

TEXT SUMMARIZATION MENGGUNAKAN FUZZY SYSTEM BERDASARKAN TERM FREQUENCY DAN RELATED TREMS

Nyoman Adi Pradnyana Saputra¹, Ade Romadhony², Untari Novia Wisey³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Text Summarization merupakan suatu metode untuk melakukan peringkasan teks secara otomatis agar mendapatkan teks atau bacaan yang lebih pendek atau ringkas. Dengan melakukan peringkasan teks atau bacaan sangat membantu dalam mendapatkan informasi yang dikutip pada suatu teks atau bacaan misalnya teks dalam bidang olahraga. Hasil ringkasan terhadap teks dalam bidang olahraga tetap mempunyai kata - kata yang kemunculannya banyak dan berkaitan dalam bidang olahraga. Oleh karena itu sudah banyak dilakukan peringkasan dengan metode - metode seperti TF (Term Frequency), RT (Related Term) dan Fuzzy System.

Pada tugas akhir ini akan dilakukan pendekatan Information Retrieval (IR) dengan mengimplementasikan metode Fuzzy System yang berdasarkan Term Frequency dan Related Term. Peringkasan teks atau bacaan pada sistem akan mendapatkan nilai akhir dari sebuah kata atau term dengan Fuzzy System, berdasarkan nilai dari suatu kata atau term tersebut akan dilakukan filter kalimat untuk menentukan hasil ringkasan.

Pengujian akan dilakukan dengan metode user judgement, dimana hasil ringkasan sistem dibandingkan dengan hasil ringkasan dari responden. Berdasarkan perbandingan tersebut akan didapatkan nilai precision (P), recall (R) dan f - measure (F). Hasil pengujian akan menunjukkan fuzzy rules yang memenuhi karakteristik dari sistem sehingga menghasilkan ringkasan yang berkualitas.

Kata Kunci : Text Summarization, Term Frequency, Related Term, Fuzzy System, User Judgement

Abstract

Text Summarization is a method to perform automatic summarization in order to get a text or reading a shorter or condensed. By doing summarization of text or readings are very helpful in getting the information quoted in the text or reading a text for example in the field of sports. Summary of the results of the text in the field of sports continues to have a word - a word related appearance and in the development of many sports fields. Therefore it has a lot to do with the method of summarization such as TF (Term Frequency), RT (Related Term) and Fuzzy System.

This final project will be approached Information Retrieval (IR) to implement Fuzzy Systems method based on Related Term and Term Frequency. Text of summarization or readings on the system will get the final value of a word or term with Fuzzy System, based on the value of a word or term that will be done to determine the sentence filter the results summary.

Testing will be done by the user judgment method, the results of system summaries compared to the results of a summary of the respondents. Based on the comparison will be obtained value of precision (P), recall (R) and f - measure (F). The test results will show fuzzy rules that meet the characteristics of the system so as to produce a quality summary.

Keywords : Text Summarization, Term Frequency, Related Term, Fuzzy Systems, Information Retrieval

1. Pendahuluan

1.1 Latar Belakang

Peringkasan terhadap sebuah teks merupakan salah satu cara untuk mendapatkan informasi, dengan berkembangnya teknologi peringkasan teks dapat dilakukan secara otomatis. Peringkasan teks dilakukan untuk mendapatkan teks yang lebih pendek dari teks aslinya agar pembaca dengan mudah menangkap inti dari teks dan tetap berkaitan dalam kategori teks tersebut.

Majalah, artikel, atau teks tentang olahraga terkadang memiliki sedikit kalimat yang memang berkaitan dalam bidang olahraga, sehingga pembaca perlu membaca keseluruhan isi teks untuk mendapatkan inti dari teks tersebut. Dengan melakukan peringkasan maka berita atau teks tentang olahraga tersebut dapat diperoleh dengan mudah. Kebanyakan orang meringkas hanya melihat kata yang sering muncul tanpa memperhatikan keterkaitan dari kata tersebut, namun ada juga yang melakukan peringkasan yang hanya melihat keterkaitan kata tersebut terhadap kategori tertentu. Berdasarkan permasalahan di atas sudah mulai banyak dikembangkan peringkasan dengan metode TF (*Term Frequency*) dan RT (*Related Term*).

Peringkasan yang dilakukan dengan metode TF (*Term Frequency*) melakukan peringkasan berdasarkan nilai kemunculan dari sebuah kata saja, namun hasil dari ringkasan kemungkinan tidak sesuai dengan kategori, sedangkan metode yang lain seperti RT (*Related Term*) yang hanya melakukan peringkasan berdasarkan nilai keterkaitan saja. Untuk mendapatkan hasil ringkasan yang memiliki kemunculan kata yang banyak dan berkaitan dengan kategori tertentu harus berdasarkan metode TF (*Term Frequency*) dan RT (*Related Term*). Berdasarkan kedua metode tersebut sangat sulit untuk menentukannya karena kedua hal tersebut sangat penting dalam melakukan peringkasan. Permasalahan yang seperti ini bisa dikatakan permasalahan yang ambigu antara TF (*Term Frequency*) dan RT (*Related Term*), dan sudah ada metode yang dikembangkan untuk permasalahan yang ambigu seperti ini yaitu *Fuzzy System*. *Fuzzy System* ini sangat cocok dalam menangani masalah yang ambigu seperti masalah peringkasan yang ingin terdiri dari kata dengan kemunculan yang banyak dan memiliki nilai keterkaitan yang tinggi.

1.2 Perumusan Masalah

Dari penjelasan diatas maka masalah yang dicari penyelesaiannya dalam Tugas akhir ini adalah :

1. Bagaimana cara mendapatkan informasi pada sebuah teks yang panjang dengan mudah ?
2. Bagaimana cara mendapatkan sebuah teks yang memang berkaitan dengan bidang olahraga ?

Dalam tugas akhir ini yang dibahas adalah *Text Summarization* menggunakan *Fuzzy System* berdasarkan *Term Frequency* dan *Related Term* dengan batasan masalahnya sebagai berikut :

1. Aplikasi ini dikembangkan dengan pemrograman bahasa Java.
2. Peringkasan teks yang dilakukan adalah artikel atau teks bahasa Indonesia yang berkaitan dengan bidang olahraga dengan kategori sepak bola, basket atau bulu tangkis.
3. Peringkasan yang dilakukan bersifat *Extractive Summarization*.
4. Pengujian yang dilakukan menggunakan skenario dengan *user judgement*.

1.3 Tujuan

Berdasarkan pada masalah yang telah diidentifikasi diatas, maka tujuan yang ingin dicapai dalam tugas akhir ini adalah :

1. Merancang suatu aplikasi yang mampu melakukan peringkasan teks menggunakan metode *Fuzzy System*.
2. Mendapatkan suatu teks yang lebih ringkas dari teks yang asli.
3. Hasil ringkasan tetap berkaitan dengan kategori dalam bidang olahraga.

1.4 Metodologi Penyelesaian Masalah

Pada pengembangan aplikasi ini telah dilakukan beberapa langkah seperti :

1. Persiapan data.
Data yang dipersiapkan adalah kata-kata yang berkaitan dalam bidang olahraga yang berisikan nilai keterkaitannya dan beberapa referensi yang menjadi acuan aplikasi.
2. Pengembangan aplikasi serta pembangunan model.
Aplikasi ini dikembangkan dengan batasan yang telah ditentukan dan melalui tahapan seperti, *preprocessing*, *processing* dan proses *output*.
3. Implementasi.
Implementasi aplikasi ini disimulasikan dengan bahasa pemrograman Java.
4. Pengujian.
Pengujian dilakukan dengan melakukan peringkasan teks atau artikel yang berkaitan dengan bidang olahraga sehingga mendapatkan ringkasan yang lebih pendek dan menggunakan skenario *user judgement*.
5. Analisa hasil.
Hasil dari ringkasan tersebut telah di analisa dengan memperhatikan panjang teks yang asli dengan *output* yang dihasilkan serta menjaga inti dari kalimat agar lebih spesifik dan jelas.
6. Pembuatan laporan.
Setelah melakukan semua tahapan yang telah ditentukan tahapan berikutnya adalah pembuatan laporan mengenai *Text Summarization* berdasarkan implementasi, pengujian, analisis serta teori-teori yang terkait dalam pengembangan aplikasi ini.

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian perangkat lunak yang dilakukan dalam tugas akhir ini dapat diambil beberapa kesimpulan, yaitu:

1. *Fuzzy rules* yang paling memenuhi karakteristik dari sistem adalah *fuzzy rules 4*, hal ini menunjukkan bahwa tidak hanya nilai frekuensi kemunculan suatu kata atau *term* lebih penting tetapi perlu memperhatikan nilai keterkaitan suatu kata atau *term* terhadap kategori tertentu agar mendapatkan hasil ringkasan yang lebih berkualitas dan terkait terhadap bidang olahraga.
2. Hasil ringkasan yang memiliki kualitas yang bagus diperoleh dengan menggunakan *fuzzy rules 4* dengan nilai rata - rata *precision (P)* = 0.629, nilai rata - rata *recall (R)* = 0.593 dan nilai rata - rata *f - measure (F)* = 0.606.

5.2 Saran

Untuk pengembangan tugas akhir ini di masa mendatang, penulis menyarankan hal – hal sebagai berikut:

1. *Dataset* yang digunakan lebih banyak sehingga memaksimalkan kinerja dari metode *Fuzzy System* berdasarkan *Term Frequency* dan *Related Term*.
2. Sistem yang dibangun tidak hanya berdasarkan *Term Frequency* dan *Related Term* tapi juga berdasarkan *similarity* suatu kata atau *term* terhadap *term* yang lainnya.

Daftar Pustaka

- [1] Anon., n.d. *Apa itu Information Rertrieval ?* Diakses November 3, 2012 dari : <http://bungaimuthz.wordpress.com/2011/05/04/apa-itu-information-retrieval/>
- [2] Anon., n.d. Information Retrieval. Diakses November 5, 2012 dari : <http://kungfumas.wordpress.com/category/lecture-notes/information-retrieval/>.
- [3] Kiabod, Maryam ; Dehkordi, Mohammad Naderi; Sharaf, Sayed Mehran;, 2012. *A Novel Method of Significant Words Identification in Text Summarization*. s.l., JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE.
- [4] Ledeneva, Yulia ; Gelbukh, Alexander ; García-Hernández, René Arnulfo;, 2008. *Terms Derived from Frequent Sequences for Extractive Text Summarization*. Berlin, s.n.
- [5] Lloret, Elena., 2008. *TEXT SUMMARIZATION : AN OVERVIEW*. Spain, Dept. Lenguajes Sistemas Informaticos, Universidad De Alicante.
- [6] Mustaqhfiri, Muchammad ; Abidin, Zainal ; Kusumawati, Ririen ;, 2012. *Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance*. Malang, Universitas Islam Negeri Maulana Malik Ibrahim.
- [7] Nagwani, Naresh Kumar; Verma, DR. Shrish;, 2012. *A Frequent Term and Semantic Similarity based Single Document Text Summarization Algorithm*. s.l., International Journal of Application.
- [8] Russiana, Endang;, n.d. *Fuzzy Logic*. Diakses November 5, 2012 dari : <http://wangready.wordpress.com/2011/05/03/fuzzy-logic/>
- [9] Saggion, Horacio;, 2012. *Introduction to Text Summarization and other Information Acces Technology..* England, University of Sheffield.
- [10] Suyanto. ST. Msc, 2007. *Artificial Intelligence*. s.l.:INFORMATIKA.
- [11] Tantiny; Santoso, Budi; Hapsari, Widi;, 2011. *Klasifikasi Email dengan Menggunakan Metode Naive Bayesian Studi Kasus : Mailing List*. s.l., Universitas Kristen Duta Wacana.

- [12] Yuanhua Lv, Zhai, ChengXiang;, 2012. *Lower Bounding Term Frequency Normalization*. Urbana, University of Illinois.

