LANGUAGE IN INDIA
Strength for Today and Bright Hope for Tomorrow
Volume 9 : 10 October 2009
ISSN 1930-2940

A Survey of the State of the Art in Punjabi Language Processing

Gurpreet Singh Lehal, Ph.D.

A Survey of the State of the Art in Punjabi Language Processing

Gurpreet Singh Lehal, Ph.D.

People throughout the world have been using computers and Internet in their own languages. So far, Indian users in general and Punjabi users in particular have been compelled to use them in English despite the dominance of Indian engineers and scientists in the IT world. Unless we support our own languages on the technology environments, it is impossible to use IT or internet to uplift and improve the socio-economic environment of our country.

There is a need for language based content and technology and we have to address it. The society at large can benefit from the Information Technology effectively if people can communicate with computers in their own languages. Barely 65 % of our population is literate, of which only an elite minority (~5%) can read, write, and speak the English language. This shuts out most of the Indian population from the worldwide web and its huge potential. Therefore it is essential to have an interface that uses not only the local language but also speech, to cater to the needs of the semi-literate and illiterate sections of the population.

A few government/educational organizations and some individual researchers have initiated programs for the technological development of Punjabi language. A survey of the work done for Punjabi language processing revealed the following facts:

- Scattered work has been done and there is little unification
- Scarcity of linguistic resources and open-source software
- Main organisations working are Punjabi University, Thapar University, CDAC Noida and IIIT Hyderabad
- Many individual researchers are also working on their own without any research support and have published their research in COLING, ICDAR, ICPR and other conferences
- The main sponsoring agency is Ministry of Communications and Information Technology (MCIT), which has been sponsoring research in areas such as OCR, grammar checker, machine translation etc. In 2000 MCIT setup a Resource Centre for Technical Development of Punjabi at Thapar University, which kick-started research in technical development of Punjabi and encouraged many other researchers in the region to start working in this field.

In this paper the technological development of Punjabi has been classified under certain heads and the research works under taken and successfully completed as well as the products developed are discussed in details.

Punjabi Fonts

Dr. Kulbir Singh Thind, can rightly be called the Father of Punjabi fonts as he was the first person to develop the Punjabi fonts. Dr Thind developed the first Punjabi font for Macintosh way back in 1984. He also developed the Punjabi fonts for MS Windows 3. In November 1995, he released the first version of Gurbani CD and the Punjabi fonts distributed with the CD became very popular. These fonts such as Anmol Lipi and Amrit Lipi are being used by people all the world. Others who contributed to development of Punjabi fonts include Hardeep Singh Pannu, Janmeja Singh Johl, Kirpal Singh Pannu, CDAC, Summit Information Technologies etc. Recently Paul Alan Grosse released a collection of 108 Gurmukhi fonts in 11 main families including Gurmukhi handwritten fonts[1].

Sukhjinder Singh Sidhu from UK developed Unicode compliant font, Saab, which was later adapted for Linux OS. Later Dr. Kulbir Thind also upgraded his Punjabi fonts to Unicode. These fonts are also available for free download from [2]. Ministry of Communications and Information Technology (MCIT) released a set of more than 100 Unicode compliant Punjabi fonts developed by CDAC, CyberSpace and Cadgraf in the language CD in 2007. The CD is available for free download from[3].
According to a rough estimate more than 100 Punjabi font families have been developed. Some of the popular fonts used in different parts of the world are:

North America – Anmol Lipi, Chatrik
UK and Europe – Sukhmani, Anmol Lipi
India – Gurmukhi, Joy, Satluj, Asees, Janmeja, Anmol Lipi

Font Convertors

More than 500 ASCII based fonts are currently available for Punjabi. The availability of too many fonts makes text recognition difficult as each font makes use of different keyboard mapping. To convert text encoded in one font to another font, font convertors have been developed.

Two notable contributions have been made by Janmeja Singh Johl[4] and Dr Gurpreet Singh Lehal. Dr Lehal also developed a convertor to convert Unicode text to any of the popular Punjabi 8bit font, such as Anmol, Satluj etc., and reverse. Sukhjinder Sidhu also developed GUCA, an application that is designed to convert ASCII encoded, font-based Gurmukhi text into Unicode. This application converts text based on Dr. Thind's fonts (e.g. AnmolLipi, GurbaniLipi fonts) into Unicode. The application is available for free download from [5]. Recently Punjabi University Patiala launched Gurmukhi typing pad, which can also be used to convert documents from Satluj and AnmolLipi fonts to Unicode.[6]

Punjabi Typing Tools

Punjabi typing is much more complex as compared to English typing, as 57 characters have to be typed on the standard QWERTY keyboard. One has to memorize the Punjabi characters corresponding to the English keys and search out each character and then worry whether to type it with SHIFT or without SHIFT. But recently, Predictive Roman-Gurmukhi transliteration techniques have been developed which have greatly eased Punjabi typing.

The first one to introduce this technique for Punjabi typing is Dr G. S. Lehal, who integrated it in the Punjabi word processor Akhar. The software generates script symbols that match the sound of the phonetically spelt word. The software uses artificial intelligence techniques and a rich collection of lexical resources to cleverly guess the most appropriate word/alphabet and presents the other phonetically similar words sorted on their relevance to the user in a suggestion box.

Similar facilities for simplifying Punjabi typing using predictive transliteration techniques are also now being freely provided online by Quillpad [7]. Punjabi University, Patiala has developed Gurmukhi typing pad, which can be used to create Unicode documents using popular the phonetic and Remington based keyboard layouts.[6]

Dictionary Tools

Punjabi has many dictionary tools. Some of the paper dictionaries have been converted into electronic dictionaries, while some have been specially made as electronic dictionaries. Many online dictionaries have also been developed.
Some of the popular electronic dictionaries are:

Punjabi Kosh

Punjabi Kosh is an English-to-Punjabi and Punjabi-to-English dictionary designed by Noah Hart. It allows keyboard entry and dictionary use in both languages. There are also a number of learning games and lessons for Punjabi students. It is a very useful tool for Punjabi learners. The dictionary is freely available at [8].

Punjabi Shabdkosh

Punjabi Shabdkosh is a Punjabi to English dictionary developed by Harwinder Singh Tiwana. The dictionary is freely available at [9].

Punjabi Dictionary by CDAC

An ISCII based Punjabi-English dictionary developed by CDAC is made available on the language CD freely made available by MCIT. The GIST typing tools have to be used for typing and searching for words in Punjabi.

Gur Shabad Ratanakar Mahankosh

Gur Shabad Ratanakar Mahankosh, popularly called as Mahankosh, is the first dictionary of Sikh Scripture and books on Sikh Religion. It is also a classical reference book of Sikh history, philosophy and contemporary Sikh states. The complete Mahankosh has been digitized by Bhai Baljinder Singh Rarewal and the pdf file is available for download from[10].

The following online dictionaries are available:

English-Punjabi Topic Dictionary by Punjabi University [11]

A topic dictionary for Punjabi learners organized into more than eighty categories such as adjectives, nouns, food, fruits, animals, months etc. has been developed by The Advanced Centre for Technical Development of Punjabi Language Literature and Culture, Punjabi University. The dictionary has around 3100 entries. Each word has an associated picture, Roman transliteration and English equivalent along with the pronunciation in Punjabi.

Punjabi-English Dictionary by Punjabi University [12]

The Punjabi-English paper dictionary developed by Punjabi University Patiala has been converted into electronic form after adding sound as well as displaying the Punjabi words both in Gurmukhi and Shahmukhi scripts. The dictionary has around 31,000 entries.

Punjabi-English and English-Punjabi Dictionary by Jasjit Singh Thind [13]

An online Punjabi-English and English-Punjabi dictionary has been provided by Jasjit Singh Thind on the website www.punjabionline.com.

Punjabi Encyclopedia and Gurbani Dictionary by Dr. Kulbir Singh Thind [14]

An powerful online search facility for simultaneously searching for any Punjabi word in Mahan Kosh Encyclopedia, Gurbani Dictionaries and Punjabi/English Dictionaries is provided on the website www.srigranth.org by Dr. Kulbir Singh Thind.

Spell Checkers/Grammar Checkers

CEDTI Mohali now CDAC (Mohali), developed the earliest spell checker for Punjabi on DOS and Windows platform. The spellchecker supported three Punjabi fonts. The main dictionary contains around 40,000 Punjabi words. User can also add new words in to his own dictionary. Later on TIET, Patiala developed a bilingual spellchecker for Punjabi and English, which could spell check documents encoded in more than 60 different Punjabi fonts.

Microsoft also included Punjabi spell checker in MSOffice 2003 supporting Unicode documents. The proofing tool had to be purchased separately. The spell checker developed by Dr G.S. Lehal [15] works on more than 125 popular Punjabi fonts, as well as Unicode documents, which are converted to 8 bit fonts. The spellchecker supports multiple dictionaries including Gurbani spellings.

The first and only Grammar Checker for Punjabi has been developed by Dr Mandeep Singh Gill under the guidance of Dr Gurpreet Singh Lehal at Punjabi University

Patiala [16]. The grammar checker is available online on the University website [17], for checking the grammatical errors of simple Punjabi sentences.

Word Processor

A number of commercial word processors have been developed for Punjabi. The earliest word processors to be developed were Sampadak [18] by CEDTI Mohali and Likhari [19] by Thapar University. Likhari provided a Punjabi spell checker, Punjabi font convertor and English-Punjabi administrative terms. Likhari supported both doc and ISCII based files but had no support for Unicode. Likhari also supported 60 popular Punjabi fonts for language processing such as spell checking and font conversion. Sampadak provided a Punjabi spell checker, but had no support for ISCII files and had support for limited Punjabi fonts. Another word processor, whose main purpose is to ease Punjabi typing, is PunjabiPad.

PunjabiPad is a shareware Punjabi wordprocessor, which makes Punjabi typing easy by providing a unique type as you speak layout that allows you to type Punjabi the same way you pronounce it. PunjabiPad is compatible with most of the popular wordprocessors and designing software. Another popular commercial word processor is Akhar, which has been developed by Dr. Gurpreet Singh Lehal.

Akhar provides many Punjabi typing tools including Romanised Punjabi typing utility. Besides a bilingual Spell Checker for Punjabi and English, English-Punjabi and Punjabi-English dictionaries have also been provided. Akhar supports more than one hundred twenty Punjabi fonts and forty Punjabi keyboard layouts. A Transliteration facility to transliterate Punjabi text from Gurmukhi to Devanagri and Roman has also been provided in Akhar.

Punjabi Teaching

A number of websites have been developed for imparting online Punjabi teaching [20-27]. Most of these websites are specially designed to provide children around the world with high quality e-learning Punjabi that facilitates their development of communication skills in listening, speaking, reading and writing Punjabi.

The website developed by Punjabi University, Patiala [20] is a comprehensive Punjabi language learning audio and visual delight. The site offers verbal help and guide on pronunciation with variations on the pronunciation of words that sound alike, but have several different meanings.

The website also includes a pictorial vocabulary of more than 3,000 words along with their pronunciation that are organised into 80 related topics such as animals, birds, colours, fruits and the days of the week. The website also includes games, tongue twisters and folk tales to peak and sustain user interest. Stories that make learning easy and interesting are nicely presented along with the text in Gurmukhi script and a corresponding English translation.

A set of talking stories, in which the user can click on any word or sentence of a story to get its meaning and pronunciation is another highlight of the website. Games like crossword puzzles, hanging man, recognising a word from its pronunciation, arranging letters in correct sequence are part of this Punjabi learning website.

The website developed by Sukhvinder Singh [21] contains some nice audio recordings of Gurmukhi alphabet. Some very interesting and interactive games for Punjabi learning have been presented in the website www.sikhpoint.com [22]. Jasjeet Singh Thind [23] has used audio for teaching Punjabi alphabets. Simple reading exercises for 2 letter and 3 letter words has also been provided. A topic based vocabulary divided into 30 categories such as fruits, animals, domestic articles etc., has also been developed.

A series of 6 simple and useful lessons for Punjabi learning have been developed by R. S. Dhillon [24]. The Punjabi learning lessons provided on www.sikhlink.com [25] have been divided into seven chapters and audio has been provided for teaching Punjabi pronunciation.

The site www.rajkaregakhalsa.com [26] has divided Punjabi lessons into 19 chapters that make Punjabi learning an easy step-by-step process. Another useful site for Punjabi learning has been developed by Paul Grosse [27]. This site teaches in an interactive manner how to write each Gurmukhi character and produce the sound it represents. One can print out the resources to improve learning and learn how to read Gurmukhi quickly.

Some commercial software have also been developed for Punjabi teaching. One such software is Punjabi Guru. It is an interactive and an audio-visual delight that guides one through animated, real-life situations encountered on a trip to India. Another popular software for Punjabi learning is iPanjabi, which is specially designed for children, parents, and teachers. The software can be used to become familiar with the shapes & sounds of the letters of the alphabet, learn pronunciation of simple and advanced Panjabi words, play word games and Increase proficiency in vocabulary. Another Punjabi learning software is Punjabi Prashikshak. It is an interactive audio-visual software which facilitates learning Punjabi Language Basics. Besides number of games, it has 19 lessons incorporating all the basic elements required to learn the language.

Digitization of Sri Guru Granth Sahib

Dr. Kulbir Singh Thind has spearheaded the development of Gurbani-CD project that led to the computerization of text of Sri Guru Granth Sahib and over time, by his efforts, accuracy of text of Sri Guru Granth Sahib has been brought closer to a level of perfection. Dr. Thind has also developed databases relating to Sri Guru Granth Sahib, for the use of Gurbani text on the internet and on the computer and then he developed many other specialized files of Sri Guru Granth Sahib text for the help of Sikh scholars. He also converted Sri Guru Granth Sahib files to Devnagri (Hindi) and did phonetic transliteration of text of Sri Guru Granth Sahib. He also organized sentence by sentence translation & phonetic transliteration of Sri Guru Granth along with

Gurmukhi as well as Devnagri (Hindi), for use on the Internet and computer. He also created specialized files to allow easy searching and printing of Gurbani Kirtan. Dr. Thind has contributed to the making of web pages over many sites regarding Sri Guru Granth Sahib translation and phonetic transliteration. He has also converted Gurbani related files into Unicode text based on the international standard.

Corpus and Corpus Management Tools

A Corpus is a collection of huge text incorporating various types of textual materials, including newspaper, fictions, scientific writings, literary writings, and so on in electronic form used for linguistic research. Corpus analysis seeks to further our understanding of language through the analysis of large quantities of naturally occurring data. The need for corpus for a language is multifarious. Starting from the preparation of a dictionary or lexicon to machine translation, corpus has become an inevitable resource for technological development of languages.

The first Punjabi corpus was developed by Central Institute of Indian languages (CIIL) Mysore. A three million word corpus for Punjabi was developed in 1990s. A 6.8 million word Punjabi corpus was developed at the Advanced Centre for Technical Development of Punjabi, Punjabi University, which was used to extract the statistics for Punjabi language teaching [28]. In 2006, CDAC Noida developed a 50,000 paged parallel Punjabi-Hindi corpus under the *Gyan Nidhi* project. The parallel corpus will be useful for building example based Punjabi-Hindi machine translation system.

The vast storage of text is of no use, if there is no tool to analyse and explore it. A corpus analysis tool, has been provided in Akhar, the Punjabi word processor. The tool can sort, alphabetize and give concordance and KWIC index for Punjabi texts in any popular Punjabi font. It can also calculate the frequency of words as well as characters. It can also perform bigram and trigram analysis. Thus such useful statistics such as most commonly used Punjabi words, Punjabi character frequency as well as common Punjabi words with multiple spellings can be generated in a matter of seconds. The most common bigram and trigram pairs can also be very easily generated. It is impossible to imagine solving any NLP problem without using such statistics.

Parts of Speech Tagging in Punjabi

Parts of speech tagging scheme tags a word with its parts of speech in a sentence. This is a necessary module in development of Natural Language Processing software such as machine translation, grammar checking, etc. A rule based part of speech tagger for Punjabi has been developed by Dr Mandeep Singh Gill and Dr Gurpreet Singh Lehal at Advanced Centre for Technical Development of Punjabi Language, Literature and Culture, Punjabi University and available for free use on the University website [29]. A HMM based statistical POS tagger for Punjabi is being developed at IIIT, Hyderabad.

Morphological Analyzer/Generator

Morphological analyzer and morphological generator are two essential and basic tools for building any language processing application for a natural language. A Morphological analyzer gives the morph analysis of a word i.e. for a given word a morphological analyzer will return its root word and word class along with other grammatical information depending upon its word class. Morphological generator does exactly the reverse of it, i.e. given a root word and grammatical information it will generate the word form of that root word.

A Punjabi Morphological Analyzer and Generator has been developed by the Advanced Centre for Technical Development of Punjabi Language, Literature and Culture, Punjabi University by Dr Mandeep Singh. The software has been included in the language CD launched by MCIT. Earlier IIIT Hyderabad had also developed a Punjabi Morph under the Anusarka project.

Machine Transliteration

Transliteration is the process of converting a word written in one language into another language. Transliteration attempts to be exact, so that an informed reader should be able to reconstruct the original spelling of unknown transliterated words. To achieve this objective, transliteration may define complex conventions for dealing with letters in a source script which do not correspond with letters in a goal script.

Gurmukhi and Shahmukhi

A unique feature of Punjabi is that it is written in two mutually incomprehensible scripts. In India, Punjabi language is written in Gurmukhi script, while in Pakistan it is written in Shahmukhi (Urdu) script. This has created a script wedge as majority of Punjabi speaking people in Pakistan cannot read Gurmukhi script, and similarly the majority of Punjabi speaking people in India cannot comprehend Shahmukhi script.

To break this script barrier, transliteration systems for transliterating Punjabi text written in Gurmukhi script to Shahmukhi and reverse have been developed. In case the language is same, then transliteration has to be more precise in the sense that spellings have to be maintained in the target language.

The earliest Gurmukhi-Shahmukhi transliteration systems have been developed by CDAC Pune and Abbas [30]. The transliteration system developed by CDAC is freely available on the Urdu language CD distributed by MCIT [31]. A shortcoming of these systems have been that they have not taken care of spellings of the transliterated words, as they performed rule based letter by letter conversion and had word transliteration accuracy of around 85% for Gurmukhi-Shahmukhi transliteration.

A high accuracy Gurmukhi to Shahmukhi Transliteration system with more than 98% word accuracy was developed at Punjabi University Patiala by the team lead by Dr. G. S. Lehal. The various challenges such as multiple/zero character mappings, variations in pronunciations and orthography and transliteration of proper nouns etc. have been handled by generating special rules and using various lexical resources such

Gurmukhi spell checker, Shahmukhi corpus, Gurmukhi-Shahmukhi transliteration dictionary, etc.

A corpus based, Shahmukhi-Gurmukhi transliteration system has also been developed at Punjabi University by Tejinder Singh Saini and Dr. G. S. Lehal. The system performs statistical analysis of Punjabi corpus and combines the artificial intelligence techniques to solve the missing diacritic and multiple character mapping problems [32].

The transliteration software developed by Punjabi University are also capable of converting complete websites from Gurmukhi to Shahmukhi and reverse. They are available for free usage on the University website [33-34]. The Gurmukhi-Shahmukhi and reverse transliteration systems developed by Abbas are also available for free use at the website [35].

Gurmukhi and Other Scripts

Rajesh Kumar and Dr G S Lehal of Punjabi University have developed Gtrans1.0, a Gurmukhi-Roman transliteration software. The software converts any Gurmukhi text to Roman script along with diactritic symbols. The software is available for free download at [36]. Vishal Goyal and G S Lehal [37]developed rule based Gurmukhi-Devnagri transliteration system, which was included in the Hindi-Punjabi machine translation system.

Machine Translation Tools

The first Punjabi to Hindi direct machine translation system was developed at IIIT Hyderabad in 1990s under the Anusaraka project headed by Dr. Rajeev Sangal. The translation is at the word level as the word order of Hindi and Punjabi languages is more or less same and that the grammatical function depends more on inflection than word order.

In 2009, the team of Dr G. S. Lehal, Dr G. S. Joshan [38] and Vishal Goyal at Punjabi University Patiala developed online Hindi-Punjabi and Punjabi-Hindi machine translation systems, which are freely available on the university website[39]. IIIT Hyderabad has also recently launched an online Punjabi-Hindi machine translation system[40]. CDAC Noida is also working for development of Hindi-Punjabi and reverse machine translation systems.

Prateek Bhatia and his team at Thapar University, Patiala are working on Punjabi Language Server which includes Punjabi-UNL Enconverter and UNL-Punjabi Deconverter. [41-42] The main objective of this project is to study Punjabi Language and make the first Punjabi Deconverter. This will convert UNL expressions in to Punjabi sentences. The team of Kamaljit Batra and Dr. G S Lehal is developing a Punjabi-English machine translation system for translation of legal documents from Punjabi to English[43].

Search Engines

A Google based search engine for Punjabi (www.punjabikhoj.com) has been developed by Punjabi University, Patiala. The search engine provides the user the facility of typing Gurmukhi text in Remington and Phonetic keyboard layouts, as well as using the onscreen key board. The user can search for his queries in Hindi and Shahmukhi (Urdu) documents too. User can also search for similar spelling and similar meaning words. The system is being upgraded to cross lingual information retrieval system, supporting Punjabi, Urdu and Hindi languages.

Gurbani Search Engine

The following offline and online search engines have been developed for searching text from Sri Guru Granth Sahib.

www.srigranth.org This website, developed by Dr. Kulbir Singh Thind and his team, provides an online search engine for Sri Guru Granth Sahib. One can do advanced searches of Sri Guru Granth Sahib & read hymns of Sri Guru Granth Sahib with many English/Punjabi translations and Teekas (commentaries in Punjabi) in an interactive way. One can perform searches in Punjabi, Hindi or English language.

SikhiToThemax This is an offline and online Gurbani search engine, which provides multiple search options in English and Gurmukhi. The user has the option of viewing each shabad in several modes such as simple view, poerpoint display, akhand paath display, hukamnama display and pdf display. The results can later be sorted and grouped to view them under subheadings of the results generated. The offline search engine is available for free download at [44].

Isher Micromedia Software This search tool, developed Bhai Baljinder Singh Rarewal, Dr. Kulbir Singh Thind and Dr. Gurcharan Singh can quickly search text of Sri Guru Granth Sahib and Vaars/Kabits Bhai Gurdas with utility to find a Word or Line. The software gives an option to do searches for a particular Raag, or Author/Authors or to separate Gurbani by these attributes. The software also provides options to read the three TEEKAS of Sri Guru Granth Sahib simultaneously. The software is available for free download at [45].

Optical Character Recognition (OCR)

An OCR is a software, which enables a computer to translate character images into editable text. The advantage of OCR is that one can enter the printed and handwritten document into the computer without retyping it, the OCR will automatically read and convert the document. This results in a large saving in time and money. Recognition of Indian language scripts is a challenging problem as compared to Latin script.

A robust, multi-font Gurmukhi OCR has been developed by Dr Gurpreet Singh Lehal and his team at Punjabi University Patiala [46-47]. The OCR can recognize Gurmukhi text printed in any of the common Gurmukhi fonts with more than 97% recognition accuracy at character level. Dr. Chandan Singh, Dr. Indu Chhabra[48] and Dr. Renu Dhir have also been working on different phases related to development of Gurmukhi

OCR. Munish Jindal, Dr. R.K. Sharma and Dr. G. S. Lehal have also been working for development of Gurmukhi OCR capable of processing bad quality documents[49-50].

An online OCR system recognizes the handwritten text as it is being written on a tablet. The team of Anuj Sharma, Rajesh Kumar and R.K. Sharma [51] have been working at Thapar University for development of online Gurmukhi OCR system.

Dharamveer Sharma and his team [52] at Punjabi University has developed a form recognition system for Gurmukhi, which can automatically detect and recognize the hand written Gurmukhi text written on pre printed forms.

Speech Technology

Research for development of a Punjabi Text-to-speech synthesis (TTS) system, which can speak out Punjabi text, has been initiated at Punjabi University and Thapar University Patiala. A preliminary syllable based TTS system has been developed by Parminder Singh and Dr. G. S. Lehal [53]. The system is further being refined to improve the output quality before it will be made available to public for use.

Open Source Software in Punjabi

The team of A S Alam, Gursharn Singh Khalsa, Jaswinder Singh and Narinderpal Singh have been involved in Punlinux project to localize GNU Software into Punjabi[54]. They have been working to build a full customized Linux for use in Punjabi language including OpenOffice, Mozilla and other software.

In the last 5 years they have created the following software in Linux environment

1) First Punjabi Operating System with
- Translated Installed Installer
- Translated Desktop
- With latest Input Methods
- Web Browser

2) Translation for Open Source Applications
- OpenOffice - Windows/Mac/Linux
- Firefox - Windows/Mac/Linux
- Email Client - Thunderbird - Windows/Mac/Linux
- Linux Operating Systems (Fedora, OpenSUSE, Debian)

3) CLDR Translation (Common Locale Data Repo)

4) Locale Data Improvement for Punjabi- Linux/Unix/Mac Base Locale Data

5) Punjabi Support for Padma (Non-Unicode to Unicode Convertor plug-in for Firefox)

Bharateeya Open Office, which has been developed by CDAC Bangalore with translations support from GIST group of CDAC Pune and Punlinux, is a full-fledged office suite with options of having the user interface in Punjabi. But there is no provision for Punjabi language processing such as spell checker, typing utility etc and the help is also available only in English. The software is available on the language CD distributed by MCIT.

Conclusion

Though some steps have been taken for computerization of Punjabi but still there is a long way to go. Defining and refining standards, development of operating systems with full support for Punjabi, human machine interfaces, Internet tools and technologies, machine-aided translations and speech related efforts are some of the major thrust areas identified for attention in the near future. Standardization of technical terminology for use in regional languages as well development of lexical resources such as annotated text and speech corpus, translation dictionaries, language models, thesaurus, word net etc have to be immediately addressed.

We have tried to list down the works undertaken by the scholars and institutions for the technological development of Punjabi. The list may not be exhaustive and it may happen that there might be some more groups and organisations who are involved in technical development of Punjabi but are not referred to here. Our source of information has been mainly the online resources on internet and research papers. We request them to furnish us with the contributions not mentioned by us so that we can include them in our next version of the paper.

References

1. http://www.billie.grosse.is-a-geek.com/resources-03.html
2. http://www.gurbanifiles.org/unicode/index.htm
3. http://ildc.in/Punjabi/Pindex.aspx
4. http://www.janmejajohl.com
5. http://guca.sourceforge.net/applications/guca/
6. http://www.learnpunjabi.org/unipad
7. http://www.quillpad.com/punjabi
8. http://punjabikosh.googlepages.com/
9. http://www.4shared.com/file/39293942/9d333376/Punjabi_Shabdkosh__English_to_Punjabi_Dictionary_.html?s=1
10. http://www.ik13.com/mahan_kosh.htm
11. http://www.advancedcentrepunjabi.org/vocabulary/vocabulary1.asp?id=46
12. http://www.advancedcentrepunjabi.org/pedic/Default.aspx
13. http://www.punjabonline.com/servlet/library.dictionary?Action=English
14. http://www.srigranth.org/servlet/gurbani.dictionary
15. G S Lehal, "Design and Implementation of Punjabi Spell Checker", International Journal of Systemics, Cybernetics and Informatics, pp. 70-75 (Jan 2007).

16. M. S. Gill and G. S. Lehal, "A Grammar Checking System for Punjabi",Coling 2008: Companion volume: Posters and Demonstrations, Manchester, UK, pp. 149-152 (August 2008).
17. http://pgc.advancedcentrepunjabi.org/
18. http://www.cdacmohali.in/default.aspx?pid=29&lang=en-us
19. http://tdil.mit.gov.in/TDIL-OCT-2003/text%20editors%20&%20word%20processors.pdf
20. www.advancedcentrepunjabi.org/intro1.asp
21. http://www.maa.com.au
22. http://www.sikhpoint.com/kidscorner/kid_gurmukhi.php?level=1&module=pb
23. http://www.punjabonline.com/servlet/library.language
24. http://www.5abi.com/5ratan
25. http://www.sikhlink.com/chapter1.htm seven chapters with audio
26. http://www.rajkaregakhalsa.net/lesson1.htm
27. http://www.billie.grosse.is-a-geek.com/index.html
28. http://www.advancedcentrepunjabi.org/statistics.html
29. http://pgc.advancedcentrepunjabi.org/#Tagger
30. M.G.A. Malik, "Punjabi Machine transliteration", Proceedings of the 21st International Conference on Computational Linguistics, pp. 1137–1144 (2006).
31. http://ildc.in/Urdu/Uindex.aspx
32. T S Saini and G S Lehal "Shahmukhi to Gurmukhi Transliteration System: A Corpus based Approach", Research in Computing Science(Mexico), Volume 33, pp. 151-162 (2008).
33. http://g2s.learnpunjabi.org/default.aspx
34. http://s2g.advancedcentrepunjabi.org/login.aspx
35. http://www.puran.info/PMT/PMT.aspx
36. http://www.learnpunjabi.org/
37. V Goyal and G S Lehal, "A Machine Transliteration System for Machine Translation System : An Application on Hindi-Punjabi Language Pair", Atti Della Fondazione Giorgio Ronchi (Italy), Volume LXIV, No. 1, pp. 27-35 (2009)
38. G. S. Josan and G. S. Lehal, "A Punjabi to Hindi machine Translation System", Coling 2008: Companion volume: Posters and Demonstrations, Manchester, UK, pp. 157-160 (August 2008).
39. http://h2p.learnpunjabi.org
40. http://sampark.iiit.net/
41. Parteek Bhatia, Sandeep Singh, "Punjabi Deconverter Architecture", National Seminar on Creation of Lexical Resources for Indian Language Computing and Processing, CDAC Mumbai, March 26-28, 2007.
42. Parteek Bhatia, R K Sharma, "Role of Punjabi morphology in designing Punjabi-UNL enconverter", Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai, India , pp. 562-566, 2009.
43. K K Batra and G S Lehal, "On Translation of Verb Phrases from Punjabi to English", Atti Della Fondazione Giorgio Ronchi (Italy), Volume LXIV, No. 2, pp. 245-251 (2009)

44. http://www.sikhitothemax.com

45. http://www.ik13.com/isher_micro_media_2005.htm

46. Gurpreet Singh Lehal, Chapter Title: "A Complete Machine Printed Gurmukhi OCR System", Guide to OCR for Indic Scripts, Series: Advances in Pattern Recognition, Springer, ISBN: 978-1-84800-329-3, (2009)

47. G. S. Lehal and Chandan Singh, "A Complete Machine Printed Gurmukhi OCR System", Vivek, pp. 10-17, Vol. 16, No. 3. (2006).

48. Indu Chhabra, Chandan Singh, "Describing Character Object With Invariant Features", Journal of CSI, 36(4), 2006, 33-38.

49. M K Jindal, G S Lehal and R K Sharma, "A Study of touching characters in degraded Gurmukhi text", Transactions On Engineering, Computing And Technology, Volume 4, pp. 121-124, (2005)

50. M K Jindal, G S Lehal and R K Sharma, "Segmentation problems and solutions in printed degraded Gurmukhi text", International Journal of Signal Processing, Volume 2, Number 4, pp. 258-267, (2005)

51. Anuj Sharma, Rajesh Kumar and R. K. Sharma, "Recognizing Online Handwritten Gurmukhi Characters using Comparison of Small Line Segments", International Journal of Computer Theory and Engineering Volume 1, No. 2, 133-137 (2009).

52. D. Sharma, G.S. Lehal, "Form Field Frame Boundary Removal for Form Processing System in Gurmukhi Script", Accepted for Publication in Proceedings of 9th International Conference of Document Analysis and Recognition, Barcelona, Spain. (2009).

53. P. Singh and G S Lehal, "Text-to-Speech Synthesis system for Punjabi language", Proceedings International Conference on Multidisciplinary Information Sciences and Technologies, Merida, Spain, pp.388-391. (2006).

54. http://punlinux.sourceforge.net/index.html

Gurpreet Singh Lehal, Ph.D.
Advanced Centre for
Technical Development of Punjabi Language, Literature & Culture
Punjabi University
Patiala 147002.
Punjab, India
gslehal@gmail.com