

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan teknologi informasi, terutama internet, telah mempermudah penyebaran dokumen dalam bentuk digital. Akibatnya, jumlah dokumen yang dapat ditemukan di internet sangatlah banyak, dengan berbagai topik dan isi yang beragam. Tingginya jumlah dokumen digital tersebut dapat menimbulkan berbagai macam permasalahan. Salah satu permasalahan adalah seringkali dokumen yang ada merupakan dokumen yang serupa namun berbeda versinya. Hal ini karena adanya kebiasaan untuk menyimpan dokumen dengan versi yang berbeda-beda, bukan hanya versi terbaru. Selain itu, penyalinan dokumen dari satu tempat ke tempat lain juga semakin memperbanyak versi dokumen yang ada (Hoad, 2003).

Permasalahan berikutnya adalah sulitnya melakukan pengelompokan dokumen. Pengelompokan diperlukan untuk menempatkan dokumen-dokumen ke dalam kategori yang sesuai sehingga memudahkan pencarian. Banyaknya jumlah dokumen mengakibatkan pengelompokan menjadi proses yang sulit dan sangat menghabiskan waktu (Kurniawati dan Wicaksana, 2008). Permasalahan terakhir adalah mudahnya melakukan plagiarisme. Plagiarisme merupakan tindakan menjiplak, menyalin, bahkan menjadikan karya ilmiah orang lain seolah-olah menjadi miliknya (Khairunnisa dkk, 2012). Seringkali dokumen yang ditemukan di internet disalin, namun tanpa menyertakan referensi secara jelas dan benar sehingga timbul masalah plagiarisme.

Salah satu solusi yang dapat digunakan dalam berbagai permasalahan tersebut adalah melalui pengukuran kesamaan dokumen (Hatzivassiloglou dkk, 1999). Pengukuran kesamaan dapat digunakan untuk menentukan apakah dokumen yang dibandingkan merupakan dokumen yang sama (hanya berbeda versi), atau sama sekali berbeda. Tingkat kesamaan dokumen juga dapat

digunakan untuk mengelompokkan dokumen, di mana dokumen yang memiliki tingkat kesamaan tinggi berada pada satu kelompok. Pada pendeteksian plagiarisme, pengukuran kesamaan dilakukan dengan asumsi bahwa semakin tinggi tingkat kesamaan di antara dokumen yang dibandingkan, semakin besar kemungkinan bahwa salah satu dokumen merupakan hasil plagiat (Clough, 2003).

Pengukuran kesamaan dokumen dilakukan dengan membandingkan suatu dokumen dengan dokumen lainnya. Hal ini tentu akan sulit untuk dilakukan secara manual, terutama jika dokumen yang dibandingkan berjumlah banyak, sehingga dibutuhkan suatu aplikasi yang dapat membandingkan dokumen dan mengukur tingkat kesamaannya secara otomatis. Penggunaan aplikasi ini diharapkan dapat memudahkan dan mempercepat proses perbandingan dokumen.

## **1.2 Perumusan Masalah**

Bagaimana mengimplementasikan algoritma TF-IDF pada pengukuran kesamaan dokumen?

## **1.3 Batasan Masalah**

1. Dokumen yang dibandingkan berupa dokumen laporan skripsi dari tiga jurusan di Sekolah Tinggi Teknik Musi, yaitu jurusan teknik industri, teknik informatika dan sistem informasi.
2. Sistem pengukur kesamaan dokumen akan melakukan perbandingan dokumen dengan membandingkan pasangan dokumen yang dipilih pada kumpulan dokumen di setiap jurusan.

## **1.4 Tujuan Penelitian**

Tujuan penelitian ini adalah membangun aplikasi pengukuran kesamaan dokumen yang mengimplementasikan algoritma TF-IDF.

## **1.5 Manfaat Penelitian**

Manfaat penelitian ini adalah untuk mempermudah dalam mengukur tingkat kesamaan dokumen.

## 1.6 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini adalah model *waterfall*. Tahap-tahap dalam model ini adalah sebagai berikut (Sommerville, 2007).

### 1. *Requirement analysis and definition*

Pada tahap ini dilakukan analisa kebutuhan sistem, yang mencakup layanan, batasan, dan tujuan sistem. Kebutuhan ini selanjutnya akan menjadi spesifikasi sistem. Pada tahap ini juga dilakukan pengumpulan data-data yang diperlukan untuk penelitian. Data-data tersebut berupa *file* laporan skripsi dari jurusan teknik industri, teknik informatika, dan sistem informasi STT Musi.

### 2. *System and software design*

Pada *system design*, kebutuhan yang ada akan dipilah untuk kemudian dibentuk arsitektur sistem secara keseluruhan. *Software design* berkaitan dengan pengidentifikasian dan penjabaran sistem *software* dan hubungan antar komponennya. Tahapan ini akan dilakukan dengan menggunakan UML.

### 3. *Implementation and unit testing*

Pada fase ini, hasil dari fase desain akan diimplementasikan pada bahasa pemrograman Java dengan Netbeans sebagai lingkungan pengembangannya. Setiap unit yang telah diimplementasikan akan diverifikasi agar sesuai dengan spesifikasi yang ada dengan menggunakan metode *whitebox testing*.

### 4. *Integration and system testing*

Unit-unit program akan diintegrasikan menjadi satu dan kemudian diuji sebagai satu sistem utuh untuk memastikan kebutuhan sistem telah terpenuhi. Pengujian yang dilakukan pada fase ini menggunakan metode *blackbox testing*. Metode ini digunakan untuk menguji kesesuaian antara *input* dan *output* dari aplikasi.

### 5. *Operation and maintenance*

Fase ini berkaitan dengan pengoperasian sistem dan perbaikan sistem secara berkala. Pada penelitian ini, fase *maintenance* tidak akan dilakukan.

## **1.7 Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut.

### **BAB I PENDAHULUAN**

Pada bab awal ini akan dijelaskan mengenai latar belakang dilakukannya penelitian, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Bab ini berisikan landasan teori dan perbandingan literatur yang akan digunakan sebagai landasan dan acuan dalam penelitian.

### **BAB III ANALISIS DAN PERANCANGAN SISTEM**

Bab ini berisikan analisis dan desain sistem. Analisis sistem dilakukan terhadap algoritma TF-IDF. Desain sistem digambarkan dengan menggunakan *use case diagram*, *sequence diagram*, *class diagram*, dan *activity diagram*. Bab ini juga berisikan desain antarmuka sistem.

### **BAB IV IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisikan proses implementasi dan tampilan program hasil perancangan, serta hasil pengujian program. Pengujian dilakukan dengan menggunakan *whitebox testing* dan *blackbox testing*.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini berisikan kesimpulan dan saran yang diperoleh setelah pelaksanaan penelitian.