

## ABSTRAK

Pengumpulan korpus paralel sedang gencar dilakukan untuk keperluan studi dan pengembangan NLP. Namun, untuk pasangan kalimat beberapa bahasa, khususnya Bahasa Indonesia-Sunda, jumlah korpus paralel yang tersedia masih sangat sedikit. Sedangkan untuk mengumpulkan korpus paralel secara manual memerlukan waktu yang lama dan biaya yang mahal. Dengan alasan tersebut, pengumpulan korpus paralel akan lebih efektif dan efisien jika dikumpulkan secara otomatis. Dalam tugas akhir ini, akan dilakukan penelitian pengumpulan korpus paralel pada Wikipedia menggunakan metode *Pointwise Mutual Information* (PMI) untuk menentukan *sentence similarity*. Pengambilan data dari artikel Wikipedia bahasa Indonesia dan Sunda dengan memanfaatkan fasilitas *interlanguage link* dan MediaWiki API. Dengan metode ini, diharapkan didapat korpus paralel yang cukup baik dengan efisien.

**Kata kunci:** korpus paralel, Wikipedia, pointwise mutual information, interlanguage link, MediaWiki API