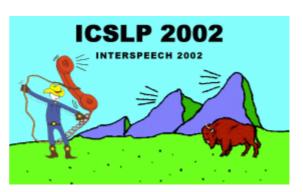# 7th International Conference on Spoken Language Processing

September 16-20, 2002
Denver, Colorado, USA

Full-Text Story Alignment Models for Chinese-English Bilingual News Corpora

Bing Zhao, Stephan Vogel

Carnegie Mellon University, USA

In this paper, we describe the full-text story alignment on Chinese-English bilingual corpora of news data to mine potential parallel data for machine translation. Several standard information retrieval methods are tested and two translation-model based alignment models are proposed and studied. Modeling the process of generating the parallel English story from Chinese story gives significant improvements over the standard information retrieval techniques. Refinements of the alignment model are also proposed and tested in detail. On one day s bilingual news collection, our methods improved the mean reciprocal rank from 0.31 to 0.68.

Full Paper