

N-gram に基づく用例対訳検索手法

田淵 裕章、坂本 廣、北村 泰彦

外国人を対象とした多言語医療通訳支援を目的として多数の医療用例対訳が収集されている。用例対訳を検索する場合に、従来のキーワード検索手法は、検索の絞り込みや用例表現のゆらぎに関して問題がある。そこで本研究では N-gram を用いた類似度計算に基づく用例検索手法を提案し、その有効性を示す。本手法は、入力された検索文が新たな用例として活用できるという利点も備えている。

キーワード 類似文章検索、N-gram、情報検索、用例対訳

Search Method for Parallel Texts Using N-gram

Hiroaki TABUCHI, Hiroshi SAKAMOTO, and Yasuhiko KITAMURA

We have collected a lot of medical parallel texts with translations to support multilingual medical interpretation services for foreign visitors. When we retrieve parallel texts, we face a problem concerning narrowing the search result and ambiguous expression of parallel texts. We propose a parallel text retrieval method based on similarity measure using N-gram, and show its performance. This method provides an advantage that we can utilize the search inputs as cues to collect new parallel texts.

Key words Similarity Search, N-gram, Information Retrieval, Parallel Text