# ANALISIS DAN IMPLEMENTASI QUERY EXPANSION PADA INFORMATION RETRIEVAL BERDASARKAN PENARIKAN KESIMPULAN DENGAN FUZZY RULES

**Tigor Maruhum[1], Yanuar Firdaus A.w.[2], Agung Toto Wibowo[3]**

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

**Abstrak**

**Information Retrieval (IR) merupakan suatu sistem yang digunakan untuk menemukan kembali(retrieve) informasi-informasi dan relevan terhadap kebutuhan user dari suatu kumpulan informasi berdasarkan query dari user. Di dalam menemukan kembali informasi tersebut, performansi pencarian dapat ditingkatkan menggunakan teknik modifikasi query yaitu Query Expansion(QE).**

**Query Expansion merupakan suatu teknik dengan menambahkan keywords baru ke dalam query awal sehingga meningkatkan performansi pencarian. Query Expansion ini kemudian dapat dikembangkan menggunakan penarikan kesimpulan dengan metode fuzzy rules. Metode ini menggunakan fungsi keanggotaan dan fuzzy rules untuk mencari nilai Relevance Degree (RD) dari calon additional keywords dan kemudian menggabungkan additional keyword yang didapat berdasarkan nilai Relevance Degree terbesar dengan query awal. Dari hasil pengujian didapat bahwa metode ini dapat meningkatkan performansi dari Information Retrieval System (IRS) dimana peningkatan performansi disini ditunjukkan dari jumlah dokumen relevan yang ditemukan dan posisi dokumen relevan di dalam perangkingan dokumen hasil pencarian.**

**Kata Kunci : Information Retrieval, Query Expansion, fuzzy rules, keywords, dan query**

**Abstract**

**Information Retrieval (IR) is a system which used for retrieving information and relevant for user from information collection based on user's query. In retrieving that information, performance in searching can be increased by using query modification technique like Query Expansion (QE). Query Expansion is a technique with adding new keywords to original user's query in order to increase performance in searching. Query Expansion can be improved using inference of fuzzy rules. These method uses membership functions and fuzzy rules to find Relevance Degree(RD) from additional keyword candidate and then combine the additional keyword which have larger Relevance Degree into the original user's query. From the result, these method can increase the performance of Information Retrieval System (IRS) where the increasing performance show from the number of relevant documents can be retrieved and the position of the relevant documents in the ranking of document results.**

**Keywords : Information Retrieval, Query Expansion, fuzzy rules, keywords, and query**

# 1. PENDAHULUAN

## 1.1 Latar belakang masalah

Perkembangan teknologi di dunia Internet semakin pesat. Dunia Internet sekarang ini memang sulit untuk dipisahkan dari dunia nyata. Sebagian besar aktivitas *user* dilakukan melalui Internet dan kadang *user* membutuhkan bantuan di dalam dunia Internet untuk mencari referensi bagi pekerjaannya. Maka dari itu,dibuat *search engine* untuk membantu *user* sehingga mempermudah *user* untuk mendapatkan bahan referensi bagi pekerjaannya.

Oleh karena itu, *search engine* dituntut agar dapat memberikan hasil pencarian yang tepat dan akurat sesuai dengan keinginan *user*. Akan tetapi, hasil pencarian yang didapat kadang tidak relevan dengan yang diinginkan *user*. *Search engine* sendiri menggunakan *Information Retrieval* (IR) sebagai konsep dasarnya dimana Information retrieval merupakan suatu sistem yang digunakan untuk menemukan kembali(*retrieve*) informasi-informasi yang relevan terhadap kebutuhan *user* dari suatu kumpulan informasi berdasarkan *query* dari *user* [11]. Sehingga dibutuhkan solusi agar meningkatkan performansi dari *Information Retrieval* yaitu dengan *query reformulation techniques* (teknik memodifikasi *query* yang ada).

*Query reformulation techniques* sendiri terdiri dari 2 yaitu *query expansion methods* dan *query reweighting methods* [1]. Untuk kasus Tugas Akhir ini adalah *query expansion method* dimana *query* awal yang diinputkan oleh *user* diperluas dengan *keyword* baru. Di dalam tugas akhir ini sendiri, akan dikembangkan suatu metode *Query Expansion* (QE) yang baru berdasarkan penarikan kesimpulan dari *fuzzy rules* dan *pseudo relevance feedback*. Metode ini menggunakan fungsi-fungsi keanggotaan untuk merepresentasikan atribut dari *term* yang digunakan (dalam kasus ini digunakan *Relevant Frequency*, *inverse document frequency*, dan *Relevance Degree*) dan mengkalkulasikan fungsi-fungsi keanggotaan tersebut menggunakan *fuzzy rules* untuk mendapatkan nilai *Relevance Degree* dari calon *additional keywords* dan menggabungkan calon *additional keywords* yang memiliki nilai *Relevance Degree* paling besar dengan *query* awal inputan dari *user*. Model dengan perluasan metode ini diharapkan mendapatkan nilai *Interpolated Average Precision* (IAP) yang lebih baik dibandingkan model awal yang menggunakan Vector Space Model (VSM).

## 1.2 Perumusan masalah

Berdasarkan uraian diatas maka permasalahan yang muncul dan yang menjadi objek penelitian pada Tugas Akhir ini :

1. Bagaimana proses pencarian calon *additional keywords* dari *relevant documents* yang ada menggunakan *pseudo relevance feedback.*

2. Bagaimana proses pencarian nilai *Relevance Degree dengan menggunakan QE+fuzzy rules*.
3. Bagaimana perbandingan nilai IAP antara model awal yang diperluas dengan *QE+fuzzy rules* dengan model awal yang menggunakan vector space model

Batasan masalah agar tidak meluasnya materi pembahasan dalam tugas akhir ini adalah sebagai berikut :

1. Analisis data dilakukan terhadap standar koleksi dokumen untuk *information retrieval* yang didapat dari ftp://ftp.cs.cornel.edu/pub/smart/med/ yang bertipe file '.txt' dimana sudah terdapat kumpulan *relevance judgement.*
2. *Query* yang diinputkan akan ditentukan dan menggunakan *query* khusus untuk pengujian dari ftp://ftp.cs.cornel.edu/pub/smart/med .
3. Aplikasi hanya melakukan *word indexing* dan tidak melakukan *phrase indexing.*
4. Algoritma *stemming* yang digunakan menggunakan algoritma *Potter*.
5. *Query* yang diuji oleh aplikasi hanya bisa berbentuk *simple query* atau tidak bisa menggunakan *Boolean operation* atau *operation* yang lain.
6. Banyak dokumen dan calon *additional keywords* yang akan digunakan hasil dari *pseudo relevance feedback* hanya diambil top 5 dokumen teratas (dengan acuan nilai *similarity* terbesar) dan top 4 kata sebagai calon *additional keywords* (dengan acuan bobot *term* terbesar).
7. Koleksi dokumen dan *query* yang digunakan menggunakan bahasa inggris.
8. Pengujian performansi akan dilakukan dengan membandingkan nilai *Precision,Recall,* dan IAP yang didapat hanya melalui model awal yang diperluas dengan *QE+fuzzy rules* dengan model awal yang menggunakan vector space model.

## 1.3 Tujuan

Secara umum tujuan penulisan yang ingin dicapai dalam tugas akhir ini adalah:

1. Merancang dan membangun suatu *Information Retrieval System* berupa *search engine* yang menggunakan VSM dengan pengembangan *QE+fuzzy rules*.

2. Menganalisa nilai *Precision, Recall,* dan IAP yang didapat dan membuktikan apakah VSM dengan *QE+fuzzy rules* mampu memberikan hasil yang memuaskan bila dibandingkan dengan VSM biasa.

## 1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini adalah :

1. Studi Literatur
   Studi literatur dari beberapa buku, jurnal, artikel yang membahas tentang *Information Retrieval, Query Expansion, Fuzzy Logic*

2. Analisis dan Desain
   Tahap ini meliputi analisis kebutuhan serta penyelesaian masalah untuk merancang perangkat lunak *search engine* dengan VSM yang dikembangkan dengan *QE+fuzzy rules.*

3. Implementasi Sistem
   Tahap ini meliputi pembangunan perangkat lunak yang telah dirancang pada tahap sebelumnya. Pembangunan perangkat lunak lebih ke arah *web-based* dengan menggunakan PHP dan database MySQL.

4. Analisis dan Pengujian
   Pada tahapan ini yang dilakukan adalah melakukan pengujian terhadap perangkat lunak yang dibangun dan sekaligus melakukan analisis terhadap hasil pemrosesan perangkat lunak. Analisa performansi dari *search engine* ini setelah digunakan *Query Expansion* akan dinilai dari IAP yang dihasilkan dari model ini dan kemudian akan dibandingkan dengan nilai IAP dari model awalnya.

5. Penyusunan dan Laporan
   Hasil penelitian akan disusun menjadi suatu laporan yang meliputi aspek-aspek dalam penelitian yaitu teori, perancangan dan implementasinya, serta membuat kesimpulan dari hasil penelitian tersebut.

## 1.5 Sistematika Penulisan

Sistematika Penulisan Tugas Akhir ini terdiri dari 5 bab yaitu :

**BAB I Pendahuluan**
Bab ini membahas kerangka penelitian dalam tugas akhir, meliputi latar belakang, perumusan masalah, batasan masalah, tujuan perancangan dan metodologi yang digunakan dalam perancangan sistem.

**BAB II Landasan Teori**
Bab ini menjelaskan seluruh teori yang menjadi landasan konseptual dan mendukung penyelesaian tugas akhir ini.

**BAB III Analisis dan Perancangan Sistem**
Bab ini membahas mengenai pengumpulan data analisis dan perancangan perangkat lunak yang terdiri dari perancangan struktur data , perancangan modul dan *interface*.

**BAB IV Implementasi dan Pengujian Sistem**
Bab ini membahas implementasi detail sistem dan pengujian terhadap sistem

**BAB V Kesimpulan dan Saran**
Berisi tentang kesimpulan dan saran yang dapat diambil dari keseluruhan sistem yang telah dibuat.

# 5. KESIMPULAN DAN SARAN

Pada bab ini akan diuraikan hal yang dapat disimpulkan dari pelaksanaan Tugas Akhir ini. Selain itu diuraikan pula beberapa saran yang dapat digunakan dalam pengembangan Tugas Akhir di masa mendatang.

## 5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian perangkat lunak yang dilakukan dalam Tugas Akhir ini dapat diambil beberapa kesimpulan yaitu :

a. Berdasarkan *query* pengujian dari ftp://ftp.cs.cornel.edu/pub/smart/med/, penggunaan metode *query expansion* berdasarkan penarikan kesimpulan dari *fuzzy rules* pada VSM dapat meningkatkan performansi pada *Information Retrieval System* (IRS)
b. Nilai peningkatan performansi dari VSM yang menerapkan metode *query expansion* berdasarkan penarikan kesimpulan dari *fuzzy rules* tidak terlalu besar. Hal ini disebabkan karena IRS yang digunakan yaitu VSM sudah baik performansinya dalam melakukan pencarian dokumen.
c. Walau nilai *precision* dan *recall* yang didapat dari VSM dengan metode *query expansion* berdasarkan penarikan kesimpulan dari *fuzzy rules* memiliki nilai lebih rendah dibandingkan dengan VSM biasa,belum tentu nilai IAP yang dihasilkan akan menurun. Hal ini disebabkan pengaruh dari posisi dokumen relevan di dalam perangkingan dokumen hasil pencarian.

## 5.2 Saran

Untuk pengembangan Tugas Akhir di masa mendatang, penulis menyarankan hal-hal sebagai berikut:

a. *Fuzzy rules* dan *membership function* yang digunakan dapat dikembangkan lebih lanjut menggunakan *optimization algorithm* seperti *genetic algorithm* untuk menghasilkan formula *query* baru yang lebih baik lagi.
b. Proses *stemming* yang digunakan dapat dikembangkan lagi.
c. Pencarian *synonym words* akan meningkatkan proses pencarian.
d. Perangkat lunak diharapkan tidak hanya menangani *word indexing* tapi juga dapat menangani *phrase indexing.*
e. Jenis dokumen yang dicari tidak hanya berupa teks saja.

# DAFTAR PUSTAKA


[1] Baeza-Yates, R., and Ribeiro-Neto, B., *Modern Information Retrieval*, 1999, ACM Press, NY, USA.

[2] Chang,Y.C., Shyi-Ming Chen, Churn-Jung Liau, *A New Query Expansion Method for Document Retrieval Based on the Inference of Fuzzy Rules*, Journal of the Chinese Institute of Engineers, Vol. 30 No. 3, 511-515, 2007, Taiwan : National Taiwan University.

[3] Grossman D, Frieder O, Lundquist C 1997, *Improving Relevance Feedback in the Vector Space Model,* Las Vegas Nevada, USA. http://www.ir.iit.edu/publications/download/97-rel_feedback_vec.pdf. Didownload pada tanggal 3 Desember 2007.

[4] Ingwersen, Peter. 2005. *The Turn* : *System-Oriented Information Retrieval,* Book Series The Information Retrieval Series Vol. 18. Springer Netherlands Publishers, Netherlands.

[5] Kusumadewi, Sri, 2003, *Artificial Intelegence Teknik dan Aplikasinya*. Jogjakarta: Graha Ilmu

[6] Lin,Hsi-Ching, Li-Hui Wang, Shyi-Ming Chen. *Query expansion for Document Retrieval Based on Fuzzy Rules and User Relevance Feedback Techniques*, Journal of the Chinese Institute of Engineers, Vol. 31 No. 3, 397-405, 2006, Taiwan : National Taiwan University.

[7] Lin, Hsi-Ching, Li-Hui Wang, Shyi-Ming Chen, *Query Expansion for Document Retrieval by Mining Additional Query Terms*, Information and Management Sciences Vol. 19, No. 1, pp. 17-30, 2008, Taiwan : National Taiwan University

[8] L. M. de Campos, J. M. Fernandez-Luna, J. F. Huete. 2003. *Implementing Relevance Feedback in the Bayesian Network Retrieval Model*. 302-313

[9] Rocchio J. J., 1971, *Relevance Feedback in Information Retrieval*. In G. Salton (Ed.), *The SMART Retrieval System. Experiments in Automatic Document Processing* (pp. 313-323). Englewood Cliffs, New Jersey: Prentice Hall.

[10] Van Rijsbergen, C.J., 1979, *Information Retrieval*. Department of Computing Science, University of Glasgow.

[11] Witten, Ian H., Moffat, Alistair, Bell,Timothy C., "*Managing Gigabytes: Compressing and Indexing Documents and Images*", second edition. Morgan Kaufmann Publishers, Academic Press,1999.