvadaksh being a consortium based project is having a hybrid approach, designed to work with the platform and technology independent modules. This system has been developed to facilitate the multi-lingual community, initially in the domain-specific expressions of Tourism, and subsequently it would foray into various other domains as well in a phase-wise manner. [27]

English to Kannada MT system

- Developed at Resource centre for Indian Language Technology Solutions (RC_ILTS), University of Hyderabad by Dr. K. Narayan Murthy. This uses a transfer based approach and it can be applied to the domain of government circulars. The project is funded by Karnataka government. This system uses Universal Clause Structure Grammar (UCSG) formalism.

4. Major Research Activities

The Department of Information Technology under Ministry of Communication and Information Technology is also putting the efforts for proliferation of Language Technology in India. Other Indian government ministries, departments and agencies such as the Ministry of Human Resource, DRDO (Defense Research and Development Organization), Department of Atomic Energy, All India Council of Technical Education, UGC (Union Grants Commission) are also involved directly and indirectly in research and development of Language Technology. All these agencies help develop important areas of research and provide funds for research, to development agencies.

4.1 TDIL

- Government of India launched TDIL (Technology Development for Indian Language) program. TDIL decides the major and minor goal for Indian Language Technology and provide the standard for language technology TDIL journal ***Vishvabharata* (Jan 2010)** outlined short-term, intermediate, and long-term goals for developing Language Technology in India. TDIL has started a consortium mode project since April 2008, for building computational tools and Sanskrit-Hindi MT. The Department has taken a major initiative ‘National Roll-Out Plan’ for wider proliferation of Indian language Software Tools and Fonts.

- The Department also promotes Language Technology standardization through active participation in International and national standardization bodies such as ISO, UNICODE, World-wide-Web consortium (W3C) and BIS (Bureau of Indian Standards) to ensure adequate representation of Indian languages in existing and future language technology standards.

- Initiatives have been taken for long term research for development of Machine Translation System, Optical Character Recognition, On-line Handwriting Recognition System, Cross-lingual Information Access and Speech Processing in Indian languages.

- The consistent initiatives of Government have fuelled the growth of industry in this sector. The spin-off of these efforts has resulted into increasing interest of MNCs to look at India as a large market for Language Technologies. India is, thus, poised to emerge as Multilingual Computing hub.

- TDIL Recently Launched Sandhan: Sandhan is a mission mode project under TDIL Programme. Its main objective is to develop a monolingual search system for tourism domain in five Indian languages viz., Bengali, Hindi, Marathi, Tamil and Telugu. The system has been developed to satisfy the user information need in tourism domain. Sandhan has the capability to process the query based on its language and retrieve results from the respective language. An additional UNL based semantic search facility has been provided for Tamil language. Many of the Indian language web pages are in custom fonts that make the system difficult for retrieving documents. Sandhan uses a font transcoder that converts the custom fonts into Unicode fonts for processing.

4.2 C-DAC

The MT revolution was kick-started by C-DAC when it started work on NLP (Natural language processing) and developed a parser which could parse Hindi, Sanskrit, Gujarati, English and German. While developing this technology, the company was looking at practical implementations of the same and suggested it to various agencies. Realizing the immense potential of MT, the Department of Official Language (DOL) Government of India began actively funding such projects. In Multilingual Computing and Allied Areas, C-DAC continues to work towards the design development and deployment of technologies /solutions for the following areas

Speech Processing and Speech Recognition: Speech corpus creation, analysis and management tools, Phoneme and grapheme mapping tools, Text conversion tools.

Speech Synthesis: Speech corpus creation, analysis and management tools, Phoneme and grapheme mapping tools, Text parsing tools, Speech synthesis tools, Learning and training modules, Speech parameter control module, Intonation and prosodic rule generation

Machine Translation: Corpus creation, analysis and management tools, Pre-processing and post-processing tools, Parsing and generation tools

Information Retrieval: English/Hindi IE/ IR System for the domains of Banking, Agriculture and Railways, Mobile Services; Cross-lingual IE/ IR system using domain specific developed translation Systems; Knowledge based as well as generic Search Engines; Summarizer for English and Hindi, etc.

Extraction and Retrieval (IR) Semantic Search: Semantic Search attempts to augment and improve traditional search results (based on Information Retrieval technology) by using data from the Semantic Web, and adding Indian languages to Semantic Search

Indian Languages OCR: Language independent components, such as image cleaning, skew adjustment, image detection, column detection, table detection, etc. Font training module, Document analysis module backed by dictionaries, spell checker and auto language detection tools. Aligning analyzer, recognition and generator modules.

XLIT is a transliteration tool developed by CDAC, Mumbai to convert words from English to Indian languages and back, without losing the phonetic characteristics. It can be used in Machine Translation systems, e-governance applications and other applications that need to enter text in any Indian language and English.

CDAC Launches U (Urdu) Trans Transliteration: The application demonstrates transliteration from Hind and

Punjabi to Urdu language. It also makes use of C-DAC's language components to depict content in Hind, Punjabi and Urdu. PARC: Perso-Arabic Resource Centre for Urdu, Sindhi, Kashmiri and Arabic Ministry of Communication and Information Technology has identified C-DAC as the "Resource Centre" for the Perso-Arabic range of languages used in India. For the last couple of years, PARC has been in continuous research and development for the scripts that are written from right to left, and is working out solutions for their technical aspect on computers. [28]

4.3 IIIT Hyderabad Machine Translation Research Center

The focus of the MT-NLP lab at LTRC is Building Machine Translation system, Translation among Indian language, Translation from English to Hindi; Example based MT Shiva & Shakti, Developing NLP tools for Language analysis Specified manpower development in Language Technology Post graduate program in Computational Linguistics [29]. Following tables summarize the ongoing major research activities in IIIT Hyderabad MTRC.

Major Research Activities	**Related Info**	
NLP & Machine Translation	Shakti system for translating English to Indian languages	Setu system for translating one Indian language to another Indian language.
Evaluating MT Systems	An online evaluation method has also been developed for evaluating MT systems	Shakti system is regularly evaluated for correctness, comprehensibility and naturalness.
English Analyzer	A sophisticated English analyzer has been built that incorporates state of the art algorithms using machine learning	Included are named entity recognition, stastical parser, PP-attachment algorithm, semantic relation labeler, etc
Dictionaries	Shabdaanjali: An English-Hindi dictionary has been linked to WordNet. The data is manually corrected and validated.	In addition to the above, annotated corpora (namely Part of Speech (POS) tagged and chunked data) for Hindi and Bengali are also created.

Table 2: Major Activities

Major Contributions of LTRC and Some Externally Funded Projects	
Shakti Machine Translation System	Text To Speech for Telugu (general purpose system)
Morphological analyzers for Indian languages	Text To Speech for Hindi (general purpose system)
Shabdaanjali: English to Hindi free e-dictionary	Hand-held Tourist Aid
Font converters and ISCII plugins for Indian scripts	Telugu to Hindi Machine Translation
Text to speech synthesizer for Telugu	FLAN Kit
Information extraction from resume	Transfer Lexicon and Grammar
Customizable Search Engine	(keyword extraction, index, search, classification, clustering)

Table 3: Major Contributions

Major Project	**Related Info** (Funded by: TDIL Program, Department Of IT Govt. Of India)	
Shallow Parser	The shallow parser gives the analysis of a sentence at various levels.	
Hindi	Bengali	Kannada
Punjabi	Tamil	Malayalam
Urdu	Telugu	Marathi

Table 4: Major Project

Major Funded Projects	**Date**	**Agency**
Indian language to Indian language machine translation	Consortium project (2006-2013)	Department of Information Technology (DIT), Govt. of India (GoI)
English to Indian language machine translation	Consortium project (2006-2009)	DIT, GoI
Multi-Representational and Multi-Layered Treebank for Hindi and Urdu	(2008-2011)	NSF, USA
Development of Sanskrit Computational Toolkit and Sanskrit-MT system	2008-2011	DIT, GoI
Discourse and Dialog Management	2008-2011	Tata Consultancy Services
Language Database Development for Example Based Machine Translation	2003-2005	DIT, GoI
Multilingual Morphological Analysis and Chunking/Phrasing modules for Text-to-Speech	2003-2004	Outside Echo Limited, UK
Indian Language to Indian Language Machine Translation System (ILMT) Phase-II	2010-2013	DIT, GoI
Dashboard Development Environment for NLP Applications	2009-2011	DIT, GoI
Development of English to Indian Language Machine	2010-2013)	DIT, GoI

Translation System Phase-II		

Table 5: Major Projects and funding Agencies

4.4 Amrita Vishwa Vidyapeetham:

The Amrita Center for Computational Engineering and Networking is highly committed to promoting quality research in various fields. Since its inception in May 2003, the Center has successfully completed seven major projects with funding from various agencies like ISRO, NPOL, DRDO, ADRIN and Ministry of IT. The Center's current research focus is in the area of Computational Linguistics and Natural Language Processing [30]. Following table summarizes the major research activities going on in Amrita Vishwa V.

Major Research Activities	**Projects are an initiative of Ministry of Human Resource Department under National Mission on Education through ICT**	
Morphological Analyzer / Generator for Tamil Description	Rule Based English to Tamil Transliterator	
Machine Learning based Morphological Analyzer	Linguistic Tree Viewer in Java	
Tamil POS Tagger	**Dravidian WordNet**	
Malayalam POS Tagger	Malayalam Wordnet	Tamil Wordnet
SVM Based English to Tamil Transliterator	Kannada Wordnet	Telugu Wordnet

Table 6: MT @ Amrita V.V

4.5 IIT Bombay Machine Translation

Center for Indian Language Technology (CFILT) was set up with a generous grant from the Department of Information Technology (DIT), Ministry of Communication and Information Technology, Government of India in 2000 at the Department of Computer Science and Engineering, IIT Bombay. Prior to this the Natural Language Processing (NLP) activity of the CSE Department, IIT Bombay took off in 1996 with a grant from the United Nations University, Tokyo to create a multilingual information exchange system for the web. [31] Following table summarizes the major research activities going on in IIT-B.

Sponsoring Agency	**Title of project**	**Time period & current status**
Ministry of IT 2012	Dravidian Wordnet: Kannad, Malayalam, Tamil, Telugu	January 3, 2012 to January 2, 2014. Consortium of 5 institutions with IIT Bombay

		leading the consortium. Ongoing.
Ministry of IT 2011	English to Indian Language Machine Translation (EILMT): Phase-II	September 1, 2011 to August 31, 2014. Consortium of 9 institutions. IITB, Word Sense Disambiguation. Ongoing
Ministry of IT 2010	Cross Lingual Information Access (CLIA): Phase-II	September 2010 to August 2013. Consortium of 12 institutions with IITB in the lead. Ongoing
Ministry of IT 2010	Indian Language to Indian Language Machine Translation (ILMT): Phase-II	May 2010 to April 2013. Consortium of 14 institutions. IITB Goal: development of Marathi-Hindi MT system. Ongoing
Ministry of IT 2010	Indradhanush: Integrated Wordnet for Bengali, Gujarati, Kashmiri,Konkani, Oriya,Punjabi & Urdu	August 2010 to July 2012. Linkage of Synsets of mentioned languages with Hindi Wordnet, establishment of semantic relations, sense disambiguation and navigation algorithms. Ongoing
Xerox Corporation 2010	Multilingual Tools and Resources for Indian Languages	Sept 2010 to August 2013. Hindi English MT in Judicial Domain, Crowd sourcing, Automatic Error Correction and Detection in MT. Ongoing.
AOL, Bangalore 2010	High Accuracy Sentiment Analysis	August 2010 to July 2011. Role of sense disambiguation on Sentiment Analysis, Sentiment Analysis for Twitter, SA for Indian Languages. Ongoing.
Ministry of Human Resource Development 2009	Tools and Resources for Machine Translation of English and Dravidian Languages	July 2009 to June 2011.Consortium of 15 institutes to create MT systems among Tamil, Telugu, Kannad, Malayalam and English. IITB Part-Linking of Hindi

		and Dravidian Wordnet. Ongoing.
Ministry of IT 2009	Indian Language Corpora	September 2010 to August 2013.Consortium of 12 institutes to create parallel corpora of Indian Languages. IITB Part- Marathi Hindi corpora. Co-PI with Prof. Malhar Kulkarni. Ongoing.
Central Institute of Indian Languages 2008	Sanskrit Wordnet	October 2008 to September 2011. Linkage of Synsets of Sanskrit with Hindi Wordnet, establishment of semantic relations, sense disambiguation and navigation algorithms. Linkage with Amarkosha. Co-PI with Prof. Malhar Kulkarni. Ongoing.
Ministry of IT 2008	Development and Integration of Wordnet of North East Languages (NE-Wordnet)	February 2009 to December 2011. Consortium of 5 institutes (IITB-lead) doing project on Linkage of Synsets of Assames, Bodo, Manipuri and Nepali with Hindi wordnet, establishment of semantic relations, sense disambiguation and navigation algorithms. Linkage with Amarkosha. Ongoing
IBM Faculty Award 2008	Tools for Indian Language Processing on UIMA Platform	February 2008 to January 2009. Integrating Indian Language processing resources and tools and semantic search capabilities with the UIMA Platform. Completed.

HP Labs, Bangalore 2008	Large Scale Application Development and Knowledge Dissemination in Natural Language Processing and Text Mining	January 2008 to December 2009. Work on Ontology, Organization of 3 day summer workshop on Ontology, NLP, IE,IR (ONII-2008). Completed

Table 7: MT @ IIT-B

5. ACTIVITIES AND ACHIEVEMENTS DURING THE 11th FIVE-YEAR PLAN (2007-2012)

The Social Sciences Division (SSD) includes the following units: Economic Research Unit (ERU), Economic Analysis Unit (EAU), Linguistic Research Unit (LRU), Planning Unit (PU), Population Studies Unit (PSU), Psychology Research Unit (PRU) and Sociological Research Unit (SRU). During the period (2007-2012) the Linguistic Research Unit of the Institute has continued with its innovative programmes of research in areas of (a) Cognitive Linguistics, (b) Corpus Linguistics and Language Technology, (c) Clinical Linguistics along with some research works in the areas of Sociolinguistics and Bengali Linguistics. A brief report of activities to be taken up during the Twelfth Plan period for each Unit is as follows.

There are at least five main research and development areas under which the Unit's future research programmes may be grouped, namely:

- **The interface between Linguistics and Cognitive Science:**

The interface between linguistics and the rapidly developing areas of cognitive science is a major focus of LRU's scientific planning for the 12th Plan Period. In particular, biaxial syntax, whole word morphology, substantives lexicology, bifocal translation theory, the semiotics and formal morpho-syntax of proper nouns are among the topics on the current wish list for the Plan Period.

- **Corpus Linguistics and Language Technology:** (CLLT)

Generation of speech and text corpora, generation of specialized text corpora, generation of domain-specific parallel corpora, processing of text and speech corpora, annotation of speech and text corpora, development of tools and techniques of language processing, generation of lexical and linguistic resources in electronic form, development of machine translation system from English to Bengali and form Hindi to Bengali, development of resources for text-to-speech in Bengali, design and develop electronic dictionary for Bengali and other Indian languages, development of usage-based grammars for Bengali, meaning recognition and word sense disambiguation, knowledge representation and machine learning, computer assisted first and second language learning, development of usage-based on-line dictionaries, development of graded vocabulary, empirical analysis of speech and text corpora for develop linguistic theories and principles, corpus-based English language teaching, developing WordNet for Bengali in parallel to Hindi and English WordNet, and development of digital corpus archive for Bengali and other Indian Languages, etc

- **Clinical Linguistics:**

Diagnostic approaches to speech pathological problems, habilitation of hearing impaired children, linguistic-cognitive tests on children with neuro-linguistic disorders and language impairment, designing appropriate methodology and upgraded assessment tools, adopting homogenized therapeutic approach is in analysis of speech sounds, development of test barriers.

- **Sociolinguistics**

Sociolinguistic dimensions of lexical and syntactic difficulties, mapping between full conceptualization system and its basic level kernel, place of English in the sociolinguistic fabric of India, study of language attitudes, language maintenance and language shift, measurement of bilingualism, language planning in multilingual society, analysis of folklores and folk language, cultivation of mother-tongue, language standardization, etc.

- **Bengali Linguistics**

Bengali phonetics, phonology, morphology, semantics, syntax, morpho-phonemics, morpho-semantics, semantic-syntax, semantic-pragmatics, lexicology, culture, Bengali sociolinguistics, field linguistics, discourse and pragmatics, conversation analysis, spoken text analysis, stylistics. [32]

6. Conclusion:

In this paper, I have presented a survey on developments of CLIR and MT systems for Indian languages and additionally given a brief idea about the ongoing current research activities, efforts put by the Government agencies, and approaches that have been used to develop machine translation systems. Research activities in Indian languages are now beyond English as the source or the target language. Most recently, a number of new projects have been started for Indian languages with Govt. funding. It is concluded that lot of research is going in the area of NLP and number of machine translation systems have been developed and also regular efforts are being made for its improvements. CLIA and Sandhan are the latest developments in CLIR which is a result of success of MT in India.

7. References:

[1] DebasisMandal, Mayank Gupta, SandipanDandapat, Pratyush Banerjee, and SudeshnaSarkar "*Bengali and Hindi to English CLIR Evaluation*" Department of Computer Science and Engineering IIT Kharagpur, India – 721302Springer Berlin Heidelberg Series Volume 5152 Series ISSN 0302-9743 Pages pp 95-102.

[2] Manoj Kumar Chinnakotla, SagarRanadive, Pushpak Bhattacharyya and Om P. Damani"*Hindi and Marathi to English Cross Language Information Retrieval*" at CLEF 2007 Department of CSE IIT Bombay Mumbai, India Advances in Multilingual and Multimodal Information Retrieval Pages 111 - 118 Springer-Verlag Berlin, Heidelberg ©2008 ISBN: 978-3-540-85759-4

[3] Prasad Pingali and VasudevaVarma"*Hindi and Telugu to English Cross Language Information Retrieval*" at CLEF 2006 Language Technologies Research Centre IIIT, Hyderabad, India.Evaluation of Multilingual and Multi-modal Information Retrieval Lecture Notes in Computer Science, 2007, Volume 4730/2007, 35-42, DOI: 10.1007/978-3-540-74999-8_4

[4] Tan Xu1 and Douglas W. Oard1 "*FIRE-2008 at Maryland: English-Hindi CLIR*" College of Information Studies and 2CLIP Lab, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA.

[5] Mallammavreddy, Hanumanthapa."*ENGLISH TO KANNADA/TELUGU NAME TRANSLITERATION IN CLIR: A STATISTICAL APPROACH*" Department of Computer Science and Applications, Bangalore University, Bangalore-560 056, INDIA IJMI International Journal of Machine Intelligence ISSN: 0975–2927 & E-ISSN: 0975–9166, Volume 3, Issue 4, 2011, pp-340-345

[6] Mallamma V Reddy, Dr. M. Hanumanthappa"*Kannada and Telugu Native Languages to English Cross Language Information Retrieval*" Department of Computer Science and Applications, Bangalore

University, Bangalore, INDIA. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5) , 2011, 1876-1880

[7] Dr.S.Saraswathi, AsmaSiddhiqaa.M, Kalaimagal.K, Kalaiyarasi.M"*BiLingual Information Retrieval System for English and Tamil*" JOURNAL OF COMPUTING, VOLUME 2, ISSUE 4, APRIL 2010, ISSN 2151-9617

[8] Pingali V.V. Prasad Rao "*Recall Oriented Approaches for improved Indian Language Information Access*" Language Technologies Research Centre International Institute of Information Technology Hyderabad - 500 032, INDIA August 2009 Source iiit.ac.in

[9] Bharat Ram Ambati, Mridul Gupta, Samar Husain, DiptiMisra Sharma "*A high recall error identification tool for Hindi Treebank Validation*" Language Technologies Research Centre, International Institute of Information Technology Hyderabad, INDIA – 500032 Source http://www.lrec-conf.org/proceedings/lrec2010/pdf/673_Paper.pdf

[10] SivajiBandhyopadhyayAmitava Das PinakiBhaskar"*English Bengali Ad-hoc Monolingual Information Retrieval Task Result at FIRE 2008*" Department of Computer Science and EngineeringJadavpur University, Kolkata-700032, India Source www.amitavadas.com/Pub/Fire_2010.pdf

[11] Ashish Almeida, Pushpak Bhattacharyya "*Using Morphology to Improve Marathi Monolingual Information Retrieval*" IIT Bombay.Source http://www.isical.ac.in/~fire/paper/Ashish_almeida-IITB-fire2008.pdf

[12] Praveen Kumar, ShrikantKashyap, Ankush Mittal "*A Query Answering System for E-Learning Hindi Documents*" SOUTH ASIAN LANGUAGE REVIEW VOL.XIII, Nos 1&2, January-June,2003.

[13] GANAPATHIRAJU Madhavi, BALAKRISHNAN Mini, BALAKRISHNAN N., REDDY Raj "*Om: One tool for many (Indian) languages*" Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA Journal of Zhejiang University SCIENCE ISSN 1009-3095

[14] Hema A Murthy, C Chandra SekharC.S.Ramalingam, SrinivasChakravarthy"*A MULTIMODAL INDIAN LANGUAGE INTERFACE TO THE COMPUTER*" Dept. of Computer Science&Engineering IIT Madras, Chennai - 600 036 source www.elda.org/en/proj/scalla/SCALLA2004/murthyv2.pdf

[15] Dinesh Kumar Gurpreet Singh Josan"*Part of Speech Taggers for Morphologically Rich Indian Languages: A Survey*" Department of Information Technology DAV Institute of Engineering & Technology Jalandhar, Punjab, INDIA International Journal of Computer Applications (0975 – 8887) Volume 6– No.5, September 2010

[16] Sujoy Das AnuragSeetha M. Kumar J.L. Rana"*Post Translation Query Expansion using Hindi Word-Net for English-Hindi CLIR System*" source www.isical.ac.in/~fire/paper_2010/sujaydas-manit-fire2010.pdf

[17] Jagadeesh Jagarlamudi, A. Kumaran "Cross-Lingual Information Retrieval System for Indian Languages" Advances in Multilingual and Multimodal Information Retrieval Pages 80 - 87 Springer-Verlag Berlin, Heidelberg 2008 ISBN: 978-3-540-85759-4

[18] A.P.SivaKumar, Dr.P.Premchand, Dr.A.Govardhan "Indian Languages IR using Latent Semantic Indexing" International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, August 2011 DOI : 10.5121/ijcsit.2011.3419

[19] Ranbeer Makin, Nikita Pandey, Prasad Pingali and Vasudeva Varma "Approximate String Matching Techniques for Effective CLIR among Indian Languages" Proceedings of the 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory
Pages 430 - 437 Springer-Verlag Berlin, Heidelberg 2007 ISBN: 978-3-540-73399-7

[20] Sethuramalingam S, Anil Kumar Singh, Pradeep Dasigi and Vasudeva Varma "Using Fuzzy String Search Based on Surface Similarity" Proceedings of the 32nd international ACM SIGIR

conference on Research and development in information retrieval Pages 682-683 ACM New York, NY, USA 2009 ISBN: 978-1-60558-483-6

[21] CDAC –Noida Web Literature http://www.cdacnoida.in/snlp/ongoing_projects/CLIR.asp

[22] R.M.K. Sinha, —An Engineering Perspective of Machine Translation: AnglaBharti-II and AnuBharti-II Architectures, Proceedings of International Symposium on Machine Translation, NLP and Translation Support System (iSTRANS- 2004), November 17-19, 2004, Tata Mc Graw Hill, New Delhi.

[23] AnglaBharti Mission: http://www.cse.iitk.ac.in/users/rmk/mission/mission.htm

[24] Durgesh Rao. 2001. Machine Translation in India: A Brief Survey. In —Proceedings of SCALLA 2001 Conference, Banglaore, India.

[25] Sivaji Bandyopadhyay. 2000. State and Role of Machine Translation in India. Machine Translation Review, 11: 25-27.

[26] Bharati, Akshar, Vineet Chaitanya, Amba P Kulkarni, and Rajeev Sangal (1997), "Anusaaraka: Machine Translation in Stages", Vivek: A Quarterly in Artificial Intelligence, Vol. 10, No.3, pp. 22-25.

[27] TDIL Resource center: Anuvadaksh (English to Indian Language Machine Translation System) http://tdil-dc.in/index.php?option=com_vertical&parentid=72&lang=en

[28] CDAC-Resource Center: Machine Translation http://pune.cdac.in/html/gist/research-areas/nlp_mt.aspx

[29] Machine Translation Research Lab: IIIT Hyderabad http://ltrc.iiit.ac.in/MachineTrans/

[30] Creation of Machine Translation Tools and Resources: Amrita Vishwa Vidyapeetham http://www.amrita.edu/cen/computational.php

[31] Machine translation: Sponsored Research Projects @IIT-B: http://www.cse.iitb.ac.in/~pb/sponsor.html

[32] Eleventh Five Year Plan (2007-2012) Activity Report And Twelfth Five Year Plan (2012-2017) Projections: Planning Commission India. Web Source planningcommission.nic.in