

## ANALISIS DAN IMPLEMENTASI RELEVANCE FEEDBACK PADA INFORMATION RETRIEVAL MENGGUNAKAN METODE IDE DEC HI

Melva Sitorus<sup>1</sup>, Yanuar Firdaus A.w.<sup>2</sup>, Toto Suharto<sup>3</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

### Abstrak

Relevance feedback adalah salah satu teknik dari information retrieval (IR) yang memformulasi ulang kata kunci yang lama menjadi kata kunci yang baru berdasarkan dokumen mana yang relevan menurut pengguna[10]. Sejak tahun 1960-an relevance feedback telah digunakan pada information retrieval dan dapat meningkatkan precision dan recall sistem. Banyak hal yang digunakan untuk meningkatkan hasil dari relevance feedback dengan menggunakan kombinasi dari beberapa metode yang ada, ide-ide dari individu meneliti dan keseluruhan dapat memperbaiki hasil pencarian.

Terdapat beberapa metode yang digunakan untuk meningkatkan hasil dari relevance feedback, dalam penelitian ini menggunakan metode Ide dec hi. Metode Ide dec hi adalah metode relevance feedback yang baik dalam meningkatkan performansi sistem, dimana menggunakan sebuah dokumen tidak relevan pada rangking teratas sebagai umpan balik[9]. Peningkatan performansi perangkat lunak mencapai 112.89 % untuk precision, 74.921 % untuk recall dan IAP 35.83%

Kata Kunci : information retrieval system, kata kunci, feedback, ide dec hi, precision, recall.

---

### Abstract

Relevance feedback is a classic information retrieval (IR) technique that reformulates a query based on documents selected by the user as relevant [10]. Relevance feedback techniques have been recently an active research area in IR. Since the use of relevance feedback in information retrieval to improve precision and recall was first proposed in the late-1960's, many different techniques have been used to improve the results obtained from relevance feedback. Since most information retrieval systems performing relevance feedback use combinations of several techniques, the individual contribution of each technique to the overall improvement is relatively unknown.

There are several methods to improve relevance feedback and in this research use the Ide dec hi method. Ide dec hi method is the best overall relevance feedback method, where terms are directly added to the queries and only one non-relevant item is used in the process[9]. By relevance feedback using ide dec hi method can increasing system performance 112.89 % for precision, 74.921 % for recall and IAP 35.83%.

Keywords : information retrieval system, feedback, query, ide dec hi, precision, recall.

---

Telkom  
University

# 1. PENDAHULUAN

## 1.1 Latar Belakang

Pada saat sekarang ini, informasi sangat cepat berkembang. Bahkan informasi menjadi suatu kebutuhan yang sangat penting bagi pengguna. Berbagai jenis media dipenuhi beragam informasi salah satunya dalam bentuk dokumen dan dengan berkembangnya teknologi yang tinggi juga sangat mendukung dalam memberikan kemudahan dan kecepatan akses dalam memperolehnya. Namun dibalik kemudahan tersebut muncul sebuah kesulitan baru, pengguna mengalami kesulitan untuk memilih dokumen yang mengandung informasi yang dibutuhkan. Pengguna bisa saja memperoleh dokumen-dokumen tersebut dengan membaca semua dokumen dalam tempat penyimpanan kemudian menyimpan informasi yang relevan dan membuang yang tidak relevan. Pemilihan informasi ini mengalami kesulitan jika dilakukan secara manual karena kumpulan informasi yang sangat besar dan terus bertambah. Untuk mengatasi kesulitan inilah *Information Retrieval(IR)* dikembangkan.

Sistem temu kembali informasi(*information retrieval system*) merupakan sistem yang digunakan untuk menemukan kembali(*retrieve*) dokumen-dokumen yang sesuai dengan kata kunci yang diberikan pengguna dari suatu koleksi informasi[13].

*Relevance feedback* adalah mekanisme untuk mendapatkan *feedback* dari pengguna, dimana pengguna akan memasukkan kata kunci ke dalam sistem, kemudian sistem akan mengembalikan sejumlah dokumen yang dianggap relevan berdasarkan kata kunci yang disebut dengan hasil pencarian awal. *Relevance feedback* merupakan proses formulasi ulang kata kunci awal berdasarkan informasi umpan balik relevansi dari pengguna terhadap dokumen-dokumen hasil pencarian awal. *Feedback* digunakan untuk menentukan kata kunci baru. Metode *Ide dec hi* adalah metode *relevance feedback* yang menggunakan sebuah dokumen tidak relevan pada rangking teratas sebagai *feedback*[11]. Berdasarkan *feedback*, sistem secara otomatis akan menentukan kata kunci baru dan melakukan pencarian berdasarkan kata kunci baru tersebut.

## 1.2 Perumusan Masalah

Permasalahan yang dijadikan sebagai objek penelitian pada tugas akhir ini adalah

1. Bagaimana menerapkan *relevance feedback* pada *information retrieval* menggunakan metode *Ide dec hi*.
2. Bagaimana pengaruh *relevance feedback* pada *information retrieval* menggunakan metode *Ide dec hi* dalam memberikan informasi yang relevan dengan kebutuhan pengguna.
3. Bagaimana penggunaan *relevance feedback* dengan metode *Ide dec hi* dapat meningkatkan performansi *Information Retrieval System(IRS)*.

Batasan masalah dari tugas akhir ini adalah:

1. Koleksi dokumen dan kata kunci yang digunakan adalah dokumen dalam teks bahasa Inggris.

2. Koleksi dokumen menggunakan koleksi dokumen yang diambil dari <ftp://ftp.cs.cornell.edu/SMART> yang terdiri dari tiga koleksi dokumen yaitu koleksi dokumen CRAN, dan MED, dan pada koleksi dokumen tersebut sudah terdapat kumpulan kata kunci beserta *relevance judgement* untuk setiap kata kunci tersebut.
3. Model sistem temu kembali informasi yang digunakan adalah model ruang vektor, dimana dokumen maupun kata kunci direpresentasikan sebagai vektor berdimensi  $n$  dengan  $n$  adalah jumlah kata atau *term* pada kata kunci. *Relevance feedback* pada model ruang vektor dapat dijelaskan sebagai penggeseran vektor kata kunci mendekati vektor dokumen relevan dan menjauhi vektor dokumen tidak relevan[2].
4. Pembobotan kata kunci akan menggunakan *term frequency* , sedangkan pembobotan dokumen akan menggunakan *term frequency inverse document frequency*.
5. Mekanisme penggunaan *relevance feedback* yang digunakan adalah secara manual

### 1.3 Tujuan

Adapun tujuan Tugas Akhir ini adalah :

1. Merancang dan membangun perangkat lunak untuk penerapan *relevance feedback* pada IR menggunakan metode Ide dec hi.
2. Melakukan analisis performansi terhadap hasil dari penerapan metode Ide dec hi, dimana performansi sistem diukur berdasarkan nilai dari *precision*, *recall*, *Interpolated Average Precision(IAP)*.

### 1.4 Metodologi Penyelesaian Masalah

Metodologi penyelesaian yang digunakan dalam menyelesaikan tugas akhir ini adalah:

1. Studi literatur  
Mempelajari sumber-sumber pustaka yang ada, yang dapat dijadikan referensi mengenai *information retrieval* khususnya model ruang vektor, *relevance feedback*, metode Ide dec hi serta sumber-sumber lain yang relevan untuk menunjang penyelesaian tugas akhir ini. Sumber-sumber pustaka dapat berupa buku, *paper*, maupun halaman *web*.
2. Analisis  
Melakukan analisis lebih lanjut mengenai metode Ide dec hi dalam *relevance feedback*. Tujuan analisis adalah untuk mengetahui karakteristik metode Ide dec hi
3. Analisis kebutuhan Perangkat Lunak  
Yaitu melakukan analisis kebutuhan mengenai perangkat lunak yang akan digunakan dan perangkat lunak yang akan dikembangkan, sehingga didapat gambaran umum mengenai perangkat lunak yang akan dikembangkan
4. Perancangan perangkat lunak  
Melakukan perancangan perangkat lunak yang akan dikembangkan. Perancangan yang dilakukan adalah perancangan arsitektur perangkat lunak

5. Implementasi  
Melakukan pembangunan perangkat lunak yang telah dirancang berdasarkan analisis dan perancangan yang telah dilakukan.
6. Pengujian dan analisis hasil.  
Melakukan pengujian perangkat lunak yang telah dikembangkan, dan kemudian menganalisis hasil performansi yang didapatkan. Pengukuran performansi adalah *precision*, *recall*, *Interpolated Average Precision(IAP)* dan waktu yang digunakan selama pemrosesan. Tujuan pengujian adalah untuk mengetahui apakah terjadi peningkatan performansi pada penggunaan metode *Ide dec hi* dalam *relevance feedback*.

## 1.5 Sistematika Penulisan

Struktur Pembahasan Tugas Akhir ini terdiri dari lima bab dengan disertai lampiran terkait pelaksanaan tugas akhir :

- BAB I           PENDAHULUAN  
Bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan pembahasan, metodologi pemecahan masalah dan sistematika penulisan.
- BAB II           DASAR TEORI  
Membahas dasar teori yang berhubungan dengan tugas akhir ini. Secara garis besar dasar teori terdiri dari *information retrieval*, *information retrieval* model ruang vektor, *relevance feedback* dalam model ruang vektor, metode *ide dec hi*, dan penggunaan metode *ide dec hi* dalam *information retrieval*.
- BAB III          ANALISIS DAN PERANCANGAN PERANGKAT LUNAK  
Bab ini berisi analisis domain masalah, analisis perangkat lunak, dan perancangan perangkat lunak. Pada bab ini akan dilakukan analisis pada pemilihan metode *relevance feedback*. Berdasarkan hasil analisis tersebut, dilakukan analisis kebutuhan perangkat lunak.
- BAB IV          IMPLEMENTASI DAN PENGUJIAN  
Bab ini berisi implementasi perangkat lunak berdasarkan hasil perancangan analisis dan perancangan pada bab sebelumnya dan berisi skenario pengujian dan hasil pengujian dari perangkat lunak yang dikembangkan dalam tugas akhir ini
- BAB V          KESIMPULAN & SARAN  
Bab ini akan menjelaskan kesimpulan dan saran sebagai hasil dari analisis dan implementasi Tugas Akhir.

## 6. PENUTUP

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.

### 6.1 Kesimpulan

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

1. Penggunaan *relevance feedback* dengan metode Ide dec hi dapat meningkatkan performansi *information retrieval system*.
2. Penambahan jumlah dokumen yang dijadikan sebagai *feedback* akan memberikan performansi yang bertambah pula jika dibandingkan dengan pencarian awal yang tanpa *feedback*.
3. Untuk pengujian parameter optimal, nilai *precision*, *recall* akan cenderung naik karena dokumen-dokumen relevan yang terambil perangkat lunak peringkatnya menaik. Sedangkan dengan memperhatikan aspek keterurutan dokumen hasil pencarian yaitu dengan menghitung nilai IAP, nilai IAP tidak selalu meningkat seiring dengan penambahan jumlah *feedback*. Pada jumlah *feedback* tertentu akan ada penurunan, hal ini disebabkan perhitungan nilai IAP yang melihat keterurutan dokumen hasil pencarian, jadi sangat tergantung kepada rangking dokumen relevan yang ditemukan. Semakin tinggi rangking dari dokumen yang relevan yang ditemukan maka nilai IAP akan meningkat.
4. waktu pencarian sangat dipengaruhi oleh jumlah term yang terdapat pada kata kunci, jumlah term pada tiap dokumen, banyak dokumen yang terdapat pada koleksi dokumen. Semakin banyak jumlah *term* pada kata kunci, jumlah term pada tiap dokumen, dan banyaknya dokumen pada koleksi dokumen akan memberikan penurunan pada waktu pencarian.

### 6.2 Saran

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

1. Penambahan jenis *file* dokumen yang dapat dicari agar tidak hanya (.txt) saja.
2. Perangkat lunak sebaiknya mampu menangani jika pengguna melakukan beberapa kali *feedback*.

## Daftar Pustaka

[1]	Berry. Michael W, and B.Murray, 1999, <i>Understanding Search Engine: Mathematical Modeling and Text Retrieval</i> . University of Tennessee.
[2]	Grossman D, Frieder O, Lundquist C 1997, <i>Improving Relevance Feedback in the Vector Space Model</i> , Las Vegas Nevada, USA. <a href="http://www.ir.iit.edu/publications/download/97-rel_feedback_vec.pdf">http://www.ir.iit.edu/publications/download/97-rel_feedback_vec.pdf</a> . Didownload pada tanggal 3 Desember 2007.
[3]	Grossman D, Frieder O, Goharian, 2002, <i>Relevance Feedback</i> , Las Vegas Nevada, USA.
[4]	Harman D, 1992, Relevance feedback Revisited. In <i>15th ACM SIGIR Conference on Research and Development in Information Retrieval</i> . Denmark.
[5]	Hiemstra D, Robertson S, 2002, Relevance Feedback for Best Match Term Weighting Algorithms in Information Retrieval. Microsoft Research, Cambridge, UK.
[6]	Ide E, 1971, New Experiments In Relevance Feedback. In G. Salton (Ed.), <i>The SMART Retrieval System. Experiments in Automatic Document Processing</i> (pp. 337-354). Englewood Cliffs, New Jersey: Prentice Hall.
[7]	Rocchio J. J., 1971, Relevance Feedback in Information Retrieval. In G. Salton (Ed.), <i>The SMART Retrieval System. Experiments in Automatic Document Processing</i> (pp. 313-323). Englewood Cliffs, New Jersey: Prentice Hall.
[8]	Orengo, Moreira V, 2004, "Assesing relevance using Automatically translated documents for cross-Language Information" . Middlesex University. Rome, Italy.
[9]	Salton G, 1971, Relevance Feedback and the Optimisation of Retrieval Effectiveness. In G. Salton (Ed.), <i>The SMART Retrieval System. Experiments Automatic Document Processing</i> (pp. 324-336). New Jersey: Prentice Hall.
[10]	Salton G and C. Buckley, 1997, Improving Retrieval Performance by Relevance Feedback. In K. Sparck Jones & P. Willet (Eds.), <i>Readings in Information Retrieval</i> (pp. 355-364). San Francisco, California: Morgan Kauffmann.
[11]	Salton G and C.Buckley , 1990, Improving Retrieval Performance by Relevance Feedback, <i>Journal of American Society for Information Science</i> <b>41</b> (4), pp.299-297, Department of Computer Science, Cornell University, Ithaca. <a href="http://www.cs.fiu.edu/vagelis/classes/cop6776/publications/jasistSalton1990.pdf">http://www.cs.fiu.edu/vagelis/classes/cop6776/publications/jasistSalton1990.pdf</a> . Didownload pada tanggal 3 Desember 2007.
[12]	Tresp Volker, Xu Xiaowei, Kai Yu, A Hybrid Relevance Feedback Approach to Text Retrieval. Siemens AG, Corporate Technology, Munich, Germany
[13]	Van Rijsbergen, C.J., 1979, <i>Information Retrieval</i> . Department of Computing Science, University of Glasgow.