

Manish Kumar

Web Based Hindi to Punjabi Machine Translation System for Translating Computer Text

Manish Kumar

(Department of Computer Science, Baba Farid College, Bathinda)

manish21578@gmail.com

Abstract – Although development of Machine Translation (MT) systems had started in late 17th Century, yet effective and actual development in this field could gather momentum only in 20th Century [3-9]. Lots of efforts and research is going on in development of various machine translation systems to translate the text or speech from one language to another language [10]. This paper will describe a machine translation system which translates the text from Hindi to Punjabi. Though there is one existing Hindi to Punjabi Machine Translation system developed by Goyel and Lehel [1], but this paper will describe the development of web based Hindi to Punjabi Translation system which works on a specific domain i.e computer text.

Index Terms – Machine Translation System, Closely related languages, Web Based Translation System, Transliteration, Natural Language Translation

Introduction - Machine Translation has evolved into the present status where its

importance is multidisciplinary and it can be used in almost all the fields. Machine translation is a bliss for those who has some language issues. With these translation systems text from one language can be easily translated to another language. It has unconcerned the division of world on the basis of language. Machine translation system can work on two languages or it can be a multi language model. Two language pair can be closely related languages [2] or it can be not closely related languages pair. Closely related languages like Hindi and Punjabi follows the same sentence structure, where as not closely languages like English and Punjabi or Hindi follows different sentence structure. It is easier to develop a MT system for closely related languages.

System described here is a domain based Hindi to Punjabi Machine translation system which can translate text related to computer subjects from Hindi to Punjabi. There are some existing MT systems which translates the text from Hindi to Punjabi, but those are general purpose systems which does not



resolve the issues which are associated with text of some specific domain like computer text. This developed Hindi to Punjabi machine translation system is integrated with certain web tools to make it a web based system which can acquire input online and can translate the text and bestow output accordingly.

System Description and Architecture

This Hindi to Punjabi Machine translation system is a domain based system which has been developed to remove all the translation errors of translating computer text from Hindi to Punjabi. There are different approaches for developing a MT system [11]. This system has been developed using direct machine translation approach, which follows word to word translation. Accuracy of Direct Machine Translation system is directly dependent on the volume of lexical resources developed for that system. This system supports a huge lexicon consisting words of source language and its corresponding target language translated word. Mainly there are three modules of this system.

Pre-Processing phase makes each sentence translatable. There are three different modules in this phase. First is text

normalization which is to normalized the word which could be written in more than one way. Second is replacing collocations. In Hindi there are certain words which are framed by combining two different words which have separate meaning individually. It is very important to store those words as a single word entity because if translated individually it can give different meaning altogether. Third module of this phase is replacing proper nouns. This module identifies the proper nouns and sends it to transliteration module instead of translation.

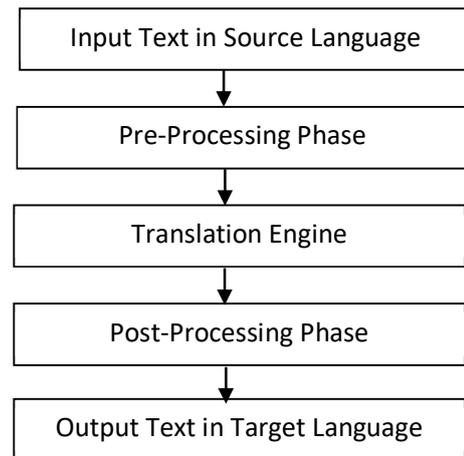


Figure 1

Phases of Hindi to Punjabi MT System

Translation engine receives the pre-processed tokens and then translates it into source language. Before word to word translation it identifies the titles and surnames as well. Titles are stored in the

ਇੰਟਰਪਟ ਪੂਰੀ ਤਰ੍ਹਾਂ ਵਲੋਂ ਆਧੁਨਿਕ ਆਪਰੇਟਿੰਗ
ਸਿਸਟਮ ਵਿੱਚ ਅਸਿੰਕਰੇਨਸ ਘਟਨਾਵਾਂ ਨੂੰ ਹੈਂਡਲ
ਕਰਣ ਨੂੰ ਅਤੇ ਕਰਨਲ ਵਿੱਚ ਸੁਪਰਵਾਇਜਰ ਮੋਡ
ਰੂਟੀਨ ਨੂੰ ਫੜਨ ਲਈ ਪ੍ਰਯੋਗ ਕੀਤਾ ਜਾਂਦਾ ਹੈ

ਜਬ ਸੰਪੂਰਨ ਸਟ੍ਰਕਚਰ ਯਾ ਊਸਕਾ ਏਕ ਮੈਂਬਰ ਫੰਕਸ਼ਨ
ਕੇ ਲਿਏ ਪਾਸ ਹੋਤਾ ਹੈ ਤੋ ਯਹ ਕਾਲ-ਬਾਇ-ਵੈਲਯੂ
ਤਕਨੀਕ ਕਾ ਪ੍ਰਯੋਗ ਕਰਤਾ ਹੈ

ਜਦੋਂ ਸੰਪੂਰਣ ਸਟਰਕਚਰ ਜਾਂ ਉਸਦਾ ਇੱਕ ਮੈਂਬਰ
ਫੰਕਸ਼ਨ ਲਈ ਪਾਸ ਹੁੰਦਾ ਹੈ ਤਾਂ ਇਹ ਕਾਲ - ਬਾਈ -
ਵੈਲਿਓ ਤਕਨੀਕ ਦਾ ਪ੍ਰਯੋਗ ਕਰਦਾ ਹੈ

Web Based Tools

To enhance to usability of the system and to enable the system to reach to maximum users, it has been developed as web based system. This system is freely accessible for all on www.h2p.learnpunjabi.org. Following features make this system as a web based system.

- I. First most problem faced by users in using any MT system is the variation in fonts used in Hindi. This system allows typing the text in Unicode. So instead of typing

some text in other font and then converting it into a Unicode, user can use this facility for expedient working.

- II. Option of browsing a file is available in the system. Instead of typing a text, already typed text in a file can also be inserted using this particular option.
- III. A very powerful feature is available in this system and that is to send the translated text directly through mail to any mail address. User can select the text to be sent and also can select that whether he wants to send it in original form or in translated form.
- IV. The most important feature of this web based system is its ability to translate any website which is available in Hindi. This feature enables the users to read any Hindi website in Punjabi without any change in the original setup and structure of the website.
- V. Major concern of any web based translation system is its speed. Response time of this particular Hindi to Punjabi MT system is

very less. Whether we type the text, select any file or even if we want to translate the whole web site, within few seconds translated version gets accessible.

Application of System

This is web based system and can used in different aspects.

- I. This system is a domain based system which is developed to translate the computer text from Hindi to Punjabi. This system can be helpful to confiscate the communication barrier for students, who want computer education in Punjabi.
- II. As a web based system, online resources like Newspapers in Hindi, websites and other technical text available in Hindi can be unswervingly translated into Punjabi.
- III. Due to the email facility available in the system. User can type the text in Hindi or can select a file and can send the mail in Punjabi.

Future Scope and Conclusion

- I. This system can work as a mock-up system for developing MT system for other language pair.
- II. This system can be a baseline system for developing Hindi to Punjabi MT system for other domains like Medical, legal etc.

This web based machine translation system has extended the prospectively and applicability of MT systems to the next level. An all-purpose MT system can be perfect for general text but every domain has their own set of troubles or vocabulary issues which needs to be resolved vigilantly. This system can be a model and consummate system for the upcoming researchers.

References

1. Goyal, V.,(2010),” *Development of Hindi to Punjabi Machine Translation System*”, Ph.D. Thesis, Department of Computer Science, Punjabi University Patiala.
2. Altintas K., “*A Machine Translation System between a pair of closely related languages* ”.Internet:

- <http://www.cs.bilkent.edu.tr/~ilyas/PDF/iscis2002.pdf>.
3. Hutchins, J., (1992), "General Introduction and brief History of Machine Translation," from web site <http://www.hutchinsweb.me.uk/IntroMT-1.pdf> accessed on 12/10/2015.
 4. Hutchins, W.J., (1997), "Looking back to 1952: the first MT conference TMI-97," in proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation, July 23-25, St. John's College, Santa Fe, New Mexico, USA, pp.19-30.
 5. Ping, K., (1996), "A Socio-semiotic Approach to Meaning in Translation," *BABLE* (42) 2, pp 74-83.
 6. Toma, P., (1976), "An operational Machine Translation system Brislin," R.W. ed. *Translation: applications and research* (New York: Gardner P.), pp. 247-259.
 7. Naskar, S., Bandyopadhyay, S., (2005), "Use of Machine Translation in India: Current Status," In the Proceedings of MT SUMMIT X; September 13-15, 2005, Phuket, Thailand, pp. 465-470.
 8. John, H., (2007), "Machine translation: a concise history, Computer aided translation: Theory and practice," Chan Sin Wai. Chinese University of Hong Kong, 2007. Citesser, pp. 1-21.
 9. Slocum, J., (1985), "A survey of machine translation: its history, current status, and future prospects," *Journal of Computational Linguistic - Special issues on Machine Translation*, 11(1), pp. 1-17.
 10. John, H., (1993), "Latest developments in machine translation technology: beginning a new era in MT research," *Fourth Machine Translation Summit: MT Summit IV, Proceedings: International cooperation for global communication*, July 20-22, 1993, Kobe, Japan. [Tokyo: AAMT, 1993], pp.11-
 11. Antony, P.J., (2013), "Machine Translation Approaches and Survey for Indian Languages," *Computational Linguistics and Chinese Language Processing*, 18(1), pp. 47-78.

