

Abstract

Machine translation (MT) is a study of automatic translation of one human language to another, which is one of the sub-areas and applications of natural language processing (NLP). This thesis describes an approach to build a practical and useful corpus-based machine translation (CBMT) system which employs a statistical approach with automatic bilingual alignment support. Our improved algorithm for aligning bilingual parallel texts can achieve 97% of accuracy, which acts as the strong back-end of our CBMT system to build up the Portuguese-Chinese corpus-base.

Keywords: machine translation, corpus, alignment, nearest example retrieval.