University of Macau

Abstract

Portuguese – Chinese Machine Translation Based on Translation Corresponding Tree (TCT) Annotation Scheme

by Tang, Chi Wai

Thesis Supervisors: Prof. Li YiPing and Dr. Wong Fai

Master of Science in E-Commerce Technology

From the compendium by EAMT, there is no any practical machine translation system between Portuguese and Chinese. In Structure Based EBMT system, it has a problem that it relies on two syntax parsers to analyze the translation examples, but robust syntax parsers are not always available. On the other hand, Portuguese and Chinese belong to two different language families and there exist the grammatical deviation problem between them. In order to resolve the weakness of the Structure Based EBMT system and the linguistic problems between Portuguese and Chinese, a new Portuguese – Chinese Machine Translation system is designed and implemented. The system is based on a novel technology call Translation Corresponding Tree (TCT) which is an example base knowledge annotation method. Based on this approach, it solves the connatural problem of the Structure Based EBMT system and owns an advantage that it only relies on a single language syntax parser. Another

critical characteristic of TCT annotation method is that it is flexible enough to extend for capturing extra linguistic information that may help to the processing task. Therefore, TCT annotation can overcome with the structure derivation problem between Portuguese and Chinese, such as the sentence constituents crossing dependence problem. Moreover, this research also proposes a conversion algorithm to reuse the existing translation knowledge represented in terms of TCT trees by transforming the translation trees of the Portuguese to Chinese translation into that of the Chinese to Portuguese translation. By this conversion method, existing knowledge can be easily reused without having to re-construct the knowledge from zero. Based on the research result of the TCT annotation method, a Portuguese-Chinese machine translation system has been implemented at the end of this thesis work. The empirical results show that the implemented MT system can achieve the translation accuracy of 85% in the domain of Macau law statements.