

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kata kunci adalah kata yang mewakili konsep atau garis besar dari suatu dokumen [1]. Kata kunci digunakan sebagai pernyataan informasi penting mengenai isi pembahasan dan ide dalam dokumen. Hal tersebut dapat membantu pembaca dalam mencari informasi-informasi yang diinginkan dari kumpulan dokumen dengan lebih efektif. Dengan adanya kata kunci, pembaca akan dibantu dalam memutuskan untuk membaca dokumen tersebut atau tidak sesuai dengan kebutuhan. Pembaca cenderung akan melihat kata kunci dari dokumen terlebih dahulu sebelum memutuskan untuk membaca keseluruhan isi dari dokumen atau tidak.

Kata kunci dapat dibentuk dari satu atau lebih kata. Gabungan dari satu atau lebih kata disebut dengan frasa. Dalam Kamus Besar Bahasa Indonesia (KBBI) frasa adalah gabungan dua kata atau lebih yang bersifat nonpredikatif (tidak berkaitan pada predikat) [2]. Kata kunci yang terdiri dari dua atau lebih kata disebut dengan frasa kunci atau *keyphrase*.

Meskipun artikel ilmiah yang telah dipublikasikan dalam jurnal umumnya terdapat beberapa frasa kunci, terdapat artikel ilmiah tanpa frasa kunci [3]. Frasa kunci yang terdapat pada dokumen publikasi ilmiah sering kali terbatas sehingga tidak dapat merepresentasikan isi dari dokumen [3]. Terkadang terdapat frasa kunci yang tidak sesuai dengan dengan isi dokumen, sebagai contoh publikasi ilmiah mengenai jaringan komputer tetapi terdapat kata kunci sistem informasi.

Dokumen yang tidak memiliki frasa kunci mengakibatkan sulitnya menemukan informasi yang relevan pada dokumen, lamanya waktu untuk mencari dokumen, sulitnya melakukan klasifikasi dokumen dan kesalahan pemahaman pembaca mengenai isi pembahasan dokumen. Untuk itu frasa kunci menjadi hal yang penting dalam dokumen.

Frasa kunci yang tidak relevan terhadap isi dokumen berimbas pada hasil pencarian dokumen yang tidak relevan atau tidak sesuai dengan keinginan. Kesalahan pemberian frasa kunci sering kali disebabkan oleh kesalahan penulis.

Kesalahan pemberian frasa kunci pada dokumen disebabkan karena terkadang bagi sebagian orang untuk menentukan frasa kunci yang baik dan relevan terhadap dokumen adalah hal yang sulit. Pengetahuan penulis yang minim dan kesalahan penulisan dapat menjadi faktor dalam kesalahan pemberian frasa kunci. Semua hal tersebut berlaku pula pada publikasi ilmiah.

Berdasarkan permasalahan-permasalahan yang telah diuraikan, maka pada Tugas Akhir ini akan dibangun sistem ekstraksi *keyphrase* terhadap publikasi ilmiah. Penelitian ini difokuskan pada publikasi ilmiah yang ditulis memakai bahasa Inggris.

Dalam pembangunan sistem ekstraksi frasa kunci, akan digunakan fitur sentimen dan *deep learning* yang akan membedakan dengan sistem sejenis. Fitur sentimen digunakan dengan alasan bahwa penulis publikasi ilmiah cenderung untuk mengeksploitasi kelebihan dari metodologi yang dibahas. *Deep learning* akan digunakan dengan alasan algoritma pembelajaran tersebut dapat memodelkan data yang bersifat kompleks sehingga dapat meningkatkan akurasi model pembelajaran.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan pada bagian 1.1., maka dapat diambil rumusan masalah sebagai berikut:

1. Bagaimana cara mengotomatisasi pemilihan frasa kunci?
2. Bagaimana meningkatkan akurasi model pembelajaran ekstraksi frasa kunci?
3. Bagaimana meningkatkan akurasi klasifikasi frasa kunci pada publikasi ilmiah?

1.3 Tujuan Pembahasan

Berdasarkan rumusan masalah yang telah dirumuskan pada bagian 1.2., maka dapat dituliskan tujuan sebagai berikut:

1. Membangun aplikasi yang dapat mengekstraksi frasa kunci secara otomatis dari publikasi ilmiah berekstensi *pdf*.
2. Melibatkan analisa sentimen sebagai salah satu fitur model pembelajaran dengan mengasumsikan bahwa penulis publikasi ilmiah cenderung mengeksploitasi kelebihan dari metodologinya dibanding kekurangannya.

3. Menerapkan *deep learning* untuk mengklasifikasi frasa kunci dengan mempertimbangkan bahwa metode tersebut umumnya menghasilkan akurasi lebih tinggi dibandingkan algoritma klasifikasi standar.

1.4 Ruang Lingkup

Ruang lingkup yang dirumuskan dari Tugas Akhir adalah:

1. Dokumen hanya terfokus pada publikasi ilmiah berupa *paper* dalam bidang ilmu komputer.
2. Bahasa dalam publikasi ilmiah hanya terfokus pada bahasa Inggris.
3. Pendekatan yang akan diteliti adalah pendekatan statistik dan pendekatan pembelajaran.
4. Masukan program dibatasi hanya dalam bentuk *softcopy* dengan ekstensi *pdf*.

1.5 Sumber Data

Sumber data yang digunakan dalam Tugas Akhir ini adalah sebagai berikut:

1. Sumber data primer
Data diperoleh dari koleksi publikasi ilmiah milik National University of Singapore (NUS) terkait publikasi ilmiah yang berjudul “*Keyphrase Extraction in Scientific Publications*” [4].
2. Sumber data sekunder
Data diperoleh dengan melakukan survei literatur/studi kepustakaan, berupa buku teks dan referensi ilmiah lain seperti *paper* yang memuat materi yang berhubungan dengan penelitian.

1.6 Sistematika Penyajian

Sistematika pembahasan penyusunan Tugas Akhir direncanakan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang diperlukannya sistem ekstraksi *keyphrase* terhadap publikasi ilmiah, rumusan masalah berdasarkan latar

belakang, rumusan tujuan sesuai dengan rumusan masalah yang telah dibuat, ruang lingkup kajian penelitian, serta sistematika pembahasan dari Tugas Akhir.

BAB II KAJIAN TEORI

Pada bab ini akan diuraikan teori-teori berkaitan yang akan dipakai dalam pembuatan sistem ekstraksi *keyphrase*. Teori-teori diambil dari berbagai buku teks maupun publikasi ilmiah. Pada akhir bab ini akan dijelaskan tinjauan pustaka mengenai penelitian terkait sesuai dengan topik Tugas Akhir. Akan dijelaskan pula kontribusi penelitian yang menjadi pembeda dari penelitian sebelumnya.

BAB III ANALISIS DAN RANCANGAN SISTEM

Pada bab ini akan dijelaskan mengenai rancangan dan cara kerja sistem ekstraksi *keyphrase*. Akan diuraikan pula desain UML sebagai gambaran akan sistem ekstraksi *keyphrase* yang akan dibuat. UML terdiri dari *use case*, *class diagram*, dan *activity diagram*. Rancangan antarmuka sistem akan dibahas pada bab ini.

BAB IV IMPLEMENTASI

Pada bab ini akan dijelaskan mengenai hasil implementasi akan rancangan sistem yang telah diuraikan sebelumnya. Implementasi yang akan dibahas dibagi menjadi empat bagian, yaitu implementasi *class*, implementasi antarmuka, implementasi algoritma, dan implementasi metode.

BAB V PENGUJIAN

Pada bab ini akan diuraikan mengenai pengujian-pengujian sistem ekstraksi *keyphrase* yang telah dibuat. Pengujian yang akan dilakukan adalah pengujian *blackbox*, pengujian fitur-fitur pembelajaran, pengujian hasil konversi pdf ke teks, pengujian jumlah *neuron* dan *layer*, *benchmarking*, dan pengujian keseluruhan sistem.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari pembahasan pada perancangan serta analisa pengujian sistem yang dibuat untuk perkembangan aplikasi kedepannya.

