

COSINE SIMILIARITY DAN GOOGLE TRANSLATE API UNTUK TEMU KEMBALI INFORMASI LINTAS BAHASA INDONESIA DAN INGGRIS

Penulis

Zulfi Osman

150411100033

Dosen Pembimbing 1

Sigit Susanto Putro, S.Kom., M.Kom.

19790313 200604 1 002

Dosen Pembimbing 2

Husni, S.Kom., M. T.

19790722 200312 1 001

ABSTRAK

Cross-Language Information Retrieval (CLIR) merupakan subbidang Temu Kembali Informasi yang berkaitan dengan pengambilan informasi yang disimpan dalam bahasa yang berbeda dari bahasa permintaan yang diberikan pengguna. Pada CLIR terdapat berbagai metode terjemahan, diantaranya adalah *Dictionary-based*, *Parallel corpora based*, *Comparable corpora based*, *Machine translator based*, *Ontology based*, dan *Transitive based*. Langkah-langkah menerjemahkan query ke dalam bahasa target meliputi stemming, kemudian perhitungan kemiripan dokumen. permasalahannya terdapat pada efisiensi waktu dan akurasi dalam penerjemahan *query* dari banyaknya metode di CLIR pada Bahasa Indonesia dan Inggris. Pada tahap stemming mempunyai metode masing-masing dari kedua Bahasa tersebut. Untuk stemming bahasa Indonesia terdapat kata dasar dan kata imbuhan. Ada variasi imbuhan termasuk *prefix*, *suffix*, *infiks*, dan *confixes*, sedangkan stemming bahasa Inggris hanya kata imbuhan *suffix*. Pada pemrosesan stemming membutuhkan waktu yang lebih lama. Untuk mengatasi masalah tersebut, di usulkan menggunakan *Google Translate API* dalam penerjemahan query dan menggunakan satu stemming bahasa Inggris dengan algoritma *Porter Stemmer* yang merupakan metode stemming terbaik dari beberapa penelitian dengan akurasi terbaik. Untuk perhitungan kemiripan menggunakan *Cosine Similarity*. penelitian ini diharapkan dapat menghasilkan sistem yang lebih akurat serta dapat mempercepat waktu dalam proses dan pencarian dokumen lintas bahasa Indonesia dan Inggris.

Kata Kunci : *Cross-Lingual Information Retrieval*, *Google Translate*, *Stemming*, *Porter Stemmer*

COSINE SIMILIARITY AND GOOGLE TRANSLATE API FOR INDONESIAN AND ENGLISH CROSS LANGUAGE INFORMATION RETRIEVAL

Penulis

Zulfi Osman

150411100033

Dosen Pembimbing 1

Sigit Susanto Putro, S.Kom., M.Kom.

19790313 200604 1 002

Dosen Pembimbing 2

Husni, S.Kom., M. T.

19790722 200312 1 001

ABSTRACT

Cross Language Information Retrieval (CLIR) is a sub-field of Information Retrieval relating to the retrieval of information stored in a language that is different from the language the request is given by the user. In CLIR there are various translation methods, including Dictionary-based, Parallel corpora based, Comparable corpora based, Machine translator based, Ontology based, and Transitive based. Steps to translate the query into the target language include stemming, then calculating the similarity of the document. the problem is the time efficiency and accuracy in query translation of the many methods in CLIR in Indonesian and English. In the stemming stage, each method has two methods. For Indonesian stemming there are basic words and affix words. There are variations of affixes including prefixes, suffixes, infixes, and confixes, whereas English stemming is only suffix affixes. In stemming processing requires a longer time. To overcome this problem, it is proposed to use the Google Translate API in query translation and use one English stemming with the Porter Stemmer algorithm which is the best stemming method from several studies with the best accuracy. For similarity calculations use Cosine Similarity. this research is expected to produce a more accurate system and can speed up the time in the process and search for documents across Indonesian and English.

Kata Kunci : *Cross-Lingual Information Retrieval, Google Translate, Stemming, Porter Stemmer*