

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ekstraksi kata kunci (*keywords extraction*) adalah kumpulan frasa pendek yang menangkap topik utama yang dibahas pada suatu dokumen. Ekstraksi kata kunci pada penelitian ini difokuskan pada ekstraksi kata kunci *website*.

Ekstraksi kata kunci pada *website* seringkali kata kunci yang dimasukan terlalu luas dalam artian kata yang dimasukan terlalu umum dan tidak spesifik serta seringkali kata yang dimasukan tidak sesuai dengan *content website*. Efek yang terjadi ketika kata kunci yang dimasukan tidak sesuai dengan *content website*, maka ketika dalam pencarian sebuah informasi, *website* yang diindex tidak sesuai dengan *content website*. *Website* termasuk pada jenis teks yang tidak terstruktur berbanding terbalik dengan jurnal atau bisa juga disebut dokumen ilmiah yang memiliki bagian - bagian yang jelas seperti judul, abstrak, isi dan kesimpulan[18], sehingga memiliki ketersulitan tersendiri ketika mencari sebuah informasi yang penting pada *website*. Selain teks atau kalimat yang tidak terstruktur, *website* memiliki banyak halaman yang artinya memiliki banyak teks atau kalimat sehingga kemungkinan juga pada halaman – halaman tersebut memiliki kata kunci yang dapat diekstrak.

Meskipun *website* termasuk pada frase atau teks yang tidak terstruktur, akan tetapi dalam ekstraksi kata kunci pada *website* bisa didapatkan melalui pencarian pada *meta tag*[9] seperti seperti *meta title*, *meta content body* dan pada *link - link* terkait dengan *website* tersebut, maka dari itu *keyword extraction* khususnya untuk *website* harus melakukan *filtering*, karena posisi teks pada *website* sangat menentukan betapa pentingnya kata atau kalimat pada suatu *website*[1]. Adapun hal lain yang membuat suatu teks itu penting yaitu betapa seringnya kata itu disebutkan berulang - berulang pada kalimat atau bagian yang lain. Akan tetapi frekuensi kata

bukan menjadi parameter utama, tetapi ada hal lain seperti keterhubungan antar setiap kata dan apakah kata itu sesuai dengan topik atau tidak.

Sistem rekomendasi secara umum adalah teknik dalam perangkat lunak yang bisa memberikan saran - saran untuk item yang sekiranya bermanfaat bagi pengguna. Sistem rekomendasi memerlukan model rekomendasi yang tepat agar yang direkomendasikan sesuai dengan keinginan *user*.

Ada berbagai banyak teknik yang digunakan untuk memberikan rekomendasi ekstraksi kata kunci dari sebuah *website*, baik teknik statistik, semantik, *supervised learning* ataupun *unsupervised learning*. Ekstraksi kata kunci dengan teknik statistik yaitu dengan cara mendeteksi kata kunci hanya dengan frekuensi kata yang muncul[2]. Ekstraksi kata kunci dengan menggunakan teknik semantik yaitu dengan cara membuat relasi antar kata sehingga dapat menghasilkan kata kunci yang memiliki makna sama[6]. Semantik sendiri berarti studi tentang makna bahasa, namun secara umum bisa diartikan sebagai bagian struktur bahasa yang berhubungan dengan makna ungkapan dan juga dengan struktur makna suatu bahasa ataupun struktur penulisan. Pada teknik semantik, data yang digunakan adalah *glove* berisi *dataset* dalam Bahasa Inggris. Masalah *dataset glove* dalam Bahasa Indonesia adalah data *glove*-nya masih sedikit dan terbatas. Ekstraksi kata kunci pada ANN yaitu dengan cara mengesktrak kata kunci berdasarkan data latih [7]. ANN adalah *feedforward* yang dilatih untuk menghasilkan perkiraan identitas antara *input* dan *output* jaringan menggunakan *backpropagation*.

Pada penelitian ini dibuat teknik gabungan antara *semantic relatedness* dan ANN, sehingga diharapkan dengan gabungan dari *semantic relatedness* dan ANN dapat menghasilkan kata kunci yang baik sesuai konteks dari suatu *website* tertentu.

1.2. Identifikasi Masalah

Seringkali kata kunci yang dimasukkan pada *website* tidak akurat, dalam artian kata kunci yang dimasukkan terlalu luas (umum) atau tidak spesifik dan tidak relevan dengan *content* dari *website* yang dibuat.

1.2.1. Rumusan Masalah

Rumusan masalah dalam penelitian ini yaitu :

Bagaimana menghasilkan kata kunci yang baik sesuai dengan *content* dari suatu *website* tertentu?

1.2.2. Batasan Masalah

Adapun batasan masalah yang akan diuraikan dalam penelitian ini yaitu sebagai berikut :

1. Ekstraksi kata kunci hanya terbatas sampai frase.
2. Hanya mengenali salah satu bahasa, yaitu bahasa inggris karena hanya memiliki data latih untuk bahasa inggris.
3. *Dataset* yang digunakan yaitu *Glove* dari Hulth (2003).

1.3. Tujuan dan Manfaat Penelitian

Berikut diuraikan tujuan dan manfaat dari penelitian yang dilakukan, yaitu sebagai berikut :

1.3.1. Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menghasilkan kata kunci yang sesuai dengan *content website*.
2. Merekomendasikan hasil dari ekstraksi kata kunci otomatis untuk pengindeks.
3. Melihat akurasi dari hasil rekomendasi ekstraksi kata kunci.

1.3.2. Manfaat Penelitian

Berdasarkan maksud dan tujuan dari penelitian ini kontribusi yang dapat diberikan adalah :

1. Pada sisi praktis, manfaat penelitian yang dihasilkan adalah mampu melakukan *generate keywords* dari sebuah *website* secara otomatis

- sehingga meringankan pada pemilik *website* dan mampu membantu dalam pemfilteran *website* sesuai dengan konteks pencarian
2. pada sisi akademis, penelitian ini memberikan kontribusi sebagai pelopor untuk penelitian-penelitian selanjutnya.
 3. Pada sisi ilmiah mampu memberikan akurasi yang lebih baik untuk ekstraksi kata kunci pada *website* dengan mengkombinasikan antara *Semantic Relatedness* dan *ANN*.

1.4. Sistematika Penelitian

Laporan penelitian ini disusun dengan sistematika penelitian sebagai berikut :

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang masalah, maksud dan tujuan, premis dan hipotesis serta bagaimana solusi dari masalah yang ada.

BAB II LANDASAN TEORI

Bab ini membahas tentang pengertian dari teknik-teknik yang akan digunakan dalam penelitian.

BAB III METODE PENELITIAN

Bab ini membahas tentang membahas tahap teknik yang diusulkan, tahapan-tahapan proses yang akan dilakukan dan implementasi teknik penggabungan untuk ekstraksi kata kunci.

BAB IV HASIL PENELITIAN

Bab ini membahas tentang hasil penelitian yang telah dilakukan dan bagaimana cara melakukan pengujiannya.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan akhir dan saran untuk perbaikan penelitian selanjutnya.