

PENERAPAN *INFORMATION RETRIEVAL* PADA *SEARCH ENGINE*

NURAINI AHMAD, ARIENDA ADDIS PRASETYO, ANIS MASRURI

Pascasarjana UIN Sunan Kalijaga Yogyakarta

E-mail: nuraini.ahmad@uin-suka.ac.id, 20200012061@student.uin-suka.ac.id, anis.masruri@uin-suka.ac.id

ABSTRAK

Information retrieval (temu kembali informasi), saat ini telah diterapkan dalam berbagai bidang kehidupan manusia. Perpustakaan, museum, manajemen dokumen, manajemen pengetahuan, manajemen koleksi, manajemen arsip telah dikenal sebelum web search engine (mesin pencari website, sering disebut dengan search engine saja) diterapkan. Web sebagai media yang belakangan muncul, membutuhkan pula model sistem temu kembali tersendiri. Sebelum digunakannya mesin pencari website yang dikenal saat ini, para ahli perpustakaan telah mencoba menerapkan prinsip-prinsip kategorisasi dan klasifikasi sebagaimana mereka mengatur koleksi perpustakaan. Namun, ternyata website mempunyai karakteristik sendiri, sehingga model yang mereka kembangkan dengan cepat ketinggalan zaman. Website mempunyai sifat cepat berubah dan sementara. Dalam tulisan ini, penulis mendiskusikan bagaimana penerapan *information retrieval* pada search engine. Hasil penelitian menunjukkan bahwa search engine terus berkembang untuk menyediakan informasi yang relevan dengan pengembangan proses evaluasi, meningkatkan kemampuan, memperbarui data, terus beradaptasi dengan kebutuhan pengguna, meningkatkan kemampuan kinerja perangkat, dan berusaha menyelesaikan masalah yang dihadapi.

Kata Kunci: *Information Retrieval*, Temu Kembali Informasi, *Search Engine*

ABSTRACT

Information retrieval (information retrieval), while it has been applied in various fields of human life. Library, museum, document management, knowledge management, collection management, records management has been known before the web search engines (search engine website, often referred to as search engines only) applied. The Web as a medium that would later appear, it requires also a model retrieval system of its own. Before the use of a search engine website that is known at this time, experts of the library has been trying to apply the principles of categorization and classification as they arrange the collection of the library. However, it turns out that the website has its own characteristics, so the model that they develop quickly outdated. The Website has the properties of fast-changing and temporary. In this paper, the authors discuss how the implementation of *information retrieval* in search engines. The results showed that search engines continue to evolve to provide information that is relevant to the development of the evaluation process, improving the ability, update data, continue to adapt to users' needs, improve the ability of the performance of the device, and try to solve the problem encountered.

Keywords: *Information Retrieval*, Temu Kembali Informasi, *Search Engine*

PENDAHULUAN

Sistem temu kembali informasi digunakan untuk mengurangi apa yang disebut kelebihan informasi. Sistem temu kembali adalah sistem perangkat lunak yang menyediakan akses ke buku, jurnal, surat, catatan, film, rekaman audio, surat kabar, buletin, poster seni, lagu, dan dokumen lainnya; menyimpan dan mengelola dokumen-dokumen tersebut.

Model pencarian *web search engine* (mesin pencari website) – selanjutnya lebih disebut dengan *search engine* saja, atau mesin pencari saja –, merupakan model temu kembali informasi yang paling nyata diaplikasikan dan digunakan manusia saat ini. Sebelumnya, para ahli telah

memiliki beragam data dan informasi yang disimpan, kemudian diorganisasikan, selanjutnya dibuatkan sistem temu kembali informasinya, seperti: perpustakaan, kearsipan, manajemen dokumen, manajemen pengetahuan, manajemen informasi pribadi, dan museum. Hingga selanjutnya data dan informasi beredar dan berkembang pada media website di atas jaringan internet, sehingga kebutuhan akan sistem temu kembali yang mumpuni berbentuk mesin pencari mutlak dibutuhkan.

Aktifitas pencarian website adalah aktifitas harian yang dilakukan oleh hampir semua orang di dunia yang menggunakan internet. Istilah ‘Googling’ menjadi kata kerja baru yang sangat dikenal orang sebagai usaha melakukan pencarian tentang topik (seseorang atau sesuatu) tertentu di internet menggunakan mesin pencari Google. Google (merujuk pada halaman Google Search) menangani pencarian 40 ribu *query* tiap detik, lebih dari 3,5 miliar pencarian per hari, 1,2 triliun pencarian tiap tahun. (Internetlivestats, 2013). Menurut StatCounter Global Stats (2021) Google memiliki pangsa pasar 92 persen dari seluruh pasar mesin pencari seluruh dunia, ditambahkan oleh Alexa (2021) serta merupakan situs web yang paling banyak dikunjungi di dunia. Google dapat diakses dengan mudah baik menggunakan komputer, laptop, telpon pintar, tablet, dsb.

Pada pertengahan 1990-an, beberapa pustakawan dan peneliti LIS dengan antusias dan agak naif mencoba merancang proyek untuk membantu mengatur Internet, tepatnya seluruh Internet. (Oder, Norman, 1998) Upaya ini dimulai dengan sungguh-sungguh setelah World Wide Web (atau web) sepertinya akan menjadi wajah Internet. Namun, para profesional ini dengan cepat menyadari bahwa ada begitu banyak perubahan yang begitu cepat, sehingga upaya yang mereka lakukan menjadi ketinggalan zaman hanya dalam waktu singkat. Pada awalnya, perpustakaan berusaha menggunakan cara pengorganisasian tradisional, tetapi segera menjadi kelihatan jelas, bahwa katalogisasi perpustakaan skala penuh tidak terlalu cocok untuk menggambarkan website yang sering berubah dan lebih bersifat sementara.

Beberapa pustakawan mulai menyusun bibliografi website, beberapa di antaranya akhirnya menjadi direktori atau gerbang yang lebih formal ke Internet. Direktori Internet adalah kumpulan tautan (*link*) yang terorganisir ke situs web tentang topik tertentu. Perkembangan selanjutnya, beberapa penyedia layanan mulai mengumpulkan tautan melalui cara otomatis. INFOMINE, didirikan pada tahun 1994, adalah contoh direktori Internet yang dibuat oleh pustakawan yang bekerja sama dari beberapa institusi akademik. Dalam INFOMINE, berisi halaman yang berisi berbagai topik (misalnya, sumber daya keanekaragaman budaya, informasi pemerintah, peta dan data geografis), yang masing-masing berisi daftar tautan ke situs web yang direkomendasikan. Tautan diurutkan sesuai daftar abjad atau dicari menggunakan kata judul, nama pencipta, subjek, atau kata kunci. Organisasinya mirip dengan Yahoo! direktori, yang juga mulai dibuat pada tahun 1994. Yahoo! menggunakan metode penelusuran penelusuran yang lebih hierarkis, di mana seseorang memulai dengan salah satu dari 14 kategori tingkat atas, yang dibagi menjadi banyak subtopik.

Direktori situs web yang terstruktur secara hierarkis, selanjutnya tidak dipakai oleh mesin pencari umum, seperti Alta Vista (1995) dan Google (1998). Kedua mesin pencari tersebut menjadi model mesin pencari seperti yang digunakan hingga saat ini. Menurut Richard Nieva (2014) dalam Web Archive (2016) Setelah dua dekade beroperasi, Yahoo! dan INFOMINE menutup direktori mereka pada bulan Desember 2014 setelah mengalami penurunan penggunaan yang signifikan.

Upaya lain untuk mengontrol informasi di Internet adalah pengembangan salah satu skema metadata yang paling dikenal luas di dunia. Pustakawan adalah bagian dari tim yang mengembangkan standar metadata untuk menggambarkan sumber daya online yang disebut Dublin Core (DC). (Dublin Core Metadata Initiative, 2021) Skema metadata tersebut dirancang sebagai kumpulan elemen metadata dasar yang relatif kecil, karena sebagian besar situs web dan objek seperti dokumen lainnya di Internet tidak memerlukan banyak detail seperti yang biasanya ditemukan di katalog perpustakaan. Pada akhir 1990-an, OCLC membangun cara bagi

perpustakaan untuk membuat katalog sumber daya online secara kooperatif dan memiliki akses ke database metadata yang menjelaskan sumber daya web penting. Sistem ini, awalnya bernama CORC (Cooperative Online Resource Catalog), kemudian berkembang menjadi antarmuka berbasis web OCLC untuk katalogisasi—Connexion. Saat ini, di Connexion, deskripsi sumber daya dapat dilakukan dalam format MARC tradisional atau DC. (Connexion, 2021)

Pemakaian metadata sebagai dasar pembuatan katalog menemukan kelemahannya. Metadata sering disalahgunakan oleh pembuat website tertentu untuk memanipulasi hasil pencarian. Penyalahgunaan ini dapat dilihat sebagai upaya lokal yang tidak etis dalam pengoptimalan mesin pencari (*search engine optimization/SEO*) yaitu aktivitas yang bertujuan meningkatkan visibilitas situs web (membuat situs berperingkat lebih tinggi dalam hasil mesin pencari). Aktivitas SEO yang benar seharusnya berfokus pada serangkaian praktik terbaik dalam pemasaran, desain web, dan organisasi konten untuk mencapai tujuan bukan menipu pengguna.

Beberapa mesin pencari paling populer di dunia (yaitu: Google, Bing, Yahoo!, dan Yandex) telah mulai menggunakan teknologi semantik tertentu yang menyematkan sejumlah kecil metadata ke dalam kode Hypertext Markup Language (HTML), untuk membuat data yang ditemukan dalam situs web lebih banyak. Jika digunakan dengan benar, metadata dapat mencakup, antara lain: informasi tentang pembuat, judul, subjek, dan aspek deskripsi tradisional lainnya; bagian nontekstual dari sebuah situs; bagian tekstual dari situs; entitas, peran, dan hubungan; dan tujuan serta sejarah situs. (Joudrey, Daniel N. and Taylor, Arlene, 2018)

Keinginan awal di antara pustakawan dan profesional informasi lainnya untuk menggambarkan setiap bagian dari keseluruhan Internet menghilang dengan cepat. Tidak realistis untuk berpikir bahwa setiap sumber daya web dapat dikatalogkan sebagaimana buku atau DVD di perpustakaan. Sebaliknya profesional dan masyarakat umum telah menerima mesin pencari untuk mencari, menemukan, dan mengambil materi di Internet. Meskipun hasilnya terkadang luar biasa, seringkali penuh dengan kesalahan, mesin pencari sangat baik untuk memenuhi sebagian besar kebutuhan pencarian umum. Dalam tulisan ini, penulis mendiskusikan tentang bagaimana penerapan temu kembali informasi pada mesin pencari website.

METODE PENELITIAN

Metode penulisan menggunakan kualitatif deskriptif dengan pendekatan *literature review*, penulis mengikuti pedoman yang diberikan oleh Kitchenham et al. (B.A Kitchenham & Charters,2007)., di mana pedoman ini dimulai dengan definisi informasi sehingga sampai menemukan informasi menjadi kebutuhan. Kemudian, dilangkah kedua, mengelompokan dan mengidentifikasi kata kunci '*information retrieval*' dan '*search engine*'. Pada langkah ketiga, kami melakukan ekstraksi data untuk '*information retrieval*' dan '*search engine*'.

Tabel 1. Study Exclusion and Inclusion Criteria

| NO | EXCLUSION CRITERIA |
|-----|---|
| EC1 | Abstrak tak menggambarkan satu atau beberapa Internet Retrieval dan Search Engine |
| EC2 | Penelitian menggunakan, tetapi penelitian tidak ditargetkan pada sebuah IR dan SE |
| EC3 | Buku tidak ditulis dengan bahasa inggris |
| EC4 | Tidak Berhubungan dengan IR dan SE |
| NO | INCLUSION CRITERIA |
| EC1 | Search Engines: Information Retrieval in Practice |
| EC2 | Introduction to Information Retrieval |
| EC3 | Modern Information Retrieval |
| EC4 | Information Retrieval Interaction |

Tabel 2. Quality assessment criteria.

| NO | QUESTION |
|----|--|
| Q1 | Sesuai dengan tujuan mata kuliah? |
| Q2 | Lingkup dan konteks dan desain penelitian dengan jelas didefinisikan? |
| Q3 | Proses penelitian yang didokumentasikan dengan baik? |
| Q4 | Menjawab semua pertanyaan penelitian? |
| Q5 | Apakah temuan utama menyatakan dengan jelas? mengenai kredibilitas, validitas, dan kehandalan? |
| Q6 | Apakah kesimpulan itu berhubungan dengan tujuan penelitian? Dapat dipercaya? |

HASIL DAN PEMBAHASAN

Information Retrieval (Temu Kembali Informasi)

Pengertian *information retrieval* (IR, temu kembali informasi) menurut Gerard Salton (1968) dalam Arguillo (2013) adalah: *Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information.* (Arguillo, Jaime, 2013) Pengertian tersebut dapat diartikan sebagai berikut, temu kembali informasi adalah bidang yang berkaitan dengan struktur, analisis, organisasi, penyimpanan, pencarian, dan pengambilan kembali informasi.

Informasi yang dimaksud disini bisa berupa dokumen, email, buku, berita, paper ilmiah, pesan teks, PDF, *tweets*, berkas Microsoft Word, *spreadsheet*, pesan singkat, dsb. Beberapa dari dokumen tersebut mempunyai ciri khas masing-masing misal: email, pasti mempunyai bagian pengirim, tujuan, judul, dan isi pesan; atau sebuah paper ilmiah memiliki bagian judul, penulis, tanggal terbit.

Tujuan dari temu kembali informasi adalah untuk menyediakan informasi yang terbaik sesuai dengan yang dibutuhkan pengguna dari database yang telah disimpan. Masalah pencarian informasi adalah menentukan informasi mana yang terbaik relevansinya sesuai dengan permintaan pengguna tertentu.

Untuk memenuhi permintaan informasi yang sesuai dengan yang dibutuhkan pengguna, sebuah sistem temu kembali informasi menggunakan metode, teknik, untuk memprediksi apa yang sebenarnya diinginkan pengguna. Informasi yang relevan ditampilkan di hasil pencarian paling atas, semakin ke bawah semakin kurang relevan, sehingga pengguna dapat dengan cepat menemukan informasi yang sesuai dengan yang dibutuhkannya.

Arguillo (2013) menjelaskan, dalam sebuah sistem temu kembali informasi setidaknya memiliki 4 komponen: (1) ada kebutuhan informasi dari pengguna, (2) ada sumber informasi, (3) ada sistem yang terdiri dari perangkat keras, perangkat lunak, dan teknik temu kembali, (4) ada strategi penelusuran.

Proses dari temu kembali informasi, menurut Arguillo (2013), dimulai dari: (1) Adanya kebutuhan informasi dari pengguna. Pengguna melakukan analisa kebutuhan informasi dengan menyusun *query* atau *keyword* atau *pertanyaan*; (2) Pemilihan sistem temu kembali informasi, sebagai contoh pengguna menggunakan Google. Pengguna melakukan penelusuran di Google dengan *query* dan strategi *query* yang telah dipersiapkan; (3) Sistem kemudian memproses *query* dan menuju ke database manager, apabila informasi yang dicari ada dan *query* tepat, maka informasi akan ditemukan. Akan tetapi apabila informasi ada, tetapi *query* tidak tepat, maka user harus melakukan pembuatan kembali *query* sehingga informasi yang dikehendaki dapat ditemukan, (4) Informasi yang ditemukan kemudian dievaluasi apakah benar-benar sesuai kebutuhan atau tidak. Untuk mengukur tingkat ketepatan hasil penelusuran digunakan metode *recall* dan *presition*.

Tingkat ketepatan dari hasil yang diperoleh ini akan dipengaruhi oleh ketrampilan dari pengguna. Jenis pengguna dapat dibedakan menjadi tiga level yaitu *novice*, *intermediary* dan *expert*.

Sistem temu kembali informasi telah diterapkan pada perpustakaan, bisnis, akademik, arsip, manajemen dokumen, manajemen pengetahuan, dalam bentuk: bibliografi, katalog, indeks, finding aids, register.

a. Dimensi *Information Retrieval*

Information Retrieval lebih dari sekedar teks dan lebih dari sekedar pencarian web, meskipun hal-hal tersebut adalah yang paling banyak digunakan manusia saat ini. IR digunakan dalam berbagai media yang berbeda, berbagai jenis pencarian aplikasi, dan berbagai tugas. Tiga dimensi IR meliputi: isi (konten), aplikasi, dan tugas.

1) Dimensi Isi (konten),

Berupa data tekstual; media baru: video, foto, musik, percakapan, ceramah, audio, dokumen hasil pemindaian, dst. Pendekatan IR untuk pencarian dan evaluasi apakah sesuai.

2) Dimensi Aplikasi,

a) Pencarian web;

b) Pencarian vertical, pada domain web yang dibatasi;

c) Pencarian enterprise: corporate intranet basis data, surel, halaman web, dokumentasi, kode, halaman wiki, tag, direktori, presentasi lembar kerja;

d) Pencarian desktop;

e) Pencarian peer to peer;

f) Pencarian literature;

g) Pencarian forum.

3) Dimensi Tugas.

a) Pencarian oleh pengguna/pencarian ad-hoc, jangkauan query sangat besar, tidak di-spesifikkan sebelumnya

b) Penyaringan (filter),

c) Klasifikasi / kategori,

d) Jawaban pertanyaan. Mirip dengan pencarian. Secara otomatis menjawab pertanyaan yang diajukan dalam bahasa aslinya. Menyediakan jawaban apa adanya, bukan daftar dokumen. (Naumann, Felix, 2011)

b. Aspek pengembangan *Information Retrieval*

Ada tiga masalah besar yang harus dijawab oleh *Information Retrieval* yaitu: relevansi, evaluasi, dan kebutuhan pengguna. Relevansi: dokumen yang relevan berisi informasi yang dicari pengguna saat dia mengirimkan kueri. Evaluasi: seberapa bagus perangkan dokumen hasil pencarian sesuai dengan harapan pengguna. Kebutuhan pengguna: pengguna adalah hakim juri penilai kualitas sistem.

Aspek relevansi bergantung dari subjektivitas pengguna. Banyak faktor yang mempengaruhi keputusan seseorang tentang apa yang relevan, bisa jadi: konteks, kebaruan, gaya, atau kebetulan saja.

Evaluasi adalah serangkaian prosedur dan pengukuran untuk membandingkan hasil yang dikeluarkan sistem dengan harapan pengguna. Evaluasi berawal dari proses percobaan Cranfield pada tahun 1960an. Recall dan Precision adalah dua contoh pengukuran keefektifitasan sistem IR. Mengenai perihal pengguna dan kebutuhan informasi, penggunaan dan evaluasi sistem IR adalah berbasis pengguna.

Search Engine (Mesin Pencari)

Bagi sebagian besar pencari informasi, mesin pencari mungkin yang paling familiar dari semua alat pencarian yang telah dibuat dan dipakai manusia. Mesin pencari adalah alat yang dikembangkan untuk sistem komputer, khususnya Internet, untuk menemukan contoh kata atau frasa yang diminta yang dapat ditemukan dalam dokumen yang tercakup dalam ruang lingkup alat tersebut. Mereka dikembangkan untuk tujuan mencari dokumen teks lengkap (atau indeks dokumen tersebut) untuk kata-kata atau frase tertentu.

Beberapa mesin pencari digunakan untuk mencari hanya satu sumber daya tertentu, seperti situs web untuk organisasi seperti OCLC. Mesin pencari lainnya adalah untuk mencari di seluruh Internet (walaupun hanya sebagian kecil dari Internet yang benar-benar dicari oleh masing-masing mesin pencari). Menggunakan program yang disebut spider (juga dikenal sebagai *crawler* atau robot), setiap mesin pencari secara otomatis mengumpulkan informasi dari sumber daya web dan menemukannya ke dalam database catatan atau dalam indeks teks lengkap. *Crawler* biasanya diprogram untuk masuk ke web untuk mengambil dan mengunduh salinan halaman web dan semua yang terkait dengannya, semua yang terkait dengan tautan, dan seterusnya. Indeks kolektif besar—dibuat dari *database crawler*—disimpan di pusat data di seluruh dunia. Indeks ini (beberapa terdiri dari lebih dari 100 juta gigabyte data) adalah apa yang dicari ketika pengguna memasukkan istilah ke dalam kotak pencarian.

Tidak semua sumber daya web telah diindeks. Beberapa sumber mungkin diindeks oleh hanya beberapa mesin pencari. Selain itu, tidak semua mesin pencari mengindeks setiap jenis materi yang tersedia secara online. Hanya dalam beberapa tahun terakhir sumber daya dalam format PDF dan PowerPoint telah dapat dicari secara online. Dalam beberapa kasus, sumber daya mungkin dilindungi kata sandi, dan akibatnya, sumber daya tersebut mungkin tidak diindeks di mesin telusur karena tidak dapat diakses oleh perayap. Secara umum, hanya web permukaan yang disertakan dalam indeks mesin pencari; *deep web* (web-web yang tersembunyi) tidak diindeks di mesin pencari.

Tidak seperti katalog dan alat bantu pencarian, mesin pencari telah menjadi bagian dari interaksi rutin pengguna dengan dunia (setidaknya bagi mereka yang cukup diuntungkan untuk memiliki akses Internet yang konsisten dan andal). Karena sudah terbiasa dengan mesin pencari dan relatif mudah digunakan (cukup masukkan beberapa kata kunci ke dalam kotak pencarian), alat pencarian lainnya seringkali terlihat kuno dan rumit. Meskipun mesin pencari mungkin atau mungkin tidak, memberikan hasil yang memuaskan secara intelektual seperti hasil dari alat pencarian lainnya, pengguna sering melaporkan kepuasan karena mereka menemukan sesuatu yang berhubungan dengan apa yang mereka cari, dan menemukannya dengan cepat. Tetapi sebagian besar pengguna tidak tahu apakah yang mereka temukan asli, otoritatif, atau yang terbaik yang tersedia tentang topik mereka. Beberapa pengguna tidak selalu menyadari bahwa mesin pencari dapat mendorong iklan dan tautan sponsor ke bagian atas halaman hasil, atau bahwa informasi yang lebih mendalam dapat ditemukan lebih jauh di bawah daftar hasil.

Pada kebanyakan mesin pencari, tampilan hasil pencarian biasanya diatur menurut relevansinya, yang dapat dihitung dengan berbagai cara. Mesin pencari dapat menghitung relevansi dengan memberikan bobot yang berbeda untuk faktor-faktor seperti berapa banyak istilah pencarian yang ditemukan di setiap halaman web, seberapa sering setiap istilah ditemukan di halaman, apakah istilah tersebut berdekatan atau tersebar, apakah istilah tersebut ada di kepala dokumen atau terkubur lebih dalam di tubuhnya, dan seberapa umum (atau jarang) istilah pencarian tertentu.

Google menggunakan sistem Pagerank ketika pertama kali diperkenalkan Larry Page dan Sergey Brin pada 1995. Pagerank mengukur halaman website mana yang lebih penting dengan menghitung jumlah halaman lain yang merujuknya. Selanjutnya, Mesin pencari terus berusaha untuk mengembangkan metode yang lebih canggih dalam menghitung relevansi dan untuk

menampilkan hasil. Misalnya, Google menggunakan rumus yang memeriksa lebih dari 200 faktor berbeda dalam menentukan relevansi. Google menggunakan faktor umum, seperti penempatan kata dan frekuensi kata, dalam formulanya, tetapi juga kriteria yang kurang diharapkan seperti kualitas situs, wilayah geografis, riwayat pencarian, kebaruan, dan popularitas situs (yaitu, seberapa sering halaman ditautkan oleh halaman web lain) sebagai elemen dalam peringkat hasil pencarian. Google, Bing, dan mesin pencari lainnya terus memperbarui, menyempurnakan, dan menguji algoritma dan formula mereka dalam pencarian mereka untuk meningkatkan pengalaman pencarian Internet.

Ketika metadata pertama kali dikembangkan untuk menggambarkan dokumen web, sebagian besar mesin pencari tidak mengambil metadata yang disediakan pengguna menjadi pertimbangan dalam peringkat relevansinya karena kesalahan dan penyimpangan metadata (misalnya, termasuk istilah populer yang tidak memiliki hubungan dengan situs yang sedang dijelaskan). Dengan perkembangan data tertaut, bagaimanapun, mesin pencari utama sekarang merangkul dimasukkannya data yang lebih terstruktur sebagai bagian dari markup tekstual yang mendasari sumber daya web. Dengan dukungan untuk data tertaut dan untuk memproses metadata kontekstual yang ditemukan dikodekan dalam dokumen web, mesin pencari diharapkan terus meningkat selama beberapa dekade mendatang. (Joudrey, Daniel N. and Taylor, Arlene, 2018)

Pengembangan mesin pencari, setidaknya mempertimbangkan aspek-aspek berikut:

- a. *Performance: Efficient search and indexing*
 Perlunya mengukur dan meningkatkan efisiensi pencarian dengan cara mengurangi waktu respon, meningkatkan waktu membuat index, meningkatkan hasil query. Index adalah struktur daya yang didesain untuk meningkatkan efisiensi pencarian.
- b. *Incorporating new data: Coverage and freshness*
 ‘Koleksi’ dalam sebuah aplikasi adalah yang paling sering dan selalu berubah dalam hal penambahan, pengurangan, pembaruan. Demikian pula dalam koleksi data website yang disimpan dalam mesin pencari. Akuisisi data baru menggunakan crawler akan selalu menambah koleksi. Aspek yang harus diperhatikan dalam lingkup ini adalah: cakupan (*coverage*) yang telah diindeks dan seberapa segar/baru koleksi baru itu diindeks.
- c. *Scalability: Growing with data and users*
 Perlunya menjamin semua proses dapat dilakukan dengan baik, dengan miliaran proses query tiap hari, miliaran pengguna, dan memproses data mencapai 100 ribu TB.
- d. *Adaptability: Tuning for applications*
 Penting untuk terus memperhatikan kebutuhan pengguna. Mengubah dan menyetel ulang komponen mesin seperti algoritma perangkian, strategi peng-indeks-an, mengatur antar muka dengan menyesuaikan perangkat yang terus berubah, dan beradaptasi untuk membuat prasyarat yang berbeda di aplikasi atau pengguna yang berbeda.
- e. *Specific problems: e.g., Spam*
 Masalah yang khas di mesin pencari adalah spam. Pengguna akan jengah dengan hasil pencarian yang tidak tepat, namun ternyata dapat tampil terus di halaman pertama hasil pencarian. Meski sudah lama menjadi bahasan dan dicari solusinya, sampai saat ini spam masih menjadi hal yang harus terus dicari jalan keluarnya. (Naumann, Felix, 2011)

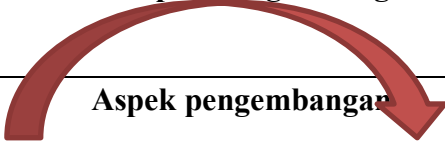
KESIMPULAN

Penerapan IR pada *search engine* (mesin pencari) adalah contoh sempurna bagaimana manusia sangat terbantu untuk memperoleh informasi yang dikehendaknya dengan memberi permintaan kata kunci atau bahkan sekumpulan kalimat tertentu. *Search engine*, berbeda dengan model sistem IR lain sebelumnya, yang cenderung kaku, tidak berkembang, dan kuno, memperlihatkan keunggulannya dalam memberi yang terbaik untuk pengguna. *Search engine* yang

ada saat ini membuktikan bahwa pengembangan dan penggunaan teknologi yang terus diperbarui dapat memenuhi kebutuhan pengguna yang mencapai miliaran manusia. Suatu hal yang tidak terbayangkan dalam beberapa dekade yang lalu.

Meski penggunaannya tidak dimaksudkan untuk menggantikan sistem-sistem IR yang telah ada sebelumnya, *search engine* berfokus untuk meningkatkan relevansi, memenuhi kebutuhan pengguna, memberikan yang terbaik untuk pengguna. Tambahan *aspek search engine* daripada *information retrieval* adalah sebagai berikut:

Tabel.3 Aspek Pengembangan



| Aspek pengembangan | |
|--|---|
| Information Retrieval System | Search Engines |
| ✓ Relevance: <i>Effective ranking</i> | ✓ Performance: <i>Efficient search and indexing</i> |
| ✓ Evaluation: <i>Testing and measuring</i> | ✓ Incorporating new data: <i>Coverage and freshness</i> |
| ✓ Information needs: <i>User interaction</i> | ✓ Scalability: <i>Growing with data and users</i> |
| | ✓ Adaptability: <i>Tuning for applications</i> |
| | ✓ Specific problems: <i>e.g., Spam</i> |

kompatibilitas. Selain itu, dapat juga ditambahkan prospek pengembangan hasil penelitian dan prospek aplikasi penelitian selanjutnya ke depan (berdasarkan hasil dan pembahasan).

DAFTAR PUSTAKA

Alexa. (2021). *The top 500 sites on the web, Global*. Alexa.com, <https://www.alex.com/topsites> (diakses 3 Oktober 2021)

Arguello, Jaime (2013). *INLS 509: Introduction to Information Retrieval*. https://ils.unc.edu/courses/2021_fall/inls509_001/ (diakses 1 Oktober 2021)

B.A Kitchenham & Charters. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Technical report, Ver. 2.3 EBSE Technical Report. EBSE 1: 1–54.

Connexion. (2021). *Connexion Overview*. Oclc.org, <https://www.oclc.org/en/connexion.html> (diakses 3 Oktober 2021)

Dublin Core Metadata Initiative. (2021). *DCMI History*. Dublincore.org, <https://dublincore.org/about/history/> (diakses 3 Oktober 2021)

Internetlivestats. (2013). *Google Search Statistics*. Internetlivestats.com, <https://www.internetlivestats.com/google-search-statistics/> (diakses 3 Oktober 2021)

Joudrey, Daniel N. and Taylor, Arlene. (2018). *The Organization of Information, Fourth Edition, with the assistance of Katherine M. Wisser. Library and Information Science Text Series*. Libraries Unlimited: Santa Barbara, California.

Naumann, Felix. (2011). *Search Engines, Chapter 1 – Introduction*. https://hpi.de/fileadmin/user_upload/fachgebiete/naumann/folien/SS11/Search_Engines/SearchEngines_01_Introduction.pdf (diakses 1 Oktober 2021)

Nieva, Richard. (2014). *Yahoo Directory -- once the firm's cornerstone -- to shut down*. Cnet.com, <https://www.cnet.com/tech/services-and-software/yahoo-directory-once-the-companys-cornerstone-to-shut-down/> (diakses 3 Oktober 2021)

Oder, Norman. (1998). *Cataloging the Net: Can We Do It?* Library Journal 123, no. 16: 47–51.

StatCounter Global Stats. (2021). *Search Engine Market Share Worldwide*. Statcounter.com, <https://gs.statcounter.com/search-engine-market-share> (diakses 3 Oktober 2021)

Web Archive. (2016). *INFOMINE: Scholarly Internet Resource Collections*. University of California, Riverside,
<https://web.archive.org/web/20160311214940/http://library.ucr.edu/view/infomine>;
(diakses 3 Oktober 2021)