

[CICLING 2010: 11th International Conference on Intelligent Text Processing and Computational Linguistics, March 21-27, 2010, Iași, Romania]

Abstract

Filip Graliński. Mining Parenthetical Translations for Polish-English Lexica.

Documents written in languages other than English sometimes include parenthetical English translations, usually for technical and scientific terminology. Techniques had been developed for extracting such translations (as well as transliterations) from large Chinese text corpora. This paper presents methods for mining parenthetical translation in Polish texts. The main difference between translation mining in Chinese and Polish is that the latter is based on the Latin alphabet and it is more difficult to identify English translations in Polish texts. On the other hand, some parenthetical translated terms are preceded with the abbreviation "ang." (=English), a kind of an "anchor", allowing for querying a Web search engine for such translations.