

**RANCANG BANGUN SKEMA *BACKEND* DENGAN
PENERAPANNYA PADA APLIKASI TEKNOLOGI PERIKANAN
MODERN DENGAN FITUR *MULTI USER* DAN
INVENTARISASI BERBASIS *MULTI PLATFORM***

Proposal Skripsi

**Disusun untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Komputer**



*Mencerdaskan dan
Memartabatkan Bangsa*

**Oleh:
Akbar Maulana Alfatih
1313619003**

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA**

2022

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan proposal skripsi yang berjudul **"Rancang Bangun Skema *Backend* Dengan Penerapannya Pada Aplikasi Teknologi Perikanan Modern Dengan Fitur *Multi User* dan Inventarisasi Berbasis *Multi Platform*".**

Keberhasilan dalam menyusun proposal skripsi ini tidak lepas dari bantuan berbagai pihak yang mana dengan tulus dan ikhlas memberikan masukan guna sempurnanya proposal skripsi ini. Oleh karena itu dalam kesempatan ini, dengan kerendahan hati penulis mengucapkan banyak terima kasih kepada:

1. Yth. Para petinggi di lingkungan FMIPA Universitas Negeri Jakarta.
2. Yth. Ibu Ir. Fariani Hermin Indiyah, M.T selaku Koordinator Program Studi Ilmu Komputer.
3. Yth. Bapak Muhammad Eka Suryana, M.Kom selaku Dosen Pembimbing I yang telah membimbing, mengarahkan, serta memberikan saran dan koreksi terhadap proposal skripsi ini.
4. Ayah dan Ibu penulis yang selama ini telah mendukung dan membantu menyelesaikan proposal skripsi ini.
5. Teman-teman Program Studi Ilmu Komputer 2019 yang telah mendukung dan membantu proposal skripsi ini.

Penulis menyadari bahwa penyusunan proposal skripsi ini masih jauh dari sempurna karena keterbatasan ilmu dan pengalaman yang dimiliki. Oleh karenanya, kritik dan saran yang bersifat membangun akan penulis terima dengan senang hati.

Akhir kata, penulis berharap tugas akhir ini bermanfaat bagi semua pihak khususnya penulis sendiri. Semoga Allah SWT senantiasa membalas kebaikan semua pihak yang telah membantu penulis dalam menyelesaikan proposal skripsi ini.

Jakarta, xx xxxx xxxx

Akbar Maulana Alfatih

DAFTAR ISI

KATA PENGANTAR	iv
DAFTAR ISI	iv
DAFTAR GAMBAR	v
I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah	6
C. Pembatasan Masalah	6
D. Tujuan Penelitian	6
E. Manfaat Penelitian	7
DAFTAR PUSTAKA	8

DAFTAR GAMBAR

Gambar 1.1	<i>High Level Google Architecture</i> (Brin dan Page, 1998)	5
------------	---	---

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Saat ini internet merupakan suatu kebutuhan bagi masyarakat secara umum yang berimplikasi terhadap jumlah situs yang terdapat pada *World Wide Web*. Berdasarkan survei yang dilakukan NetCraft pada bulan April tahun 2022, terdapat 1 miliar situs yang terdapat di web dengan 202 juta situs yang aktif. Dengan banyaknya situs tersebut, maka pengguna internet memerlukan sebuah mesin pencari.

Mesin pencari atau yang biasa disebut *Search Engine* merupakan sebuah program komputer yang berguna untuk membantu pengguna dalam mencari situs web berdasarkan permintaan pencarian pengguna. Mesin pencari sebenarnya tidak berbeda dengan *website* pada umumnya, hanya saja perannya lebih terfokus pada pengumpulan dan pengorganisasian berbagai informasi di internet sesuai dengan kebutuhan penggunanya. Selain untuk memudahkan pencarian, mesin pencari juga berguna untuk meningkatkan pengunjung sebuah situs web.

Mesin pencari pertama kali diperkenalkan oleh Alan Emtage pada tahun 1990 dengan nama *Archie*. *Archie* melakukan pengindeksan *file* pada web dan hanya dapat menampilkan daftar nama situs, tidak dengan isi atau kontennya (Seymour dkk., 2011). Setelah kemunculan *Archie*, di tahun berikutnya mulai bermunculan mesin pencari baru. Pada tahun 1993 *Aliweb* muncul dengan memberikan kesempatan pengguna untuk mengunggah halaman situsnya agar terindeks dan memberikan deskripsi untuk halaman tersebut. Kemudian *Altavista* muncul pada tahun 1995 yang menggunakan teknik dan algoritma yang lebih maju. Setelah

Altavista, pada 2004 muncullah *Yahoo* sebagai salah satu mesin pencari yang maju di mana pengelola dan pemilik situs dapat menambah dan mengedit informasi dalam web tanpa mengeluarkan biaya. Namun, perkembangan mesin pencarian *Yahoo* sangat lambat sehingga pengguna internet perlahan mulai mencari mesin pencari yang lain (Seymour dkk., 2011).

Search engine yang saat ini paling banyak digunakan adalah *Google*. Mesin pencari yang diluncurkan pada tahun 1998 ini berkembang jauh berbeda dengan mesin pencari yang lain. Mereka mampu mengindeks halaman-halaman dan melakukan perangkingan, yang berarti bahwa halaman yang paling sering dikunjungi atau diklik akan berada paling atas dalam pencarian dengan *keyword* terkait.

Search engine dalam perkembangannya tidak hanya memuat halaman berisi informasi, namun juga dapat menjadi sarana iklan. Pada mesin pencari *Google*, terdapat layanan iklan yang disebut dengan *Adwords*. Layanan ini memungkinkan untuk pemilik usaha untuk mengiklankan situs web usahanya, sehingga saat pengguna *Google* mencari suatu barang, hasil pencarian *Google* akan menampilkan situs web pemilik usaha di bagian teratas dan tertulis sebagai "Ad" atau iklan.

Dalam arsitektur pembuatan *search engine*, terdapat salah satu bagian yang disebut *crawler*. *Crawler* berfungsi untuk mengumpulkan data-data yang akhirnya data-data tersebut akan diproses dan ditampilkan ke pengguna. *Crawler* ini perlu dirancang sebaik mungkin untuk mengumpulkan data yang dapat mendukung *search engine*.

Web crawler pertama kali dimunculkan pada tahun 1944 bernama *RBSE* dengan dasar dua program yaitu *Spider* dan *Mite* (Eichmann, 1994). Dua program dasar ini berguna untuk membentuk antrian data dalam *database* serta untuk mengunduh halaman dari sebuah web. Guna dari *web crawler* sendiri adalah sebagai

perangkat lunak yang secara otomatis bekerja untuk menjelajahi *website* dengan cara terorganisir.

Web crawler berperan penting pada *search engine* karena berfungsi untuk mengumpulkan data sebanyak-banyaknya. Tanpa *web crawler*, *search engine* tidak akan mempunyai data dan tidak bisa melakukan *indexing*. Data yang dihasilkan dari sebuah *web crawler* ini adalah data yang terbaru dan memiliki nilai keakuratan yang tinggi. *Web crawler* yang dimiliki *Google* bernama *Googlebot*, *web crawler* *Bing* bernama *Bingbot*, dan *web crawler* *Yahoo* bernama *Slurp bot*.

Dalam jurnal tahun 2020 oleh Kaur dan Geetha, mereka membuat sebuah *web crawler* untuk *hidden web* menggunakan *SIM+HASH* disertai dengan *Redis Server* (Kaur dan Geetha, 2020). *Web crawler* ini mampu menghasilkan hasil halaman dengan tingkat kecocokan yang maksimum. Meski hasil yang ditampilkan sangat relevan dengan *keyword* yang dimasukkan, namun dalam pembangunannya, *web crawler* ini sangatlah rumit meski komponennya hampir sama dengan *web crawler* pada umumnya.

Penelitian yang dilakukan Muhammad Fathan menghasilkan *web crawler* yang dapat berjalan dengan baik di situs yang menggunakan *html* versi 4 maupun *html* versi 5. Algoritma *crawler* yang digunakan pada penelitian tersebut adalah algoritma *breadth first search crawling* dan algoritma *modified similarity based crawling* yang dikembangkan oleh Google (Qoriiba, 2021).

Dalam menghasilkan informasi berdasarkan *keyword* yang dimasukkan pengguna pada *search engine*, diperlukan sebuah sistem temu kembali informasi atau yang biasa disebut dengan *Information Retrieval*. Temu kembali informasi ini dimaksudkan untuk menemukan material atau data-data (sekumpulan dokumen) pada penyimpanan yang tak beraturan (biasanya berbentuk teks) di mana menjadi kebutuhan akan sebuah informasi dari kumpulan data berukuran besar (dalam

penyimpanan komputer) (Manning dkk., 2009).

Information retrieval mampu menemukan informasi yang tersimpan dalam sistem sesuai dengan *keyword* tertentu yang dimasukkan pengguna. Teknologi ini mencakup 2 hal, yaitu pembobotan dan perangkingan. Peringkat tertinggi akan diisi oleh data yang paling relevan dengan *keyword* masukan pengguna dan akan terus diurutkan berdasarkan relevansinya.

Google PageRank adalah salah satu algoritma perangkingan situs (*page ranking*) yang berfungsi untuk menentukan situs web mana yang lebih populer. Algoritma ini diperkenalkan oleh Sergey Brin dan Larry Page pada tahun 1998 dalam jurnalnya yang berjudul *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Konsep dari perangkingan *PageRank* adalah semakin banyak situs lain yang memasang *link* ke situsnya, maka situs tersebut akan semakin populer, dengan anggapan bahwa konten situs tersebut lebih bermanfaat daripada konten situs lain.

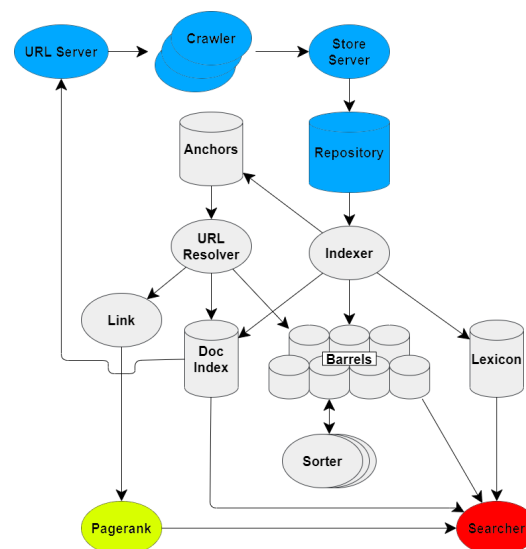
Dalam pengembangan *search engine*, selain perangkingan situs web diperlukan sebuah metode *document ranking*, yaitu untuk memverifikasi apakah konten dalam situs web tersebut relevan atau tidak. Salah satu metode yang paling umum digunakan adalah metode *TF-IDF*. *TF-IDF* atau *Term Frequency-Inversed Document Frequency* adalah suatu metode untuk memberikan bobot kata yang berhubungan terhadap suatu dokumen. Metode ini menggunakan algoritma yang menggabungkan 2 nilai yaitu *Term Frequency (TF)* yang merupakan nilai frekuensi kemunculan kata dan *Inverse Document Frequency (IDF)* yang merupakan nilai untuk mengukur seberapa penting sebuah kata. Dengan menggunakan metode ini pencarian pada *search engine* akan memiliki hasil yang relevan sesuai dengan *keyword* masukan pengguna.

Dari penelitian yang dilakukan oleh Juan Ramos yang berjudul "*Using*

TF-IDF to Determine Word Relevance in Document Queries", terdapat kesimpulan bahwa *TF-IDF* merupakan algoritma sederhana dan efisien yang dapat menghasilkan kumpulan dokumen dengan tingkat kesesuaian yang tinggi sesuai dengan *keyword* yang dimasukkan.

Terdapat beberapa kekurangan dari algoritma *TF-IDF*, salah satunya adalah algoritma ini tidak bisa mengidentifikasi kata di bahasa inggris sesuai dengan *tense*-nya. Sebagai contoh, algoritma ini akan memperlakukan kata "*play*", dan "*plays*" sebagai kata yang berbeda. Namun, hal ini dapat diatasi dengan melakukan proses *stemming* sebelum *TF-IDF*, yaitu mengubah kata yang tadinya menggunakan berbagai bentuk tertentu seperti "*play*", "*plays*", atau "*played*" menjadi sebuah kata dasar "*play*" (Qaiser dan Ali, 2018).

Search engine yang populer saat ini telah memiliki teknik dan algoritma yang lebih maju dan berbeda-beda. Perusahaan-perusahaan besar pemiliknya memiliki tekniknya masing-masing dan enggan membeberkan detail algoritma yang ia pakai dalam mengembangkan *search engine* karena bersifat komersial. Oleh karena itu, pengembangan *search engine* ini menarik untuk dilakukan.



Gambar 1.1: *High Level Google Architecture* (Brin dan Page, 1998)

Penelitian ini akan merancang dan membangun *search engine* berdasarkan arsitekturnya dengan mengintegrasikan penelitian dari Muhammad Fathan yang berfokus pada perancangan *crawler* menggunakan algoritma *modified similarity based crawling* yang digambarkan sesuai arsitektur pada gambar 1.1 dengan arsiran berwarna biru, penelitian Mohammad Riza yang mengimplementasikan algoritma perangkingan *Google PageRank* yang terdapat pada arsiran kuning, dan penelitian Savira Rahmayanti yang mengembangkan sistem pencarian teks menggunakan metode *TF-IDF* yang terdapat pada arsiran merah.

B. Rumusan Masalah

Dari uraian latar belakang di atas, perumusan masalah pada penelitian ini ialah “Bagaimana perancangan arsitektur *search engine* berbasis terminal menggunakan *web crawler*, *page ranking*, dan *document ranking* berbasis *TF IDF*?”

C. Pembatasan Masalah

Pembatasan masalah pada penelitian ini antara lain:

1. Situs yang menjadi titik awal saat proses *crawling* adalah indosport.com dan curiouscuisiniere.com.
2. Proses *crawling* dilakukan dengan jangka waktu 1 hari.
3. Mengkombinasikan metode relevansi pencarian berbasis *TF IDF* dan *Google PageRank*

D. Tujuan Penelitian

1. Membuat *search engine* berbasis terminal yang akan menerima *keyword* pencarian dari pengguna menggunakan *crawler*, *Google PageRank*, dan

pencarian teks berbasis *TF IDF*.

2. Membuat *API* dari *REST search engine* yang akan dibangun.

E. Manfaat Penelitian

1. Bagi penulis

Memperluas pengetahuan tentang *search engine*, menambah pengalaman dalam *programming*, memperoleh gelar sarjana di bidang Ilmu Komputer, serta menjadi media untuk penulis dalam mengaplikasikan ilmu yang didapatkan dari kampus.

2. Bagi Program Studi Ilmu Komputer

Penelitian ini dapat menjadi pembuka untuk penelitian di masa depan, dan dapat memberikan panduan bagi mahasiswa program studi Ilmu Komputer tentang rancang bangun aplikasi *search engine*.

3. Bagi Universitas Negeri Jakarta

Menjadi evaluasi akademik program studi Ilmu Komputer dalam penyusunan skripsi sehingga dapat meningkatkan kualitas pendidikan dan lulusan program studi Ilmu Komputer di Universitas Negeri Jakarta.

DAFTAR PUSTAKA

- Brin, S. dan Page, L. (1998). The anatomy of a large-scale hypertextual web search engine.
- Eichmann, D. (1994). The rbse spider - balancing effective search against web load. *Proceedings of the first World Wide Web Conference*.
- Kaur, S. dan Geetha, G. (2020). Simhar-smart distributed web crawler for the hidden web using sim+ hash and redis server. *IEEE Access*, 8:117582–117592.
- Manning, C., Raghavan, P., dan Schutze, H. (2009). An introduction to information retrieval (online edition). *Cambridge: Cambridge University Press*.
- Qaiser, S. dan Ali, R. (2018). Text mining: Use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications* (pp. 25-29).
- Qoriiba, Muhammad, F. (2021). Perancangan crawler sebagai pendukung pada search engine.
- Seymour, T., Frantsvog, D., dan Kumar, S. (2011). History of search engines. *International Journal of Management & Information Systems (IJMIS)*, 15(4):47–58.