

QUERY EXPANSION MENGGUNAKAN METODE *PSEUDO RELEVANCE FEEDBACK* DENGAN DATASET WIKIPEDIA DALAM BAHASA INDONESIA

Penulis

Hadi Yulian Dharmanto

150411100027

Dosen Pembimbing 1

Dosen Pembimbing 2

Dr. Yeni Kustiyahningsih, S.Kom., M.Kom.

Husni, S.Kom., M.T.

19770921 200812 2 002

19770722 200312 1 006

ABSTRAK

Proses pencarian dokumen yang sangat banyak, tentunya memerlukan sebuah teknik khusus dalam mengambil / mencari dokumen yang relevan dengan permintaan pengguna. Pada proses pencarian dokumen dapat dilakukan dengan menggunakan sebuah *query*. *Query* adalah sebuah kata kunci yang digunakan untuk menampilkan data dari *database* kemudian akan diolah lebih lanjut. Data tersebut diambil dari tabel-tabel dalam *database* yang memudahkan pengguna dalam mengolah banyak data. Permasalahan dari penggunaan *query* secara umum adalah dari sudut pandang pengguna atau *user*. Pengguna memiliki karakter yang berbeda-beda. *Query* yang terlalu singkat dapat diselesaikan dengan cara *Query Expansion*. *Query Expansion* adalah proses merancang kembali *query* awal untuk meningkatkan kinerja sistem dalam proses pengambilan informasi. Dalam penelitian ini *query expansion* akan menggunakan metode *pseudo relevance feedback* dengan dataset Wikipedia bahasa Indonesia. Untuk melakukan perhitungan persamaan dokumen, digunakanlah metode *cosine similarity* dan pembobotan menggunakan TF-IDF. Bahasa Indonesia memiliki banyak sekali kata-kata yang ambigu, sehingga wikipedia bahasa Indonesia perlu digunakan sebagai acuan agar kata-kata yang ambigu atau kata yang mengandung istilah dapat memiliki suatu makna tertentu. Dalam penelitian ini, memiliki hasil rata-rata presisi dalam penelitian menggunakan 100 artikel pariwisata Madura tanpa *query expansion* adalah **75,5%**, sedangkan dengan *query expansion* hanya **36,67%**. Jika menggunakan 75 artikel pariwisata Madura tanpa *query expansion* adalah **70,5%**, sedangkan dengan *query expansion* hanya **31,67%**. Hal ini terjadi karena hasil dari *query expansion* tidak berpusat pada pariwisata di Madura. Sebab abstrak wikipedia bahasa Indonesia yang digunakan tidak berfokus pada pariwisata di Madura, melainkan abstrak artikel wikipedia bahasa Indonesia secara umum

Kata kunci : *Information Retrieval, Query, Query Expansion, Pseudo Relevance Feedback*

QUERY EXPANSION USING METHOD PSEUDO RELEVANCE FEEDBACK WITH THE WIKIPEDIA DATASET IN INDONESIAN

Penulis

Hadi Yulian Dharmanto

150411100027

Dosen Pembimbing 1

Dosen Pembimbing 2

Dr. Yeni Kustiyahningsih, S.Kom., M.Kom.

Husni, S.Kom., M.T.

19770921 200812 2 002

19770722 200312 1 006

ABSTRACT

The process of searching many documents, requires a special technique in retrieving / finding documents that are relevant to the user's request. In the process of searching documents can be done by using a query. A query is a keyword that is used to display data from a database which will then be processed further. The data is taken from tables in a database that makes it easy for users to process a lot of data. The problem with using queries in general is from the user's perspective. Users have different characters. Queries that are too short can be completed by means of Expansion Queries. Query Expansion is the process of redesigning the initial query to improve system performance in the information retrieval process. In this research, query expansion will use the pseudo relevance feedback method with the Indonesian Wikipedia dataset. To calculate document equations, cosine similarity and weighting methods are used using TF-IDF. Indonesian has a lot of ambiguous words, so Indonesian Wikipedia needs to be used as a reference so that ambiguous words or words that contain terms can have a certain meaning. In this study, the average precision in research using 100 Madura tourism articles without query expansion was 75.5%, while with query expansion only 36.67%. If using 75 Madura tourism articles without query expansion is 70.5%, while with query expansion only 31.67%. This happens because the results of query expansion are not centered on tourism in Madura. Because the Indonesian wikipedia abstract used does not focus on tourism in Madura, but abstracts on Indonesian wikipedia articles in general.

Keyword : Information Retrieval, Query, Query Expansion, Pseudo Relevance Feedback