

ABSTRACT

A software designed to translate a natural language into another natural language is a machine translation system. There are various machine translation approaches to develop a complete system. The current trend in the research of machine translation is Statistical Approach to Machine Translation. We have presented incremental statistical machine learning algorithm for Urdu to Punjabi machine translation. Asian languages are resources inferior languages, therefore, it is a challenging task to collect parallel corpus for training these statistical models. There are many machine translation systems which have been developed for Indo-Aryan languages. Most of the work has been done using rule-based or hybrid approaches because of the non-availability of resources. Efforts have been made to develop parallel phrase corpus in place of parallel sentence corpus. Collecting parallel phrases were more convenient as compared to the parallel sentences. In preprocessing phase, the system has used rules for segmentation, tokenization and text classification system to translate given text according to a preferred domain which also helped translation system to improve overall accuracy. the system is able to achieve 0.87 BLEU score which is relatively high compared to other statistical translation systems. This system has been developed as two different applications. On Windows platform, it can work as a standalone system without the requirement of Internet connection. It has become the part of Akhar 2016, which is a Windows based Word Processor for Indic languages. Another version of this system has been deployed for the Web server.