

**MENGUKUR KINERJA *SEARCH ENGINE*:
SEBUAH EKSPERIMENTASI PENILAIAN *PRECISION AND RECALL* UNTUK
INFORMASI ILMIAH BIDANG ILMU PERPUSTAKAAN DAN INFORMASI
Oleh: Wishnu Hardi ^{*)}**

PENDAHULUAN

Bagi mereka yang bekerja di sektor informasi, *search engine* merupakan bagian dari aktivitas kerja sehari-hari. Seorang pustakawan rujukan, misalnya, akan memprioritaskan penggunaan *search engine* sebagai sarana penelusuran informasi. Pustakawan yang bergelut dengan penelusuran sumber daya informasi *online*, setidaknya harus memiliki dua keahlian ketika bekerja dengan *search engine*. Pertama, dapat mengartikulasikan strategi penelusuran dari sisi pemilihan istilah, penggabungan konsep, maupun sintaksisnya. Yang kedua, mengetahui pilihan *search engine* yang tepat sesuai dengan kebutuhan informasi yang dicari.

Internet adalah belantara informasi yang luas dan seakan telah menjadi “dunia paralel” dari dunia fisik kita sendiri. Harriet Shalat, seorang pustakawan rujukan di New York Public Library, Amerika Serikat, mengungkapkan, saat ini sudah terbentuk pandangan di publik Amerika yang mengasumsikan bahwa sesuatu yang tidak dapat ditemukan di internet berarti memang tidak eksis.

Ketergantungan yang sangat tinggi terhadap *search engine* sering tidak diimbangi dengan pemahaman bahwa sebenarnya *search engine* tidak dilengkapi filter dalam pengumpulan informasi yang diinginkan oleh pengguna. Kenyataan itu semakin ironis, ketika Google melakukan penelitian pada tahun 2002 dan mendapatkan hasil bahwa hampir 85% pengguna hanya melihat hasil perolehan *search engine* pada halaman pertama saja.

Dalam dunia sistem temu kembali informasi, cara yang paling efektif untuk menguji kehandalan sebuah *search engine* adalah dengan meneliti tingkat *precision* (ketepatan) *and recall* (ketepatan)-nya. Tulisan ini bermaksud mengevaluasi kinerja enam *search engine* yang cukup dikenal, yakni Google, Yahoo, Scirus, Sciseek, Askjeeves, dan America Online (AOL). Evaluasi difokuskan pada efektivitas temu kembali informasi ilmiah untuk bidang Ilmu Perpustakaan dan Informasi dengan cara meneliti *precision and recall* dari hasil pencariannya. *Precision* adalah jumlah kelompok dokumen relevan dari total jumlah dokumen ditemukan oleh sistem. Dalam hal ini, *precision* lebih mendefinisikan tingkat “gangguan” dari informasi yang ditampilkan. Sedangkan, *recall* diartikan sebagai jumlah dokumen relevan yang ditemukan oleh sistem. Dalam tulisan ini, penulis tidak menerjemahkan istilah *precision and recall* ke dalam bahasa Indonesia, dengan alasan bahwa istilah tersebut adalah konsep yang sudah cukup dipahami dan umum dipakai dalam dunia sistem temu kembali informasi.

TINJAUAN LITERATUR

a. *Search engine* dan Konsep Relevansi

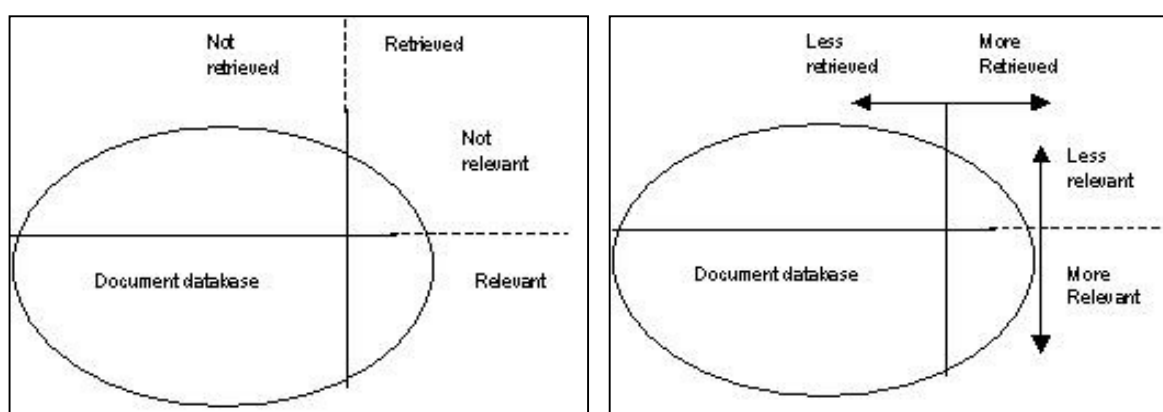
The *American Heritage Dictionary* mendefinisikan *search engine* sebagai sebuah program perangkat lunak (*software*) yang menelusur, menjaring, dan menampilkan informasi dari pangkalan data. Informasi yang ditampilkan mengandung atau berhubungan dengan suatu istilah spesifik. Lancaster mendefinisikan temu kembali informasi sebagai proses pencarian

^{*)} Peneliti pada Veritas Document Management, Jakarta
(wishnu.hardi@gmail.com)

dokumen dengan menggunakan istilah luas untuk mengidentifikasi dokumen yang berhubungan dengan subjek tertentu.

Diskusi mengenai efektivitas kinerja dari sebuah *search engine* selalu dikaitkan dengan tingkat relevansi hasil pencarian. Meskipun demikian, definisi konsep relevansi telah menjadi perdebatan bertahun-tahun dalam Ilmu Informasi. Schamber menyatakan bahwa konsep relevansi adalah sebuah fenomena yang multidimensional dan dinamis. Sementara itu, Saracevic menjelaskan tentang bagaimana seharusnya konsep relevansi itu dimaknai, apakah relevansi yang dimaksud dilihat dari aspek algoritma, topik, kognitif, situasi, atau motivasi.

Konsep penilaian biner pada hasil pencarian yang mendikotomikan antara yang relevan dan tidak relevan; yang ditemukan (retrived) dan tidak ditemukan, seperti yang digagas oleh Salton dan McGill kini mengalami pergeseran. Mizarro mengatakan seiring dengan meningkatnya penggunaan *search engine*, maka penilaian biner tidak lagi mencukupi.



Gambar 1. Perubahan konsep penilaian biner ke penilaian kontinu

Dari gambar di atas dapat dilihat bahwa efektivitas kinerja *search engine* tidak semata-mata dilihat dari relevan dan tidak relevan atau ditemukan dan tidak ditemukannya informasi yang diinginkan. Akan tetapi, semua hasil pencarian *search engine* akan ditampilkan secara kontinu dari hasil yang lebih relevan sampai yang kurang relevan dengan metode pemerinkatan. Dengan demikian, konsep penilaian biner bergeser menjadi penilaian kontinu.

Semakin banyaknya sumber informasi *online* memunculkan berbagai penelitian tentang efektivitas sistem temu kembali informasi. Chu dan Rosental (1996) melakukan penilaian *precision* pada Alta Vista, Excite, dan Lycos. Leighton (1997) mengevaluasi *precision* pada Alta Vista, Excite, HotBot, dan Infoseek dengan menggunakan 15 query dan mengambil 20 hasil temuan di peringkat teratas. Kekäläinen (2002) melakukan evaluasi sistem temu kembali informasi dengan menggunakan metode penilaian relevansi bertingkat (*graded relevant assessment*) yang mengubah penilaian biner ke penilaian kontinu. Mizarro (2004) melakukan eksperimen uji efektivitas sistem temu kembali informasi dengan menggunakan metode *Average Distance Measure (ADM)*. Shafi (2005) melakukan penilaian *precision and recall* pada lima *search engine* untuk bidang bioteknologi.

PRECISION AND RECALL 6 SEARCH ENGINE UNTUK ILMU PERPUSTAKAAN DAN INFORMASI

Penelitian ini dilakukan untuk mengidentifikasi bagaimana kinerja enam *search engine* dalam temu kembali informasi ilmiah untuk bidang Ilmu Perpustakaan dan Informasi. Selain

itu, penelitian ini juga bertujuan memberikan penilaian *precision and recall* pada masing-masing *search engine* tersebut.

Metode evaluasi dilakukan melalui tiga tahapan. Pertama, mengumpulkan literatur-literatur berhubungan baik tercetak maupun elektronik. Kedua, menyeleksi *search engine* dan menentukan *query* yang akan digunakan untuk penelusuran. Ketiga, penelusuran dengan *search engine*. Setelah semua data diperoleh, langkah selanjutnya adalah melakukan analisis.

a. Seleksi Search engine

Seleksi *search engine* dilakukan dengan mengacu kepada rilis yang dibuat oleh Danny Sullivan, *Editor-in-chief*, Search Engine Watch (www.searchenginewatch.com) pada bulan Maret 2006. Kategori *search engine* yang menjadi referensi utama penelitian ini adalah *top choices category* dan *strongly consider category*. Penulis juga menyertakan *search engine* untuk informasi ilmiah yang umum dipakai yakni Scirus dan Sciseek. Dengan demikian, *search engine* yang akan diuji adalah Google (Umum), Yahoo (Umum), Askjeeves (Umum), America Online (AOL) (Umum), Scirus (Sains dan Teknologi), dan Sciseek (Sains dan Teknologi).

b. Istilah pencarian

Penulis menggunakan lima belas istilah bidang Ilmu Perpustakaan dan Informasi yang diambil dari *Library of Congress Subject Headings 2003*. Istilah pencarian yang diperoleh kemudian diklasifikasi kedalam tiga kelompok:

1. Konsep tunggal, dengan istilah pencarian: *cybernetics; librarianship; entropy; telematics; dan bibliometrics*.
2. Konsep gabungan, dengan istilah pencarian: “*information highway*”; “*information resource management*”; “*library automation*”; “*comparative librarianship*”; dan “*semantic networks*”.
3. Konsep kompleks, dengan istilah pencarian: “*information superhighway*” OR “*data superhighway*” AND *research*”; “*information measurement*” AND *research*”; “*public libraries*” AND *copyright*”; “*bibliographical centers*” AND “*library cooperation*” AND *research*; *cataloging* OR “*analytical entry*”.

c. Test environment

Pencarian dokumen yang dilakukan pada enam *search engine* menggunakan *advanced mode*. Metode ini dipilih agar penjangkauan informasi memiliki tingkat *precision* yang lebih tinggi. Untuk konsep tunggal dan kompleks digunakan metode “*match all words*” dan “*exact phrase*”, sedangkan untuk konsep majemuk digunakan pilihan “*exact phrase*”. Hasil temuan dibatasi pada periode tahun 2004-2006 dan dikontrol dengan hanya menggunakan bahasa Inggris.

Dari masing-masing *search engine* tersebut diperoleh hasil pencarian yang berjumlah sangat besar. Penelitian ini membatasi pada 10 hasil temuan yang muncul pada halaman pertama saja. Pengujian dilakukan pada hari yang sama untuk menghindari terjadinya variasi hasil pencarian yang disebabkan oleh *updating* algoritma pengindeksan dari masing-masing *search engine*. Pengujian *search engine* berlangsung pada bulan Maret – April 2006.

d. Kalkulasi Precision and Recall

Rumus penilaian *precision* yang digunakan dalam penelitian ini adalah:

$$\text{Precision} = \frac{\text{Jumlah skor dokumen ilmiah ditemukan oleh search engine}}{\text{Total hasil temuan yang dievaluasi}}$$

Relevansi hasil temuan *search engine* diukur dengan menggunakan metode penilaian relevansi bertingkat (*Graded Relevant Assessment*) 0-3 yang kemudian dijadikan dasar kalkulasi tingkat *precision*. Skala penilaian yang dimaksud adalah sebagai berikut:

- Halaman yang menampilkan dokumen-dokumen makalah penelitian, prosiding seminar/konferensi atau paten mendapat skor 3.
- Halaman yang menampilkan abstrak makalah penelitian, prosiding seminar/konferensi atau patent mendapat skor 2.
- Halaman yang menampilkan buku atau pangkalan data mendapat skor 1.
- Halaman yang menampilkan selain poin-poin di atas (misalnya *website* perusahaan, kamus, ensiklopedi, organisasi, dsb.) mendapat skor 0.
- Halaman yang tidak bisa ditampilkan karena server yang tidak merespons setelah tiga kali penelusuran secara berurutan mendapat skor 0.

Sedangkan nilai *relative recall* dihitung dengan rumus:

$$\text{Relative Recall} = \frac{\text{Total dokumen ilmiah yang ditemukan search engine}}{\text{Jumlah dokumen yang ditemukan oleh enam search engine}}$$

Jika terjadi *overlap* pada hasil temuan *search engine*, maka hanya temuan yang mengalami *overlap* yang dimasukkan ke dalam kalkulasi dari enam *search engine* tersebut (misalnya *search engine* a, b, c, d, e, f, g). Apabila tidak ada *overlap* di antara *search engine* ($a \cap b, a \cap c, a \cap d, a \cap e, a \cap f, \text{ dan } a \cap g$) adalah nol) maka *relative recall* dari sebuah *search engine* dikalkulasikan sebagai $a1/(a1+b1+c1+d1+e1+f1+g1)$.

Nilai tengah *precision* dan *relative recall* diperoleh dari perhitungan rata-rata mikro. Dalam penelitian ini, rata-rata skor untuk tiap *search engine* suatu *query* adalah hasil penjumlahan keseluruhan lima belas *query*, dan nilai tengah diperoleh dari jumlah masing-masing konsep tunggal, gabungan, dan kompleks.

e. Hasil dan Pembahasan

Dari hasil pengujian terhadap enam *search engine*, diperoleh hasil nilai tengah untuk *precision* dan *relative recall* adalah sebagai berikut:

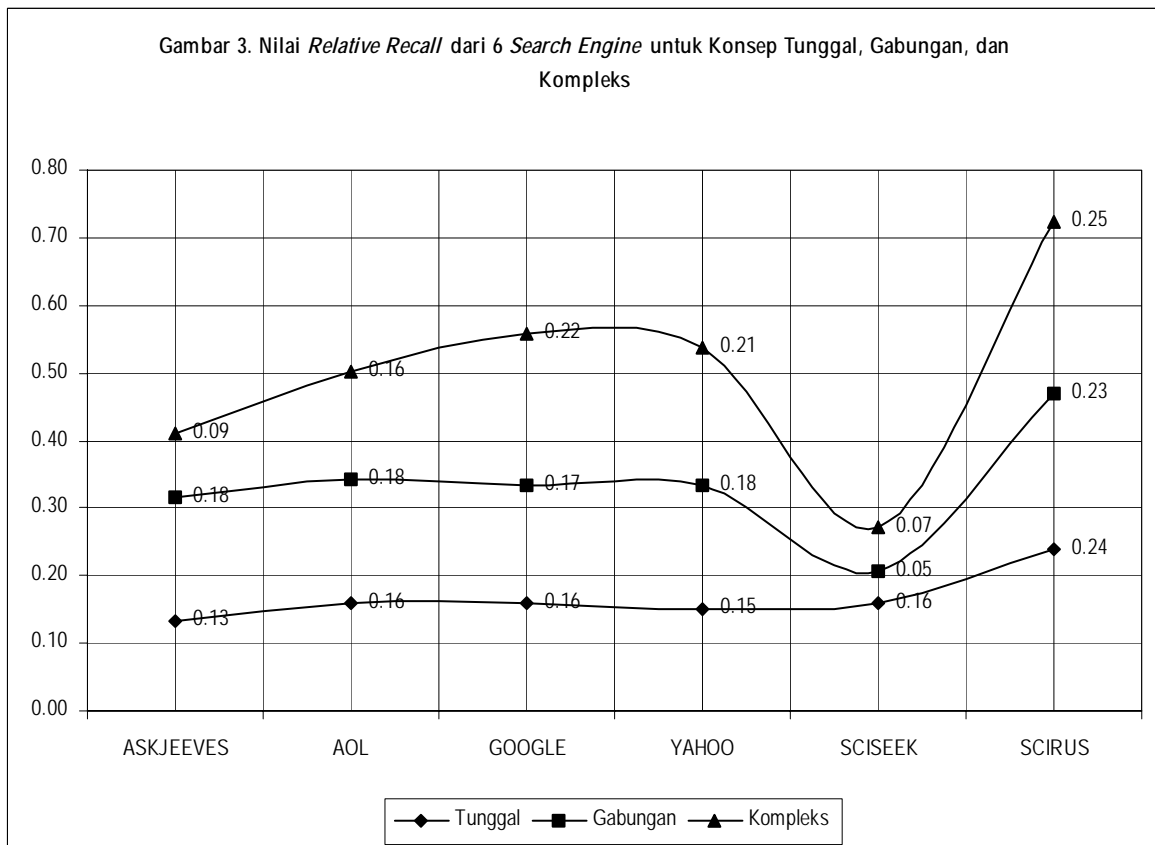
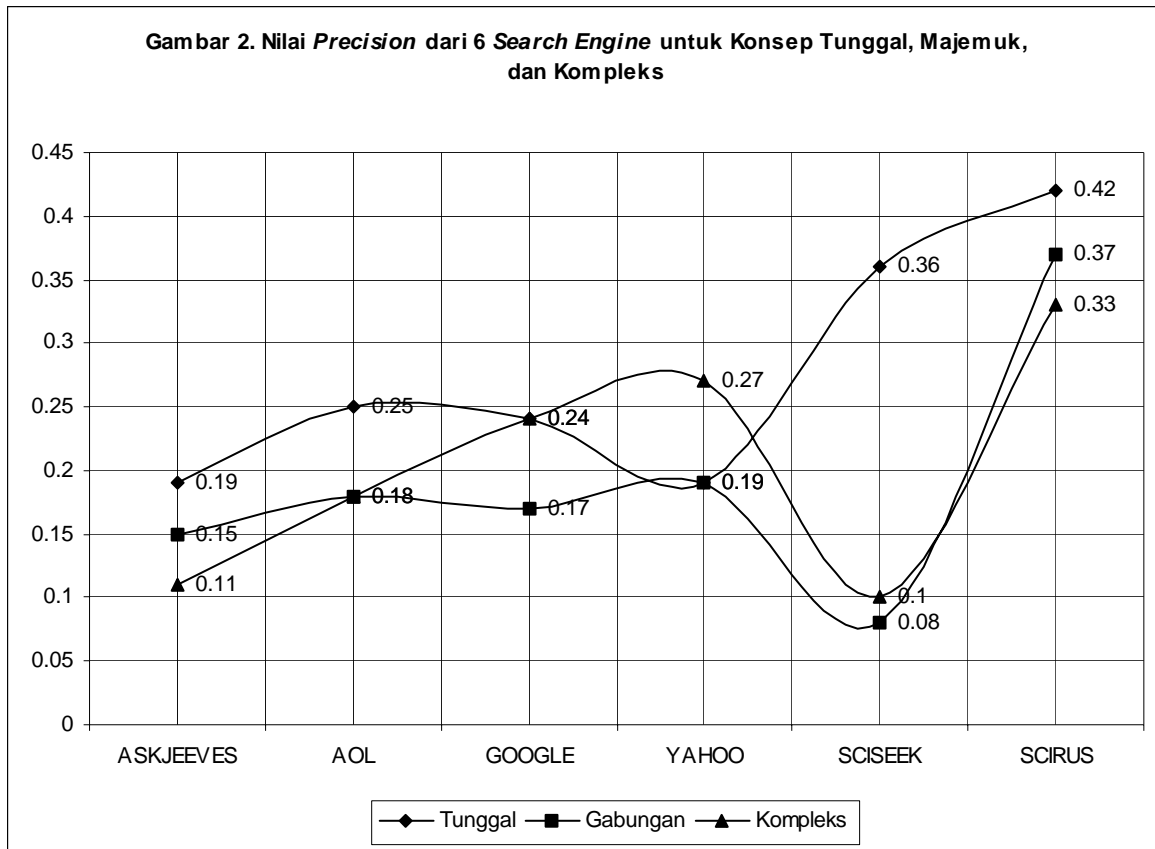
Tabel 1. Nilai tengah *precision* dan *relative recall* periode Maret – April 2006

Nilai Tengah	Askjeeves	AOL	Google	Yahoo	Sciseek	Scirus
Precision	0.15	0.20	0.22	0.22	0.18	0.37
Relative Recall	0.14	0.17	0.19	0.18	0.09	0.24

Dari tabel tersebut dapat dilihat, untuk nilai tengah *precision*, Scirus (0.37) berada di peringkat teratas. Kemudian diikuti Google dan Yahoo (0.22), AOL (0.20), Sciseek (0.18), dan Askjeeves (0.15). Scirus mendapat nilai *precision* tertinggi karena memang *search engine* ini didesain untuk pencarian informasi ilmiah. Dalam penelusurannya, Scirus telah melakukan filterisasi penjaringan informasi yang cukup handal baik untuk konsep tunggal, gabungan, maupun kompleks. Google dan Yahoo mendapat nilai tengah *precision* yang sama yaitu 0.22. Jika kedua *search engine* itu dibandingkan, maka terlihat Google memiliki keunggulan dalam pengelolaan konsep tunggal (0.24) sedangkan Yahoo mampu mengatasi Google untuk pengelolaan konsep yang kompleks (0.27). Sementara itu, Sciseek yang juga merupakan *search engine* bidang sains dan teknologi, dalam pengujian ini kurang memiliki algoritma pengindeksan yang memuaskan.

Untuk nilai tengah *relative recall*, Scirus (0.37) peringkat tertinggi. Kemudian diikuti Google (0.19), Yahoo (0.18), AOL (0.17), Askjeeves (0.14), dan Sciseek (0.09). Scirus

mengungguli *search engine* yang lain dalam pencarian konsep tunggal, gabungan, dan kompleks. Google dan Yahoo memperlihatkan performa yang baik untuk pencarian konsep yang kompleks. Askjeeves dan AOL cukup menonjol untuk konsep gabungan. Sedangkan, Sciseek cukup handal dalam mengelola pencarian konsep tunggal.



KESIMPULAN DAN SARAN

Untuk penelusuran informasi ilmiah, Scirus menjadi *search engine* yang cukup handal. Algoritma pengindeksannya telah cukup baik dalam mengelola konsep tunggal, gabungan, maupun kompleks serta efektif menjangkau informasi ilmiah dengan menampilkan akses ke berbagai jurnal *online* dan pangkalan data. Google mampu memperlihatkan performanya untuk pencarian konsep tunggal. Google scholar versi beta juga dapat dijadikan referensi untuk informasi ilmiah. Sementara itu, Yahoo cukup bersaing dengan Google, namun Yahoo menunjukkan kinerjanya yang lebih baik dalam pengelolaan konsep majemuk. Kelemahan kedua *search engine* ini adalah seringnya memunculkan hasil pencarian yang *redundant*. Askjeeves meskipun direkomendasi sebagai *top list search engine* oleh Search Engine Watch, dalam penelitian ini, kurang menunjukkan hasil yang maksimal. Fenomena yang sama juga terjadi pada America Online (AOL). Sedangkan, Sciseek nampaknya harus bekerja keras untuk membuktikan dirinya sebagai spesialis *search engine* bidang sains dan teknologi. Algoritma pengindeksan harus ditingkatkan sehingga dapat menjangkau informasi ilmiah seluas mungkin yang dibarengi dengan peningkatan kualitas pengelolaan konsep penelusuran.

Penelitian lebih lanjut dapat difokuskan pada pengukuran *search engine* bidang sains dan teknologi. Dengan demikian, dapat diperoleh gambaran aktual kualitas pencarian *search engine* bidang sains dan teknologi yang terus bermunculan, khususnya untuk bidang Ilmu Informasi dan Perpustakaan.

BIBLIOGRAFI

- Chu, H., & Rosenthal M. (1996). "Search engines for the World Wide Web: a comparative study and evaluation methodology" dalam *Proceedings of the ASIS 1996 Annual Conference*, October, 33. Hlm.127-35. <http://www.asis.org/annual-96/ElectronicProceedings/chu.html>. Akses tanggal 4 April 2006.
- Clarke, S., & Willett, P. (1997). "Estimating the recall performance of search engines" dalam *ASLIB Proceedings*, 49 (7). Hlm. 184-189.
- Kekäläinen, Jaana dan Kalervo Jarvelin. 2002. "Using graded relevance assesment in IR evaluation", *Journal of the American Society for Information Science and Technology*, Nov 2002 : 53, 13; Academic Research Library. Hlm.1120.
- Lancaster, F.W., and Fayen, E.G. 1973. *Information Retrieval On-Line*. Los Angeles, CA: Melville Publishing Co.
- Leighton, H., & Srivastava, J.1997. *Precision among WWW search services (serach engines): Alta Vista, Excite, HotBot, Infoseek and Lycos*. <http://www.winona.edu/library/webind2.htm>. Akses tanggal 11 April 2006.
- Mea, Vincenzo Della and Stefano Mizzaro. 2004. "Measuring retrieval effectiveness: a new proposal and a first experimental validation", dalam *Journal of the American Society for Information Science and Technology*; Apr. 2004; 55, 6; Academic Research Library. Hlm.530.
- Mizzaro, Stefano. 1997. "Relevance: The whole history" dalam *Journal of the American Society for Information Science*, 48. Hlm. 810-832.
- Salton, G and M.I McGill. 1984. *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Saracevic, T. 1996. "Relevance reconsidered" dalam *Proceedings of the international conference on conceptions of library and information science: integration in perspective*. Copenhagen: The Royal School of Librarianship. Hlm. 201-218.
- Schamber, I. 1994. *Relevance information behaviour* dalam *Annual review of information science and technology* 29. NJ: Information Today. Hlm. 3-48.
- Selingo, Jeffrey. "When search engine isn't enough, call a librarian", *New York Times*, February 5, 2004.

Shafi, S.M and Rafiq A. Rather. 2005. "Precision and Recall of Five Search Engines for Retrieval Scholarly Information in the Field of Biotechnology". *Webology*, Vol.2 No.2, August. 2005.

Sullivan, Danny. *Major search engines and directories*. <http://www.searchenginewatch.com>. Akses tanggal 13 Maret 2006.