Million Books to Web

Technological Challenges and Research Issues

N. Balakrishnan[1], Raj Reddy[2],
Madhavi Ganapathiraju[2] and Hemant R Gogineni[3]
[1]*Indian Institute of Science, Bangalore, 560012, India.*
[2]*Carnegie Mellon University, Pittsburgh 15213, USA.*
[3]*International Institute of Information Technology, Hyderabad, India.*
balki@serc.iisc.ernet.in, ***rr+@cmu.edu***, ***madhavi@cs.cmu.edu***, ***hemant@cs.cmu.edu***

Abstract

The fires of Alexandria irrevocably severed our access to many of the works of the ancients. The introduction of printing technology made several Indian and Chinese knowledge disseminated by word of mouth and on palm leaves virtually disappear or inaccessible. We have learnt from the past and more recently from Afghanistan, Iran and more recently in India, that the new cultural revolutions are edifices built by destroying the past irrevocably. Later revolutions seek solace in attempting to preserve what was destroyed. We need to preserve our heritage independent of the political and social ups and downs. A single wanton act of destruction can destroy an entire line of heritage. The Universal Digital Library Initiative is our humble attempt in preserving our heritage from intentional and unintentional acts of destruction.

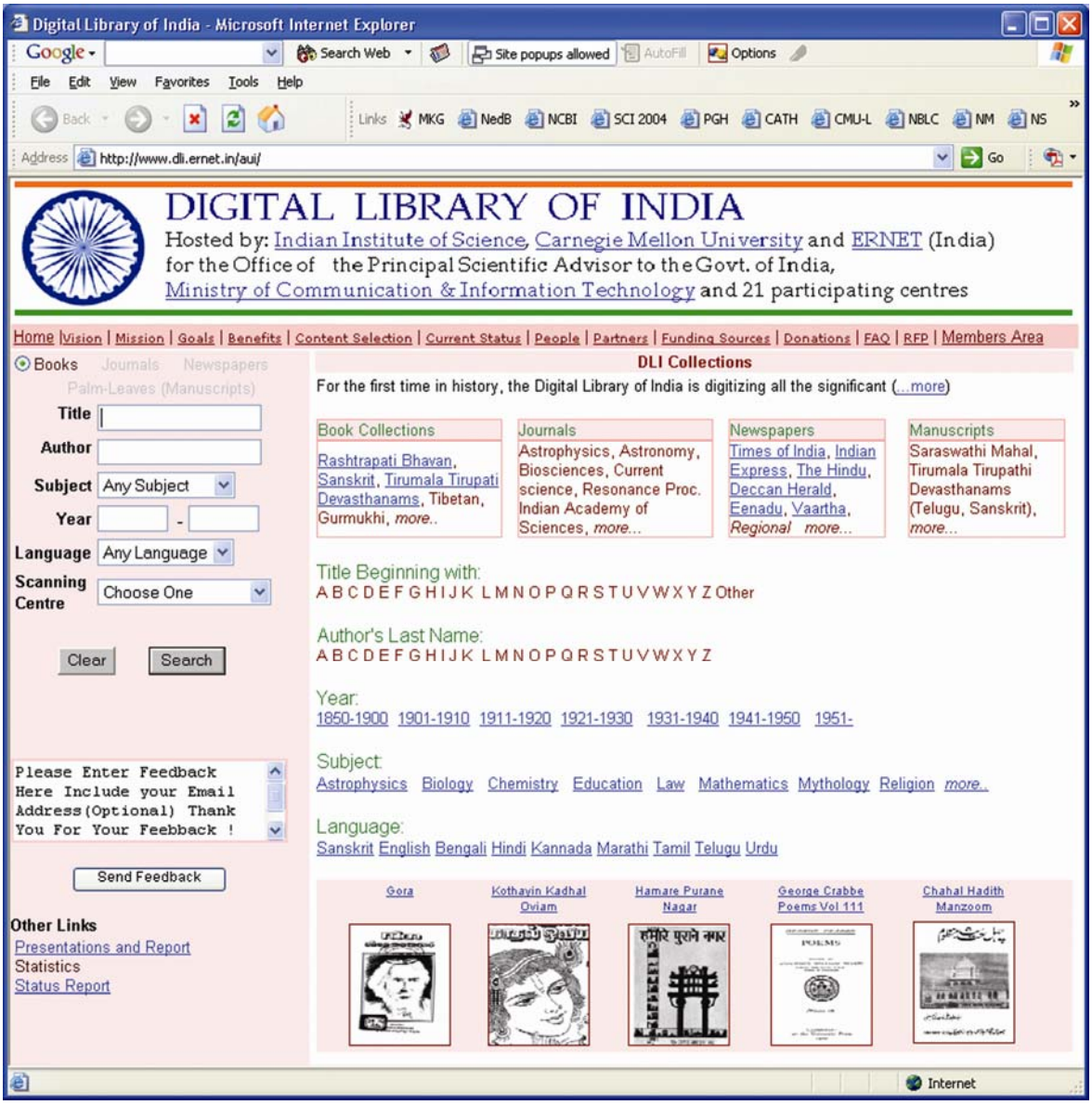
In this paper, we address the technological challenges and the research issues in the Million Books to the Web Project (MBP) that was initiated due to the grandiose vision of Prof Raj Reddy.

1. Introduction

In a thousand years from now, only a few of the paper documents available today will survive the ravages of deterioration, loss, and outright destruction. Many other works still in existence today are rare, and only accessible to a small population of scholars and collectors at specific geographic locations. Contrary to the popular beliefs, the libraries, museums, and publishers do not routinely maintain broadly comprehensive archives of the considered works of man. No one can afford to protect our heritage and make them available to every one, unless the *archive is digital*.

With this as a back drop, the Carnegie Mellon University (CMU) and the Indian Institute of Science (IISc) embarked upon a Technology Driven Mission. It was also decided to involve all stake holders at every stage and make the list of stake holders inclusive rather than exclusive. With this vision, many academic, religious and Government organizations, totaling to about 21 “Content Creation Centres” have been started under the Digital Library of India

Initiative- supported first by the Office of the Principal Scientific Advisor to Government of India and thereafter by the Ministry of Communication and Information Technology. A first pilot project with to scan around 10,000 books was initiated at CMU and then followed at IISc so that all the processes involved could be perfected.



Insert 1: Digital Library of India portal. http://www.dli.ernet.in/. Books are searchable by meta fields and full text. With the integration of information and language technologies, books from across different languages will be seamlessly searchable and readable. See a sample of search results in Insert 2.

There are more than 641 Digital Library Projects initiated across the world. Though India has the distinction of having held many well-attended Digital Library conferences and also many initiatives funded in this direction, almost all of them had a half-life of less than two years. The major reason for this is the fact that in India much of the discussions and energy were spent on deciding what to Digitize without realizing that the technology today makes it

possible to digitize and store almost all of Human Race's knowledge. This lesson was embedded in the planning of the CMU-IISc Digital Library of India Initiative.

This project has emanated from the visionary ideas and the missionary zeal of Prof Raj Reddy of Carnegie Mellon University, the first Asian to win the prestigious Turing Award and also a recipient of Padma Bhushan. The vision is to use the disruptive technologies like the ICT to preserve all the knowledge of the human race in Digital Form and make them searchable in a language-independent and location-independent way and ensure that rich cultures like India do not loose all their preserves during the transition from paper to bits and bytes, just like they lost out during the transition from palm leaves to paper. Prof Raj Reddy of Carnegie Mellon University is the Overall International Coordinator for this project. This project involves participation from many countries including USA and China. Prof N. Balakrishnan, Chairman, Division of Information Sciences at the Indian Institute of Science, is the Coordinator for the Indian efforts.

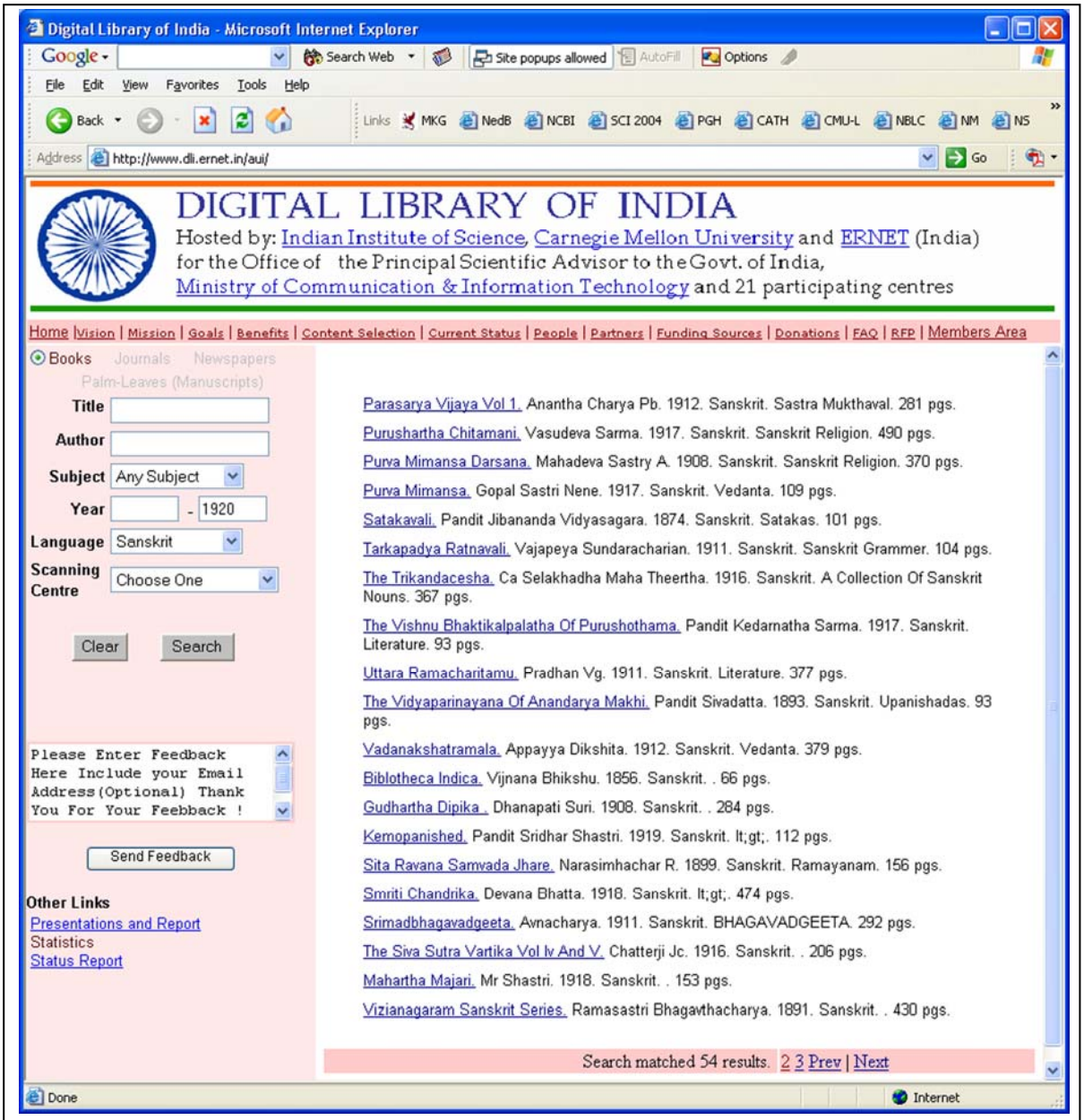
The approximate number of books ever produced by the human race would be around 100 Million. Each book (stored in the TIFF format) would require storage of around 50 Mbytes of storage space. All others form of knowledge in the form of voice, video and pictures would account for around 10 times the storage needed for the books. . Hence, on a rough estimate, it can be seen that the storage of the entire human race's knowledge is around 550 Tera bytes which in today's technology levels is not insurmountable. A single drive that can be attached to a laptop today can have a storage capacity of around 250GB, or almost 50,000 books. It is worth noting that a medium sized College Library will have only around 50,000 books-almost ten times more than what a human being would refer to in one's life time. With the storage capacity doubling every year, we could expect that in ten year's time, the capacity of a single drive in the laptop will be adequate to store 50 Million books. Hence, it is not the technology that would limit the Digitization, but the shear act of scanning and taking the books and manuscripts and other forms of knowledge to the web.

Today, most of the work of humans – be it books or music or movies, are born digital. Hence if we make the Digital Library with a proper frame work and architecture, it would become possible to make the library current any time. Only the fragility of technology will be an issue and proper planning and phased investments can sort this out.

Another important ingredient enshrined in the planning of the Digital Library is the fact that the cost of selection of books is more expensive than the scanning and storage cost. In order to make sure that we scan books and materials of relevance, instead of selecting individual items, active libraries are chosen and everything in the chosen library beyond the copyright has been taken up for digitization.

The vision of the Digital Library is thus to store everything that the human race ever produced. As part of this vision, a mission to digitize 1 Million Books and make them freely available was taken up by CMU and IISc with a large number of Indian partners.

This Digital Library is also intended to be a test bed for Indian language Research since more than 10,000 books will be available in all major Indian languages. The Digital Libray in India is also intended to be a leading and contributing partner to worldwide efforts in making knowledge available for free.



Insert 2: **Results of a search for Sanskrit books.** The search criteria on the books may be given in the left pane. The design of the search system allows a "Browsable Searchable" front end to the digital library repository. When the user is not aware of what he/she is looking for, the search can begin with a broad category such as the language or the subject. The results may gradually (incrementally) be refined by adding more criteria for search in the left pane. Clicking on any of the books from the results opens metadata, and the contents of the book for reading (See Insert 3).

Carnegie Mellon University, USA and the Indian Institute of Science, Bangalore have developed a complete solution comprising of scanning using the state-of the-art planetary

scanners, cropping, image processing to clean up the image, and software for OCR conversion of English documents, document format converters and search engines. With this technology, a full book of around 500 pages can be scanned without having to unbind the book into sheets, can be digitized and taken to the web in two hours. Presently, this technology is limited to bi-tonal black and white pictures scanned at 300 or 600 DPI. This programme in India was initiated by a seed grant from the Office of the Principal Scientific Advisor to the Government of India to the Indian Institute of Science. Subsequently, with the support from the Ministry of Communication and Information Technology and the Carnegie Mellon University, so far 21 Centres across the country spanning academia, religious Institutions and Government agencies have been created and more than 100 scanners are in operation. So far more than 90,000 books have been taken to the scanner, of which nearly 30,000 are in Indian languages. More than 41,000 books are available on the web. The books can be accessed from http://www.dli.ernet.in/. Recently, the Ministry of Communication and Information Technology had initiated actions to create four Mega Content Creation Centres in Hyderabad, Kolkatta, Allahabad and Noida. With this, the country will have a capacity to scan more than a million pages a day.

With the success of the joint efforts of the IISc- CMU collaboration, many other nations including China, Egypt, Poland Turkey and Mauritius have shown interest in participating in this effort, making it truly global effort in knowledge sharing. China has made significant progress and has taken the Million Books to the Web as national initiative.

In order to take a million books to the web, it is estimated that around 1000 man-years would be needed. Such massive efforts and the costs involved in selecting the books, in manpower and hardware make it prohibitively expensive for any one organization to carry out in isolation. Further, many research issues for the development of standards for language independent search engines, speech and character recognizers and data storage and interchange formats for scholarly communications are yet to be fully addressed. The Million Books to the Web Project (MBP) in fact acts today as a catalyst for research in Indian Languages. These research efforts are described below.

Digital Library of India is an effort to bring the advances in communication and information technologies towards preserving the rich Indian heritage present in the form of literature, art and manuscripts by converting them into digital form. This would not only extend the life of the material medium of storage, but would also be accessible to anyone anytime anywhere. There are more than twenty centres participating in the digitization efforts, each centre bringing its own unique collection of literature into the digital library. Authors are forthcoming in contributing their books to the digital library that is available for free to anyone. DLI has demonstrated the use of {\it book mobile} that is more like {\it sanchara grandhalaya} with the difference that it can carry a hundred thousand books and does not require the borrower to return the book. The DLI has now made available close to ninety thousand books online that are retrievable with the meta data fields. English books are

available for full text search. The primary goal of DLI, apart from making the books available online, is to make them available in fully functional form. The DLI is not simply a static repository of books---it has made possible bringing home the language and information processing technologies for Indian languages. Some of a major impact contributions to the Indian language information technologies are: (1) a machine translation system that we call Good-Enough Translation (GET-across) system (2) Om transliteration system that is an integral component of all the other systems (3) text editor for Indian language available for everyone (4) OCR for Indian languages (5) Search engine for Indian language texts.

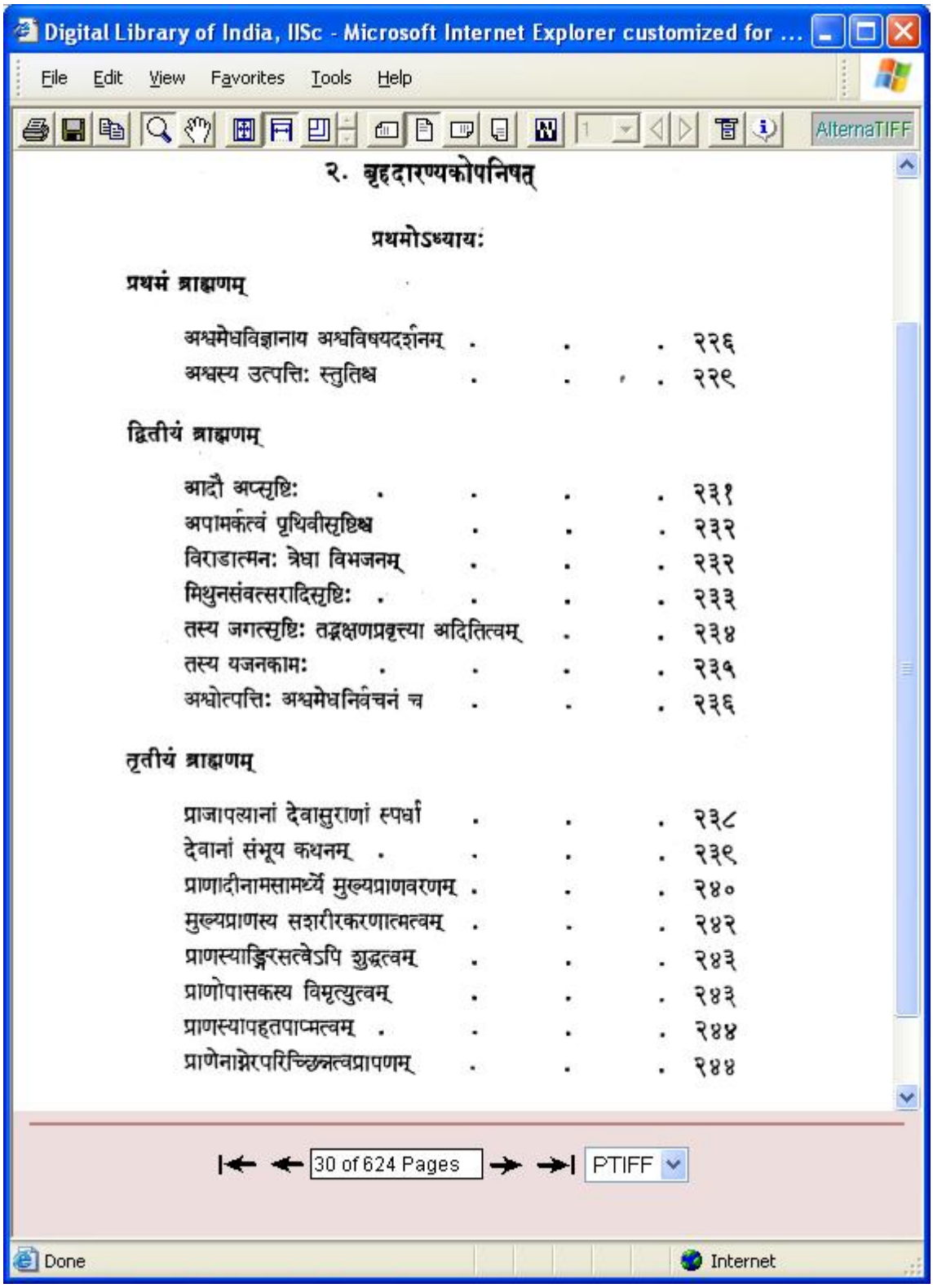
2. Mission

The mission of Universal Digital Library is to create a portal which will foster creativity and free access to all human knowledge. As a first step in realizing this mission, it is proposed to create the Universal Digital Library with a free-to-read, searchable collection of one million books, available to everyone over the Internet. Digital Library of India which is a partner of the UDL, aims to make at least 1 Millon books predominantly in Indian languages by the end of 2005. DLI portal is an aggregation of all the digital contents created by other digital library initiatives in India. Very soon it is expected that this portal would provide a gateway to Indian Digital Libraries in science, arts, culture, music, movies, traditional medicine, palm leaves and many more.

Typical large high-school libraries house fewer than 30,000 volumes. Most libraries in the world have less than a million volumes. The total number of different titles index in OCLC's WorldCat is about 48 million. One million books therefore, is more than the holdings of most high-schools and is equivalent to the libraries at many universities and represents a useful fraction of all available books.

One of the goals of DLI is to provide support for full text indexing and searching based on optical character recognition (OCR) technologies where available. The availability of online search allows users to locate relevant information quickly and reliably, thus enhancing the users success in their research endeavors. The resource, which is available 24 hours a day 7 days a week (24x7) would also provide an excellent testbed for language processing research in areas such as machine translation, OCR, summarization, speech and handwriting recognition, intelligent indexing and information retrieval in Indian languages.

DLI is soon to be mirrored at several locations worldwide, and would also partner with many country specific digital libraries such as Universal Library spearheaded by Professor Raj Reddy and Carnegie Mellon University (http://www.ulib.org).



Insert 3: **Book reader interface.** All scanned books are available for reading on the web. They are available in processed (denoised, deskewed) TIFF format. Where OCR is available, the books are available in HTML and plain text formats besides the image format available for books of all kinds.

3. Book to the Web: The Process

3.1. Content Selection and copyright

The Million Book Project envisages to develop a collection of one million digital books by adopting a staged approach as described below. The Million Book project will adhere to the copyright law.

Most of the centres that have been chosen for content creation have access to many active libraries and also contents that relate to the Indian Heritage. In order to avoid duplication across the centres, the centres select the books and check on a centrally located data base to see any other centre had scanned the book. Next, the search is done to see if the copy right on this book is renewed or still active. A database meticulously prepared by Dr Michael Lesk♦ is used for the copy right verification. Only those books that pass the above two tests are scanned [http://dli.ernet.in/mindex.htm and http://revati.dli.ernet.in/BOOK-SEARCH/copyrenew.html]

The OCLC data base is then searched for the availability of meta data for these books. This is useful mostly for English books. A meta data file format that is specially tailor-made for the UDL Project and has fields that can be entered in any Indian language has been developed. This meta data format is in fact a superset of the Dublin Core Format, making it possible for meta data mining from the publicly available software developed for English.

Preliminary discussions with OCLC as a host for a registry of scanned items and the meta data are underway. Certain key projects, such as the Making of America project, are already represented in the OCLC database as digital books. Other large digitization projects may require some data entry of their content in order to avoid duplication.

Best books approach

It is well known that more than 90%, which are within copyright, are also out of print. In a way, the copy right law has become a dog in the manger. However, many publishers and authors agree to give up their copy right for not for profit use like the Million Books Project. A previous study done at Carnegie Mellon University Libraries indicates that 22% of publishers granted permission for scanning and mounting their works on the web. The materials in the study at CMU were a random sample of Carnegie Mellon Libraries' books and included a broad range of dates, publishers, and books in and out of print status. Numerous difficulties from out of business publishers, lack of publisher records, return of copyright to authors, and other circumstances were encountered and ways to overcome these have been evolved. This experience has also been exploited at the State and City Central Libraries in the state of Andhra Pradesh in India with much success. In India as well, the project sought the permission of the publishers to scan in copyright.

♦ Dr. Micheal Lesk is a Professor at Rutgers University, and one of the pioneers in Digital Library and a Director of UDL Project.

OCLC owns a database of books from the latest edition of Books for College Libraries (BCL). It contains about 50,000 titles. A 22% success rate in clearing copyright would result in 10,000 of the best books for college students being included in the project. Clearing copyright is labor intensive and expensive. Bradd Burningham's recent article estimated those costs ("Copyright Permissions" in Journal of Interlibrary Loan, Document Delivery, and Information Supply, 11:2 (2000), 95-111). The BCL database, however, will allow for sorting by publisher so that permission requests can contain the names of several books. A quick sample indicates that as many as 25,000 publishers may be represented there. Despite the expense, this commitment to quality should be attempted. Carnegie Mellon University Libraries will seek private foundation funding to undertake this project.

In India, there is also a move to create a Consortium for Compensating for Creative Contents through which a balanced view of the copy right law is being evolved to benefit both the creator and the consumer of scientific and other literary work.

Publishers increasingly see that digital presentation of their works can attract buyers. They are interested in exploring ways in which their out of print titles may be returned to profitability. Continued work with publishers through the course of this project may attract many of them to it. That would be most beneficial in enriching the content to be made available.

3.2. Scanning, Cropping, Quality Control, OCR and Indexing

The selected books are scanned using any of the three types of scanners that are available to the UDL centres depending on the size, color and type of manuscripts. The Minolta 7000 scanner is used for text books in black and white and is capable of scanning upto 10,000 pages a day. The Zeutschel Omniscan TT 5000 Scanners are used for high volume scanning while the AVA3+ Scanners are used for low speed color manuscripts. The scanned images are in the TIFF format which are then processed for deskewing, speckle removal and image cropping using SCANFIX. The processed images are then OCRed using ABBY Fine Reader software if the text is in English. An in house developed distributed search engine which is extremely fast for searching the text and displaying the corresponding scanned images is used to for indexing. The entire architecture for the naming of the files and the location of books has been developed exclusively for this project and this is scalable easily to several tera bytes of data and millions of books, journals and manuscripts.

4. Om: Indian language text processing

4.1. Om transliteration scheme

Om is a transliteration scheme for typing Indian languages using the standard keyboard. It has been designed with phonetic mappings such that it is easy to remember. Om transliteration is largely based on mapping scheme developed for Indian Language Transliteration (ITRANS) package. ITRANS is a carefully designed scheme that has been in use for many years now. Om mapping is meant to add many more features to enhance the usability and readability, and

has been designed on the following principles: (i) easy readability (ii) case-insensitive mapping: while preserving readability, this feature allows the use of standard natural language processing tools for parsing and information retrieval to be directly applied to the Indian language Texts and (iii) phonetic mapping, as much as possible. This makes it easier for the user to remember the key combinations for different Indian characters ASCII representation may be used simply as a means of typing the text with standard keyboard, which is then mapped to the Indian language fonts for display or converted to any other format such as Unicode for storing. In addition, the case insensitive phonetic mapping is also highly readable by itself in the English script. This is of particular importance since often people can only speak fluently and understand the language but cannot read the script. This holds true for many people of the current young generation who study in schools where medium of instruction is English, or are educated in a country or state other than their own. India being a multi-lingual country, and inter-mixed population, often the people can speak and understand more than one Indian language and also English. Hence even in the absence of Om to native font converters, people around the globe can type and publish texts in Om scheme which can be read and understood by many even when they cannot read native script. The readability criterion that is benefitted from the case-insensitive phonetic mapping proves very useful. The Om mapping tables for many Indian languages are shown at www.dli.ernet.in/Om/. The table also shows the ITRANS mapping for the characters, and some sample Om texts.

4.2. Transliteration tool

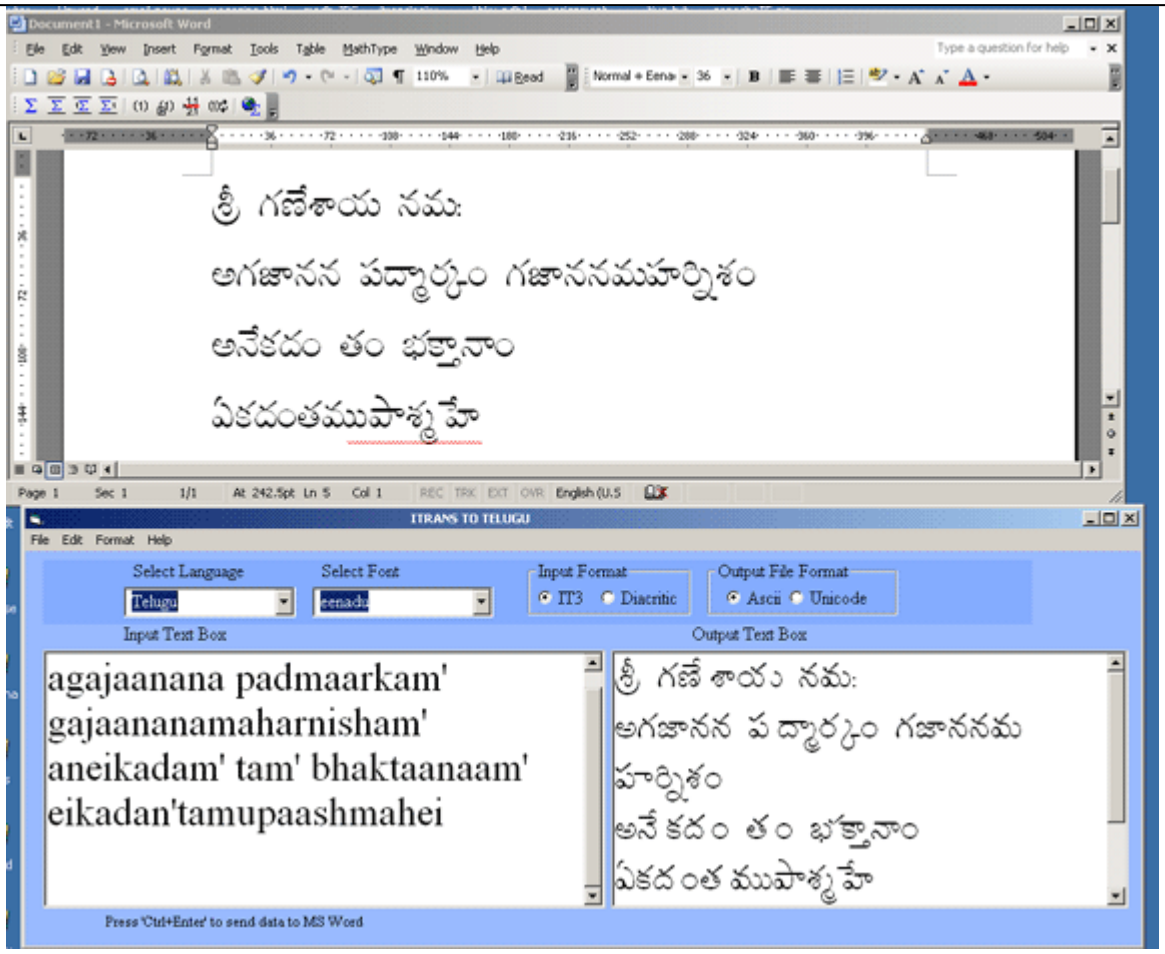
An integrated transliteration package that accepts Om ASCII keystrokes as input and maps them to native fonts has been developed. The script in any one of the chose true type fonts is sent to MS word for further formatting and layout options. Since the Om scheme is common to all the Indian languages, the display of the text can be converted between the supported languages by a choosing it on the menu. The text may also be saved in plain ASCII and Unicode formats. The tool also integrates with email clients on the windows platform. A web-interface with similar functionality has also been developed. The text may be saved as Om text, native font text or in Unicode. This does not support formatting explicitly but can be independently opened in MS Word like applications for such functionality.

A snapshot of the tool is shown in Insert 4.

4.3. Features

Easy support for new languages

A mapping table between Om symbols and the glyphs of the font of the new language is required. Once this is provided, it is only a matter of a few minutes to integrate this new language into the package. All the other features of transliteration to other languages and use of word-editing features of Microsoft word are avialble after the integration of the new font into the package. Currently, the Om transliteration package supports eight Indian languages.



Insert 4: **Om transliteration tool**. Allows typing Indian language text with ASCII keyboard following Om transliteration mapping. Currently supports 8 Indian languages Assamese, Hindi, Kannada, Malayalam, Marathi, Sanskrit, Tamil, and Telugu. Tool integrates with Microsoft Winword unifying its capabilities for Indian language digital documents. The tool is in use content creation and is available for download as well as hosting for free at www.dli.ernet.in/Om/

Key in the input as we speak

The most notable feature of the OM transliteration package is we can key in the input data just the way it sounds when we speak. For example if we have to key in ‘Bharat’ just type ‘bhaarat’.

Uses lowercase English alphabets and some special characters

The use of lowercase letters provides awesome power to language modeling tools such as stemmer, translation etc. The special characters used in OM are ‘, *, ~,

Switch between the languages at the click of a mouse

The option to choose any language and font is incorporated in the interface of OM by which switching from one language or font to the other is made easy.

Saves the output in ASCII and Unicode format

The file menu of the interface provides an option to save the input as well as the output, so that the user can import it later for future use.

Exchange email in Indian languages

This feature lets the user to send electronic mail in plain text in Indian languages

Integration with Microsoft Winword (MSWord)

The output can be exported to MSWord allowing users to take advantage of all the features in MSWord provided this application is present in the user's computer.

Web Interface

For those who wish to create content using a web interface, without the need to install the package locally, a java based web interface is also available. ({\tt http://swati.dli.ernet.in/om} and {\tt http://www.cs.cmu.edu/madhavi/Om}). The web interfce creates the output in plain text format, which may be opened in MSWord with the appropriate font selection, thereby using the full functionality of MSWord for the Indian language text editing.

Freely available for download and hosting

The Om transliteration mapping and the integrated editor have been used extensively for data entry for applications such as machine translation and optical character recognition. It has also been used purely for content creation by outside community. An example may be seen at the magazine section of www.telugumn.org/ where the story of Ramayanam and also some Slokas have been created using this software. The mapping scheme an the integrated editor are available in open source at www.dli.ernet.in/Om/, and will also be provided for hosting on any other site free of cost or use, such as done at {\tt http://www.telugumn.org/}.

5. OCR in Indian Languages- Kannada:

The MBP Project provides a phenomenally large amount of data for training and testing of OCRs in Indian Languages. Many of the contents have been manually entered besides scanned images for this purpose. Using this extremely large repertoire of data, under the MBP Project , a Kannada OCR had been developed. The Indian language OCRs poss a grand challenge since the number of characters to be recognized is an order of magnitude higher than the mere 26 needed in English. Many Indian languages have more than 300 characters.

The first block of the OCR is the segmentation algorithm that segments lines, characters and within the character a 32 X 32 block to identify the different key strokes that make up the character. These take care of the morphological dilation, base character, vowel modifiers and consonant conjuncts. Base characters are then normalized to 32X32 and the consonant conjuncts as well as modifiers are resized to a 16X16 matrix. Through a series of signal processing algorithms using DCT and KLT, the features are extracted. Structural features

include aspect ratio, stroke at different orientations, height of the segment in the top zone and the width of the character in the middle zone

A Neural Network based classifier is then used for training with the extracted feature vectors and testing. The current level of accuracy that we get is around 96-97% on clean documents scanned at 400 dots per inch. This accuracy falls to 40-50 % if the image is of bad quality. Efforts are underway to improve the accuracy of the OCR further with better segmentation and enhanced training. This OCR can easily be extended to any other Indian Language.

6. Machine translation: GET-Across

Seed database creation

The Example Based Machine Translation Developed under the MBP, is "Good Enough Translation Scheme" which can be built in any language in less than a month. This is not a perfect translation which will require enormous amount of development time and is never perfect. This translation assumes that the reader has certain level of intelligence and is able to understand the context and the meaning even if the translation is grammatically incorrect. This is in fact a semantic translator. This is good enough if one is using the present day search engines such as Google which in any case look for key words and remove any semblance of grammer by removing stop words and by stemming.

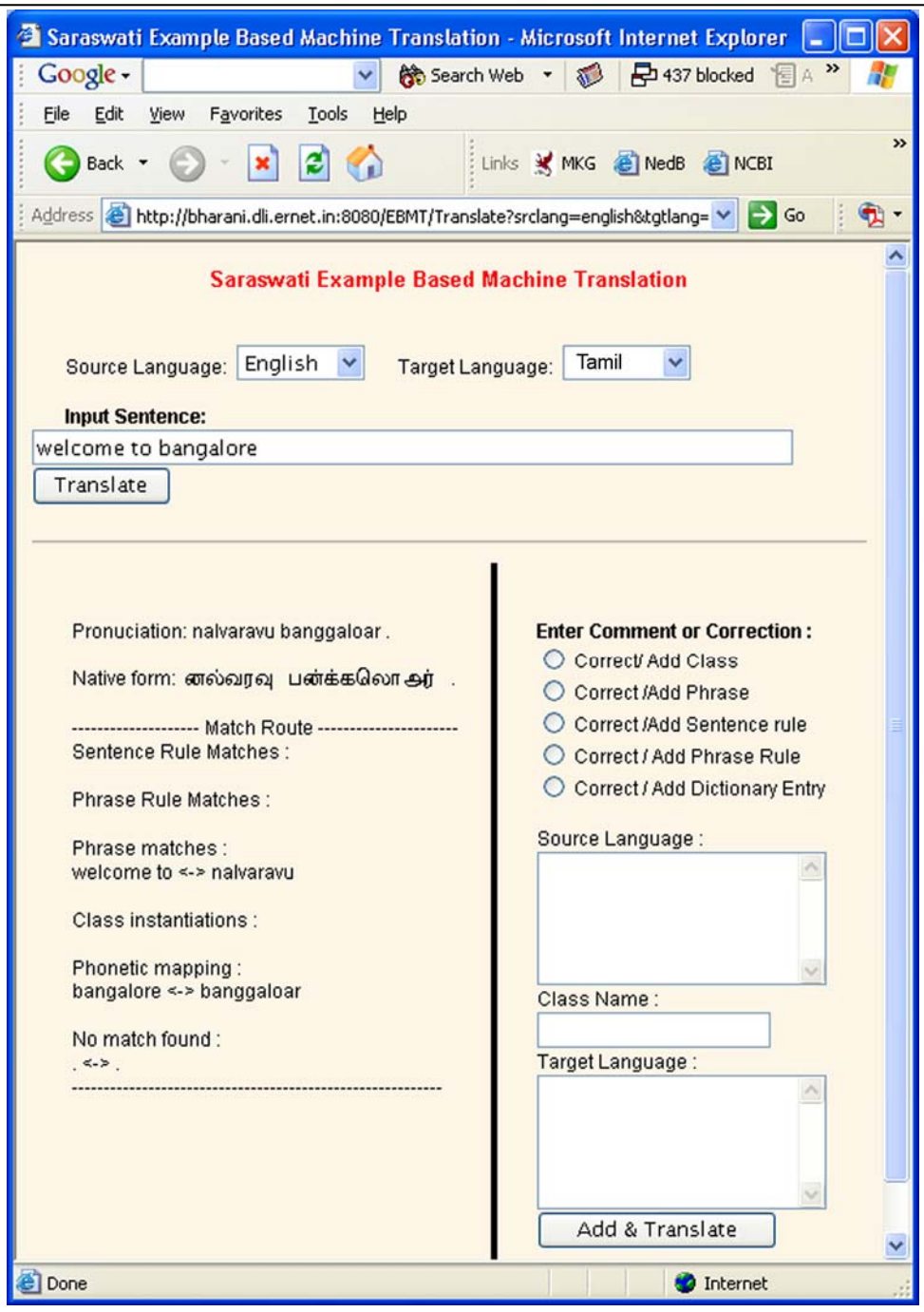
We first created a Universal Dictionary which has the words and their meanings in all the European and 6 Indian Languages. This dictionary base can be used as an aid for translation by merely replacing key words and frequently occurring words by their meaning in the target language. This, in our experience is a first level input to decide if the document is of use for further reading.

The Example Based Machine Translation has a set of 75000 most commonly spoken sentences that are originally available in English. These sentences have been manually translated into three of the target Indian languages, namely Hindi, Kannada and Tamil. Bilingual word and phrase dictionaries between these target languages and English of over 25,000 entries were also created manually. The Artificial Intelligence engine learns from these examples and provides a good enough translation by looking for the logest match at the sentence, phrase and word level.

While Indian languages are phonetic languages, English is not phonetic. In order to display English words in native language where required, a pronunciation dictionary is used. This pronunciation dictionary is created from CMU Dictionary by mapping its phonetic table to Om Transliteration Scheme

.

Users are encouraged to provide feedback to the Saraswati - the Machine Translation system to add or correct: **Dictionary entries, phrase translations, word order, pronunciation and translation rules.**



Insert 5: **Web interface of the Saraswati:** GET-across (Good enough translation) system, named Saraswati. The translated sentence is shown in target language, along with a transliteration in English (or the source language). The rules of translation that map the sentence in source language to target language are also shown to the user. This allows the user to be able to give a feedback to improve the correction. The corrections or feedback may be entered in the right side pane. The web-entered rules go through a moderator's approval to be included the main translation database.

Corpus entries such as dictionary, pronunciation and phrase translations are added to the corpus through human mediation. In case of indirect information such word-ordering rules, automatic rule inferring capability is being built in order to increase the accuracy of the machine translation. The EMBT is available currently for English to Hindi, Kannada and Tamil at http://bharani.dli.ernet.in/ebmt/

6.1. Translation

The translation system is developed in an iterative method. To begin with, translation was done by looking up the sentences in the database. The system is then interactively improved by asking users for feedback on specific mistakes it made. Rules are inferred automatically from the user feedback.

6.2. Key features

Good enough translation: provide at least fragments of translation, where perfect complete translation is not available, so that the user can communicate something in the target language Internet based corpus creation: Language corpus in digital form is a rare resource for Indian languages. Particularly rare is a parallel corpus across English and Indian languages Infer generalized rules in order to be able to go from any-language to any-language from among Indian languages and English Training requirements are reasonable.

7. Conclusion

The MBP besides giving an opportunity to create freely browsable and searchable documents of value to the humanity, had also become a test bed for Indian Language research as well as in the development of search engines, clusters for data base storage and retrieval. It had also helped in creating a tight bondage in research in India in around 21 centres, spanning academia, government and the religious institutions. It is heartening to note that India – the Government and the Scientists have put India at the top of the world map in digital library. It is also fortunate that the First Citizen of India, His Excellency Dr APJ Abdul Kalam who himself is one of the contributors to this vision has personally taken interest to make the Rashtrapathi Bhavan as one of the major centres of the Digital Library- MBP. In fact, Rashtrapathi Bhavan alone has contributed to more than 2 million pages of the 30 Million pages that we have scanned so far.

8. Acknowledgments

Funding for the Digital Library of India is coming from multiple sources. The Office of the Principal Scientific Advisor to the Government of India is funding the project in the Indian Institute of Science. The Ministry of Communication and Information Technology (MCIT) is funding the project at various partner centres of the Digital Library of India. Various Centres have also pledged their local resources to make the Digital Library of India a reality. The individual centres would acknowledge and provide more details about the agencies from which they have received support in their websites. National Science Foundation (USA) is

providing funding for Scanners and Software research and development. The Carnegie Mellon University and the Indian Institute of Science have provided valuable human and other resources for the entire development of the programme and the training of people from other centres.

The entire MBP has grown with the active involvement of many people. Though it is difficult to acknowledge every one, it is our pleasure to acknowledge the contributions of the Directors of MBP from CMU- Dr Gloriana St Clair, Dr Mike Shamos, Dr Jaime Carbonell, researchers from CMU – Ed, Eric, Walt, Krishna, and from IISc- Srinivas, Mini, Pradeepa, Sheik, Anand and Jiju and Kiran.