Jinhua Qian

L555

Abstract

Converting Sino-Korean words into Chinese characters

This program is going to detect Sino-Korean words written in Hangul (Korean characters) and convert them into Chinese characters. In addition, this program will calculate the frequency of the appearance of Sino-Korean in certain sentences. Sino-Korean refers to Korean words of Chinese origin. In Korean, a large portion of the words is composed of Sino-Korean. The Standardized Korean Language Dictionary published by the National Institute of the Korean Language (NIKL) in 2004 contains approximately 57% of Sino-Korean words. In addition, the Survey of Korean Vocabulary frequency (1956), has shown that about 70% of the frequently used words are Sino-Korean (Lu et al. 2015). Due to such a great number of shared vocabularies, converting Sino-Korean words into Chinese characters plays an important role in the machine translation process. It will provide better source sentences for later machine translation model training and improve the translation quality (Park & Zhao 2019). Although Sino-Korean words are written in Hangul, we can convert them into Chinese characters because there are correspondences between those two languages. This program will be tested on a corpus of 100 Korean sentences. The goal of this program is to provide the source sentences for improving the performance of machine translation between Chinese and Korean by converting Sino-Korean which is the bridge between these two languages.

Bibliography

Park, Jeonghyeok & Zhao, Hai (2019) *Korean-to-Chinese Machine Translation using Chinese Character as Pivot Clue,* 33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 558-566.

Chu, Chenhui et al.(2013) *Chinese-Japanese machine translation exploiting Chinese characters,* ACM Transactions on Asian Language Information Processing, Volume 12, Issue 4. https://doi.org/10.1145/2523057.2523059

Huang, Jinxia & Choi, KeySun (2000). *Using Bilingual Semantic Information in Chinese-Korean Word Alignment*. Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation, pages 121–130.

Lu, Yuanmei et al.(2015), *Korean-to-Chinese Word Translation using Chinese Character Knowledge,* Proceedings of MT Summit, 15(1). page 256–269.