

Desain Konseptual Personalisasi Pencarian Pada Perpustakaan *Online*: Kasus Perpustakaan Pusat IPB

Firman Ardiansyah

Program Studi Ilmu Komputer Sekolah Pascasarjana IPB

Kudang B. Seminar

Kepala Perpustakaan Pusat IPB, staf pada Departemen Teknik Pertanian, Fakultas Teknologi Pertanian, Kampus IPB

Badollahi Mustafa

Pustakawan Madya pada Perpustakaan IPB

Abstrak

Current IPB Center Library search engine has been developed to serve all users, independent of the special needs of any individual user. Personalized search is to carry out retrieval for each user incorporating his/her interest. This research propose a novel technique to learn user profile from user lending history to represent long-term interest and user search history to represent short-term interest. The user profiles are then used to improve retrieval effectiveness in online search. A general profile are learned from a category hierarchy. These two profiles are combined to map a user query into a set of categories which represent the user's search intention and serve as context to disambiguate the words in the user's query. Online search is conducted based on both the user query and the set of categories. Several profile learning and category mapping algorithms and a merging algorithm are evaluated. Experimental results indicate that this technique to personalize search is effective.

Kata Kunci: *information filtering, personalization, retrieval effectiveness, search engine*

1. PENDAHULUAN

Pertumbuhan koleksi yang semakin tinggi pada perpustakaan mengakibatkan berkurangnya tingkat kedekatan pengguna dengan perpustakaan tersebut. Pada perpustakaan yang besar, layanan personal semakin sulit dilakukan. Salah satu layanan yang diberikan oleh perpustakaan adalah fasilitas pencarian. Saat ini mesin pencarian (*search engine*) pada perpustakaan umumnya dibangun untuk semua pengguna tanpa mempertimbangkan kebutuhan khusus dari

tiap individu pengguna. Ketika kueri yang sama dikirim oleh pengguna yang berbeda, mesin pencarian umumnya akan mengembalikan hasil yang sama tanpa mempertimbangkan siapa yang mengirim kueri. Sebagai contoh, untuk kueri "*apple*" beberapa pengguna mungkin tertarik pada dokumen yang berhubungan dengan apel sebagai buah, sementara pengguna lain menginginkan dokumen yang berkaitan dengan komputer Apple.

Seorang pengguna sebenarnya dapat mengasosiasikan satu atau lebih

istilah/kategori pada kuerinya secara manual. Namun menurut Liu (2004), pengguna sering tidak sabar untuk mengidentifikasi kategori yang tepat sebelum mengirimkan kueri. Pendapat tersebut dikuatkan dengan hasil survei yang dilakukan Richard (2001) yang menyatakan bahwa panjang kueri pada sistem pencarian OPAC (*Online Public Access Catalogue*) dan Web hanya berkisar pada satu sampai dua kata. Selain itu mesin pencarian yang masih menggunakan operator Boolean akan menghasilkan kegagalan lebih tinggi seiring dengan semakin banyaknya operator yang dipakai (Chia & Garcia, 2002). Oleh karena itu dibutuhkan mesin pencarian yang handal yang dapat mengakomodasi beberapa permasalahan tersebut. Mesin pencarian yang dapat mengembalikan dokumen yang sesuai dengan kebutuhan dan minat pengguna akan dapat meningkatkan nilai layanan yang diberikan.

Tujuan dari penelitian ini adalah menyajikan sebuah strategi dalam membangun mesin pencarian, khususnya untuk perpustakaan, yang dapat memberikan layanan personal berdasarkan pada minat dan kebutuhan penggunanya. Penelitian yang dilakukan dibatasi pada kasus yang berlaku pada Perpustakaan Pusat IPB

2. METODE PENELITIAN

Penelitian ini dilakukan dengan menggunakan data yang dimiliki oleh Perpustakaan Pusat IPB. Data yang digunakan adalah data sirkulasi dari Mei hingga Juli 2005 dan data koleksi buku hingga bulan Juli 2005. Pada percobaan ini, enam set data dikumpulkan dari enam pengguna yang berbeda. Penelitian ini dilakukan selama enam bulan, semenjak Oktober 2005 hingga Maret 2006, bertempat di Laboratorium Temu Kembali Informasi, Departemen Ilmu Komputer, FMIPA, IPB.

Secara singkat tahapan yang dilakukan dalam penelitian ini meliputi empat tahapan utama yaitu:

1. memodelkan dan mendapatkan sejarah tindakan pengguna,

2. membangun profil pengguna berdasarkan sejarah tindakan pengguna dan membangun profil umum berdasarkan pada hierarki kategori UDC
3. mendeduksi kategori yang tepat untuk tiap kueri pengguna berdasarkan pada profil pengguna dan profil umum,
4. meningkatkan keefektifan pencarian dengan menggunakan kategori tersebut sebagai konteks untuk tiap kueri.

Pembahasan lebih rinci dijelaskan pada beberapa subbab berikut.

Memodelkan dan Mendapatkan Sejarah Tindakan Pengguna

Sejarah tindakan pengguna dalam penelitian ini diambil dari dua hal. Pertama adalah sejarah peminjaman pengguna yang nantinya digunakan dalam profil sebagai minat jangka panjang dan bersifat tetap. Kedua adalah sejarah pencarian yang dilakukan pengguna pada suatu sesi tertentu yang akan mewakili minat jangka pendek dan bersifat sementara. Kedua data tersebut akan digunakan untuk membentuk profil pengguna.

Sejarah peminjaman pengguna didapatkan dari tabel sirkulasi dalam basisdata Perpustakaan Pusat IPB. Dalam penelitian ini hanya dipilih enam pengguna dengan intensitas peminjaman tertinggi pada kurun waktu Mei–Juli 2005. Dari sejarah peminjaman pengguna dibentuk sebuah Tabel Minat seperti yang terlihat pada Tabel 1.

Tabel 1. Contoh tabel minat dari seorang pengguna pada transaksi ke-4

<i>Keyword</i>	Pertama	Terakhir	Frekuensi	Dukungan
<i>vaccination</i>	1	1	1	N/A
<i>pathology</i>	2	4	2	0,67
<i>disease</i>	1	4	3	0,75
<i>fish</i>	3	4	2	1,00
<i>breeding</i>	1	3	4	0,75
<i>genetic</i>	2	3	4	0,67

Tabel 1 dibentuk dari empat komponen yaitu kata kunci (*keyword*), transaksi pertama kali munculnya *keyword* (Pertama), transaksi terakhir kali munculnya *keyword* (Terakhir), frekuensi kemunculan, dan nilai

dukungan yang dihitung menggunakan rumus berikut yang dimodifikasi dari Wu *et al* (2001). Variabel *sekarang* adalah identitas transaksi yang sedang berjalan.

$$D = \frac{F}{S - P + 1}$$

D = Dukungan(keyword)

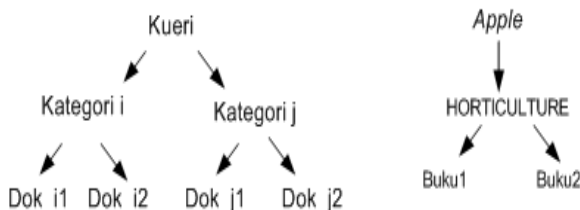
F = Frekuensi (keyword)

S = Sekarang (keyword)

P = Pertama (keyword)

Tabel minat akan selalu berubah seiring dengan bertambahnya transaksi. Dari nilai dukungan yang didapat ditetapkan tiga nilai ambang, yaitu α , γ , dan λ yang dalam penelitian ini ditetapkan sebesar 75%, 2 dan 4. α adalah ambang nilai dukungan apakah sebuah *keyword* akan dimasukkan sebagai profil atau tidak. γ adalah nilai ambang untuk frekuensi. Jika frekuensi masih sedikit dapat dianggap bahwa minat pengguna pada *keyword* tersebut belum terbentuk. Terakhir adalah λ yang merupakan nilai ambang waktu kadaluarsa untuk membatasi interval antara transaksi terakhir dan transaksi saat ini. Selanjutnya diambil informasi dokumen yang mengandung *keyword* terpilih dari daftar peminjaman pengguna tersebut untuk membentuk profil pengguna.

Untuk mendapatkan sejarah pencarian pengguna dipertimbangkan penggunaan informasi kueri, dokumen relevan dan kategori terkait untuk mewakili sejarah pencarian pengguna. Sebuah cantuman pencarian dibangun untuk tiap sesi pencarian pengguna. Model *tree* dari cantuman pencarian ditunjukkan pada Gambar 1.



Gambar 1 Contoh model cantuman pencarian.

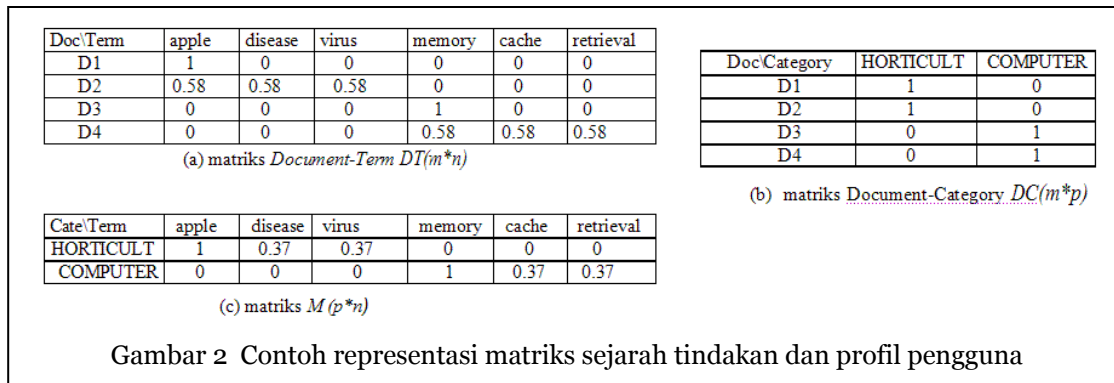
Membangun Profil Pengguna

Profil pengguna digunakan untuk menampilkan minat pengguna dan menduga arah minat pengguna untuk kueri berikutnya. Pada penelitian ini, profil pengguna mengandung seset kategori dan untuk tiap kategori, seset istilah dengan bobotnya. Tiap kategori mewakili minat pengguna dalam kategori tersebut. Bobot sebuah istilah dalam kategori mencerminkan pentingnya istilah tersebut dalam mewakili minat pengguna dalam sebuah kategori. Sebagai contoh, jika istilah “*apple*” memiliki bobot yang tinggi dalam kategori “*horticulture*” maka kejadian kata “*apple*” dalam kueri selanjutnya memiliki tendensi bahwa kategori “*horticulture*” merupakan minat pengguna tersebut. Profil pengguna dipelajari secara otomatis dari sejarah peminjaman dan pencarian pengguna.

Representasi Matriks

Matriks digunakan untuk mewakili sejarah peminjaman, sejarah pencarian dan profil pengguna. Gambar 2 menunjukkan sebuah contoh representasi matriks dari sebuah sejarah tindakan dan profil untuk pengguna tertentu, yang tertarik pada kategori “*horticulture*” dan “*computer*”. Sejarah tindakan pengguna ini ditampilkan dengan matriks *DT* (Gambar 2a) dan *DC* (Gambar 2b). *DT* adalah matriks dokumen-istilah (*document-term*) yang dibangun dari kueri pengguna dan dokumen relevan. *DC* adalah sebuah matriks *document-category* yang dibangun dari keterhubungan antara kategori dengan dokumen. Profil pengguna diwakili oleh matriks *category-term* *M* (Gambar 2c). Dalam contoh ini, *D1, ..., D4* adalah dokumen, kata dengan huruf kecil seperti “*apple*” adalah istilah, sedangkan kata dengan huruf besar seperti “*HORTICULTURE*” adalah kategori. Pembangunan matriks *DT* dan *DC* dilakukan sebagai berikut:

Matriks *DT(m*n)*. *DT* dibangun dari kueri (*root node* dalam model *tree*) dan dokumen relevan (*leaf node*) pada sejarah tindakan pengguna. *m* adalah jumlah dokumen dalam sejarah tindakan pengguna



dan n adalah jumlah istilah *distinct* yang ada dalam dokumen. Tiap kueri atau dokumen relevan merupakan vektor baris dari bobot istilah dalam DT . Jika suatu istilah, misalkan istilah j , muncul pada kueri/dokumen relevan ke- i , Bobot $DT(i,j) > 0$; selainnya bernilai 0. Nilai dari $DT(i,j)$ didapatkan dengan menggunakan skema pembobotan *normalized TF*IDF* (Grossman, 1998). Sebelum membangun DT , daftar *stop word* digunakan untuk menghilangkan kata-kata umum. *Porter stemmer* juga diterapkan dalam tiap istilah.

Matriks $DC(m \times p)$. Untuk tiap baris dalam matriks DT , terdapat baris koresponden dalam matriks DC . Kolom dari DC adalah set kategori terkait. Karena sebuah baris dalam DT mewakili kueri/dokumen, baris koresponden dalam matriks DC mengindikasikan set kategori yang berhubungan dengan kueri/dokumen. Lebih tepat lagi. Jika terdapat *edge* antara kategori ke- j dan kueri/dokumen ke- i dalam model *tree* sejarah peminjaman, maka entri $DC(i,j)=1$. selainnya 0.

Matriks $M(p \times n)$. Dari DT dan DC , dapat dipelajari sebuah matriks M , yang mewakili profil pengguna. Tiap baris dalam matriks M , yang mewakili sebuah kategori minat pengguna, merupakan sebuah vektor istilah terboboti. Oleh karena itu, kategori dan dokumen ditampilkan pada ruang vektor istilah dan kemiripan antara keduanya dapat dihitung. Metode pembelajaran untuk mendapatkan M adalah menggunakan algoritma bRocchio, LLSF, pLLSF dan kNN

(Frakes & Yates, 1992). Dalam penelitian ini keempat metode tersebut dibandingkan untuk mendapatkan algoritma terbaik.

Membangun Profil Umum

Sebagai tambahan terhadap matriks DT , DC dan M yang dijelaskan sebelumnya, dibentuk pula profil umum yang berlaku untuk semua pengguna. Alasan untuk menggunakan informasi tambahan ini adalah bahwa pengetahuan yang diambil dari pengguna sering terbatas dan tidak cukup untuk menjelaskan minat pengguna ketika kueri pengguna baru dipertimbangkan. Pengetahuan umum yang digunakan oleh sistem disarikan dari UDC (*Universal Decimal Classification*).

Penelitian ini hanya menggunakan tiga level pertama UDC. Kategori pada dua level pertama digunakan untuk menampilkan set dari semua kategori. Istilah yang muncul dalam level ketiga kategori tersebut digunakan untuk mewakili kategori-kategori pada dua level pertama. Dari hierarki kategori, digunakan proses yang serupa yang digunakan untuk mempelajari profil pengguna untuk membentuk tiga matriks koresponden yang berhubungan dengan pengetahuan umum, dinyatakan dengan DTg , DCg dan Mg (profil umum).

Memetakan Kueri pada Kategori Terkait

Pada penelitian ini, langkah pertama untuk melakukan personalisasi pencarian adalah memetakan kueri pengguna terhadap

seset kategori yang mencerminkan minat pengguna dan menjadikannya sebagai konteks kueri berdasarkan pada profil pengguna dan profil umum.

Kemiripan (*similarity*) antara vektor kueri q dengan tiap vektor kategori c dalam profil pengguna M dihitung menggunakan fungsi kosinus (Salton & McGill, 1983). Terdapat tiga proses pemetaan yang akan diuji dalam penelitian ini yaitu pemetaan hanya menggunakan profil pengguna, pemetaan hanya menggunakan profil umum dan pemetaan menggunakan kedua profil.

Pada pemetaan yang menggunakan kedua profil, dilakukan tiga metode kombinasi untuk dibandingkan dengan kedua kasus dasar. Misalkan c^u dan c^g adalah vektor kategori untuk profil pengguna dan profil umum. Komputasi kemiripan berikut ini dilakukan untuk tiap kategori:

1. hanya menggunakan profil pengguna : $Sim(q,c)=Sim(q,c^u)$
2. hanya menggunakan profil umum : $Sim(q,c)=Sim(q,c^g)$
3. metode kombinasi1: $Sim(q,c)=(Sim(q,c^u)+Sim(q,c^g))/2$
4. metode kombinasi2: $Sim(q,c)=1-(1-Sim(q,c^u))*(1-Sim(q,c^g))$
5. metode kombinasi3: $Sim(q,c)=\max(Sim(q,c^u),Sim(q,c^g))$

Meningkatkan Keefektian Temu Kembali Menggunakan Kategori

Tujuan dari penelitian ini adalah meningkatkan efektifitas temu kembali. Untuk mencapainya, penelitian ini melakukan beberapa medel temu kembali.

1. Kueri pengguna dikirim pada mesin pencarian tanpa menspesifikasikan

kategori apapun. Mode ini bukan personalisasi pencarian dan akan digunakan sebagai mode *Baseline* dalam penelitian ini.

2. Sebagaimana terlihat dalam Liu (2001), pengguna umumnya menemukan kategori yang diinginkan dalam tiga kategori pertama yang ditampilkan oleh sistem sehingga dibentuk mode otomatis yang secara otomatis mengambil kategori teratas (Otomatis1) atau dua kategori teratas (Otomatis2) atau tiga kategori teratas (Otomatis3) tanpa melibatkan pengguna.

Dalam model terakhir, kueri pertama pengguna akan mengambil kategori dari profil pengguna yang dibentuk dari sejarah peminjaman. Kemudian untuk selanjutnya kueri dikirim dengan menspesifikasikan kategori yang terpilih sebagai konteks. Daftar dokumen yang dikembalikan digabung menggunakan algoritma *weighting voting-based merging* (Liu, 2001).

Pengukuran kinerja dilakukan dengan menghitung tingkat akurasi pemetaan queri pengguna terhadap kategori dan mengukur keefektifan temu kembali. Nilai keakuratan dihitung dengan menggunakan pengukuran kinerja yang diusulkan Liu (2001), sedangkan pengukuran keefektifan temu kembali, dilakukan dengan menggunakan evaluasi TREC (Vorhees, 2001) yang disebut sebagai "*Precision at 11 standard recall levels*"

3. HASIL DAN PEMBAHASAN

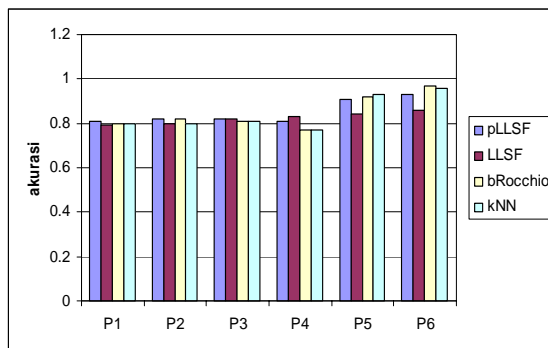
Hasil Pemetaan Kueri Pengguna pada Kategori

Pertama, penelitian ini menyelidiki keefektifan empat algoritma pembelajaran yang berbasis hanya pada profil pengguna. Gambar 3 dan Tabel 2 menunjukkan hasil keakuratannya. Seperti yang dapat dilihat pada Gambar 3, pLLSF, kNN, dan bRocchio memiliki keefektifan yang serupa dan semua menyajikan hasil yang baik. Keakuratan keempat algoritma tersebut berkisar dari 0,72 hingga 0,97. Ketiga algoritma tersebut

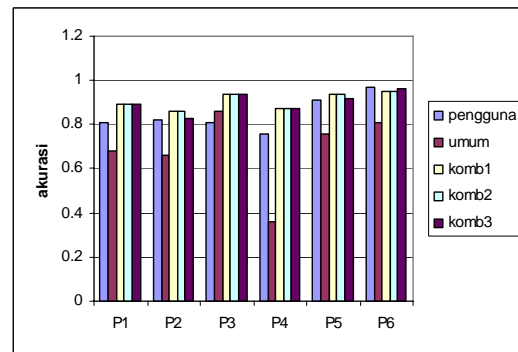
Gambar 4 dan Tabel 3 menunjukkan bahwa tiga metode kombinasi memiliki rata-rata kinerja yang hampir sama dan semuanya mengalahkan dua metode dasar. Hal ini menjelaskan bahwa lebih baik mengkombinasikan profil pengguna dan profil umum untuk mendapatkan keakuratan yang lebih tinggi dibandingkan dengan hanya menggunakan salah satu dari dua profil tersebut.

Tabel 2. Perbandingan rata-rata empat algoritma

Algoritma	pLLSF	LLSF	bRocchio	kNN
Rataan	0,850	0,823	0,848	0,845



Gambar 3 Perbandingan empat algoritma.



Gambar 4 Perbandingan metode pemetaan.

mengalahkan LLSF seperti yang terlihat pada Tabel 2.

Penelitian ini juga mengamati efek dari mengkombinasikan profil pengguna dengan profil umum dan membandingkan tiga metode kombinasi dengan dua metode dasar.

Hal lain yang didapatkan dari Tabel 3 adalah bahwa dengan hanya menggunakan profil pengguna saja akan memberikan kinerja yang lebih baik dibandingkan dengan hanya menggunakan profil umum. Hal ini menunjukkan bahwa personalisasi pencarian

Tabel 3. Perbandingan rata-rata metode pemetaan

Metode	pengguna	umum	komb1	komb2	komb3
Rataan	0,847	0,688	0,908	0,908	0,902

Karena pLLSF, bRocchio dan kNN telah ditunjukkan memiliki keakuratan yang serupa, maka dalam penelitian ini bRocchio dipilih untuk membentuk profil pengguna dan pLLSF untuk membentuk profil umum. Alasan lain memilih bRocchio adalah bahwa metode tersebut dapat dibuat adaptif.

dapat memperbaiki kinerja pencarian jika dilakukan.

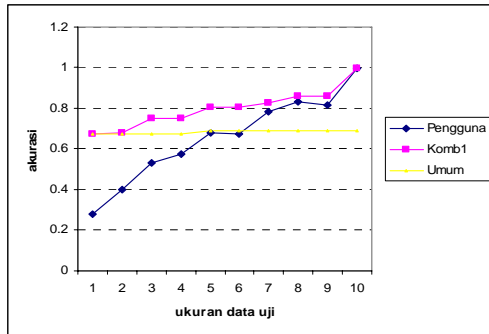
Akhirnya, penelitian ini menguji keakuratan metode pembelajaran adaptif aRocchio dengan semakin banyaknya data yang diberikan (misalkan ukuran sejarah pencarian menjadi lebih besar). Hanya

metode kombinasi pertama yang digunakan dengan alasan tidak ada perbedaan yang signifikan di antara tiga metode kombinasi.

Dengan semakin banyaknya subset uji yang diberikan, keakuratan menggunakan hanya profil pengguna, hanya menggunakan profil umum dan menggunakan keduanya dapat diuji. Berikut ini disampaikan beberapa

baik daripada hanya menggunakan salah satu profil saja.

Dengan semakin banyak data uji yang diberikan, keakuratan menggunakan profil pengguna meningkat. Hal ini juga meningkatkan keakuratan penggunaan kedua profil. Ketika semua data digunakan sebagai data uji, hampir 100% keakuratan dapat



Gambar 5 Hasil aRocchio pada Pengguna3.

Tabel 4 Hasil metode pLLSF dengan hanya menggunakan profil umum

	P1	P2	P3	P4	P5	P6
Rataan	0,691	0,663	0,762	0,813	0,855	0,561

Tabel 5 Hasil aRocchio dengan hanya menggunakan profil pengguna

Ukuran	P1	P2	P3	P4	P5	P6
1	0,346	0,284	0,398	0,326	0,422	0,433
2	0,424	0,405	0,592	0,419	0,675	0,601
3	0,512	0,531	0,611	0,438	0,774	0,794
4	0,512	0,576	0,609	0,482	0,865	0,831
5	0,731	0,678	0,611	0,513	0,898	0,877
6	0,735	0,673	0,703	0,647	0,913	0,903
7	0,736	0,772	0,682	0,647	0,914	0,983
8	0,787	0,821	0,752	0,679	0,916	0,984
9	0,805	0,816	0,806	0,767	0,917	0,978
10	0,966	1	1	1	1	1

Tabel 6 Hasil aRocchio menggunakan dua profil – Metode komb1

Ukuran	P1	P2	P3	P4	P5	P6
1	0,742	0,673	0,879	0,653	0,807	0,837
2	0,751	0,677	0,896	0,697	0,867	0,839
3	0,773	0,752	0,896	0,672	0,897	0,828
4	0,766	0,753	0,896	0,682	0,913	0,881
5	0,825	0,803	0,897	0,726	0,933	0,913
6	0,831	0,805	0,941	0,803	0,933	0,915
7	0,833	0,826	0,941	0,805	0,937	0,976
8	0,858	0,862	0,942	0,834	0,951	0,952
9	0,891	0,863	0,942	0,879	0,943	0,941
10	0,966	1	1	1	1	1

observasi untuk hasil yang ditunjukkan pada Gambar 5, Tabel 4 hingga Tabel 6.

Ketika data uji berukuran kecil, keakuratan menggunakan profil pengguna saja lebih buruk dibandingkan dengan hanya menggunakan profil umum. Bagaimanapun, meskipun dengan set data uji yang kecil, keakuratan menggunakan kedua profil lebih

dicapai.

Hasil Efektifitas dan Efisiensi Temu Kembali.

Berdasarkan kategori yang didapatkan pada langkah pertama (digunakan hasil dari metode kombinasi (komb1) yang ditunjukkan pada Gambar 4 dan Tabel 3), penelitian ini mengamati perbaikan dalam

efektifitas temu kembali menggunakan algoritma *weighted voting-based merging*. Hasilnya (*Precision at 11 standard recall levels* dan perbaikan dari dua mode personalisasi terhadap *baseline*) untuk enam pengguna diberikan pada Tabel 7 dan Gambar 6. Pengamatan yang dilakukan dari hasilnya adalah sebagai berikut:

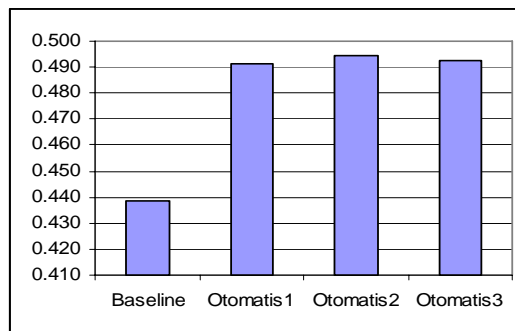
Perbaikan efektifitas temu kembali menggunakan hanya satu dari tiga metode otomatis memiliki hasil yang serupa yang

(sejarah peminjaman dan pencarian) dikumpulkan oleh sistem tanpa adanya keterlibatan pengguna. Selain itu profil pengguna dibuat secara otomatis. Kategori yang dianggap diminati pengguna dideduksi secara otomatis berdasarkan kueri dan dua profil.

Percobaan personalisasi pencarian dilakukan terhadap enam pengguna terhadap data buku Perpustakaan Pusat IPB sampai Juli 2005. Percobaan ini menghasilkan

Tabel 7. *Precision at 11 standard Recall* pada enam pengguna

Pengguna\Mode	Baseline	Otomatis1	Otomatis2	Otomatis3	Akurasi
1	0,556	0,583 (+4,78%)	0,592 (+6,06%)	0,593 (+6,06%)	0,947
2	0,315	0,383 (+21,57%)	0,382 (+21,55%)	0,383 (+21,54%)	0,862
3	0,440	0,476 (+8,22%)	0,483 (+9,27%)	0,472 (+7,23%)	0,891
4	0,402	0,469 (+16,38%)	0,467 (+16,34%)	0,468 (+16,34%)	0,879
5	0,358	0,432 (+20,26%)	0,438 (+22,23%)	0,438 (+22,23%)	0,951
6	0,542	0,607(+12,11%)	0,608 (+12,14%)	0,608 (+12,14%)	0,942
Rataan P +	0,436	0,492 (+12,72%)	0,494 (+13,64%)	0,493 (+13,18%)	



Gambar 6 Rataan *precision at 11 standard recall*

memiliki kisaran 12-13 persen. Karena Otomatis1 hanya perlu mengkombinasikan hasil dari dua daftar dokumen, maka Otomatis1 menjadi lebih efisien dibanding Otomatis2 dan Otomatis3. Oleh karena itu Otomatis1 lebih dipilih. Pada semua kasus tersebut, perbaikan dalam meningkatkan efektifitas temu kembali ditampakkkan ketika personalisasi pencarian digunakan.

4. KESIMPULAN DAN SARAN

Penelitian ini memberikan strategi personalisasi pencarian pada kasus perpustakaan *online*. Kelebihan yang dimiliki adalah bahwa sejarah tindakan pengguna

efektifitas sekitar 13% dibandingkan menggunakan metode *baseline*. Selain itu penelitian ini juga memberikan masukan bahwa metode pembelajaran aRocchio memberikan hasil yang paling baik dalam mempelajari profil pengguna.

Penelitian ini hanya menggunakan enam pengguna, beberapa ratus kueri, dan dokumen relevan yang terbatas. Penelitian selanjutnya dapat menggunakan skala yang lebih luas dengan perubahan pada parameter. Selain itu dapat dilakukan metode pengujian secara semiotomatis yang melibatkan pendapat pengguna tentang akurasi dokumen yang dikembalikan.

5. DAFTAR PUSTAKA

- [1] Chia C, Garcia J. 2002. The Personalization Challenge in Public Libraries: Perspectives and Prospects. Bertelsmann Foundation. Gutersloh.
- [2] Frakes WB, Yates RB, 1992. Information Retrieval: Data Structures and Algorithm. Prentice Hall
- [3] Grossman DA, Frieder O. 1998. Information Retrieval: Algorithm and Heuristics. McGraw Hill
- [4] Liu F, Yu C, Meng W. 2004. Personalized Web Search for Improving Retrieval Effectiveness. IEEE Transaction on Knowledge and Data Engineering vol 16, no.1 January 2004.
- [5] Richard, D. 2001. Personalizing Your Web. CNET Builder.com – Web Programming.
<http://builder.cnet.com/webbuildig/pages/Business/Personal/?tag=st.bl.3881.dir2.7273> [25 Januari 2006]
- [6] Salton G, McGill M. 1983. Introduction to Modern Information Retrieval. McGraw Hill. New York
- [7] Voorhees, EM. 2001. Overview of TREC2001, Proc. Text Retrieval Conf. (TREC-10).
http://trec.nist.gov/pubs/trec10/papers/overview_10.pdf [25 Januari 2006]
- [8] Wu YH, Chen YC, Chen ALP. 2001. Enabling Personalized Recommendation on the web Based on User Interests and Behaviors. 11th International Workshop on Research Issues in Data Engineering. April 2001.