

# **RELEVANCE FEEDBACK PADA TEMU-KEMBALI TEKS BERBAHASA INDONESIA DENGAN METODE *IDE-DEC-HI* DAN *IDE-REGULAR***

Julio Adisantoso, Ahmad Ridha, Andika Wahyu Agusetyawan

## **ABSTRAK**

*Tujuan penelitian ini adalah mengimplementasikan dan menganalisis kinerja perluasan kueri dengan relevance feedback pada sistem temu kembali informasi untuk dokumen berbahasa Indonesia. Metode relevance feedback yang digunakan adalah Ide-Dec-Hi dan Ide-Regular. Untuk kepentingan pengujian, penelitian ini juga melengkapi corpus yang digunakan dengan 30 kueri disertai gugus jawabannya. Evaluasi kinerja relevance feedback dilakukan menggunakan test and control group. Masing-masing group terdiri atas 500 dokumen yang berupa artikel-artikel pertanian berbahasa Indonesia dari berbagai situs media massa. Sistem dasar yang digunakan adalah sistem temu kembali berbasis vector space model hasil penelitian Ridha (2002). Sistem ini melakukan rule-based stemming sekaligus memakai stoplist untuk bahasa Indonesia. Variasi jumlah dokumen yang diperiksa yakni lima dan sepuluh.*

*Hasil penelitian menunjukkan bahwa relevance feedback secara keseluruhan meningkatkan kinerja sistem temu kembali. Siklus relevance feedback dalam penelitian ini telah menunjukkan hasil memuaskan pada iterasi pertama. Peningkatan kinerja terbesar diperoleh ketika menggunakan formula Ide-Dec-Hi. Hasil ini sesuai dengan hasil penelitian Ruthven & Lalmas (2003). Kinerja sistem tanpa relevance feedback adalah 0.447 sedangkan dengan Ide-Dec-Hi mencapai 0.516, meningkat 15.44%. Sementara menggunakan Ide-Regular peningkatan yang diperoleh adalah 14.54%, menjadi 0.512. Dari perbandingan query-by-query dapat disimpulkan bahwa penggunaan relevance feedback tidak terlalu membantu pada kueri yang kinerja awalnya memang sudah tinggi. Sebaliknya, untuk kueri-kueri yang memberikan hasil buruk pada pencarian awal, relevance feedback sangat cocok untuk digunakan dan menjanjikan peningkatan kinerja yang cukup tinggi.*

**Kata kunci:** Temu Kembali Informasi, *Relevance Feedback*, formula *Rocchio*, *Ide-Dec-Hi* dan *Ide-Regular*.

## **1. PENDAHULUAN**

### **Latar Belakang**

Sejalan dengan semakin populernya penggunaan Internet dan Perpustakaan Digital, informasi dalam jumlah yang luar biasa besar kini bisa diakses secara luas oleh masyarakat, suatu hal yang tidak tersedia di masa lalu. Akan tetapi bersamaan dengan itu, muncul masalah baru. Pengguna kini harus menyaring seluruh kumpulan informasi tersebut untuk menemukan kebutuhan informasinya. Bahkan setelah melalui bantuan *search engine* pun, yang rata-rata kini telah mengindeks milyaran halaman *Web* dari beragam kategori, setelah disaring melalui penggunaan kata kunci, pengguna seringkali belum bisa menemukan dokumen yang relevan dengan informasi yang dicarinya.

Dalam temu kembali informasi, jumlah dokumen relevan yang ditemukembalikan

akan dipengaruhi oleh jumlah kata kunci yang digunakan untuk pencarian. Dalam *Web*, pencarian pada umumnya dilakukan menggunakan kueri pendek, hanya kurang dari tiga kata (Spink *et al.* 2000). Hal ini cukup menyulitkan karena kueri pendek hanya menyediakan sedikit informasi bagi sistem temu kembali mengenai kebutuhan pencarian pengguna. Sebagai hasilnya adalah meskipun jumlah dokumen yang ditemukembalikan banyak akan tetapi hanya sedikit yang dianggap relevan oleh pengguna.

Masalah lain yang sering terjadi dalam *Web* dan Perpustakaan Digital adalah pilihan kata yang digunakan pengguna dalam kueri seringkali berbeda dengan pilihan kata yang digunakan penulis dalam dokumen. Bahkan ketika beberapa pengguna memiliki kebutuhan informasi yang sama, jarang mereka menggunakan pilihan kata yang sama untuk menggambarannya dalam kueri (Xu &

Croft 2005). Beberapa pendekatan dalam temu kembali informasi telah dilakukan untuk menanggapi perbedaan ini.

Salah satu pendekatannya adalah penggunaan *relevance feedback*. Dalam pendekatan ini sistem menemukembalikan dokumen-dokumen yang mungkin relevan bagi kueri pengguna, selanjutnya pengguna memeriksa beberapa dokumen tersebut dan menandai dokumen yang dianggap relevan. Informasi ini digunakan oleh sistem untuk memodifikasi atau menambahkan kata dalam kueri dan diharapkan meningkatkan hasil pencarian. Siklus ini bisa terus diulang sampai pengguna merasa kebutuhan informasinya telah terpenuhi.

### Tujuan

Penelitian ini bertujuan mengimplementasikan dan menganalisis kinerja perluasan kueri dengan *relevance feedback* pada sistem temu kembali informasi untuk dokumen berbahasa Indonesia.

### Ruang Lingkup

Penelitian ini dibatasi pada analisis kinerja *relevance feedback* pada sistem temu kembali informasi berbahasa Indonesia menggunakan formula *Ide-Dec-Hi* dan *Ide-Regular*. Model sistem temu kembali yang digunakan adalah *vector space model*.

## 2. DESKRIPSI ALGORITME RELEVANCE FEEDBACK (RF)

Teknik RF ditemukan pertama kali oleh Rocchio dan diterapkan dalam *vector space model*. Rocchio memandang temu kembali sebagai permasalahan mencari sebuah kueri optimal, yaitu kueri yang memaksimalkan selisih antara vektor rata-rata dokumen relevan dengan vektor rata-rata dokumen tak-relevan.

RF berguna untuk mendekatkan vektor kueri ke vektor rata-rata dokumen relevan dan menjauhkan dari vektor rata-rata dokumen tak-relevan. Ini bisa dilakukan melalui penambahan istilah kueri dan penyesuaian bobot istilah kueri sehingga sesuai dengan kegunaan istilah tersebut dalam fungsinya membedakan dokumen relevan dan tak-relevan (Ruthven & Lalmas 2003).

Formula awal *Rocchio* adalah sebagai berikut:

$$Q_1 = Q_0 + \frac{1}{n_1} \sum_{i=1}^{n_1} R_i - \frac{1}{n_2} \sum_{i=1}^{n_2} S_i,$$

dengan  $Q_0$ =vektor kueri awal,  $Q_1$ =vektor kueri baru,  $n_1$ =jumlah dokumen relevan,  $n_2$ =jumlah dokumen tak-relevan,  $R_i$ =vektor dokumen relevan ke- $i$ ,  $S_i$ =vektor dokumen tak-relevan ke- $i$ . Kueri yang telah termodifikasi akan memiliki istilah-istilah baru (dari dokumen-dokumen relevan). Selain itu penyesuaian bobot diterapkan terhadap istilah-istilah kueri. Jika bobot istilah kueri turun hingga nol atau di bawah nol, istilah tersebut dihilangkan dari kueri.

Ide (1971), diacu dalam Ruthven & Lalmas (2003), kemudian melakukan modifikasi terhadap formula *Rocchio*, dinamakan *Ide-Dec-Hi*, dengan menggunakan hanya dokumen tak-relevan pertama yang ditemukembalikan. Formula tersebut adalah sebagai berikut:

$$Q_1 = Q_0 + \sum_i^{n_r} r_i - s_i,$$

dengan  $n_r$ =jumlah dokumen relevan,  $r_i$ =vektor dokumen relevan ke- $i$ ,  $s_i$ =vektor dokumen tak-relevan pertama.

Modifikasi lain yang lazim dilakukan adalah dengan memboboti kontribusi relatif kueri awal, dokumen relevan dan dokumen tak-relevan terhadap proses RF. Dalam persamaan di bawah ini, nilai  $\alpha$ ,  $\beta$  dan  $\gamma$  merupakan derajat pengaruh tiap komponen dalam RF.

$$Q_1 = \alpha \cdot Q_0 + \frac{\beta}{n_1} \sum_{i=1}^{n_1} R_i - \frac{\gamma}{n_2} \sum_{i=1}^{n_2} S_i,$$

Versi yang lebih baru lagi dari skema Ide adalah *Ide-regular*, yang menggunakan semua dokumen non-relevan yang ditemukembalikan. *Ide-regular* juga berdasar formula *Rocchio* namun tidak melakukan normalisasi terhadap vektor-vektor dokumen. Berikut adalah formula *Ide-regular*:

$$Q_1 = Q_0 + \sum_{i=1}^{n_1} R_i - \sum_{i=1}^{n_2} S_i.$$

## 3. METODOLOGI Algoritme RF

Penelitian ini akan menginvestigasi kinerja relatif dua algoritme RF yaitu *Ide-Dec-Hi* dan

*Ide-Regular*. Keduanya dipilih karena meskipun formula *Ide-Dec-Hi* dan *Ide-Regular* tidak memberikan peningkatan kinerja yang signifikan namun lebih konsisten dibanding formula *Rocchio* (Ruthven & Lalmas 2003).

### Sistem Temu Kembali dan Evaluasinya

Sistem dasar yang digunakan untuk pengujian RF adalah sistem temu kembali hasil penelitian Ridha (2002). Sistem yang dihasilkan menggunakan *vector space model* dengan pembobotan *tf\*idf*. Dalam penelitian ini diterapkan *rule-based stemming* dan penggunaan daftar kata buang dalam Bahasa Indonesia.

Cara yang paling umum untuk mengevaluasi sistem temu kembali adalah dengan menentukan kerjanya dalam *recall* dan *precision*. Hal ini dilakukan dengan menggunakan koleksi pengujian yang terdiri dari kumpulan dokumen yang disertai gugus kueri dan penilaian relevansinya (gugus jawaban).

Dari hasil penemuan kembali yang dilakukan sistem, dihitung *precision* pada berbagai tingkat *recall*. Tingkat *recall* yang digunakan adalah sebelas tingkat *recall* standar terinterpolasi. Hasilnya kemudian dirata-rata untuk mendapatkan *average precision* (AVP).

### Evaluasi RF

Chang *et al.* (1971), diacu dalam Ruthven & Lalmas (2003) menunjukkan bahwa evaluasi algoritme RF memiliki beberapa masalah berkaitan dengan *recall* dan *precision*. RF bertujuan untuk meningkatkan *recall* dan *precision* berdasar informasi dari dokumen-dokumen relevan yang diidentifikasi pengguna, hal ini biasanya akan membuat dokumen-dokumen relevan yang telah diidentifikasi (yaitu dokumen relevan yang digunakan oleh RF) terdorong menempati ranking teratas. *Ranking effect* ini seolah-olah membuat kurva *recall-precision* meningkat tajam karena sistem meranking ulang dokumen relevan yang telah diidentifikasi. Seberapa banyak RF meningkatkan penemuan kembali dokumen relevan yang belum teridentifikasi, *feedback effect*-nya, justru tidak terdeteksi.

Untuk mengatasi masalah pengukuran *feedback effect*, digunakan teknik *test and control groups*. Dalam teknik ini, koleksi dokumen dibagi menjadi dua bagian, *test*

*group* dan *control group*. Modifikasi kueri dilakukan oleh RF pada *test group* dan kueri baru dijalankan pada *control group*. Penilaian *recall* dan *precision* hanya dilakukan pada *control group* sehingga tidak terjadi *ranking effect*. *Test and control* pada dasarnya mengukur kinerja relatif kueri baru pada tiap iterasi (Ruthven & Lalmas 2003).

Dalam penelitian ini pembagian koleksi untuk *test* dan *control group* dilakukan dengan cara sebagai berikut:

1. Dokumen-dokumen dikelompokkan berdasar sumbernya
2. Pada masing-masing kelompok, dokumen diurutkan berdasar ID dokumen
3. Untuk setiap kelompok, separuh dokumen teratas dimasukkan ke dalam *test group* dan sisanya masuk ke *control group*

Untuk melihat pengaruh penggunaan RF terhadap kinerja sistem dilakukan pengukuran *average precision* pada lima jenis penemuan kembali, yakni:

1. NoRF: penemuan kembali awal pada *control group*
2. DH5: menggunakan *Ide-Dec-Hi* dengan memeriksa 5 dokumen teratas
3. DH10: menggunakan *Ide-Dec-Hi* dengan memeriksa 10 dokumen teratas
4. RG5: menggunakan *Ide-Regular* dengan memeriksa 5 dokumen teratas
5. RG10: menggunakan *Ide-Regular* dengan memeriksa 10 dokumen teratas

Perbandingan *query-by-query* juga dilakukan untuk melihat secara lebih detail pengaruh RF pada tiap kueri. Selanjutnya dilakukan pengujian statistik uji *Wilcoxon Signed Rank Test* dengan selang kepercayaan 95% terhadap:

1. DH5 - NORF dan DH10 - NORF: untuk pengaruh formula *Ide-Dec-Hi*
2. RG5 - NORF dan RG10 - NORF: untuk pengaruh formula *Ide-Regular*
3. RG10 - RG5 dan DH10 - DH5: untuk pengaruh jumlah dokumen yang diperiksa.
4. DH10 - RG5, DH5 - RG10, DH5 - RG5 dan DH10 - RG10: untuk perbedaan kinerja *Ide-Dec-Hi* dan *Ide-Regular*.

Pengujian dilakukan oleh dua mahasiswa Fakultas Pertanian dengan skenario sebagai berikut:

1. Penguji mengeksekusi kueri yang telah disediakan pada gugus kueri pada *test group*.
2. Dari hasil penemuan kembali diperiksa dokumen-dokumen teratas. Dokumen yang dianggap relevan ditandai dan seluruh

- dokumen yang tidak ditandai akan dianggap tidak relevan.
- RF kemudian dilakukan untuk memodifikasi kueri.
  - Kueri yang telah dimodifikasi lalu dijalankan pada *control group*.

Pengujian RF dilakukan pada kedua formula, *Ide-Dec-Hi* dan *Ide-Regular*. Untuk masing-masing formula dilakukan dua kali iterasi dan untuk setiap kali iterasi variasi jumlah dokumen yang diperiksa adalah lima dan sepuluh.

### Corpus

Penelitian ini menggunakan *corpus* yang merupakan hasil penelitian Adisantoso & Ridha (2004). Koleksi dokumen yang dimiliki *corpus* ini terdiri dari 1000 artikel berbahasa Indonesia yang seluruhnya mempunyai domain yang sama, yaitu pertanian. Artikel-artikel ini dikumpulkan dari berbagai situs *web* Indonesia, yang sebagian besar merupakan situs-situs media massa. Sumber-sumber tersebut antara lain:

- Gatra
- Indosiar
- Kompas
- Media Indonesia
- Republika
- Situs Hijau
- Suara Karya
- Suara Merdeka

Sebagian besar dokumen yang terkumpul bersifat artikel media umum sedangkan hanya 22 dokumen yang merupakan tulisan ilmiah. Seluruh sumber artikel menggunakan Bahasa Indonesia semi-formal/formal (Adisantoso & Ridha 2004).

Untuk kepentingan evaluasi RF koleksi dokumen dibagi dua secara acak menjadi *test* dan *control group*. Langkah berikutnya adalah pembentukan gugus kueri untuk pengujian beserta gugus jawabannya karena keduanya belum tersedia dalam *corpus*. Pembentukan gugus kueri dan gugus jawaban dilakukan oleh dua mahasiswa Fakultas Pertanian yang dianggap kompeten untuk menentukan penilaian relevansi dokumen-dokumen dengan domain pertanian.

### Asumsi-asumsi

Asumsi-asumsi yang digunakan dalam penelitian ini adalah sebagai berikut:

- Jumlah dokumen yang diperiksa untuk *feedback* dalam setiap iterasi adalah tetap.

- Jumlah dokumen relevan untuk tiap kueri telah diketahui sebelumnya.
- Penilaian relevansi yang dilakukan akurat.

### Lingkungan Pengembangan

Lingkungan pengembangan yang digunakan adalah sebagai berikut:

- Perangkat lunak: Windows XP Professional, Visual Basic 6.0, Microsoft Access 2004.
- Perangkat keras: Duron 1.2 GHz, 256 MB RAM.

## 4. HASIL DAN PEMBAHASAN

### Gugus Kueri dan Gugus Jawaban

Kueri yang dibuat untuk pengujian penelitian ini berjumlah 30. Sebagian besar berupa frasa, seperti “gagal panen” atau “tanaman obat”, yang panjangnya kurang dari lima kata. Hal ini dilakukan untuk mensimulasikan keadaan di dunia nyata yang pada umumnya pengguna jarang menggunakan kueri panjang.

Setelah kueri tersedia, hal selanjutnya yang harus dilakukan adalah pembentukan gugus jawaban pada *test* dan *control group*. Pembentukan gugus jawaban merupakan proses yang penting dan sangat mempengaruhi hasil evaluasi RF sekaligus sistem temu kembali pada umumnya. Jika gugus jawaban yang dibentuk tidak akurat maka kinerja sistem akan tampak buruk karena ada dokumen-dokumen yang ditemukembalikan sistem, yang sebenarnya relevan, akan dianggap tidak relevan sebab tidak ada dalam gugus jawaban dan berakibat mengurangi nilai *recall* dan *precision*.

Pembentukan gugus jawaban menggunakan metode *pooling* dan dilakukan sebanyak dua kali. Pada kali pertama, sistem-temu kembali yang digunakan untuk mendapatkan *pool* adalah *Google Desktop Search 20050325* dan *Copernic Desktop Search* versi 1.63. Keduanya merupakan perangkat lunak *freeware* yang dikenal luas sebagai mesin pencari dengan kinerja yang cukup baik. Hasil *pool* yang didapat ternyata tidak terlalu akurat karena setelah diperiksa masih cukup banyak dokumen-dokumen yang sebenarnya relevan namun tidak termasuk dalam *pool*. Alasannya adalah karena memang kedua sistem temu kembali tersebut tidak dikhususkan untuk mencari dokumen-dokumen berbahasa Indonesia. Keduanya

tidak menggunakan *stemming* terhadap kueri maupun dokumen-dokumen berbahasa Indonesia yang terindeks.

Pada kali kedua, sistem temu kembali hasil penelitian Ridha (2002) diikutsertakan untuk mendapatkan calon dokumen relevan tambahan. Karena sistem ini memang dikhususkan untuk Bahasa Indonesia, hasil yang didapat cukup memuaskan dengan didapatkannya tambahan dokumen relevan hampir untuk setiap kueri yang diujikan. Tabel 1 menunjukkan deskripsi *corpus* yang digunakan sedangkan di Tabel 2 dapat dilihat contoh hasil *pooling* beberapa kueri menggunakan ketiga sistem yakni *Google Desktop Search* (GDS), *Copernic Desktop Search* (CDS) dan sistem temu kembali hasil penelitian Ridha (2002) (R2002).

Tabel 1 Deskripsi koleksi pengujian

Koleksi	Test	Control
Jumlah kueri	30	30
Ukuran (KB)	16.048	23.972
Jumlah dokumen	500	500
Rataan kata tiap kueri	2,6	2,6
Rataan kata tiap dokumen	166	177
Rataan dokumen relevan tiap kueri	13	18,8

Tabel 2 Contoh hasil *pooling* beberapa kueri

ID	Tingkat Recall *		
	GDS	CDS	R2002
1	28/48 (8)	21/48 (4)	25/48 (8)
5	21/39 (6)	17/39 (2)	26/39 (9)
11	34/57 (7)	30/57 (5)	41/57 (9)

\*Angka di dalam tanda kurung menunjukkan jumlah dokumen unik yang hanya ditemukan pada sistem tersebut dan tidak ditemukan di kedua sistem lainnya.

#### Evaluasi RF

Hasil pengujian RF dengan perhitungan *average precision* dapat dilihat pada Tabel 3 sedangkan hasil *Wilcoxon Signed Rank Test* dirinci pada Tabel 4.

Dari Tabel 3 terlihat bahwa hampir secara keseluruhan RF memberikan peningkatan kinerja pada sistem. Pada DH5 dan DH10 iterasi ke-1, terlihat bahwa dengan RF kinerja sistem meningkat 15.44% dibanding tanpa RF. Peningkatan kinerja ini secara statistik signifikan (*Sig.* < 0.05). Pada *Ide-Regular*, peningkatan juga signifikan, RG5 mencapai

14.54%, akan tetapi kemudian turun pada RG10 tinggal 12.75%. Ini menunjukkan bahwa meskipun secara statistik tidak ada perbedaan signifikan antara *Ide-Dec-Hi* dan *Ide-Regular* (*Sig.* > 0.05), *Ide-Dec-Hi* mampu memberikan hasil yang lebih konsisten daripada *Ide-Regular*. Hal ini sesuai dengan hasil penelitian Ruthven dan Lalmas (2003). Kekonsistenan ini karena *Ide-Dec-Hi* hanya menggunakan satu dokumen tak-relevan teratas. Sehingga meskipun terjadi ketidakakuratan penilaian relevansi saat pengguna memberikan *feedback*, yaitu dokumen relevan tidak ditandai oleh pengguna dan dihitung sebagai dokumen tak-relevan, hal itu tidak akan terlalu memperburuk kueri yang baru.

Tabel 3 Perbandingan nilai *average precision* sebelum dan setelah menggunakan RF

Metode	NoRF	Iterasi I	Iterasi II
DH5	0.447	0.516	0.492
% *	0	15.44%	10.07%
DH10	0.447	0.516	0.492
% *	0	15.44%	10.07%
RG5	0.447	0.512	0.477
% *	0	14.54%	6.71%
RG10	0.447	0.504	0.437
% *	0	12.75%	-2.24%

\*Persentase perubahan kinerja dibandingkan dengan NoRF

Tabel 4 Daftar nilai *Sig.* pada *Wilcoxon Signed Ranks Test* ( $\alpha=0.05$ )

Pair	Iterasi I	Iterasi II
RG5 - NORF	0.008*	0.220
RG10 - NORF	0.008*	0.876
DH5 - NORF	0.009*	0.102
DH10 - NORF	0.009*	0.102
RG10 - RG5	0.157	0.088
DH10 - DH5	1.000	1.000
DH10 - RG5	0.553	0.987
DH5 - RG10	0.078	0.020*
DH5 - RG5	0.553	0.987
DH10 - RG10	0.078	0.020*

\**Sig.* < 0.05

Pada iterasi ke-2, DH5 dan DH10 keduanya hanya memberikan peningkatan masing-masing sebesar 10.07%, lebih kecil dari iterasi ke-1. RG5 juga mengalami penurunan kinerja dibanding iterasi ke-1 menjadi 6.71%. RG10 bahkan menurun hingga -2.24%. Pada iterasi kali ini, DH5 dan DH10 secara signifikan lebih baik bila dibandingkan RG10 namun tidak jika

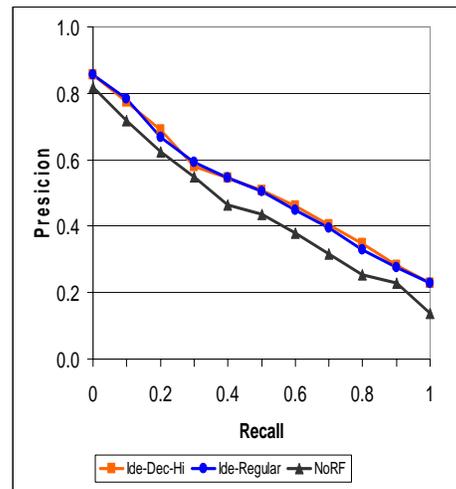
dibandingkan RG5. Melalui pemeriksaan lebih lanjut, alasan penurunan kinerja ini adalah karena pada iterasi ke-1 RF sudah bekerja dengan sangat baik sehingga pada hampir semua kueri, seluruh dokumen relevan untuk kueri tersebut telah ditemukan oleh sistem. Lebih jauh lagi untuk kueri-kueri yang belum seluruh dokumen relevannya ditemukan tingkat *recall*-nya sudah mencapai di atas 90%. Hal lain yang berkaitan erat dengan penurunan ini adalah kinerja sistem dasar yang ternyata sudah cukup tinggi.

Melalui hasil perbandingan *query-by-query*, ditemukan pula bahwa kueri yang tidak berhasil ditingkatkan kinerjanya adalah kueri-kueri yang seluruh atau setidaknya nyaris seluruh dokumen relevannya telah muncul di penemukembalian awal. Hal ini mengindikasikan bahwa pada keadaan-keadaan seperti ini RF sudah tidak cocok lagi digunakan untuk membantu menemukan tambahan dokumen relevan.

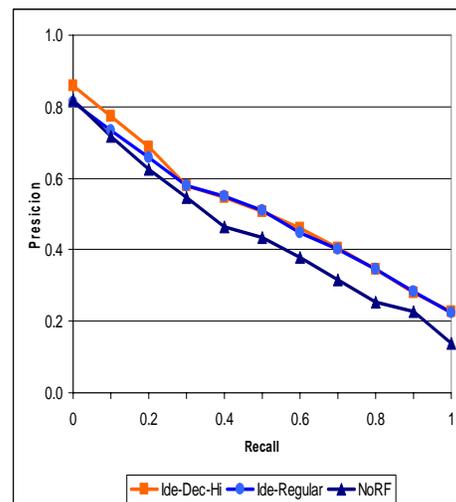
Melalui perbandingan *query-by-query* terlihat bahwa kueri ke-27 yakni “upaya peningkatan pendapatan petani” menunjukkan nilai AVP nol pada NoRF. Ini terjadi karena memang tidak ada satu pun dokumen relevan yang ditemukan. Ketiga *stem* istilah dalam kueri, yaitu “upaya”, “peningkatan” dan “pendapatan” hanya muncul di tiga dokumen di *control group*, semuanya tidak relevan. Di lain pihak, *stem* istilah “petani” muncul di semua dokumen sehingga nilai *idf*-nya nol, akibatnya bobotnya juga nol. Setelah menggunakan RF nilai AVP kueri ini bisa ditingkatkan mencapai di atas 0.3. Penambahan istilah-istilah dari *test group* terlihat sangat membantu kinerja kueri ini.

Keseluruhan hasil yang ada pada penelitian ini menunjukkan bahwa RF memberikan peningkatan kinerja terbesar ketika menggunakan *Ide-Dec-Hi*. Penambahan jumlah dokumen yang diperiksa dari lima menjadi sepuluh ternyata tidak berpengaruh terhadap hasil yang diperoleh. Ini menunjukkan bahwa pemeriksaan lima dokumen sudah optimal untuk koleksi pengujian ini. Iterasi ke-2 tampaknya tidak diperlukan karena kinerja RF telah maksimal pada iterasi ke-1 dengan menemukembalikan nyaris seluruh dokumen relevan untuk semua kueri, sekali lagi ini hanya berlaku untuk koleksi pengujian dalam penelitian ini. Hal ini sangat dipengaruhi oleh ukuran koleksi dan gugus jawaban yang relatif kecil.

Selain menunjukkan kinerja yang sangat baik, secara komputasional *Ide-Dec-Hi* juga lebih efektif daripada *Ide-Regular* karena hanya menggunakan satu dokumen tak-relevan teratas untuk perhitungan, berbeda dengan *Ide-Regular* yang menggunakan seluruh dokumen tak-relevan.

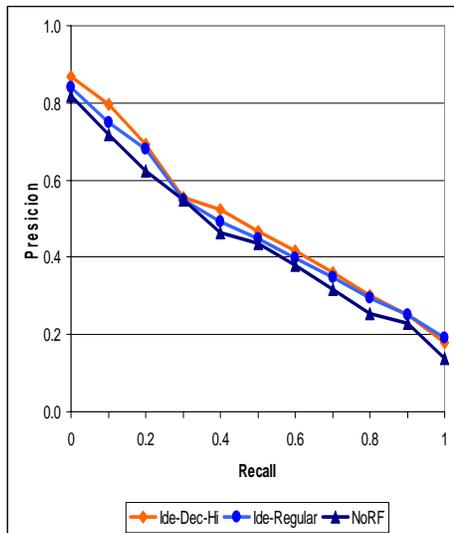


Gambar 1 Kurva *recall-precision* RG5 dan DH5 pada iterasi ke-1.

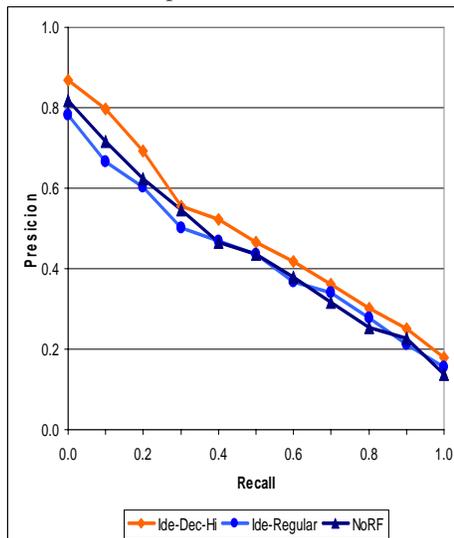


Gambar 2 Kurva *recall-precision* RG10 dan DH10 pada iterasi ke-1.

RF umumnya dipandang sebagai suatu mekanisme untuk meningkatkan *recall*. Akan tetapi melalui penelitian ini, terlihat dari kurva *recall-precision* pada Gambar 1-4, *precision* setelah menggunakan RF ternyata lebih tinggi pada hampir semua tingkat *recall* dibanding sebelum menggunakan RF. Ini mengindikasikan bahwa RF juga bisa menjadi mekanisme untuk memperbaiki *precision*.



Gambar 3 Kurva *recall-precision* RG5 dan DH5 pada iterasi ke-2.



Gambar 4 Kurva *recall-precision* RG10 dan DH10 pada iterasi ke-2.

## 5. KESIMPULAN DAN SARAN

### Kesimpulan

Melalui penelitian ini dapat ditarik kesimpulan sebagai berikut:

1. Menggunakan penilaian relevansi yang diberikan oleh pengguna, RF teruji mampu meningkatkan kinerja sistem temu kembali informasi. Pengguna juga cukup dimudahkan karena hanya perlu mengenali dokumen yang dianggapnya relevan, tanpa perlu mendeskripsikannya.

2. Pada kasus kali ini, peningkatan kinerja yang optimal dicapai melalui penggunaan formula *Ide-Dec-Hi* dan cukup melalui pemeriksaan lima dokumen teratas. Iterasi ke-2 juga tidak perlu dilakukan karena pada iterasi ke-1, RF telah memberikan hasil yang sangat baik dengan tingkat *recall* mendekati 100%.
3. Penggunaan *relevance feedback* tidak terlalu membantu pada kueri yang kinerja awalnya memang sudah tinggi. Sebaliknya, untuk kueri-kueri yang memberikan hasil buruk pada pencarian awal, RF sangat cocok untuk digunakan dan menjanjikan peningkatan kinerja yang cukup tinggi.

### Saran

Untuk pengembangan penelitian disarankan hal-hal sebagai berikut:

1. Penggunaan koleksi pengujian dan gugus jawaban dengan ukuran maupun jumlah lebih besar sehingga bisa memberikan gambaran yang lebih akurat mengenai kinerja RF pada umumnya.
2. Penggabungan formula *Ide-Dec-Hi* dengan teknik pemilihan istilah diperkirakan akan memberikan peningkatan kinerja yang cukup signifikan (Selberg EW 1997).
3. Penggunaan teknik *Automatic Query Expansion*, seperti *Local Analysis* dan *Local Context Analysis*. Hasilnya kemudian bisa diperbandingkan dengan RF yang merupakan teknik *Interactive Query Expansion*.

## REFERENSI

- Adisantoso J, Ridha A.** 2004. *Corpus Dokumen Teks Bahasa Indonesia untuk Pengujian Efektivitas Temu Kembali Informasi*. Laporan Akhir Hibah Penelitian SP4. Departemen Ilmu Komputer FMIPA IPB, Bogor.
- Baeza-Yates R, Ribeiro-Neto B.** 1999. *Modern Information System*. Addison-Wesley.
- Belkin NJ et al.** 2000. *Relevance Feedback versus Local Context Analysis as Term Suggestion Devices*. Rutgers' TREC-8 Interactive Track Experience.
- Buckley C, Salton G, Allan J.** 1994. The Effect of Adding Relevance Information in a Relevance Feedback Environment

*Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval.*

- Chang YK, Cicirillo C, Razon J.** 1971. Evaluation of Feedback Retrieval using Modified Freezing, Residual Collection & Test and Control Groups. *The SMART Retrieval System - Experiments In Automatic Document Processing* 17: 355-370.
- Hiemstra D, Leeuwen D.van.** 2001. *Creating a Dutch Information Retrieval Test Corpus.*
- Ide E.** 1971. New Experiments in Relevance Feedback. *The SMART Retrieval System - Experiments In Automatic Document Processing* 16:337-354.
- Liddy E.** 2001. How a Search Engine Works. *Searcher* 9(5). Information Today, Inc.
- McEnery T, Wilson A.** 2001. *Corpus Linguistic 2<sup>nd</sup> Edition.* Edinburgh University Press.
- Ridha A.** 2002. *Pengindeksan Otomatis dengan Istilah Tunggal untuk Dokumen Berbahasa Indonesia.* Skripsi. Departemen Ilmu Komputer FMIPA IPB, Bogor.
- Ruthven I, Lalmas M.** 2003. A Survey on the Use of Relevance Feedback for Information Access Systems. *Knowledge Engineering Review* 18(2): 95-145.
- Selberg EW.** 1997. *Information Retrieval Advances Using Relevance Feedback.* Department of Computer Science and Engineering University of Washington, Seattle.
- Spink A, Jansen BJ, Ozmultu HC.** 2000. Use of Query Reformulation and Relevance Feedback by Excite Users. *Internet Research: Electronic Networking Applications and Policy* 10(4):317-328.
- Xu J, Croft WB.** 1996. Query Expansion Using Local and Global Document Analysis. *Proceedings of the 19<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.*
- Xu J, Croft WB.** 2000. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Transactions on Information Systems* 18(1):79-112.