

Review Jurnal tentang Peringkasan Dokumen

Jurnal 1

Judul:

Peringkasan Teks Otomatis pada Modul Pembelajaran Berbahasa Indonesia Menggunakan Metode Cross Latent Semantic Analysis (CLSA)

Penulis:

- Yunita Maulidia Sari
- Nenden Siti Fatonah

Review:

Salah satu solusi untuk mendapatkan inti informasi dari keseluruhan modul dengan cepat dan menghemat waktu adalah dengan membaca ringkasannya. Penelitian ini bertujuan untuk menghasilkan peringkasan teks otomatis pada modul pembelajaran berbahasa Indonesia dan mengetahui hasil akurasi peringkasan teks otomatis yang menerapkan metode Cross Latent Semantic Analysis (CLSA).

Jumlah data yang digunakan pada penelitian ini sebanyak 10 file modul pembelajaran yang berasal dari modul para dosen Universitas Mercu Buana, dengan format. Penelitian ini menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk pembobotan kata dan metode Cross Latent Semantic Analysis (CLSA) untuk peringkasan teks. Pengujian akurasi pada peringkasan modul pembelajaran dilakukan dengan cara membandingkan hasil ringkasan manual oleh manusia dan hasil ringkasan sistem.

Hasil pengujian peringkasan teks otomatis menggunakan algoritma Cross Latent Semantic Analysis, menghasilkan rata-rata nilai akurasi f-measure, precision, dan recall pada compression rate 20% secara berturut-turut yaitu 0. Sedangkan pengujian untuk nilai f-measure, precision, dan recall tertinggi terdapat pada Doc 2 yang secara berurutan yaitu 0.

Pengujian ini menghasilkan nilai akurasi yang cukup rendah dibandingkan dengan penelitian sebelumnya dengan menggunakan algoritma yang sama Cross Latent Semantic Analysis. Selain dikarenakan perbedaan banyaknya jumlah kata atau kalimat dalam dokumen yang diproses, hasil pemilihan ringkasan antara sistem dan secara manual oleh manusia juga dapat memengaruhi hasil akhir dari akurasi peringkasan teks otomatis.

Untuk penelitian yang akan datang diharapkan bisa menerapkan algoritma Cosine Similarity atau metode lainnya yang sesuai kebutuhan. Hal tersebut bertujuan untuk mencocokkan antara judul dengan kalimat hasil ringkasan.

Pada proses Cosine Similarity, dokumen yang sudah berhasil diringkas menggunakan metode Cross Latent Semantic Analysis. Selanjutnya akan dilakukan pengecekan kemiripan antara judul dengan kalimat yang sudah diringkas. Dengan menambahkan algoritma Cosine Similarity diharapkan dapat menghasilkan akurasi yang lebih baik lagi.

Jurnal 2

Judul:

Peringkasan Teks Otomatis Menggunakan Metode Maximum Marginal Relevance Pada Hasil Pencarian Sistem Temu Kembali Informasi Untuk Artikel Berbahasa Indonesia

Penulis:

- Nirmala Fa'izah Saraswati
- Indriati
- Rizal Setya Perdana

Review:

Query-focused summarization adalah salah satu kategori dari peringkasan teks otomatis. Queryfocused summarization adalah peringkasan yang diberikan query untuk menghasilkan ringkasannya sehingga, peringkasan teks ini menampilkan ringkasan yang relevan terhadap query yang diberikan.

Redundansi adalah permasalahan yang muncul pada sejumlah kalimat yang berulang atau ganda, sehingga dibutuhkan metode yang dapat mengaitkan teks dengan query, mengurangi tingkat redundansi dan ambiguitas dari peringkasan teks tersebut.

Maximum Marginal Relevance (MMR) merupakan salah satu metode ekstraksi ringkasan (extractive summary) yang digunakan untuk meringkas dokumen tunggal atau multi dokumen. MMR meringkas dokumen dengan menghitung kesamaan (similarity) antara kalimat dengan kalimat dan antara kalimat dengan query.

Terdapat dua kemiripan yang dicari, yaitu kemiripan antar dokumen atau kalimat dan kemiripan antara dokumen atau kalimat dengan query. Dalam memilih sebuah query perlu dipertimbangkan aspek kerelevanan dokumen atau kalimat dengan query, untuk mendapatkan ringkasan yang baik. Hasil dari precision dan recall pada penelitian tersebut tidak cukup baik sehingga diperlukan metode lain agar hasil ringkasan lebih baik.

Pada penelitian Indriani yang menggunakan metode Maximum Marginal Relevance mendapatkan nilai precision 62,4% nilai recall 73,6%, nilai tersebut cukup baik untuk digunakan dalam peringkasan teks otomatis. Maka dari itu, penulis melakukan penelitian ini yang dapat

membantu pembaca mendapatkan informasi yang tepat dengan menampilkan ringkasan dari sebuah dokumen.

Selanjutnya jika sudah mendapatkan hasil dokumen-dokumen yang diperoleh dari sistem temu kembali informasi maka dokumen tersebut diringkas menggunakan metode Maximum Marginal Relevance. Metode ini mencari similarity antar setiap kalimat yang ada pada dokumen dan similarity antara kalimat dengan query.

Berdasarkan hasil pengujian peringkasan teks otomatis menggunakan Maximum Marginal Relevance didapatkan hasil rata-rata precision sebesar 0,70, rata-rata recall sebesar 0,75, rata-rata f-measure sebesar 0,70 dan rata-rata akurasi sebesar 74,17% 150 data yang diuji.

Berdasarkan hasil pengujian sistem temu kembali informasi didapatkan hasil rata-rata evaluasi berperingkat yaitu precision@5 sebesar 0,96, precision@10 sebesar 0,92, precision@15 sebesar 0,89, precision@20 sebesar 0,79 dan precision@25 sebesar 0,74 dari 150 data dan 5 query yang diuji. 96 yang mana sistem temu kembali informasi sudah cukup baik untuk mendapatkan dokumen yang paling relevan pada 5 peringkat teratas.