

BAB I

PENDAHULUAN

1.1 Latar Belakang

Semakin berkembangnya jaman, teknologi di dunia ini mengalami kemajuan yang sangat pesat. Salah satunya perkembangan di dunia maya tentang kebutuhan informasi, dengan kebutuhan informasi yang meningkat menyebabkan peningkatan jumlah dokumen *digital* yang sangat pesat. Karena jumlah penyedia dokumen *digital* yang sangat banyak, dengan fakta tersebut maka seorang pengguna kesulitan dalam pencarian dan penemuan kembali dokumen *digital*.

Salah satu solusinya adalah dengan membangun sebuah daftar indeks dokumen digital. *Digital library* menggunakan *web crawler* adalah salah satu cara memanfaatkan daftar indeks, *web Crawler* merupakan program yang mengumpulkan informasi yang hasilnya akan disimpan pada sebuah *database*. Sebuah *web crawler* akan berjalan menelusuri halaman web dan mengumpulkan dokumen-dokumen kemudian melakukan proses *download* data yang telah dikumpulkan oleh *web crawler*.

Setelah data didapatkan proses berikutnya adalah melakukan proses *preprocessing* bertujuan untuk menghilangkan *noise* data atau data yang tidak diinginkan diantaranya dengan melakukan proses *stemming* dan *tokenaizing*, setelah data dihilangkan *noise* data proses berikutnya yaitu melakukan proses pembobotan dengan menggunakan metode *cosin similarity* setelah data di bobotkan proses selanjutnya adalah pengklasifikasian data menggunakan metode *K-Nearest Neighbor (KNN)*, *K-Nearest Neighbor* merupakan metode klasifikasi data berdasarkan *voting* terbanyak dari perhitungan antara *data testing* dan *data training*.

1.2 Perumusan Masalah

Permasalahan yang diangkat dalam tugas akhir ini adalah:

1. Banyaknya jumlah dokumen *digital* membuat *user* kesulitan dalam pencarian kembali dokumen.
2. Bagaimana mengklasifikasikan data *electronic book* dalam *digital library* menggunakan *K-Nearest Neighbor*.

1.3 Tujuan dan Manfaat

Tujuan dari dibuatnya tugas akhir ini adalah :

1. Menerapkan *information retrieval* untuk mempermudah pengguna mendapatkan informasi *electronic book*.
2. Meng-implementasikan perhitungan kemiripan data dengan menggunakan *data training* dengan *data testing* untuk mengklasifikasikan data menggunakan *K-Nearest Neighbor* ke dalam *Digital Library*.

Adapun manfaat Dengan adanya sistem *Digital library* ini diharapkan dapat memudahkan *user* untuk mencari *referensi* buku maupun *referensi* lainya untuk menunjang proses pembelajaran.

1.4 Batasan Masalah

Untuk menghindari pembahasan yang terlalu melebar, masalah yang dibahas pada Tugas Akhir ini dibatasi pada :

1. Pengumpulan data *electronic book* menggunakan *Web Crawler*.
2. Website penyedia dokumen *electronic book* yang akan di melakukan proses *crawling* yaitu :
<http://duniadownload.com/>, <http://www.ilmukomputer.com/>
<http://www.anesularnaga.com/>, <http://buku-hitam.ikhisan.my.id>, <http://www.ebook.mampirbro.web.id>.
3. *Electronic book* yang dimaksud adalah yang memiliki format file zip, rar, doc, docx, exe, dan pdf maupun link yang mengarah ke file tersebut.
4. Buku atau *electronic book* yang dicari dari halaman web yang berbahasa indonesia.
5. Menggunakan metode perhitungan *TF-IDF* untuk pembobotan dan *cosine similarity* untuk mengukur kemiripan dokumen *electronic book* kemudian akan dibandingkan antara *data training* dan *data testing* menggunakan metode *K-nearest Neighbor* untuk melakukan proses klasifikasi.
6. Visualisasi aplikasi ini dapat ditampilkan di web.

1.5 Metodologi Penelitian

Metode yang digunakan untuk menyelesaikan tugas akhir ini adalah:

1. Studi Literatur

Studi literatur ini didapatkan melalui buku, jurnal, internet, dan sumber referensi lainnya yang terkait dengan: Teori tentang *web crawler cosine similarity* dan *k-nearest neighbor*.

2. Pengumpulan Data

Data yang dikumpulkan dengan cara *crawling* pada website penyedia *elektronik book*.

3. Perancangan Sistem

Perancangan sistem ini berupa *flowchart diagram* dan *desain database*.

4. Implementasi Sistem

Implementasi dilakukan dengan membuat sistem yang sesuai dengan rancangan sistem.

5. Evaluasi

Evaluasi dilakukan dengan menguji sistem yang telah dibuat dan melakukan perbaikan apabila ada kesalahan.

6. Penyusunan Laporan Tugas Akhir

Penyusunan dilakukan dengan membuat laporan tentang tugas akhir yang dibuat.

1.6 Sistematika Penulisan

Sistematika pembahasan yang digunakan untuk menyelesaikan Tugas Akhir adalah sebagai berikut:

BAB I : PENDAHULUAN

Menjelaskan mengenai latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Membahas beberapa teori penunjang yang berhubungan dengan pokok pembahasan dalam Tugas Akhir ini yang secara garis besar berisi tentang *Information Retrieval*, *Preprocessing*, *Web Crawler*, *Cosine Similarity*, *K-Nearest Neighbor*, *Digital Library*.

BAB III : RANCANGAN SISTEM

Dalam bab ini akan dijelaskan rinci rancangan sistem dari aplikasi tugas akhir ini.

BAB IV : IMPLEMENTASI SISTEM

Pada bab ini akan dijelaskan mengenai implementasi pembuatan aplikasi berdasarkan hasil rancangan yang telah dibuat sebelumnya.

BAB V : PENUTUP

Pada bab ini akan dijelaskan mengenai kesimpulan hasil akhir dari proses pembelajaran dan uji coba sistem serta berisi saran atas pembuatan tugas akhir ini.