



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penerapan teknologi informasi saat ini telah menyebar hampir di semua bidang tidak terkecuali di institusi pendidikan. Institusi pendidikan sebagai institusi yang selain memberikan pengajaran juga memiliki koleksi karya ilmiah ataupun buku didalam perpustakaan. Perkembangan dari penerapan teknologi informasi bisa kita lihat dari perkembangan teknologi yang digunakan perpustakaan pada institusi pendidikan yang berkaitan dengan teknologi informasi, seperti *digital collection*.

Digital collection adalah bahan koleksi dalam bentuk *digital* terdiri dari foto-foto digital, manuskrip, peta, rekaman suara, film, dan buku (Library of Congress, 1994). Pembuatan *digital collection* dimaksudkan untuk memudahkan anggota perpustakaan mengakses perpustakaan dan sekaligus juga memancing minat baca. Dengan adanya *digital collection*, anggota perpustakaan tidak harus pergi ke perpustakaan untuk memperoleh bacaan. Bahan koleksi perpustakaan yang dapat dijadikan koleksi *digital collection* yaitu berbagai jenis *e-book*.

Dengan beragam *e-book* dalam *digital collection*, pembaca akan informasi lebih mudah sesuai dengan topik yang dibutuhkan. Namun dari perkembangan *digital collection* tersebut terdapat permasalahan dalam pengklasifikasian (Herwansyah, 2009). Pengklasifikasian yang tidak baik dan tidak teratur akan membuat struktur dari kategori *e-book* dalam suatu *digital collection* menjadi

tidak rapi. Serta jika pengklasifikasian dilakukan secara manual oleh manusia akan menyebabkan kesalahan dalam proses pengklasifikasian kategori jenis *e-book* tersebut.

Sebelumnya telah dikembangkan “Aplikasi Pengkategorian Dokumen dan Pengukuran Tingkat Similaritas Dokumen Menggunakan Kata Kunci pada Dokumen Penulisan Ilmiah Universitas Gunadarma” dengan algoritma *tf-idf* dan *vector space model* (Herwansyah, 2009), namun aplikasi tersebut dirasa kurang akurat dalam mengkategorikan dokumen dimana keakuratan klasifikasi sekitar 80%. Salah satu solusi yang dapat ditambahkan adalah dengan penambahan proses *stemming* dalam diharapkan dapat meningkatkan keakuratan klasifikasi.

Untuk mengklasifikasi sebuah koleksi secara otomatis, dibutuhkan sebuah standar yang banyak digunakan dalam klasifikasi buku. Salah satu standar klasifikasi yang banyak digunakan oleh perpustakaan dunia adalah *Library of Congress*. Standar *Library of Congress* ini yang akan dijadikan struktur kategori standar dalam pengembangan aplikasi. Penggunaan *Library of Congress* digunakan karena dapat diimplementasikan dengan baik pada perpustakaan yang memiliki jumlah koleksi yang besar (Library of Congress, 1994).

Berangkat dari permasalahan yang ada, maka akan dilaksanakan penelitian yang bertujuan mengimplementasikan algoritma *tf-idf* dan *vector space model* untuk klasifikasi *e-book* berbasis *library of congress*” dengan penambahan proses *stemming*.

1.2 Rumusan Masalah

Masalah yang dapat dirumuskan dalam penelitian ini adalah

- a. Bagaimana data *e-book* dapat diklasifikasikan dengan baik dan akurat?
- b. Bagaimana membangun aplikasi pada komputer khususnya berbasis *Web* yang dapat memberikan fasilitas untuk melakukan proses klasifikasi jenis suatu data *e-book* sehingga *e-book* tersebut dapat dengan mudah ditemukan dengan hanya melihat klasifikasi jenisnya?
- c. Bagaimana teknologi klasifikasi *e-book* dimanfaatkan untuk membangun suatu aplikasi yang tidak hanya dapat mengklasifikasikan *e-book* dengan akurat tetapi juga dapat mengklasifikasikan *e-book* secara otomatis dan cepat tanpa harus diklasifikasikan oleh manusia?
- d. Bagaimana penambahan proses *stemming* akan mempengaruhi hasil akhir dari klasifikasi?

1.3 Batasan Masalah

Ada pun batasan yang harus ditentukan dalam melakukan penelitian ini.

- a. Data *digital* yang akan dilakukan proses klasifikasi adalah *e-book* berbahasa inggris.
- b. Data *digital* yang akan dilakukan proses klasifikasi memiliki format PDF, doc, docx, dan TXT.
- c. Aplikasi dibuat menggunakan bahasa pemrograman PHP.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah

- a. Mengimplementasikan algoritma TF-IDF dan *Vector Space Model* untuk aplikasi pengklasifikasian *e-book*.
- b. Membuktikan keakuratan antara klasifikasi yang dilakukan oleh aplikasi dengan klasifikasi standar *Library of Congress*.

1.5 Manfaat Penelitian

Diharapkan penelitian ini memberikan manfaat sebagai berikut.

- a. Mempermudah dalam proses pencarian suatu *e-book* yang terdapat didalam suatu aplikasi.
- b. Pengklasifikasian jenis *e-book* dapat dilakukan secara otomatis dan cepat tanpa harus dilakukan secara manual.
- c. Meminimalisasi kesalahan dalam proses pengklasifikasian jenis *e-book*.

UMMN