

## **BAB 1**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Berdasarkan UUD 1945 Pasal 36 Bahasa Indonesia sebagai bahasa nasional yang berfungsi sebagai alat komunikasi formal di negara Indonesia. Selain sebagai alat komunikasi, Bahasa Indonesia juga digunakan sebagai pengantar ilmu pengetahuan.

Di zaman sekarang teknologi sudah berkembang dengan pesat sehingga penyebaran informasi juga semakin mudah untuk diakses. Informasi tidak hanya mencakup domestik namun juga informasi yang lebih luas. Penyebaran informasi yang luas ini memiliki dampak penggunaan kata asing dalam penyebaran informasi domestik. Penelitian dan karya ilmiah mengenai bidang komputer dan teknologi informasi mengalami sedikit kendala dalam menggunakan Bahasa yang baik dan benar dengan masih banyaknya penggunaan istilah asing. Sebelum tahun 2001, beberapa kata dalam bidang komputer dan teknologi informasi masih dalam bentuk asing dan belum memiliki padanan dalam Bahasa Indonesia.

Pada tahun 2001, pemerintah menerbitkan suatu regulasi untuk pembakuan peristilahan perkomputeran dalam Bahasa Indonesia melalui Instruksi Presiden No. 2 Tahun 2001. Awalnya instruksi presiden ini merupakan himbuan untuk perancangan aplikasi komputer berbahasa Indonesia namun pada perkembangannya padanan kata berbahasa Indonesia bidang komputer dan teknologi informasi ini juga digunakan untuk penulisan karya ilmiah untuk kalangan akademisi. Bersama dengan instruksi tersebut, pemerintah juga mengeluarkan Senarai Padanan Istilah (SPI) yang berisi daftar substitusi Bahasa Indonesia untuk menggantikan istilah asing pada bidang komputer. Sesuai dengan isi regulasi tersebut, salah satu kementerian yang wajib melaksanakan regulasi ini adalah Kementerian Negara Riset dan Teknologi.

Namun, sampai saat ini belum ada mekanisme untuk mengevaluasi penggunaan substitusi Bahasa Indonesia bidang komputer pada karya ilmiah akademisi di Indonesia sekarang tidak diketahui apakah regulasi ini berjalan baik atau tidak implementasinya.

Pada penelitian ini akan dilakukan pengevaluasian penggunaan kata baku bahasa indonesia dan istilah asing dengan menggunakan algoritma string

matching *Boyer-Moore* pada karya ilmiah akademisi di Indonesia. Beberapa tahapan seperti *file harvesting* pada OAI-PMH dan selanjutnya dilakukan proses *Text Pre-processing* yang akhirnya dijalankan algoritma *Boyer-Moore* yang akan melakukan pencocokan kata pada dokumen dengan kata baku dan kata asing pada kamus SPI yang berisi 629 kata baku dan istilah asing bidang ilmu komputer.

Penelitian mengenai Algoritma *Boyer-Moore* telah banyak dilakukan, antara lain oleh (Sheshasayee & Thailambal, 2015) yang melakukan perbandingan antara algoritma *Boyer-Moore*, *Knuth Moris Pratt* dan *Naïve String* dalam text mining yang diimplementasikan didalam Bahasa pemrograman Python untuk membandingkan waktu eksekusi ketiga algoritma berdasarkan perbedaan panjang teks dan pola teks. Didapatkan hasil pada percobaan panjang teks dengan parameter 182 dan 475, *Boyer-Moore* menjadi algoritma dengan hasil kedua tercepat pada kedua parameter. Selain itu, Yuan (2011) melakukan penelitian pada pengutipan informasi Chinese dengan menggunakan *Boyer-Moore* dengan tujuan untuk meminimalkan waktu untuk mencocokkan string dan membandingkan. Hasil yang didapatkan dengan menggunakan *Boyer-Moore* cukup baik

Dalam penelitian yang dilakukan (Jaiswal, 2014) menggunakan algoritma *Boyer-Moore* pada *Multicore GPU* untuk *Network Security*. Dilakukan lah perbandingan dengan *serial version* dan *multithreaded version* pada CPU dan didapatkan hasil 10 kali lebih cepat dibandingkan dengan versi CPU dan 9 kali lebih cepat dibandingkan versi *multithreaded*. Selanjutnya pada penelitian (Choudhary *et. al*, 2012) dilakukan berbagai macam bentuk peningkatan algoritma *Boyer-Moore* dengan hasil cukup baik.

Dilakukan penelitian mengenai *Harvesting Reliability Data from the Internet* (Dussault, *et al*. 2008) yang mana mempanen data-data dari banyak sumber. Yang mana dilakukan panen data yang didapatkan *Portable Document Format (PDF)*, memparsing laman dengan benar, termasuk tanda baca dan format angka, dan untuk mengekstrak data dari table sangatlah penting untuk pengumpulan data yang baik. Dengan hasil memuaskan yang didapat dari *parse* laman web dengan baik dan benar, lalu pengambilan pdf, dan kemampuan untuk mengarsip laman web yang dipilih digunakan sebagai sumber data andalan. Penelitian mengenai *metadata harvesting* melalui OAI-PMH juga pernah dilakukan oleh (Simek P, *et al*. 2011) dimana mereka melakukan *harvesting* jurnal internasional *agris on-line* Ekonomi dan Informatika yang diterbitkan oleh Fakultas

Ekonomi dan Manajemen di Universitas Ceko *Life Science Prague (CULS)*. *Metadata format* yang digunakan adalah *Dublin core (dc)* dengan hasil yang cukup baik.

Berdasarkan penelitian terdahulu diatas, penulis akan membangun sebuah sistem untuk melakukan evaluasi rasio penerapan kata substitusi Bahasa Indonesia bidang komputer dengan algoritma *Boyer-Moore* dengan tahapan awal mengekstraksi karya ilmiah yang dihasilkan direpositori beberapa universitas, file yang diekstraksi dalam bentuk pdf, untuk lebih memudahkan ekstraksi, pdf dikonversi dulu menjadi bentuk *.txt*. lalu dilakukan beberapa metode *text pre-processing (tokenization, stop word removal, case folding dan stemming)* untuk meningkatkan akurasi dan kecepatan *string matching* yang akan menggunakan algoritma *Boyer-Moore*.

## **1.2 Rumusan Masalah**

Penerbitan regulasi pemerintah Republik Indonesia yaitu Inpres no. 2 tahun 2001 mengenai pembakuan istilah komputer ke dalam Bahasa Indonesia, wajib dilaksanakan oleh seluruh lapisan masyarakat dan pemerintahan. Himbauan ini seperti yang tercantum dalam instruksi presiden terutama harus dilaksanakan oleh Menteri Negara Riset dan Teknologi dan Menteri Pendidikan Nasional. Oleh karena itu, untuk melihat keberhasilan dari regulasi ini perlu dilakukan mekanisme untuk mengecek karya ilmiah akademisi yang bernaung dibawah kementerian yang dimaksud.

## **1.3 Batasan Masalah**

Dalam melakukan penelitian ini, peneliti membatasi ruang masalah yang akan diteliti.

Batasan-batasan masalah yang digunakan adalah :

1. Penelitian dilakukan untuk repositori universitas negeri di Indonesia
2. Hanya repositori yang dapat diakses melalui OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting* ).
3. Hanya mengobservasi tulisan ilmiah untuk bidang ilmu komputer
4. Tidak menggunakan PDF yang berbentuk *Scanned Image*

#### 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah

1. Mengevaluasi tingkat penggunaan istilah asing bidang ilmu komputer pada karya ilmiah akademisi universitas di Indonesia dengan menggunakan algoritma *Boyer-Moore*.

#### 1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Sebagai bahan evaluasi untuk Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan Indonesia.
2. Dapat menjadi pilihan acuan untuk penelitian terkait *string matching* algoritma *boyer-moore*.
3. Dapat menjadi acuan untuk penelitian mengenai *metadata harvesting* pada OAI-PMH.
4. Hasil evaluasi dapat menjadi masukan untuk pihak terkait seperti Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan Indonesia.

#### 1.6 Metode Penelitian

Tahapan-tahapan yang akan dilakukan dalam pelaksanaan penelitian ini adalah sebagai berikut :

1. Studi Literatur

Tahap ini dilakukan untuk mengumpulkan dan mempelajari informasi – informasi yang diperoleh dari buku, jurnal dan berbagai sumber referensi lain yang berkaitan dengan penelitian seperti *Harvesting Metadata*, *Optical Character Recognition*, *Tokenizing*, *Stop Word Removal*, *Case Folding* algoritma *Boyer-Moore*.

## 2. Analisis Permasalahan Sistem

Pada tahap ini dilakukan analisis terhadap berbagai informasi yang telah diperoleh dari berbagai sumber yang terkait dengan penelitian agar didapatkan metode yang tepat untuk menyelesaikan masalah dalam permasalahan ini.

## 3. Perancangan Sistem

Tahap ini dilakukan perancangan sistem untuk menyelesaikan permasalahan yang terdapat didalam tahap analisis. Kemudian dilanjutkan dengan mengimplementasikan hasil analisis dan perancangan ke dalam sistem.

## 4. Implementasi dan Pengujian

Pada tahap ini dilakukan implementasi kedalam kode sesuai dengan analisis dan perancangan yang telah dilakukan pada tahap sebelumnya dan dilakukan pengujian terhadap hasil yang didapatkan melalui implementasi algoritma *Boyer-Moore* terhadap data yang sudah di dapatkan melalui tahap *Harvesting*.

## 5. Analisis dan Pengambilan Kesimpulan

Pada tahap ini dilakukan analisis data yang didapatkan dari implementasi algoritma *Boyer-Moore* terhadap data yang sudah di dapatkan melalui tahap *Harvesting* dan menyimpulkan hasil analisis tersebut.

### 1.7 Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri atas lima bagian utama sebagai berikut :

#### **Bab 1 : Pendahuluan**

Bab ini berisi latar belakang dari penelitian yang dilaksanakan, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, metodologi penelitian, serta sistematika penulisan.

#### **Bab 2 : Landasan Teori**

Bab ini berisi teori-teori yang diperlukan untuk memahami permasalahan yang dibahas pada penelitian ini. Teori-teori yang berhubungan dengan *Metadata Harvesting*, *Optical Character Recognition*, *Tokenizing*, *Stop Word Removal*, *Case Folding* dan *Stemming* dan algoritma *Boyer-Moore* akan dibahas pada bab ini.

**Bab 3 : Analisis dan Perancangan**

Bab ini menjelaskan tentang analisis dan perancangan sistem yang bangun untuk Pencarian kata baku dan tidak baku istilah bidang komputer pada karya ilmiah akademisi yang diperoleh dari *Harvesting Metadata* dan algoritma *Boyer-Moore*. Adapun dua tahapan yang dibahas pada bab ini yaitu tahap analisis dan perancangan sistem. Pada analisis sistem meliputi kebutuhan perangkat lunak dan perangkat keras dan pada perancangan sistem meliputi tahapan untuk perancangan sistem terdistribusi dan juga tahapan percobaan yang dilakukan.

**Bab 4 : Implementasi dan Pengujian**

Bab ini berisi pembahasan tentang implementasi dari perancangan yang telah dijabarkan pada bab 3. Selain itu, hasil yang didapatkan selama proses yang terjadi pada penelitian juga dijabarkan pada bab ini.

**Bab 5 : Kesimpulan dan Saran**

Bab ini berisi ringkasan serta kesimpulan dari rancangan yang telah dibahas pada bab 3, serta hasil penelitian yang dijabarkan pada bab 4, serta pada bagian akhir bab ini akan berisi saran-saran yang diajukan penulis untuk pengembangan penelitian selanjutnya.