

**TEMU KEMBALI INFORMASI BIBLIOGRAFI
DENGAN BAHASA ALAMI PADA FIELD JUDUL DAN SUFJEK
(Studi Efektivitas Katalog Induk Terpasang Perpustakaan UGM)**

(Bibliographic Information Retrieval with Natural Language on Title and Subject:
The Effectiveness Study at UGM Union Catalog Online)

Janu Saptari, Purwono***

ABSTRACT

This Research aim to know the retrieval effectiveness on Online Union Catalog of UGM Library, index pattern at data bases, knowing which's more effective of searching at entri of title and subjec and also to know the cause of effectiveness difference retrieval at the both entri.

Along with growth and information accretion, the main problem have shifted from way of accessing information become to chosen the relevant information with its requirement. Information retrieval is not possible to be done in the manual system, because very big information corps and non-stoped growing larger. The information retrieval system is very needed to to assist the consumer in finding information. One of the information retrieval system is online union catalog. With the online union catalog, by users easily can look for the book title, pickings of research and other documentation of library where from and any time.

Getting a number of relevant document with the requirement represent the crux in the searching activity, and effectiveness from online union catalog is this represent the key. Effectiveness from a system influenced by a lot of component which each other related/relevant like: quality of input metadata in data bases, index, searching strategy, ability of system application and keyword election. A system told effective if the system can find more amount document/ appropriate information of request by precision / high accuracy.

Research done by testing search at online union catalog to use the natural language keyword. Keyword inclusion done at entri of title and entri subjek. Keyword of taken as sampel come from one of title matakuliah of each; every faculty in UGM. From each; every the topic then translated / formulated into 4 natural language keyword. Data obtained is grouped by relevant level. From the acquirement data; then analysed with the test of nonparametrik Mann Whitney.

From processing data got by conclusion that ratio of retrieval at title entries of equal to 66,6% and subject entries of equal to 58,3%. Data got by mean at title entries of equal to 85,9 document with the detail: very relevant 42,4%, less be relevant 24,2% irrelevant and also 33,5%. While at subject entries found by a mean data of equal to 62,5 document with the detail: very relevant 31,9%, less be relevant 26,45 and irrelevant 30,6%.

Pursuant to this research result is suggested to increase performance from Online Union Catalog of UGM library. Is things required to improved by technique of searching and ability of searching system, the importance of taking care of quality of data input, similarity of metadata and always the existence of renewal / updating of catalogue data from all member library.

Keywords: *effectiveness encounter of retrieval information retrieval system online union catalog.*

* Perpustakaan UGM

** Perpustakaan Fakultas Teknik UGM

1. Latar Belakang

Perpustakaan sebagai salah satu pusat informasi, dalam menyajikan sumber-sumber informasi baik buku maupun dokumen lainnya yang dimiliki menggunakan sarana temu kembali berupa katalog, bibliografi dan indeks. Katalog sebagai salah satu sarana temu kembali bisa berupa katalog tercetak maupun dalam bentuk digital dan berbasis web.

Semakin besar jumlah koleksi yang dimiliki perpustakaan, maka kecepatan dan ketepatan perolehan informasi sangat penting bagi pencari informasi. Suatu sistem otomatis diperlukan untuk membantu pengguna dalam menemukan informasi. Sistem temu kembali informasi (information retrieval system) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunaannya secara otomatis dari suatu koleksi informasi. Salah satu aplikasi dari Sistem Temu Kembali Informasi adalah katalog induk terpasang Perpustakaan UGM. Dengan katalog induk terpasang inilah pengguna dengan mudah dapat mencari/menelusur judul buku, hasil-hasil penelitian dan dokumen perpustakaan lainnya dari mana dan kapan saja.

Pada dasarnya ada dua pendekatan penelusuran yang lazim digunakan dalam sistem temu kembali informasi yaitu bahasa alamiah (natural language), dan kosa kata terkontrol yang sering juga disebut controlled vocabulary (Hasugian: 2003). Kedua pendekatan ini sejak semula telah digunakan secara luas dalam sistem temu kembali informasi. Banyak database yang telah dibangun untuk digunakan sebagai sarana penelusuran eksperimen dalam rangka pembuktian efektifitas dan efisiensi dari kedua pendekatan tersebut.

Dengan pertimbangan itulah maka dalam penelitian ini akan diuji keefektifan temu kembali dengan menggunakan bahasa alamiah pada indeks judul dan indeks subjek. Dengan penelitian ini nantinya diharapkan akan ditemukan cara penelusuran Katalog Induk Terpasang Perpustakaan UGM yang lebih efektif

dan efisien serta menghasilkan temuan dan ketepatan yang tinggi.

2. Rumusan Masalah

Dari uraian diatas, dapat dirumuskan permasalahan yang akan dibahas dalam penelitian ini sebagai berikut:

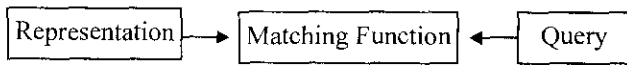
1. Berapa rasio keefektifan nilai temu kembali katalog induk terpasang menggunakan bahasa alami pada entri judul dan subjek ?
2. Bagaimana pola pengindeksan pada katalog induk terpasang Perpustakaan UGM?
3. Manakah yang lebih efektif dalam temu kembali menggunakan bahasa alami pada entri judul atau subjek ?
4. Adakah perbedaan hasil temu kembali dengan bahasa alami pada entri judul dan subjek ?

3. Landasan Teori

3.1. Temu Kembali Informasi

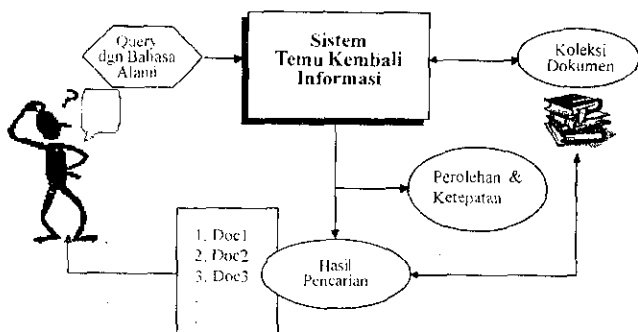
Zainab (2002: 41) menjelaskan bahwa temu kembali sebagai suatu proses pencarian dokumen dengan menggunakan istilah-istilah pencarian untuk mendefinisikan dokumen sesuai dengan subjek yang diinginkan. Sementara itu Salton (1983:1) menjelaskan bahwa secara sederhana temu kembali informasi merupakan suatu sistem yang menyimpan informasi dan menemukan kembali informasi tersebut. Secara konsep bahwa ada beberapa dokumen atau kumpulan *record* yang berisi informasi yang diorganisasikan ke dalam sebuah media penyimpanan untuk tujuan mempermudah ditemukan kembali. Dokumen yang tersimpan tersebut dapat berupa kumpulan *record* informasi bibliografi maupun data lainnya.

Ingwersen (2002: 49) secara sederhana memberikan ilustrasi model temu kembali informasi seperti gambar berikut.



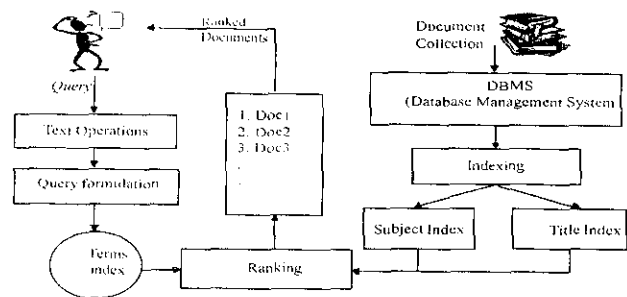
“Representation” dari gambar di sebelah kiri menunjukkan representasi dokumen, data dan informasi. “Query” pada komponen sebelah kanan merupakan representasi dari pertanyaan pengguna, serta “matching function” komponen yang di tengah merupakan fungsi pencocokan antara representasi data/dokumen dengan pertanyaan.

Kemudian dalam “Temu Balik Informasi”, ilustrasi dari sistem temu kembali informasi dapat digambarkan seperti di bawah ini.



Gambar 2.1. Ilustrasi Sistem Temu Kembali Informasi

Selanjutnya dalam “Sistem Temu Balik Informasi”, sebagai suatu sistem, sistem temu kembali informasi memiliki beberapa bagian yang membangun sistem secara keseluruhan. Gambaran bagian-bagian yang terdapat pada suatu sistem temu kembali informasi digambarkan pada gambar 2.2. berikut ini.



Gambar 2.2. Bagian-bagian Sistem Temu Kembali Informasi

Sementara itu Houghton (1977:19) menjelaskan bahwa pada prinsipnya dalam temu kembali informasi adalah penelusuran yang merupakan interaksi antara pemakai dan sistem

dan pernyataan kebutuhan pengguna diekspresikan sebagai suatu istilah tertentu. Selanjutnya dinyatakan bahwa komponen fundamental dari sistem temu kembali informasi adalah penyimpanan (storage), dan proses temu kembali (retrieval). Penyimpanan (storage) menyangkut analisis subjek oleh pengindeks dan penerjemahan dari istilah ke dalam bahasa pengindeksan oleh sistem. Proses temu kembali (retrieval) berkaitan dengan analisis dan pernyataan penelusuran; penerjemahan pertanyaan ke dalam bahasa pengindeksan oleh sistem; serta formulasi dari strategi penelusuran

Sedangkan menurut Lancaster yang dikutip Salton (2002:3) Sistem Temu Kembali Informasi terdiri dari 6 (enam) subsistem, yaitu:

1. Subsistem dokumen
2. Subsistem pengindeksan
3. Subsistem kosa kata
4. Subsistem pencarian
5. Subsistem antarmuka pengguna-sistem
6. Subsistem penyesuaian.

Sistem Temu Kembali Informasi didesain untuk menemukan dokumen atau informasi yang diperlukan oleh masyarakat pengguna. Selanjutnya Salton (1983:2) menjelaskan bahwa Sistem Temu Kembali Informasi bertujuan untuk menjembatani kebutuhan informasi pengguna dengan sumber informasi yang tersedia dalam situasi seperti dikemukakan sebagai berikut:

1. Penulis mempresentasikan sekumpulan ide dalam sebuah dokumen menggunakan sekumpulan konsep.
2. Terdapat beberapa pengguna yang memerlukan ide yang dikemukakan oleh penulis tersebut, tapi mereka tidak dapat mengidentifikasi dan menemukannya dengan baik.
3. Sistem temu kembali informasi bertujuan untuk mempertemukan ide yang dikemukakan oleh penulis dalam dokumen dengan kebutuhan informasi pengguna yang dinyatakan dalam bentuk key word query/istilah penelusuran.

Selanjutnya Salton (1983: 3) juga mengemukakan fungsi utama Sistem Temu Kembali Informasi adalah sebagai berikut:

1. Mengidentifikasi sumber informasi yang relevan dengan minat masyarakat pengguna yang ditargetkan.
2. Menganalisis isi sumber informasi (dokumen)
3. Merepresentasikan isi sumber informasi dengan cara tertentu yang memungkinkan untuk dipertemukan dengan pertanyaan pengguna.
4. Merepresentasikan pertanyaan (query) pengguna dengan cara tertentu yang memungkinkan untuk dipertemukan sumber informasi yang terdapat dalam basis data.
5. Mempertemukan pernyataan pencarian dengan data yang tersimpan dalam basis data.
6. Menemu-kembalikan informasi yang relevan.
7. Menyempurnakan unjuk kerja sistem berdasarkan umpan balik yang diberikan oleh pengguna.

3.2. Temu Kembali Informasi Bibliografi Terpasang

Pada dasarnya sistem ini merupakan sarana menemukan kembali informasi bibliografi/katalog, yang meliputi data perpustakaan yang ada di dalam daerah deskripsi bibliografi. Dengan sistem ini semua data yang berupa kosa kata/istilah yang ada dalam setiap entri katalog/field dapat ditemukan secara cepat dan akurat. Teknik pencarian yang dilakukan dengan logika Boolean. Pada awalnya sistem ini bersifat desktop, artinya hanya terpasang pada tiap-tiap komputer. Namun dalam perkembangan selanjutnya akses sistem ini bisa melalui jaringan internet atau web based, dengan demikian maka akses temu kembali informasi bibliografi terpasang dapat dilakukan dimana saja, kapan saja dan oleh siapa saja.

Sehubungan dengan kemudahan yang

didapatkan bila menggunakan sistem temu kembali bibliografi terpasang, Houghton (1977:430) menyatakan sebagai berikut:

1. Waktu yang diperlukan sedikit
2. Hasil penelusuran yang ditampilkan dalam bentuk baku/tetap
3. Masalah lokasi geografi/tempat dapat teratasi
4. Memungkinkan akses lebih banyak user/pengguna
5. Penelusuran bibliografi terpasang memungkinkan mendapatkan sitasi/sitiran yang lebih banyak.

3.3. Bahasa Penelusuran

Bahasa penelusuran dapat berupa kosa kata/bahasa alami maupun kosa kata terkontrol. Bahasa alami berasal dari setiap istilah yang ada pada *field* judul, sementara bahasa terkontrol berasal dari istilah khusus yang berasal dari *field* subjek.

Bahasa Alamiah

Allen (1987:6) menjelaskan bahasa alami sebagai kata yang digunakan dalam temu kembali, akan dipengaruhi oleh faktor-faktor: fonetik, morfologi, sintaksis, semantik, pragmatik dan pengetahuan perkembangan dunia

Kelebihan bahasa alami dalam penelusuran:

1. Bahasa alamiah dapat dengan mudah dimengerti oleh pengguna.
2. Bahasa alamiah memiliki spesifikasi yang tinggi. Spesifikasi istilah ini muncul karena dapat menggunakan seluruh istilah yang terdapat dalam setiap judul dan subjek sebagai query.
3. Bahasa alamiah memiliki kedalaman yang tinggi. Artinya, banyak tema atau subjek baru yang dihasilkan dokumen yang dapat dijadikan sebagai istilah baru dalam penelusuran. Pada prinsipnya bahwa semua kata terkecuali stop word dapat dijadikan sebagai kata kunci dalam penelusuran.
4. Penelusur yang merupakan praktisi dalam

bidangnya, dapat melakukan penelusuran dengan bahasa alamiah dengan lebih efektif.

5. Ketuntasan memungkinkan perolehan yang tinggi.
6. Selalu terbaru/up date, istilah-istilah baru langsung bisa ditelusur
7. Penelusuran dapat dilakukan dengan menggunakan kata dan frasa secara lebih leluasa.
8. Biaya untuk pemasukan/entri data lebih murah
9. Pertukaran data antar pangkalan data mudah dilakukan, tidak terhambat oleh perbedaan bahasa indeks.

Akan tetapi bahasa alamiah juga memiliki kekurangan sebagai berikut:

1. Bahasa alamiah kurang ringkas. Query yang digunakan penelusur sering berupa kata atau istilah tidak standar sehingga sering terjadi kehilangan informasi saat penelusuran.
2. Mempunyai ambiguitas yang tinggi. Ambiguitas adalah kata atau istilah yang dapat memiliki lebih dari satu arti sehingga mengakibatkan kerancuan. Ambiguitas dapat terjadi karena sinonim atau homograf.
3. Kesulitan komputer untuk menginterpretasikan teks. Kelemahan ini terjadi karena ketidakmampuan sistem menyerap atau menangkap makna dari suatu pernyataan.
4. Beban intelektual ada pada penelusur, yaitu masalah istilah dengan banyak sinonim dan beberapa *species*/variasi istilah.
5. Masalah sintaksis, terjadinya *false drops* karena penggabungan istilah yang tidak tepat. Contoh kata *library school* dengan *school library* akan menghasilkan temuan yang berbeda.

Righ yang dikutip Azhari dan Kasiyadi (1994:3) mengemukakan kelemahan sistem komputer dalam menginterpretasikan masukan, arena faktor-faktor: struktur kalimat, leksikon, morfologi, aturan percakapan dan pelaku. elanjutnya Azhari dan Kasiyadi (1994:5)

menambahkan bahwa agar sistem dapat menginterpretasikan bahasa alami dibutuhkan pengetahuan yang luas seperti: analisis morfologi, sintaksis, semantik, pragmatik dan keterkaitan wacana.

Menurut Aitchison (1997:6) dan Hasugian (2003:6) masing-masing kosa kata tersebut mempunyai kelebihan dan kekurangan. Dengan mempertimbang- kan kelebihan dan kekurangan penggunaan bahasa alamiah dan bahasa terkontrol dalam penelusuran tersebut di atas, Harter dalam Zaenab (1998:76) merumuskan penelusuran dengan bahasa alami dilakukan bila :

1. Pencari informasi memerlukan spesifikasi dan ekspresi dalam merepresentasikan konsep pencarian.
2. Istilah yang diperlukan untuk merepresentasikan konsep tidak terdapat pada bahasa terkontrol
3. Pencari informasi menginginkan pencarian topik yang menyeluruh termasuk hal-hal yang tidak berhubungan langsung dengan topik yang dimaksud.
4. Subjek yang dicari bukan termasuk koleksi inti dari pangkalan data yang dipergunakan.

Bahasa Terkontrol

Keunggulan kosa kata terkontrol adalah sebagai berikut:

1. Proses penelusuran dan temu kembali informasi lebih efisien
2. Mempunyai representasi dokumen yang konsisten. Istilah yang digunakan dalam pengindeksan dokumen pada saat input data adalah istilah yang terkontrol dan standar.
3. Memudahkan penelusuran komprehensif dengan menyatukan istilah terkait secara makna, artinya suatu istilah di dalam indeks subjek tertentu mempunyai hubungan makna dengan indeks yang lain,
4. Memiliki ambiguity (kerancuan) yang relatif kecil.
5. Sangat bagus untuk pangkalan numerik dan

multi bahasa

6. Mengatasi masalah sintaksis dengan istilah majemuk dan sarana lain.
7. Mengurangi beban pada tahap penelusuran seperti: sinonim, homograf dan menyatakan konsep yang sulit dinyatakan dengan bahasa alami.

Selain itu juga ada beberapa kelemahan dari kosa kata terkontrol tersebut sebagai berikut:

1. Kosa kata terkontrol harus selalu diperbaharui. Perkembangan ilmu dan teknologi menyebabkan munculnya berbagai subjek baru yang sekaligus juga berdampak terhadap pemunculan atau penghilangan suatu istilah atau kosa kata.
2. Kosa kata terkontrol sering dihadapkan kepada ketidak cocokan istilah diantara satu basis data dengan basis data yang lainnya pada bidang ilmu yang sama.
3. Kurangnya spesifikasi dalam kosa kata. Berbeda dengan bahasa alamiah, dimana penelusur dapat menggunakan secara bebas kosa kata yang spesifik. Akan tetapi pada kosa kata terkontrol, spesifikasi istilah ditentukan oleh ketersediaannya pada indeks subjek atau tesaurus.
4. Kosa kata terkontrol memiliki struktur yang tidak lengkap. Artinya rincian subjek adalah sangat terbatas untuk pencarian atau penelusuran komprehensif.
5. Kosa kata terkontrol memerlukan biaya dan upaya yang besar pada waktu input data ke sistem.
6. Kurang tuntas, yang biasanya dibatasi pada konsep yang penting saja
7. Penelusur harus menguasai bahasa indeks
8. Biaya proses pemasukan yang tinggi dan lama.

3.4. Efektivitas Sistem Temu Kembali Informasi

Lancaster (1980:140) menyatakan bahwa efektifitas dari suatu sistem temu kembali informasi adalah kemampuan dari sistem itu

untuk memanggil berbagai dokumen dari suatu basis data sesuai dengan permintaan pengguna. Ada dua hal penting yang biasanya digunakan dalam mengukur kemampuan suatu sistem temu kembali informasi yaitu rasio atau perbandingan dari perolehan (*recall*) dan ketepatan (*precision*).

Tabel .1. Perolehan dokumen dalam penelusuran

	Relevan	Not Relevan	Total
Retrieve	a	b	a + b total retrieve
Not Retrieve	c	d	c + d total information retrieve
Total	a + c total relevan	b + d total not retrieve	a + b + c + d total collection

Perolehan (*recall*) berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan.

$$\text{Recall (R)} = \frac{\text{Jumlah dokumen relevan yang diambil (a)}}{\text{Jumlah dokumen relevan yang ada dalam database (a + b)}}$$

Ketepatan (*precision*) berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan. Rasio dari tingkat perolehan (*recall*) dan ketepatan (*precision*) yang dicapai dalam kegiatan penelusuran dapat diungkapkan sebagai berikut :

$$\text{Precision (P)} = \frac{\text{Jumlah dokumen relevan yang diambil}}{\text{Jumlah dokumen yang diambil dalam pencarian}}$$

Kondisi ideal dari keefektifan suatu sistem temu kembali informasi adalah apabila rasio perolehan dan ketepatan sama besarnya (1:1). Boyce (1994:197) mengemukakan beberapa pengukuran efektivitas temu kembali informasi yang ditawarkan oleh para pakar informasi seperti Meadow, Vickery, Heine, Van Rijsbergen, Shaw dan Goffman-Newill. Akan tetapi karena rasio dari *recall* sebenarnya sulit diukur karena jumlah seluruh dokumen yang relevan dalam database sangat besar. Oleh karena itu presisi-lah (*precision*) yang menjadi salah satu

ukuran yang digunakan untuk menilai keefektifan suatu sistem temu kembali informasi. (Rowley dalam Hasugian, 2003:5).

3.5. Katalog Terpasang dan Katalog Induk Terpasang

Barbara (2001) menyatakan bahwa katalog terpasang (online catalog) merupakan katalog perpustakaan yang memuat informasi data bibliografi berbasis komputer, dimana data disimpan pada suatu web server, sehingga data tersebut bisa diakses langsung secara terpasang dari komputer terminal (workstation) baik lokal maupun global.

Sedangkan Katalog Induk terpasang merupakan katalog gabungan dari beberapa Perpustakaan unit/anggota, berbasis komputer dan bisa diakses secara langsung terpasang dari komputer terminal (workstation) baik akses lokal maupun akses global (internet). Contoh katalog induk terpasang seperti Katalog Induk Terpasang Perpustakaan UGM dengan alamat: <http://www.lib.ugm.ac.id>.

4. Metode Penelitian

4.1. Tipe Penelitian

Penelitian ini adalah kuantitatif. Bryman dalam Pendit (2003:195) mengemukakan bahwa penelitian kuantitatif sebagai penelitian yang terutama mengandung upaya mengumpulkan data numerik dan menggunakan logika deduktif dalam pengembangan dan pengujian teorinya.

4.2. Subjek dan Objek Penelitian

Subjek penelitian di sini yang dimaksudkan adalah Katalog Induk Terpasang Perpustakaan UGM pada situs resmi Perpustakaan UGM dengan alamat: <http://lib.ugm.ac.id/index.html>. Sedangkan objek penelitiannya adalah pilihan penelusuran melalui indeks judul dan indeks subjek.

4.3. Populasi dan Sampel.

Populasi dalam penelitian ini adalah database katalog perpustakaan-perpustakaan di lingkungan UGM yang tergabung dalam katalog induk yang terdiri dari 29 perpustakaan. 29 perpustakaan tersebut meliputi Perpustakaan Pusat, Perpustakaan fakultas, jurusan, program studi, pusat studi dan lembaga penelitian. Sementara sample yang dimaksud dalam penelitian ini adalah istilah pencarian yang digunakan untuk menelusur. Sampel diambil dari salah satu mata kuliah dari setiap fakultas yang ada di UGM. Dari tiap sample akan dijabarkan ke dalam 4 bahasa alamiah sebagai kata kunci spesifik yang terdiri dari 2 berbahasa Indonesia dan 2 berbahasa Inggris.

4.4. Metode Pengumpulan Data

Untuk mendapatkan data, maka dilakukan penelusuran pada web site katalog induk perpustakaan. Setiap istilah pencarian/kata kunci yang digunakan dalam proses penelusuran hasilnya berupa sejumlah data katalog. Dari data katalog yang diperoleh kemudian dianalisa dan dicatat data mana yang sangat relevan, kurang relevan dan tidak relevan dengan kebutuhan informasi (topik) tersebut. Masing-masing data dicatat dalam sebuah kolom dalam tabel perolehan dari kegiatan penelusuran katalog induk terpasang

4.5. Instrumen Penelitian

Instrumen penelitian diperlukan sebagai alat untuk memperoleh data. Data yang didapatkan berasal dari pangkalan data bibliografi/katalog dari perpustakaan-perpustakaan yang tergabung dalam katalog induk terpasang. Sehingga untuk bisa mendapatkan data tersebut maka diperlukan kegiatan penelusuran melalui web site katalog induk tersebut dengan alamat <http://lib.ugm.ac.id/index.html>. Untuk itu agar bisa melakukan penelusuran secara terpasang (online) diperlukan fasilitas komputer yang bisa terhubung ke jaringan internet.

4.6. Teknik Pengolahan Data

Dokumen yang ditemukan dikelompokkan menjadi tiga kriteria, yaitu dokumen sangat relevan, dokumen kurang relevan dan dokumen sama sekali tidak relevan. Dalam perhitungannya, jumlah dokumen yang sangat relevan dapat digabungkan dengan yang kurang relevan.

$$P = \frac{\text{Jumlah dokumen sangat relevan} + \text{kurang relevan}}{\text{Jumlah total dokumen yang ditemukan}} \times 100\%$$

Uji statistik yang digunakan yaitu menggunakan uji Mann Whitney atau Uji U pada taraf nyata $\alpha = 0,05$, Uji U dirumuskan sebagai berikut:

$$U1 = n1 n2 + n1 (n1 + 1) / 2 - R1$$

$$U2 = n1 n2 + n2 (n2 + 1) / 2 - R2$$

$$U = n1 n2 - U'$$

di mana:

U1 dan U2: nilai hasil uji U pada sampel pertama dan kedua

U' : nilai hasil uji U yang lebih besar dari sampel lainnya

U : nilai hasil Uji U yang sebenarnya

n1 dan n2 : jumlah sampel kelompok 1 dan 2

R1 dan R2 : jumlah peringkat pada kelompok yang ukuran sampelnya n1 dan n2

Kriteria pengambilan keputusannya adalah:

Ho diterima apabila $U \geq U\alpha$

Ho ditolak apabila $U < U\alpha$

5. Hasil Penelitian dan Pembahasan

5.1. Komponen Sistem Katalog Induk Terpasang Perpustakaan UGM

a. Kumpulan Dokumen

Kumpulan dokumen yang dimaksudkan dalam penelitian ini merupakan kumpulan data katalog (katalog induk) dari

perpustakaan-perpustakaan di lingkungan UGM yang tergabung dalam kerja sama pembentukan katalog induk. Data katalog dalam bentuk digital, yang dibuat dengan program aplikasi bernama CDS/ISIS baik versi DOS maupun versi Windows.

b. Pengindeksan

Berdasarkan penelitian yang dilakukan terhadap database server pada Katalog Induk Terpasang Perpustakaan UGM, maka teknik pengindeksan dengan teknik pengindeksan bahasa alamiah, artinya pengindeksan yang didasarkan pada bahasa yang digunakan dalam dokumen, seperti setiap istilah yang terdapat pada judul, subjek, pengarang dan penerbit.

c. Kebutuhan Informasi Pemakai

Kebutuhan informasi yang diperlukan dalam hal ini untuk mendukung usaha penelitian, referensi mata kuliah maupun untuk pengembangan ilmu pengetahuan lainnya. Dalam hal ini, penelitian mengambil sampel topik/pertanyaan berdasarkan salah satu judul mata kuliah dari tiap-tiap fakultas yang terdiri dari 18 fakultas.

d. Strategi Pencarian

Dari 18 topik yang merupakan representasi dari salah satu mata kuliah yang ada pada tiap-tiap fakultas yang ada, kemudian masing-masing diterjemahkan ke dalam bentuk kata kunci. Dari tiap-tiap topik dibuat menjadi 2 kata kunci dengan bahasa Indonesia dan 2 kata kunci berupa bahasa Inggris. Setiap kata kunci diujicobakan/dimasukkan ke dalam query judul dan subjek. Dengan demikian maka setiap topik akan diwakili oleh 4 (empat) kata kunci yang berbeda.

Berdasarkan hasil penelitian yang dilakukan selama proses penelusuran, ternyata jumlah kata akan sangat mempengaruhi hasil penelusuran, baik jumlah dokumen yang ditemukan maupun dokumen yang relevan dan tidak relevan. Semakin sedikit jumlah kata dalam kata kunci, maka semakin besar dokumen yang ditemukan, tetapi tingkat relevansinya rendah. Sebaliknya, semakin banyak kata dalam kata kunci maka

perolehan dokumen semakin sedikit, tetapi tingkat relevansinya semakin besar.

Berdasarkan hasil penelitian yang dilakukan maka ditemukan bahwa teknik penelusuran yang dipakai pada katalog induk ini adalah sebagai berikut:

1. Teknik penelusuran alamiah/bebas (free text searching).

Teknik ini berlaku pada masing-masing field dalam satu istilah penelusuran/kata kunci, pada penelusuran sederhana maupun penelusuran mahir. Ini berarti bahwa setiap penggunaan kata kunci pada setiap field tersebut maka sistem akan memperlakukan proses pencarian ke dalam basis data secara bebas dan alami sesuai dengan kata kunci yang dimasukkan.

2. Teknik penelusuran Boolean "AND"

Teknik Boolean "AND" ini berlaku untuk penggunaan field judul, subjek, pengarang dan penerbit pada penelusuran mahir saja.

e. Kumpulan Dokumen yang Ditemukan

Berdasarkan hasil penelitian ini bahwa perbedaan jumlah dokumen yang ditemukan menunjukkan banyaknya ketersediaan literatur/koleksi tersebut dan banyak pemakai/peminat dengan topik tersebut. Sebagai contoh pertanyaan nomor 5 dengan topik "penginderaan jauh dasar" mendapatkan 337 dokumen untuk field judul dan 288 dokumen untuk field subjek, sementara pertanyaan nomor 4 dengan topik "metode penelitian filsafat" mendapatkan 5 dokumen untuk field judul dan pada field subjek tidak menghasilkan temuan dokumen. Hal ini disebabkan karena topik penginderaan jauh dasar banyak dipakai di Perpustakaan Teknik, Perpustakaan Geografi, Perpustakaan Pertanian, Perpustakaan Kehutanan dan Perpustakaan MIPA. Sementara topik metode penelitian filsafat hanya dipakai di Perpustakaan Filsafat.

f. Penilaian Relevansi

Dokumen relevan artinya dokumen-dokumen yang didapatkan dapat memenuhi kebutuhan akan informasi yang sedang dibutuhkan. Penilaian relevansi mengacu pada pendapat Burgin yaitu sangat relevan, kurang relevan dan tidak relevan. Sementara definisi dan interpretasi dari masing-masing tingkat relevansi dapat dilihat pada tabel berikut:

Tabel 4.1.
Tingkat relevansi, definisi, dan interpretasinya

Kategori relevansi	Definisi	Interpretasi
Sangat relevan	Dokumen merupakan tanggapan langsung dari pertanyaan	Saya kecewa bila sistem gagal menemukan dokumen
Kurang relevan	Topik dari dokumen relevan, tetapi bukan tanggapan langsung dari pertanyaan	Dokumen ditemukan atau tidak, saya tetap merasa senang
Tidak relevan	Dokumen tidak relevan dengan pertanyaan	Saya kecewa bila sistem menemukan dokumen ini

Untuk menentukan dokumen yang ditemukan sangat relevan, kurang relevan atau tidak relevan dilakukan dengan cara menganalisis setiap dokumen berdasarkan nomor kalsifikasi, judul dan subjek yang ada.

a). Pada Field Judul

Berdasarkan dari data yang dihasilkan diperoleh bahwa pada field judul ditemukan jumlah dokumen relevan (sangat relevan dan kurang relevan) yang terbanyak pada pertanyaan nomor 13 dengan topik kimia organik dasar ada 303 dokumen. Sementara itu dokumen yang ditemukan dengan tingkat relevansi paling kecil ditemukan pada pertanyaan nomor 4 dan 15 dengan topik metode penelitian filsafat dan teknologi pengawetan kulit masing-masing ada 4 dan 7 dokumen. Akan tetapi jika dokumen relevan yang ditemukan berdasarkan prosentase maka ditemukan pada pertanyaan nomor 16 dengan topik psikologi perkembangan ada 90,3%. Perbedaan jumlah dokumen yang

relevan dan jumlah keseluruhan dokumen yang ditemukan disebabkan karena cakupan dan ketersediaan sumber data di basis data yang ada.

b). Pada Field Subjek

Berdasarkan dari data yang dihasilkan diperoleh bahwa pada field subjek ditemukan jumlah dokumen relevan yang terbanyak pada pertanyaan nomor 5 dengan topik penginderaan jauh dasar ada 228 dokumen. Sementara itu dokumen yang ditemukan dengan tingkat relevansi paling kecil ditemukan pada pertanyaan nomor 4 dan 15 dengan topik metode penelitian filsafat dan teknologi pengawetan kulit masing-masing tidak ditemukan dokumen. Akan tetapi jika dokumen relevan yang ditemukan

berdasarkan prosentase maka ditemukan pada pertanyaan nomor 1 dengan topik pengantar anatomi tumbuhan ada 85,7%. Perbedaan jumlah dokumen yang relevan dan jumlah keseluruhan dokumen yang ditemukan disebabkan karena cakupan dan ketersediaan sumber data di basis data.

5.2. Pengolahan dan Analisis Data

Setelah dilakukan pengujian berupa penelusuran dengan berbagai kata kunci pada field judul dan field subjek, kemudian data yang diperoleh dikelompokkan berdasarkan tingkatan: Sangat Relevan (SR), Kurang Relevan (KR) dan Tidak Relevan. Data-data ini kemudian dimasukkan ke dalam tabel 4.2 berikut:

Tabel 4.2.
Jumlah dokumen yang ditemukan pada penelusuran field judul dan field subjek

NO	FIELD JUDUL									FIELD SUBJEK								
	JML	SR		KR		TR		P	JML	SR		KR		TR		P		
	DOK	JML	(%)	JML	(%)	JML	(%)	(%)	DOK	JML	(%)	JML	(%)	JML	(%)	(%)		
1	56	33	58.9	13	23.2	10	17.9	82.1	28	12	42.9	12	42.9	4	14.3	85.7		
2	124	75	60.5	15	12.1	31	25.0	72.6	209	106	50.7	70	33.5	33	15.8	84.2		
3	29	8	27.6	4	13.8	17	58.6	41.4	60	6	10.0	18	30.0	36	60.0	40.0		
4	5	4	80.0	0	0.0	1	20.0	80.0	0	0	0.0	0	0.0	0	0.0	0.0		
5	337	81	24.0	194	57.6	65	19.3	81.6	288	94	32.6	134	46.5	60	20.8	79.2		
6	45	18	40.0	9	20.0	10	22.2	60.0	10	2	20.0	2	20.0	6	60.0	40.0		
7	29	9	31.0	7	24.1	13	44.8	55.2	32	14	43.8	8	25.0	10	31.3	68.8		
8	130	19	14.6	58	44.6	56	43.1	59.2	82	15	18.3	32	39.0	35	42.7	57.3		
9	52	26	50.0	7	13.5	29	55.8	63.5	24	8	33.3	3	12.5	13	54.2	45.8		
10	67	23	34.3	17	25.4	27	40.3	59.7	56	21	37.5	15	26.8	20	35.7	64.3		
11	34	13	38.2	11	32.4	10	29.4	70.6	11	4	36.4	5	45.5	2	18.2	81.8		
12	88	29	33.0	21	23.9	38	43.2	56.8	76	17	22.4	31	40.8	28	36.8	63.2		
13	398	203	51.0	100	25.1	95	23.9	76.1	116	43	37.1	33	28.4	40	34.5	65.5		
14	29	15	51.7	7	24.1	7	24.1	75.9	7	3	42.9	1	14.3	3	42.9	57.1		
15	13	4	30.8	3	23.1	6	46.2	53.8	0	0	0.0	0	0.0	0	0.0	0.0		
16	31	17	54.8	10	32.3	4	12.9	87.1	90	45	50.0	21	23.3	24	26.7	73.3		
17	26	12	46.2	5	19.2	9	34.6	65.4	28	11	39.3	5	17.9	12	42.9	57.1		
18	53	19	35.8	10	18.9	24	45.3	54.7	14	8	57.1	4	28.6	2	14.3	85.7		
	85.9	33.8	42.4	27.3	24.1	25.1	33.7	66.4	62.8	22.7	31.9	21.9	26.4	18.2	30.6	58.3		

Keterangan: SR= sangat relevan, KR= kurang relevan, TR= tidak relevan, P= Presisi

Pengujian statistik nonparametrik dengan Uji Mann Whitney pada taraf nyata $\alpha = 0,05$, maka hasil yang didapatkan dari pengujiannya adalah sebagai berikut:

1. Hasil penggabungan dua sampel dan penjenjangan/ranking (tabel 4.3.)
2. Nilai uji U
 $U_1 = 146$ dan $U_2 = 178$
 Karena $146 < 178$, maka ditetapkan $U' = 178$
 Maka nilai uji $U = 146$
3. Di dalam tabel pada taraf nyata $\alpha = 0,05$ pada sampel $n_1 = 18$ dan $n_2 = 18$ didapat angka 109.

Karena $U \geq U'_{\alpha}$ yaitu $146 \geq 109$, maka hipotesis H_0 diterima. Hal ini artinya ada perbedaan tingkat efektivitas temu kembali antara pada field judul dan field subjek. Dari tabel 4.2 terlihat bahwa rata-rata rasio presisi pada field judul sebesar 66,4%, sementara pada field subjek presisi sebesar 58,3%. Dengan demikian terbukti bahwa proses temu kembali pada field judul lebih efektif daripada field subjek pada Katalog Induk Terpasang UGM.

5.3. Keefektifan Sistem Katalog Induk Terpasang Perpustakaan UGM

Dokumen yang ditemukan dalam penelusuran pada Katalog Induk Terpasang UGM merupakan respon dari interaksi penggunaan kata dengan pangkalan data. Pengujian penelusuran yang dilakukan dengan menggunakan topik salah judul mata kuliah dari tiap-tiap fakultas menggunakan pendekatan bahasa alami dengan kosa kata terkontrol pada

Tabel 4.3.
Penggabungan dua sampel dan penjenjangan

No.	Field Judul		Field Subjek	
	Nilai	Jenjang	Subjek	Jenjang
1	41.4	5	0	2
2	53.8	7	0	2
3	54.7	8	40.0	3
4	55.2	9	40.0	3
5	56.8	10	45.8	6
6	59.2	14	57.1	11.5
7	59.7	15	57.1	11.5
8	60.0	16	57.3	13
9	63.5	18	63.2	17
10	65.4	20	64.3	19
11	70.6	23	65.5	21
12	72.6	24	68.8	22
13	75.9	26	73.3	25
14	76.1	27	79.2	28
15	80.0	29	81.8	31
16	81.6	30	84.2	33
17	82.1	32	85.7	34
18	87.1	36	85.7	35
Jumlah	349			317

field judul dan field subjek didapatkan bahwa ketepatan pada field judul 66,6%. Sementara ketepatan/presisi yang didapatkan dari pengujian pada field subjek adalah 58,3%.

Berdasarkan hasil pengujian yang didapatkan di atas ternyata penggunaan bahasa alami pada field judul lebih efektif dari pada field subjek, meskipun perolehan dokumen "tidak relevan" yang didapatkan pada field judul lebih besar (33,5%) dari pada field subjek (30,6%). Hal ini disebabkan karena pada field judul semua kata dipergunakan/dapat ditelusur. Sedangkan pada field subjek hanya kata-kata tertentu yang dikendalikan daftar tajuk subjek saja di pergunakan.

6. Kesimpulan

Berdasarkan hasil pembahasan penelitian, maka penulis memberikan kesimpulan sebagai berikut:

1. Pola teknik pengindeksan di dalam sistem

- basis data server Katalog Induk Terpasang Perpustakaan UGM adalah bahasa pengindeksan alami/bebas (natural indexing languages).
2. Rasio efektivitas nilai temu kembali Katalog Induk Terpasang Perpustakaan UGM dengan topik salah satu judul mata kuliah dari setiap fakultas di UGM pada field judul adalah 66,6%. Sementara rasio efektivitas nilai temu kembali informasi katalog induk Perpustakaan UGM pada field subjek adalah 58,3%.
 3. Penggunaan bahasa alamiah dengan topik salah satu judul mata kuliah dari tiap-tiap fakultas di UGM sebagai buku ajar dan acuan perkuliahan akan lebih efektif jika dilakukan penelusuran dengan menggunakan entri judul.
 4. Perbedaan perolehan data dan nilai ketepatan penelusuran dengan bahasa alamiah pada field judul dan subjek disebabkan antara lain karena interpretasi setiap deskriptor katalog berbeda di dalam merumuskan subjek. Sehingga meskipun pada judul yang sama kemungkinan deskriptor katalog lain akan membuat rumusan subjek yang berbeda.

DAFTAR PUSTAKA

- Aitchison, Jean; Gilchrist Alan and David Bawden. 1997. *Thesaurus Construction and Use: A practical Manual*. London: Aslib.
- Allen, James. 1987. *Natural Language Understanding*. California: Benjamin/Cummings Publishing.
- Azhari dan Kasiyadi. (1994). *Analisis Sintaksis pada Pengolahan Bahasa Alami: Laporan Penelitian*. Yogyakarta: Fakultas MIPA UGM.
- Barbara, Preece and Peters Thomas. 2001. "Union and Virtual Catalog in Co xortial Environment" dalam *Journal of Academic Libraries*, vol. 27 issue 6, 2001
- Beni, Romanus. 1998. "Peran Pustakawan sebagai Intermediary dalam Penelusuran Terpasang (*Online Searching*)". Dalam *Jurnal Ilmu Informasi, Perpustakaan dan Kearsipan*, volume I No. 1, September 1998.
- Boyce, Bert R., Charles T. Meadow dan Donald H. Kraft. 1994. *Measurement in Information Science*. New York: Academic Press.
- Diet. (t.t.). "Natural Language". Dalam <http://diet.die.net/natural%20language>, tanggal 21 Maret 2006, pukul 14.02.
- Dirjen Dikti. 2005. *Perpustakaan Perguruan Tinggi*. Ed. 3. Jakarta: Direktorat Jenderal Pendidikan Tinggi, Departemen Pendidikan Nasional RI.
- Djarwanto. 2001. *Statistik nonparametrik*. Yogyakarta: Badan Penerbit Fakultas Ekonomomi UGM
- Hamakonda, Towa P. dan JNB Tairas. 1988. *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta: Gunung Mulia
- Hardi, Wishnu. "Kajian Koleksi Bidang Linguistik dengan Metode Conspectus di Perpustakaan Fakultas Ilmu Budaya Universitas Indonesia". (Skripsi). Jakarta: Jurusan Ilmu Perpustakaan dan Informasi Fakultas Ilmu Budaya UI.
- Hasugian, Jonner. 2003. "Penggunaan Bahasa Alamiah dan Kosa Kata Terkontrol dalam Sistem Temu Kembali Informasi Berbasis Teks". Dalam *USU Digital Library*. Medan: Perpustakaan Universitas Sumatera Utara
- Houghton, Bernard and John Convey. 1977. *Online Information Retrieval Systems: An Introductory Manual to Principles and Practice*. London: Clive Bringley
- Huang, Jie. 2004. "Retrieval of Chinese Languages Titles in Pinyin: A Comparative Study". Dalam *Information Technology and Libraries*. No. 1 September 2004.
- Jurusan Ilmu Fisika ITB. 2005. "Sistem Temu Balik Informasi". Dalam <http://mahasiswa.if.itb.ac.id/~if11016/STBI/>

- vektor.pdf*, tanggal 14 Nopember 2005, pukul 11.10
- Lasa HS. 1998. *Kamus Istilah Perpustakaan*. Yogyakarta: Kanisius
- Lancaster, F.W. and M.J. Joncich. 1980. *The Measurement and Evaluation of Library Services*. Arlington: Information Resources Press.
- Nawawi, Hadari dan Mimi Martini. 1994. *Penelitian Terapan*. Yogyakarta: Gadjah Mada Univeristy Press.
- Pendit, Putu Laxman. 2003. *Penelitian Ilmu Perpustakaan dan Informasi*. Jakarta: JIP-FSUI.
- Perpustakaan Nasional. 1992. *Format Marc Indonesia (INDOMARC) untuk Buku*. Jakarta: Perpustakaan Nasional RI.
- Perpustakaan UGM. 2005. *Buku Panduan Perpustakaan Univeistas Gadjah Mada*. Yogyakarta: Perpustakaan UGM
- Praptono. 1986. *Buku materi pokok metode statistika nonparametrik*. Jakarta: Universitas Terbuka.
- Reitz, Joan M. (t.t.). "ODLIS Online Dictionary for Library and Information Science". Dalam http://lu.com/odlis/odlis_u.cfm, tanggal 14 Mei 2005, pukul 13.11
- Saleh, Abdul Rahman dkk. *CDS/ISIS Panduan Pengelolaan Sistem Manajemen Basis Data untuk Perpustakaan dan Unit Informasi*. Jakarta: Saraswati Utama.
- Suryabrata, Sumadi. 1992. *Metode Penelitian*. Jakarta: Rajawali.
- Universitas Petra. (t.t.). "Penelusuran Informasi secara Online (*Online Search*)". Dalam <http://incuvl.petra.ac.id/learn/learn-2.htm#tigadua>, tanggal 21 Oktober 2005; pukul 13.47
- Universitas Indonesia. "Metodologi Penelitian" dalam [http://telaga.cs.ui.ac.id/ WebKuliah/](http://telaga.cs.ui.ac.id/WebKuliah/), tanggal 30 okt 2005; pukul 09.40
- Universitas Indonesia. (t. t.). "Konsep Dasar Sistem Temu Kembali Informasi". Dalam <http://telaga.cs.ui.ac.id/WebKuliah/TKSI/MIK/MIK%20Bab%20%20konsep%20IRS.doc>, tanggal: 19 Oktober 2005 pukul 07.59 WIB.
- Van Alles. (t.t.) "Perkembangan Pengindeksan Subjek dan Kosa Kata Indeks". Dalam <http://radio.weblogs.com/0140033/>, tanggal 11 Nopember 2005, pukul 07.50
- Wikipedia. (t.t.). Defining Natural Language. Dalam http://en.wikipedia.org/wiki/Natural_language, tanggal 21 Maret 2006, pukul 14.30
- Yusup , Pawit. 1997. M . "Mencoba Memahami Pengertian Pengindeksan Subjek di dunia Perpustakaan dan Informasi". Dalam http://bdg.centrin.net.id/~pawitmy/index_files/Makalah/, tanggal 20 Ok 2005, pukul 14:45
- Zaenab, Ratu Siti. 2002a. "Efektivitas Temu Kembali Informasi dengan Menggunakan Bahasa Alami pada CD-ROM Agris dan CAB Abstracts". Dalam *Jurnal Perpustakaan Pertanian*, vol. 11, no. 2.
- Zaenab, Ratu Siti. 2002b. "Bahasa Terkendali vs Bahasa Ilmiah dalam Penelusuran Bidang Ilmu Perikanan". Dalam *Sekapur Sirih Pendidikan Perpustakaan di Indonesia 1952-2002*. Jakarata: Alumni & Mahasiswa Program Studi Ilmu Perpustakaan PPS FIB-UI.