

EKSPANSI KUERI BERDASARKAN KAMUS DWIBAHASA MENGGUNAKAN PELUANG BERSYARAT

Muhammad Awet Samana¹, Julio Adisantoso²

¹Mahasiswa Departemen Ilmu Komputer, Fakultas Matematika dan IPA, Institut Pertanian Bogor

²Staf Departemen Ilmu Komputer, Fakultas Matematika dan IPA, Institut Pertanian Bogor

ABSTRACT

Query expansion is a technique that can be used to bridge vocabulary gaps between queries and documents in a collection. With query expansion, user will be helped to formulate queries well so that user will optimize the search result. In this study, the expansion term was resulted from the translation of bilingual dictionary. Subsequently, several terms of expansion were elected based on the closeness of its relationship with the original query.

This research showed that conditional probability can be used to select expansion term in bilingual translation method query expansion. The result of 1000 documents was more than 97% relevant document can be retrieved by this system. The more expansion term used, the larger number relevant document was retrieved.

Keywords: query expansion, conditional probability, query translation, bilingualism dictionary.

PENDAHULUAN

Tidak jarang pengguna suatu sistem temu kembali informasi mengalami kesulitan dalam membentuk kueri. Kueri ini ditujukan untuk menemukembalikan informasi hal ini dikarenakan mereka tidak mengetahui detail dari konstruksi koleksi dan lingkungan temu kembali (Baeza-Yates & Ribeiro-Neito 1999). Padahal, jumlah dokumen relevan yang diperoleh dipengaruhi oleh jumlah kata kunci dalam kueri. Hal ini akan mengakibatkan hasil pencarian yang dilakukan pun menjadi kurang optimal.

Ekspansi kueri merupakan salah satu teknik yang dapat digunakan dalam membantu pengguna dalam memberikan kueri yang baik. Ekspansi kueri dapat berperan sebagai penghubung karena adanya *vocabulary gaps* antara kueri dan dokumen (Fang 2008). Dalam teknik ini kueri awal akan dimodifikasi dan akan ditambahkan istilah-istilah tertentu.

Rusidi (2008) telah melakukan penelitian untuk menggunakan peluang bersyarat dan mengaplikasikannya dalam ekspansi kueri suatu sistem temu kembali informasi. Ia mengambil istilah ekspansi dari istilah yang berada dalam indeks dan memilihnya berdasarkan keeratan hubungan suatu istilah dengan istilah lainnya.

Keeratan hubungan ini diukur dengan menggunakan peluang bersyarat. Namun,

metode ini kurang mempertimbangkan apakah istilah tersebut memiliki makna yang berkaitan dan jauh dari konsep dalam kueri yang dimasukkan pengguna.

Sitohang (2009) telah menggunakan metode kamus dwibahasa dalam melakukan ekspansi pada kueri pengguna. Kamus dwibahasa digunakan untuk mendapatkan padanan kata yang merupakan sinonim dari kueri atau memiliki makna yang hampir sama dengan kueri awalnya. Ia menggunakan nilai IDF untuk memilih kueri ekspansi yang dihasilkan dari penerjemahan.

Penggunaan nilai IDF ini masih dirasa kurang optimal dalam memilih istilah ekspansi. Hal ini dikarenakan kurangnya perhatian terhadap keeratan hubungan antar istilah ekspansi dengan kueri awal. Untuk itu, ada beberapa metode yang dapat digunakan dalam menemukan keeratan hubungan ini di antaranya: korelasi, ukuran kesamaan, dan dengan menggunakan peluang bersyarat. Keeratan hubungan berdasarkan kemunculan suatu objek sebaiknya menggunakan peluang bersyarat karena korelasi maupun ukuran kesamaan lainnya umumnya diberlakukan bagi data berskala ordinal atau interval (Adisantoso 1997).

Peluang Bersyarat

Koopman (1997) mengemukakan bahwa peluang bersyarat B setelah A merupakan