

PEMBELAJARAN BAHASA INDONESIA BERBASIS WEB MENGUNAKAN METODE MAXIMUM MARGINAL RELEVANCE

Linda Marlinda¹⁾, Harsih Rianto²⁾

¹⁾Manajemen Informatika, AMIK Bina Sarana Informatika

Kramat Raya 25, Jakarta Pusat, 10450

HP : +62 813102176 55

E-mail : linda.ldm@bsi.ac.id¹⁾, harsih.hhr@bsi.ac.id²⁾

Abstrak

Pembelajaran Bahasa Indonesia bagi siswa sekolah dasar amatlah sulit, khususnya mulai kelas satu sampai kelas empat dalam hal pemahaman penalaran kalimat dalam bahasa Indonesia. Dibutuhkan suatu metode yang dapat mempermudah pemahaman kalimat bahasa Indonesia menggunakan metode maximum marginal relevance. metode ini dapat mengurangi redundansi dalam perangkaian kalimat pada dokumen dan memiliki lima tahap dalam pencarian text preprocessing yaitu pemecahan kalimat, case folding, tokenizing, filtering, dan stemming. Proses selanjutnya menghitung bobot tf-idf, bobot query relevance dan bobot similarity. Aplikasi ini telah di ujicoba secara acak pada siswa sekolah dasar kelas satu sampai kelas empat.

Learning Indonesian Language for elementary school students is very difficult, especially the starting grade one to grade four in terms of understanding the reasoning in Indonesia sentences. Need a method that can facilitate the understanding of Indonesian sentence using the maximum marginal relevance. This method can reduce the redundancies in the assembly of the sentence in the document and has the five stages in the search text preprocessing is split sentences, case folding, tokenizing, filtering, and stemming. The next process is to calculate tf-idf weighting, weighting query relevance and similarity weights. This application has been in a randomized trial in primary school grade one to grade four.

Kata kunci: *Text Bahasa Indonesia, Text Preprocessing, Maximum Marginal Relevance.*

1. PENDAHULUAN

Bahasa Indonesia merupakan salah satu mata pelajaran yang harus diajarkan di sekolah dasar. Bahasa merupakan sistem lambang bunyi yang dihasilkan dari alat ucap (artikulasi) yang bersifat sewenang – wenang dan konvensional untuk melahirkan perasaan dan pikiran. Selain itu bahasa merupakan alat komunikasi dan percakapan dengan sesama manusia^[4]. Sedangkan Bahasa Indonesia merupakan salah satu bahasa yang digunakan sehari – hari dalam kegiatan belajar mengajar dan digunakan sebagai bahasa nasional. Metode pembelajaran Bahasa Indonesia pada sekolah dasar saat ini masih menggunakan metode konvensional.

Perkembangan teknologi dalam dunia pendidikan telah menggeser penyampaian materi dengan metode ceramah ke arah penggunaan media pembelajaran. Media pembelajaran yang digunakan juga mengalami perkembangan yang cukup pesat, dari media cetak dalam bentuk buku dan media audio visual melalui jaringan internet berbasis web^[14]. Internet merupakan ruang bebas yang bisa digunakan sebagai sarana pembelajaran bagi siswa sekolah dasar lebih interaktif.

Dalam penelitian ini penulis terdorong untuk meneliti lebih lanjut antara pembelajaran Bahasa Indonesia yang konvensional ke pembelajaran interaktif berbasis web. Pembelajaran Bahasa Indonesia yang interaktif menggunakan peringkasan teks dengan metode Maksimum Marginal Relevan. Masalah penelitian yang muncul adalah bagaimana mengembangkan sistem pembelajaran berbasis web yang interaktif dan disertai latihan soal dari setiap materi yang di berikan. Tujuan penelitian adalah mengembangkan sistem pembelajaran Bahasa Indonesia yang interaktif dan disertai latihan soal.

2. TINJAUAN PUSTAKA

2.1 Peringkasan Teks Otomatis

Peringkasan teks otomatis (automatic text summarization) adalah pembuatan bentuk yang lebih singkat dari suatu teks dengan memanfaatkan aplikasi yang dijalankan dan dioperasikan pada komputer^[16]. Sedangkan menurut Hovy, ringkasan adalah teks yang dihasilkan dari sebuah teks atau banyak teks, yang mengandung isi informasi dari teks asli dan panjangnya tidak lebih dari setengah panjang aslinya (Hovy,2001).^[9]

2.2 Maximum Marginal Relevance

Algoritma maximum marginal relevance (MMR) adalah algoritma yang menggunakan metode ekstraksi ringkasan (extractive summary) yang digunakan untuk dokumen tunggal atau multi dokumen dengan menghitung kesamaan antar bagian teks. Cara kerja algoritma MMR meringkas kalimat dengan menghitung kesamaan (*similarity*) antar bagian kalimat^[16]. Cara kerja algoritma ini juga mengkombinasikan matrik *cosine similarity* untuk meranking kalimat – kalimat sebagai tanggapan pada query yang diberikan oleh user.

Metode *Maximal Marginal Relevance* untuk pemilihan kalimat atau unit teks lain yang mempertimbangkan aspek kerelevanan kalimat dengan *query* dan keterbaruan informasi (Carbonell&Goldstein 1998)^[4]. Ide dasar dari MMR ini yaitu memberikan penambahan nilai bagi kalimat yang relevan dan memberikan pengurangan nilai redundansi informasi antara kalimat tersebut dengan kalimat lain yang telah terpilih.

Sebuah kalimat dikatakan memiliki *marginal relevance* yang tinggi jika kalimat tersebut relevan terhadap isi dari kalimat dan mempunyai kesamaan bobot *term* maksimum dibandingkan dengan *query*. Peringkasan kalimat dengan tipe ekstraktif, nilai akhir diberikan pada kalimat S_i dalam MMR dihitung dengan persamaan (1).

$$MMR(S_i) = [\lambda Sim_1(V_{si}, V_{ip}) - (1 - \lambda) \max_{S_j \in R} Sim_2(V_{si}, V_{sj})] \quad (1)$$

Dimana *MMR* adalah himpunan kalimat relevan yang dipilih, *Sim₁* dan *Sim₂* adalah matriks dari kesamaan kalimat. Untuk mendapatkan kumpulan kalimat yang relevan ini, kita memberikan peringkat pertama pada kalimat yang relevan dengan memberikan presentase tertentu dari kalimat tersebut^[13].

2.3 Teks Preprocessing

Teks Preprocessing adalah pengkonversian dari input yang berupa teks menjadi gabungan dua buah fonem. Ketikamasukanyangberupateks,akronim (singkatan)ataupunangka.Strukturdatayangbaikdapat memudahkan proses komputerisasi secara otomatis.^[8]

2.4 Case folding

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter^[17].

2.5 Tokenizing

Tokenizing adalah pemotongan string input berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-scan kalimat dengan pemisah (delimiter) white space (spasi, tab, dan newline)^[17].

2.6 Filtering

Filtering adalah mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist* / *stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan bag-of-words. Contoh stopwords adalah “yang”, “dan”, “di”, “dari” dan seterusnya^[17].

2.7 Stemming

Stemming adalah mencari *root* kata dari tiap kata hasil *filtering*. Pada proses ini dilakukan pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama. Tahap ini kebanyakan dipakai untuk teks berbahasa Inggris dan lebih sulit diterapkan pada teks berbahasa Indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen^[17].

2.8 Algoritma TF/IDF

Algoritma TF/IDF (Term Frequency – Inversed Document Frequency) Pada algoritma TF/IDF digunakan rumus untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci dengan rumus yaitu:

$$W_{dt} = tf_{dt} * IDF_t \quad (2)$$

Dimana:

- d = dokumen ke-d
- t = kata ke-t dari kata kunci
- W = bobot dokumen ke-d terhadap kata ke-t
- tf = banyaknya kata yang dicari pada sebuah dokumen
- IDF = Inversed Document Frequency
- IDF = $\log_2 (D/df)$

D = total dokumen
df = banyak dokumen yang mengandung kata yang dicari

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses sorting/pengurutan dimana semakin besar nilai W, semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

Perhitungan bobot query relevance merupakan bobot hasil perbandingan kemiripan (similaritas) antara query yang dimasukkan oleh user terhadap keseluruhan kalimat. Sedangkan bobot similarity kalimat, merupakan bobot hasil perbandingan kemiripan antar kalimat^[16].

3. PERANCANGAN SISTEM

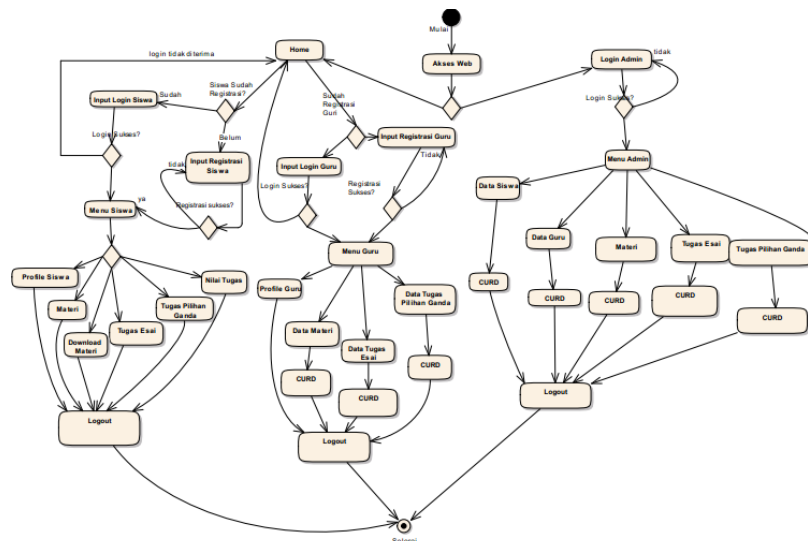
Dalam perancangan sitem Pembelajaran Bahasa Indonesia Berbasis Web ini digunakan beberapa alat bantu untuk mempermudah dalam tahapan perancangan.

3.1 Metodologi

Pada penelitian ini, Metode Maximum Marginal Relevan penulis penggunaan untuk pembobotan kalimat inputan dari user dengan kalimat yang ditampilkan pada halaman web. Dimana hasil pembobotan ini begunan untuk membandingkan hasil jawaban yang diinputkan user dengan jawaban yang benar dari system. Tahapan dalam pembobotan hasil jawaban sebagai berikut:

1. Inputan dari user berupa kalimat jawaban-jawaban di badingkan dengan kalimat-kalimat soal dan materi yang di tampilan, tahapan Teks Preprocessing, case folding, tokenizing dan filtering kalimat dijalankan pada proses ini
2. Kalimat yang sudah di filtering dangna jawaban kemudian diberikan pembobotan melalui metode MMR serta pembobotan nilai menggunakan algoritma TF/IDF.

3.2 Activity Diagram

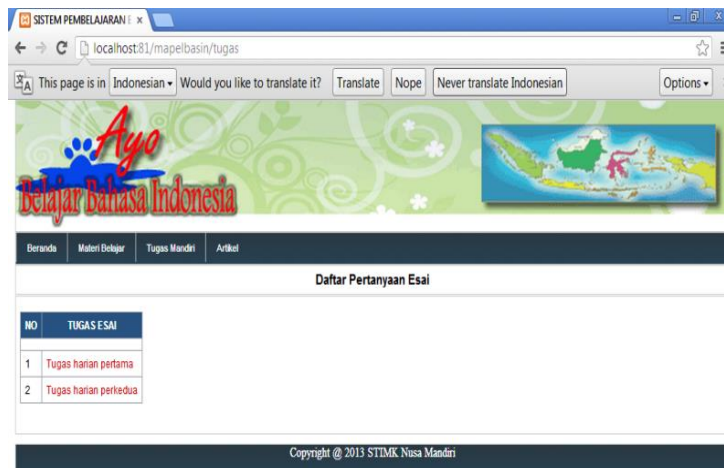


Gambar 1 Activity Diagram

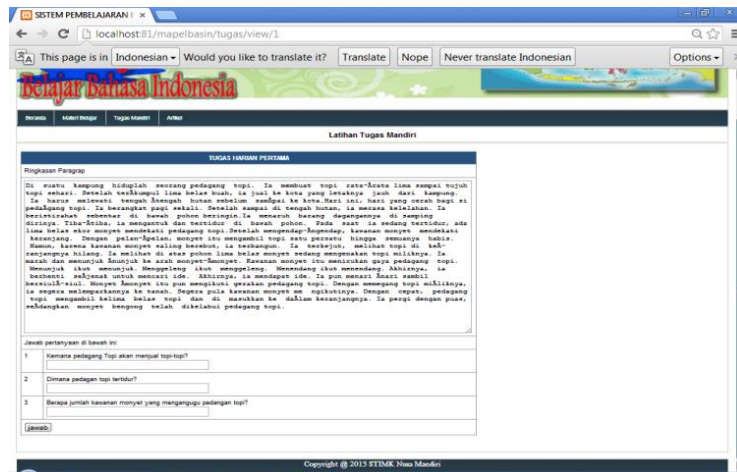
Perancangan system Pembelajaran Bahasa Indonesia yang penulis menggunakan *tools system* Activity Diagram buat seperti tampak pada gambar 2. Berikut ini alur proses dari system yang penulis buat:

1. User mengakses web adalah siswa atau guru sekolah dasar dan administrator dari web tersebut. User siswa mengakses materi pembelajaran yang ditampilkan, serta mengakses butir latihan soal dari tiap materi pembelajaran yang di tampilan. User guru dan Administrator memiliki akses untuk menambah materi dan membuat butir soal.
2. Sistem menyediakan butir soal yang diakses siswa dan melakukan penghitungan terhadap inputan jawaban dari siswa, pada proses ini dilakukan pembobotan kalimat dengan teks kalimat ringkasan dari web.
3. Hasil pembobotan dari kalimat dan jawaban dari siswa akan di olah sistem menggunakan metode maximum marginal relevan.
4. User siswa akan mendapatkan hasil penilaian dari butir-butir soal yang dikerjakan oleh siswa.

3.3Tampilan Web



Gambar 3 Tampilan Web Daftar Pertanyaan



Gambar 4 Tampilan Web Latihan soal

Tampilan web yang penulis buat seperti gambar 3 dan 4 dibangun menggunakan bahasa HTML yang disisipi bahasa PHP (*Hypertext Preprocessor*). Sebagai halaman muka (*interface*) dari web ini adalah *index.php*. File ini merupakan induk dari template web dan mengatur file-file mana yang digunakan saat mengakses link-link dalam menu-menu yang ditampilkan. Di web ini user pengguna sebagai siswa bisa mendaftar atau sebagai pengunjung biasa, dan user guru juga bisa mendaftar sebagai pengisi materi pelajaran. Siswa sebagai pengunjung web akan mengakses materi pelajaran dan butir-butir soal yang ditampilkan.

3.3Hasil Uji Coba Sistem

Pada saat peneliti melakukan riset menggunakan media kuesioner dan wawancara untuk mengetahui apakah penelitian ini bermanfaat atau diinginkan oleh siswa/i sekolah dasar dengan jumlah 10 responden dengan 5 pertanyaan yang berkaitan langsung dengan penelitian. Hasil pengamatan penggunaan web pembelajaran sebagai berikut:.

Tabel 1 Skor Pengujian

Responden	Skor Pertanyaan					Skor Total
	1	2	3	4	5	
1	3	4	3	4	2	16
2	3	4	4	2	3	16
3	4	3	2	4	3	16
4	3	4	3	4	2	16

5	4	2	4	3	4	17
6	3	4	3	4	4	18
7	4	4	4	4	4	20
8	2	3	4	3	4	16
9	3	3	3	4	2	15
10	4	4	2	4	3	17
Jumlah	33	35	32	36	31	167

Hasil kuesioner tersebut diolah dengan menggunakan SPSS metode Bivariate Pearson (korelasi produk momen pearson) adalah analisis dengan cara mengorelasi masing – masing skor item dengan skor total. Hasil pengujian berdasarkan isian responden sebagai berikut:

Tabel 2 Hasil Pengujian

		ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEMTOT
ITEM1	Pearson corelation	1	-,518	-,334	,283	,132	,337
	Sig(2-tailed)	-	,125	,346	,429	,727	,341
	N	10	10	10	10	10	10
ITEM2	Pearson corelation	-,518	1	-,161	,136	-,617	-,207
	Sig(2-tailed)	,125	-	,656	,707	,057	,566
	N	10	10	10	10	10	10
ITEM3	Pearson corelation	-,344	-,161	1	-,645*	,450	,367
	Sig(2-tailed)	,346	,656	-	,044	,191	,297
	N	10	10	10	10	10	10
ITEM4	Pearson corelation	,283	,136	-,645*	1	-,290	,207
	Sig(2-tailed)	,429	,707	,044	-	,416	,566
	N	10	10	10	10	10	10
ITEM5	Pearson corelation	,132	-,617	,450	-,290	1	,637*
	Sig(2-tailed)	,717	,057	,191	,416	-	,047
	N	10	10	10	10	10	10
ITEMTOT	Pearson corelation	,337	-,207	,367	,207	,637*	1
	Sig(2-tailed)	,341	,566	,297	,566	,047	-
	N	10	10	10	10	10	10

4. SIMPULAN

4.1 Simpulan

Algoritma Maximal Marginal Relevance sangat berguna untuk memberikan informasi pada dokumen yang bersifat multi redundansi dan dapat meningkatkan kecepatan pengukuran yang tepat dari kesamaan kalimat dalam hal pencarian informasi dibandingkan dengan pendekatan algoritma yang lain. Aplikasi pembelajaran yang dibuat oleh penulis telah di ujicoba pada 10 siswa sekolah dasar kelas satu sampai kelas empat secara random dengan kesimpulan yaitu: siswa dapat lebih memahami penalaran pelajaran Bahasa Indonesia dengan mudah terutama pada soal essay. Dari hasil analisis didapat nilai korelasi antara skor item dengan skor total. Nilai ini dibandingkan dengan nilai r tabel pada signifikan 0,05 dengan uji 2 sisi dan jumlah data responden (n) = 10, maka didapat r tabel sebesar 0,47.

5. DAFTAR RUJUKAN

- [1] Das and Martins. 2007. A Survey on Automatic Text Summarization. Language Technologies Institute Carnegie Mellon University
- [2] Erwin A.H., Muhammad. 2005. Sistem Pengidentifikasi Otomatis Pokok Kalimat Suatu Paragraf Dalam Dokumen Ekspositori Dengan Model Ruang Vektor. Laboratorium Pemrograman dan Informatika Teori. Yogyakarta: Jurusan Teknik Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
- [3] Firmin, T and M.J Chrzanowski. 1999. An Evaluation of Automatic Text Summarization System. The MIT Press: Cambrige
- [4] Flaviana Tumini. 2013. Pembelajaran Membaca Permulaan Menggunakan Kartu di Kelas I Sekolah Dasar Negeri 04 Sandai Ketapang. Penelitian, Pontianak.
- [5] Golstein, Jade and Carbonell, Jaime. 1998. Summarization: Using MMR for Diversity Based-Reranking and Evaluating Summaries. Languange Technologies Istitute. Carnegie Mellon University.
- [6] Golstein, Jade. 2008. Genre Oriented Summarization. Thesis. Pittsburgh: Language Technologies Institute School of Computing Carnegie Mellon University.

- [7] Grossman, D., dan Ophir, F. 1998. *Information Retrieval: Algorithm and Heuristics*. Kluwer Academic Publisher.
- [8] Handi, 2011. *Text Pre-Processing Pada Text To Speech Synthesis System Untuk Penutur Berbahasa Indonesia*. ITS Surabaya
- [9] Hovy, E. and Lin, C. Y. (1999). Automated text summarization in summarist. In Mani, I. and Maybury, M. T., editors, *Advances in Automatic Text Summarization*, pages 81-94. MIT Press
- [10] Jones, K.S, dan Galliers, J.R. 1996. *Evaluating Natural Language Processing System : An Analysis and Review*. New York: Springer.
- [11] Larson and Hearst. 2000. *Computing Relevance, Similarity: The Vector Space Model*. UC Berkeley. <http://www.sims.berkeley.edu/courses/is202/f00/>. Diakses tanggal 25 Maret 2011.
- [12] Mani, I. and Maybury. 1999. *Advance in Automatic Text Summarization*. The MIT Press: Cambrige.
- [13] Mani, Inderjeet. 2001. *Summarization Evaluation: An Overview*. The MITRE Corporation, W640 11493 Sunset Hills Road Reston, VA 20190-5214 USA.
- [14] Muntoha Muhammad, akhlis isa dan subali bambang. 2010. Pengembangan sistem evaluasi pembelajaran berbasis web. Page 195-199 *Prosiding Pertemuan Ilimah XXIV HFI Jateng & DIY*. Semarang
- [15] Mustaqhfirin Muhammad, dkk. 2011. *Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance*. UIN, Malang
- [16] Sum Jian, dkk 2009. *Trac-2003 Novelty and Web Track at ICT*, Institute of Computing Technolgy (ICT), Chinese Academy of Sciences, P.R. China
- [17] Triawati, Candra, 2009. *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia, Statistical Concept Based Weighting Method for Indonesian Language Document Clustering and Categorisation*, IT TELKOM Jakarta
- [18] Herwansyah, Adhit. *Aplikasi Pengkategorian Dokumen Dan Pengukuran Tingkat Similaritas Dokumen Menggunakan Kata Kunci Pada Dokumen Penulisan Ilmiah Universitas Gunadarma*