

Search Engines dan Topics Directory

Search Engines

Berdasarkan Pew Internet Project Report (2002):

- ♦ Lebih dari 25% dari pengguna Internet memperoleh informasi pada Web dengan bantuan search engines.
- ♦ Lebih dari 80% pengguna Internet telah memakai search engine paling sedikit sekali.

Desain dari sebuah search engine dapat dibedakan dengan banyak kriteria. Perbedaan yang paling penting mungkin antara kedua jenis berikut:

- ♦ Search engine yang mempunyai fungsi umum (yang ditujukan untuk menyediakan jawaban umum untuk kemungkinan yang paling besar dari user) dan
- ♦ Search engine khusus yang memfokuskan pada topik atau domain yang spesifik, misal untuk medis, hukum, biotechnology, dan lain sebagainya.

Kondisi Web 2002-2003

Pada tahun 2002-2003, jumlah web sudah mencapai 11 triliun halaman, yang ukurannya hampir mencapai 100 terabytes. Media penyimpanan data pada skala tersebut belum ada.

- ♦ Sebuah stasiun radio musik memiliki sekitar 10000 record, atau sekitar 5 terabytes dari data yang tidak dikompres
- ♦ US Library Of Congress memiliki sekitar 20 juta volume, atau diperkirakan sekitar 20 terabytes text.
- ♦ The Internet Archive tersedia untuk para peneliti, sejarawan, dan pelajar. Sebuah interface yang disebut sebagai *The Wayback Machine* memperbolehkan user untuk mengakses berkas lama dari halaman web yang sudah diarsip.

Sejarah Singkat Beberapa Search Engines

- ♦ **Lycos**
 - Ditemukan pada bulan Januari 1994, beroperasi pada bulan Juni 1994, dan diperdagangkan pada bulan April 1996.
 - Lycos dilahirkan dari suatu proyek riset pada Carnegie Mellon University oleh Dr. Michael Mauldin.
- ♦ **WebCrawler**
 - Mulai online pada musim semi tahun 1994. Itu juga dimulai sebagai proyek riset, di Universitas Washington, oleh Brian Pinkerton.
- ♦ **Alta Vista**
 - Salah satu search engine terbaik yang dikenal dengan klaim di atas 60 hak paten, sejauh ini yang tertinggi di area ini.
 - Dikembangkan sepanjang musim semi tahun 1995, Louis Monier, Joella Paquette, dan Paul Flaherty pada laboratorium riset Digital Equipment Corporation's

- Diluncurkan pada bulan Desember 1995, AltaVista memulai dengan dua juta query dalam satu hari selama tiga minggu.

♦ **HotBot dan Inktomi**

- Banyak orang lain yang mengikuti para pelopor search engine dan menawarkan berbagai inovasi.
- Kedua engine ini menawarkan suatu *distributed architecture* untuk crawling dan penyimpanan halaman-halamannya.

♦ **PageRank dan Google**

- Pada musim gugur 1996, Larry Page dan Sergey Brin, Ph.D. mahasiswa dari Stanford University, menggunakan ide yang berbeda untuk menjangkau 60 juta pages untuk menempati score bergengsi yang disebut PageRank.
- Kemudian mereka membangun sebuah sistem search disebut Backrub. Pada 1997, Backrub mulai online sebagai Google (www.google.com).

Direktori Topik

Gelombang inovasi yang pertama berhubungan dengan infrastruktur dasar yang berisikan: crawler dan search engine.

Direktori topik menjadi corak penting gelombang berikutnya.

♦ **Yahoo**

- Pada tahun 1994, Jerry Yang dan David Filo, Ph.D. mahasiswa Universitas Stanford, menciptakan direktori Yahoo! (www.yahoo.com) untuk membantu para temannya mencari lokasi Web yang bermanfaat. Ia bertumbuh sebanyak ribuan tiap harinya.
- Srinija Srinivasan, alumni Stanford yang lain, menyajikan keahlian untuk menciptakan dan memelihara hirarki bercabang yang seperti pohon yang memandu orang-orang menuju isi. Pada April 1995, proyek yang telah dimulai sebagai kegemaran disatukan sebagai sebuah perusahaan.
- Masa lalu: Pengaturan pengetahuan ke dalam ontologi adalah suatu seni, diturunkan dari filosofi dan epistemologi. Suatu ontologi menggambarkan suatu kosa kata, kesatuan yang ditunjuk oleh unsur-unsur dalam kosa kata, dan hubungan antar kesatuan tersebut. Kesatuan mungkin sangat kecil, seperti di WordNet, suatu jaringan yang berhubungan dengan kamus untuk Bahasa Inggris, atau mereka mungkin adalah topik yang secara relatif cukup besar, seperti pada topik direktori Yahoo!.
- Rata-rata pemakai komputer menjadi terbiasa dengan hirarki direktori dan file, dan kebiasaan ini secara alami berpindah ke taksonomi topik-topik.
- Yahoo! memperkerjakan tim editor untuk menjaga taksonomi.

♦ **About.com dan Open Directory Project (dmoz.org)**

- About.com dan Open Directory Project (dmoz.org) menjadi lebih didesentralisasi dan membentuk jaringan sukarelawan yang hasil pekerjaannya digabungkan.