

AUTOMATIC SEARCH IN A SINGLE DOCUMENT THE CORE SENTENCES INDONESIAN LANGUAGE

¹ Syamsul Arifin(0910651057), ² Lutfi Ali Muharom S.Si, ³ Ratih Ayuninghem, S.ST, M.Kom
Jurusan Teknik Informatika, Fakultas Teknik
Universitas Muhammadiyah Jember
Email : syamsul_arifin19@yahoo.com

ABSTRACT

Search this sentence is automatically nuclei using tetx mining and weighting document to find the core sentence of a single document. Core search process automatically sentence was originally a form of document that will then be in the form of sentences and then of the sentences will be the core of a document sentence. document in this study is a collection of sentences Obtained from a sports site on the internet. This research is the development of an application that generates an automatic search phrase core document in Indonesian. The method used in this study is the Term Frequency-Inverse Document Frequency (TF-IDF) in roomates to calculate the weight of each word. By applying Text Mining and Term Frequency-Inverse Document Frequency (TF-IDF), result obtained from this research is the application that is able to produce a summary document of the document article.

Keyword : search, core sentence, automatically, text mining, tf-idf.

ABSTRAK

¹ Syamsul Arifin(0910651057), ² Lutfi Ali Muharom S.Si, ³ Ratih Ayuninghem, S.ST, M.Kom
Jurusan Teknik Informatik,a Fakultas Teknik
Universitas Muhammadiyah Jember
Email : syamsul_arifin19@yahoo.com

Pencarian kalimat inti secara otomatis ini adalah menggunakan *tetx mining* dan pembobotan dokumen untuk menemukan kalimat inti dari sebuah dokumen tunggal. Proses pencarian kalimat inti secara otomatis ini awalnya adalah berupa dokumen yang kemudian akan di bentuk menjadi kalimat-kalimat yang kemudian dari kalimat-kalimat tersebut akan di dapat kalimat inti dari sebuah dokumen. dokumen di penelitian ini adalah kumpulan dari kalimat-kalimat yang didapat dari situs olahraga di internet. Penelitian ini adalah pengembangan sebuah aplikasi yang menghasilkan pencarian otomatis kalimat inti dokumen dalam Bahasa indonesia. Metode yang digunakan dalam penelitian ini adalah *Term Frekuensi-Frekuensi Dokumen Invers* (TF-IDF) yang mana dengan menghitung bobot setiap kata. Dengan menerapkan *Text Mining* dan *Term Frekuensi-Frekuensi Dokumen Invers* (TF-IDF) ini, hasil yang diperoleh dari penelitian ini adalah aplikasi yang mampu menghasilkan summarization dokumen dari dokumen berita.

Kata Kunci: pencarian, kalimat inti, otomatis, text mining, tf-idf.

PENDAHULUAN

Latar Belakang

Perkembangan teknologi di bidang komputasi dan telekomunikasi telah memberi kemudahan kepada setiap orang untuk menyampaikan dan mendapatkan informasi. Kemudahan ini menyebabkan informasi menjadi semakin banyak dan beragam. Informasi dapat berupa dokumen, berita, surat, cerita, laporan penelitian, data keuangan dan lain-lain.

Ketersediaan informasi yang semakin banyak menjadikan ringkasan sebagai kebutuhan yang sangat penting. Dengan adanya ringkasan, pembaca dapat dengan cepat dan mudah memahami makna sebuah teks tanpa harus membaca keseluruhan teks. Hal ini dapat menghemat waktu pembaca karena dapat menghindari pembacaan teks yang tidak relevan dengan informasi yang diharapkan oleh pembaca, terutama jika informasi tersedia di internet cukup banyak.

Berbagai metode telah diterapkan dan masih terus dikembangkan oleh para peneliti tentang peringkasan. Salah satunya adalah peringkasan dokumen yang kemudian akan menghasilkan informasi yang penting atau inti kalimat dari dokumen dengan menggunakan komputer. Tujuannya adalah mengambil sumber informasi dengan mengutip sebagian besar isi yang penting untuk mencari kalimat inti dari sebuah dokumen artikel dan menampilkan kepada pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca. Dengan demikian teknologi ini dapat membantu pembaca untuk menyerap informasi yang ada dalam artikel melalui ringkasan tanpa harus membaca seluruh isi dokumen.

Pencarian otomatis kalimat inti ini dibutuhkan pembobotan pada masing-masing term sehingga metode *term frequency – invers document frequency (tf-idf)* digunakan sebagai penentu bobot masing-masing term tersebut.

Rumusan Masalah

Dari latar belakang yang sudah dijelaskan di atas diperoleh rumusan masalah sebagai berikut

- Bagaimana mengolah dokumen dengan teks mining.
- Bagaimana menghitung bobot yang dimiliki suatu dokumen dalam keseluruhan teks menggunakan metode TF-IDF.
- Bagaimana mengurutkan bobot yang diperoleh dari metode TF-IDF.

Batasan Masalah

- Input hanya dari satu dokumen atau satu artikel saja.
- Dokumen berupa artikel dari website/situs olahraga.
- Dokumen yang digunakan sebagai data set mengambil dari situs Goal.com.
- Database untuk kata-kata yang tidak relevan mengambil dari kamus bahasa Indonesia.
- Input dan output dalam bahasa Indonesia.
- Tidak menangani kesalahan penulisan kata.
- Singkatan di anggap satu kata.
- Hanya melayani list dokumen.

Tujuan Penelitian

- Mengolah dokumen dengan teks mining dan melakukan pembobotan dengan tf-idf.
- Melakukan pencarian otomatis kalimat inti dalam sebuah dokumen tunggal berbahasa Indonesia.
- Mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada

pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca.

Manfaat Penelitian

- Dapat membantu pembaca untuk menyerap informasi yang ada dalam artikel melalui ringkasan tanpa harus membaca seluruh isi dokumen.
- Dapat menghemat waktu pembaca karena dapat menghindari pembacaan teks yang tidak relevan dengan informasi yang diharapkan oleh pembaca.

TINJAUAN PUSTAKA

Kalimat Inti

Kalimat inti sering juga di katakan sebagai kalimat utama ataupun kalimat topik. Pengertian dari kalimat sendiri ini adalah kalimat yang paling penting dalam sebuah *paragraph* karena merupakan ide utama dalam *paragraph* tersebut. Kalimat inti atau kalimat topik bertugas memberitahukan kepada pembaca gagasan pokok yang akan dibicarakan dalam alinea dimaksud. Karena karangan terdiri dari alinea-alinea yang memiliki kesatuan yang utuh, maka kalimat inti yang biasanya terletak di awal paragraf juga berfungsi menghubungkan alinea itu dengan alinea sebelumnya. Jika kalimat tersebut sengaja dihilangkan, maka isi paragraf tersebut akan hilang. Hal ini tidak terjadi dengan kalimat-kalimat lain yang memang hanya berfungsi sebagai penjelas. Kalimat utama berupa ringkasan dari sebuah paragraf yang merupakan pandangan mendalam dari ide pokok penulis dalam paragraf tersebut.

Dokumen

Kata dokumen berasal dari bahasa latin yaitu *docere*, yang berarti mengajar. Pengertian dari kata dokumen ini menurut Louis Gottschalk (1986; 38) seringkali digunakan para ahli dalam dua pengertian, yaitu pertama, berarti sumber tertulis bagi informasi sejarah sebagai kebalikan daripada kesaksian lisan, artefak, peninggalan-peninggalan terlukis, dan petilasan-petilasan arkeologis. Pengertian kedua diperuntukan bagi surat-surat resmi dan surat-surat negara seperti surat perjanjian, undang-undang, hibah, konsesi, dan lainnya. Lebih lanjut, Gottschalk menyatakan bahwa dokumen (dokumentasi) dalam pengertiannya yang lebih luas berupa setiap proses pembuktian yang didasarkan atas jenis sumber apapun, baik itu yang bersifat tulisan, lisan, gambaran, atau arkeologis.

Artikel

Artikel (*article*) adalah tulisan tentang suatu masalah berikut pendapat penulis tentang masalah tersebut yang dimuat di media *online* ataupun media cetak seperti surat kabar, majalah, atau buletin. Secara umum dapat dikatakan bahwa artikel tidak terlalu panjang, hanya beberapa halaman. Artikel termasuk tulisan kategori *views* (pandangan), yaitu tulisan yang berisi ide, opini, atau penilaian penulisnya terhadap suatu peristiwa atau masalah.

Penulis artikel bertujuan menyampaikan gagasan dan fakta guna meyakinkan, mendidik, menawarkan pemecahan suatu masalah, atau menghibur. Biasanya, tahap-tahap penulisan artikel meliputi latar belakang masalah, identifikasi masalah, analisis masalah, dan kesimpulan.

Text Mining

Text mining adalah salah satu bidang khusus dari *data mining*. Sesuai dengan buku *the text mining handbook*, *text mining* dapat didefinisikan suatu proses untuk menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan *tool* analisis yang merupakan komponen-komponen dalam data mining yang salah satunya adalah kategorisasi.

Text mining bisa dianggap subyek riset yang tergolong baru. *Text mining* dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian atau pengelompokan dan menganalisa *unstructured* teks dalam jumlah besar. Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *data mining*, *Information retrieval*, *Statistik* dan *Matematik*, *Machine Learning*, *Linguistic*, *Natural Language Processing*, dan *Visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan penyimpanan teks, *preprocessing* akan konten teks, pengumpulan data statistik dan *indexing* dan analisa konten.

Tokenizing

Penggunaan istilah kata dan token sering kali saling dipertukarkan. Sebuah token dapat di definisikan sebagai unit terkecil dari sebuah teks atau dapat juga merupakan suatu kumpulan dari string alphanumerik [Konchady,2006]. Suatu unit terkecil yang digunakan disini adalah sebuah kata tunggal. Sebuah kata tunggal dapat berisi kumpulan dari string alphanumerik. Pada proses *tokenizing* ini, akan terjadi proses pemotongan dokumen menjadi daftar kata yang berdiri sendiri sebelum dilakukan proses selanjutnya. Tahap *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Lebih tepatnya adalah proses pemecahan dokumen yang ada dalam sebuah dokumen menjadi kata.

Stoplist

Dalam sebuah dokumen sangat banyak didapatkan jenis-jenis kata seperti, kata sambung, kata depan, kata ganti, kata sifat dan lain sebagainya. Sebagian besar kata-kata itu merupakan kata yang tidak berpotensi dalam mengidentifikasi isi sebuah dokumen. Selain itu juga pendefinisian dokumen dengan berdasarkan frekuensi kata menjadi kurang efektif bila jenis kata-kata diatas tidak dilakukan proses penyaringan.

Stoplist adalah proses penyaringan (*filtering*) terhadap kata-kata yang tidak layak untuk dijadikan sebagai pembeda atau kata kunci sehingga kata-kata tersebut dapat dihilangkan dari dokumen.

Stemming

Setelah kata-kata yang terdapat dalam dokumen menjalani proses *tokenizing* dan *stoplist*, maka selanjutnya kata-kata yang tersisa akan menjalani proses stemming. Proses stemming bertujuan untuk mengubah atau mengembalikan kata menjadi kata/bentuk dasarnya dengan menghilangkan imbuhan-imbuhan pada kata dalam dokumen. Proses stemming dilakukan dengan mengecek kata apakah mengandung imbuhan atau tidak, terutama imbuhan yang bersifat *suffix*. Proses stemming kata dalam Bahasa Inggris memiliki karakteristik tersendiri [Porter, 2001], yang tidak lepas dari pengaruh tata bahasanya. Selain itu, yang perlu diperhatikan pada proses ini adalah konsistensi dalam

memproses suatu kata, meskipun tidak sepenuhnya sempurna dalam memberikan hasil.

Term Frequency-InverseDocument Frequency

Metode Term Frequency-InverseDocument Frequency (TF-IDF) adalah cara pemberian bobot hubungan suatu kata (term) terhadap dokumen. Untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu Term frequency (TF) merupakan frekuensi kemunculan kata (t) pada kalimat (d). Document frequency (DF) adalah banyaknya klaimat dimana suatu kata (t) muncul. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut.

Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Bobot kata semakin besar jika sering muncul dalam suatu dokumen dan semakin kecil jika muncul dalam banyak dokumen (Robertson, 2005). Pada Metode ini pembobotan kata dalam sebuah dokumen dilakukan dengan mengalikan nilai TF dan IDF.

METODE PENELITIAN

Desain Sistem

Model yang akan di kembangkan dalam pencarian kalimat inti dalam penelitian dapat di lihat pada gambar di bawah ini:



Gambar Desain Sistem

Adapun langkah-langkah pencarian otomatis kalimat inti adalah sebagai berikut:

1. Menginput dokumen yang akan dicari kalimat intinya
2. Melakukan *Scanning* untuk menelusuri setiap kata pada dokumen
3. Proses *casefolding* yaitu proses merubah huruf menjadi huruf kecil dan selain karakter akan dibuang menjadi dokumen
4. Memilah dokumen menjadi beberapa kalimat. Pemilahan kalimat dilakukan dengan memecah string teks dari

dokumen yang panjang menjadi kalimat-kalimat menggunakan fungsi *split()*, dengan tanda titik ".", tanda tanya "?" dan tanda seru "!" sebagai delimiter untuk memotong string dokumen

- Memilah kalimat yang terbentuk menjadi kata yaitu dengan *tokenizing*, dilanjutkan dengan proses *stoplist* yaitu membuang kata-kata yang tidak relevan dan proses mengembalikan kata ke bentuk dasar dengan menggunakan *stemming*.

- Pembobotan TF-IDF

Pembobotan diperoleh berdasarkan jumlah kemunculan term dalam kalimat (TF) dan jumlah kemunculan term pada seluruh kalimat dalam dokumen (IDF). Bobot suatu istilah semakin besar jika istilah tersebut sering muncul dalam suatu dokumen dan semakin kecil jika istilah tersebut muncul dalam banyak dokumen (Grossman, 1998). Nilai IDF sebuah term dihitung menggunakan persamaan berikut :

$$= \log \frac{1}{\text{IDF}}$$

Dengan

N : jumlah kalimat yang berisi *term(t)*

D_{ft} : jumlah kemunculan kata (*term*) terhadap D

- Menghitung bobot (W) masing-masing kata pada dokumen dengan persamaan berikut (Mustaqhfiri, 2011) :

$$W_{d,t} = TF_{d,t} * IDF_t$$

Dengan

d = kalimat ke- d

t = kata (*term*) ke- t

TF = *term frequency*

W = bobot kalimat ke- d terhadap kata (*term*) ke- t

IDF = *inverse document frequency*

- Melakukan proses pengurutan (*sorting*) nilai kumulatif dari W untuk setiap kalimat
- Tiga kalimat dengan nilai W terbesar dijadikan sebagai hasil dari ringkasan atau sebagai output dari pencarian otomatis kalimat inti.

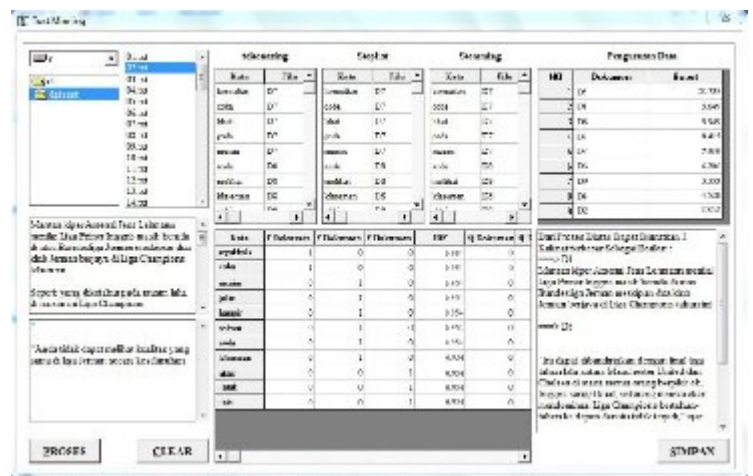
HASIL DAN PEMBAHASAN

Pencarian Otomatis Kalimat Inti

Pada Penelitian ini pengembangan aplikasi peringkasan teks otomatis digunakan

bahasa pemrograman Microsoft Visual Basic. Aplikasi yang dikembangkan dibuat untuk single dokumen secara dinamis artinya hanya membuat Pencarian dari satu dokumen tetapi dapat digunakan untuk dokumen yang berbeda.

Berikut tampilan sistem pencarian otomatis kalimat inti dokumen tunggal berbahasa Indonesia :



Gambar Sistem Pencarian Otomatis Kalimat Inti

Gambar diatas adalah sistem pencarian otomatis kalimat inti dengan menggunakan metode tf-idf. Berikut pembahasan dari sistem pencarian kalimat inti :

Proses Text Mining

Pada proses ini, memperlihatkan dimana dokumen dipecah menjadi kata dan kemudian dihitung nilai kemunculannya.

tokenizing		Stoplist		Stemming	
Kata	File	Kata	File	Kata	File
komodori	D7	komodori	D7	komodori	D7
coba	D7	coba	D7	coba	D7
lihat	D7	lihat	D7	lihat	D7
pada	D7	pada	D7	pada	D7
rumah	D7	rumah	D7	rumah	D7
anda	D8	anda	D8	anda	D8
melihat	D8	melihat	D8	melihat	D8
klasekmen	D8	klasekmen	D8	klasekmen	D8
akhir	D8	akhir	D8	akhir	D8

Gambar Proses Text Mining

Proses TF-IDF

Pada proses ini, memperlihatkan nilai IDF yang ditemukan pada proses *text mining*.

Kata	F Dokumen	F Dokumen	F Dokumen	IDF	ij Dokumen	ij I
sejauh	1	0	0	0.954	0	0
coba	1	0	0	0.954	0	0
rumah	0	1	0	0.954	0	0
jalan	0	1	0	0.954	0	0
harapir	0	1	0	0.954	0	0
selesai	0	1	0	0.954	0	0
naca	0	1	0	0.954	0	0
klasekmen	0	1	0	0.954	0	0
akhir	0	0	1	0.954	0	0
jurak	0	0	1	0.954	0	0
tan	0	0	1	0.954	0	0

Gambar Pembobotan TF-IDF

Pengurutan Bobot (W)

Pada tahap inilah yang paling penting karena tahap inilah yang menentukan hasil dari pencarian kalimat inti.

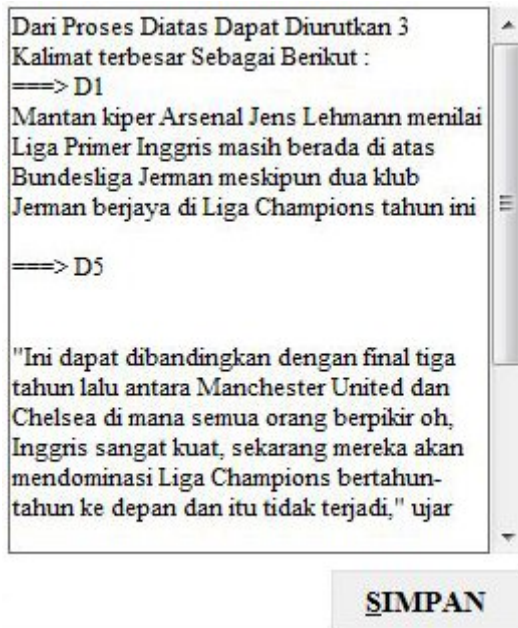
Pengurutan Data

NO	Dokumen	Bobot
1	D5	20.738
2	D1	15.649
3	D8	9.949
4	D3	8.415
5	D7	7.030
6	D6	6.206
7	D9	5.553
8	D4	4.520
9	D2	2.862

Gambar Pengurutan Bobot (W)

Hasil Pencarian

Tahap inilah yang menampilkan hasil dari pencarian otomatis kalimat inti.



Gambar hasil pencarian

KESIMPULAN DAN SARAN

KESIMPULAN

Dari uji coba dan analisa yang telah dijelaskan dalam bab sebelumnya, maka penulis mengambil beberapa kesimpulan sebagai berikut :

1. Program ini berhasil mengolah data atau dokumen artikel dengan pra-proses *Text mining* dan pembobotan TF-IDF, serta dapat melakukan pencarian secara otomatis kalimat inti dalam sebuah dokumen berbahasa Indonesia.
2. Program ini berhasil mengambil sumber informasi dengan mengutip sebagian besar isi yang penting dan menampilkan kepada pembaca dalam bentuk yang ringkas sesuai dengan kebutuhan pembaca dengan tidak mengubah produk teks yang memiliki / mengandung

bagian-bagian yang penting dari artikel asli meskipun secara tata bahasa belum baik. Selain itu, sistem yang dikembangkan ini mampu menghasilkan *summarization* dokumen dari sebuah dokumen artikel.

SARAN

Penulis ingin memberikan beberapa saran yang mungkin dapat membantu dalam pengembangan Tugas Akhir ini adalah sebagai berikut :

1. Untuk penelitian selanjutnya pengembang diharapkan dapat menggunakan dokumen dalam bahasa Inggris atau bahasa lainnya, bukan hanya menggunakan bahasa Indonesia. Sehingga dapat menghasilkan program yang lebih meluas lagi dan berguna bagi pembaca.
2. Penulis menyarankan bagi pengembang agar memadukan dengan metode lainnya bukan hanya tf-idf agar hasil yang diperoleh lebih bagus lagi.
3. Diharapkan pengembang mampu menghitung akurasi dari hasil percobaan.

DAFTAR PUSTAKA

- Arifin, A Z., (2002), *Penggunaan Digital Tree Hibrida pada Aplikasi Information Retrieval untuk Dokumen Berita*, Jurusan Teknik Informatika, FTIF, Institut Teknologi Sepuluh Nopember, Surabaya.
- Gottschalk, L., (1986), *Understanding History: A Primer of Historical Method* (terjemahan Nugroho Notosusanto). UI Press, Jakarta.
- Grossman, D., Ophir, F., (1998), *Information Retrieval : Algorithm and Heuristics*. Kluwer Academic Publisher, USA.
- Jones K, S., (1998), *Charting a New Course: Natural Language Processing and Information Retrieval : Essays in Honour of Karen Spärck Jones*. Penerbit: Springer, New York.
- Kunchady, M., (2006), *Text Mining Application Programming*. Thomson Learning Inc. ISBN 1-58450-460-9, USA.
- Mustaqhfiri, M., Abidin Z., Kusumawati, R. (2011). Peringkasan Teks Otomatis Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *Ejournal Matics*, 4, 4, 135-147. Dimbali 5 Januari 2012 dari basis data saintek.
- Robertson, S., (2004). "Understanding Inverse Document Frequency: On theoretical arguments for IDF", *Journal of Documentation*, Vol. 60, no. 5, pp. 503-520, New York.
- Widjono, H.S., (2007). "Bahasa Indonesia Mata Kuliah Pengembangan Kepribadian di PT". Penerbit : Grasindo, Jakarta.

=== HASIL INDEXING OTOMATIS ===

1. Metode TF-IDF:

- dokumen (halaman: 1, 2, 3, 4, 5)
- kalimat (halaman: 1, 2, 3, 4, 5)
- inti (halaman: 1, 2, 3, 4, 5)
- mining (halaman: 1, 2, 3, 4, 5)
- idf (halaman: 1, 2, 4, 5)
- tf (halaman: 1, 2, 4, 5)
- pencarian (halaman: 1, 2, 3, 4, 5)
- text (halaman: 1, 3, 4)
- otomatis (halaman: 1, 2, 3, 4, 5)
- document (halaman: 1, 2)
- the (halaman: 1, 3, 5)
- of (halaman: 1, 4, 5)
- bobot (halaman: 1, 2, 3, 4, 5)
- penelitian (halaman: 1, 2, 3, 5)
- hasil (halaman: 1, 2, 3, 4, 5)
- indonesia (halaman: 1, 2, 4, 5)
- pembobotan (halaman: 1, 2, 3, 5)
- in (halaman: 1, 2, 3, 4, 5)
- is (halaman: 1, 2, 3, 4, 5)
- core (halaman: 1)
- bentuk (halaman: 1, 2, 3, 4, 5)
- and (halaman: 1, 2, 3, 4, 5)
- diperoleh (halaman: 1, 2, 4, 5)
- aplikasi (halaman: 1, 4)
- this (halaman: 1)
- sentence (halaman: 1)
- menghasilkan (halaman: 1, 2, 5)
- search (halaman: 1)
- sentences (halaman: 1)
- tunggal (halaman: 1, 2, 3, 4)
- menghitung (halaman: 1, 2, 5)
- single (halaman: 1, 4)
- internet (halaman: 1, 2)
- situs (halaman: 1, 2)
- berita (halaman: 1, 2)
- arifin (halaman: 1)
- to (halaman: 1, 2, 3, 4, 5)
- application (halaman: 1)
- proses (halaman: 2, 3, 4, 5)
- informasi (halaman: 2, 3, 5)
- pembaca (halaman: 2, 5)
- teks (halaman: 2, 3, 4, 5)
- term (halaman: 2, 3, 4)
- artikel (halaman: 2, 5)
- frequency (halaman: 2, 3)
- bahasa (halaman: 2, 3, 4, 5)
- metode (halaman: 2, 4, 5)
- data (halaman: 2, 3, 5)
- gambar (halaman: 2, 3)
- tahap (halaman: 2, 3, 5)

- surat (halaman: 2)
- isi (halaman: 2, 3, 4, 5)
- mengambil (halaman: 2, 5)
- nilai (halaman: 2, 3, 4)
- penulis (halaman: 2, 5)
- ringkasan (halaman: 2, 4, 5)
- alinea (halaman: 2)
- sumber (halaman: 2, 5)
- menampilkan (halaman: 2, 5)
- kesimpulan (halaman: 2, 5)
- peringkasan (halaman: 2, 4)
- diharapkan (halaman: 2, 5)
- paragraf (halaman: 2)
- pengertian (halaman: 2)
- tulisan (halaman: 2)
- berbahasa (halaman: 2, 4, 5)
- jenis (halaman: 2, 3)
- kebutuhan (halaman: 2, 5)
- sesuai (halaman: 2, 5)
- invers (halaman: 2)
- mengolah (halaman: 2, 5)
- salah (halaman: 2, 3)
- ide (halaman: 2, 3)
- mengutip (halaman: 2, 5)
- membaca (halaman: 2)
- latar (halaman: 2)
- berisi (halaman: 2, 3, 4)
- bidang (halaman: 2, 3)
- frekuensi (halaman: 3)
- teknik (halaman: 3)
- stemming (halaman: 3, 4)
- kemunculan (halaman: 3, 4)
- muncul (halaman: 3, 4)
- tokenizing (halaman: 3, 4)
- string (halaman: 3, 4)
- retrieval (halaman: 3)
- stoplist (halaman: 3, 4)
- istilah (halaman: 3, 4)
- imbuhan (halaman: 3)
- input (halaman: 3)
- mengandung (halaman: 3, 5)
- berdasarkan (halaman: 3, 4)
- sistem (halaman: 4, 5)
- 1998 (halaman: 4, 5)
- tanda (halaman: 4)
- pengurutan (halaman: 4)
- program (halaman: 5)

2. Metode RAKE:

- core search process automatically sentence was originally a form of document that will then be in the fo
- com abstract search this sentence is automatically nuclei using tetx mining and weighting document to f
- idf), result obtained from this research is the application that is able to produce a summary document o

- this research is the development of an application that generates an automatic search phrase core document
- document in this study is a collection of sentences obtained from a sports site on the internet (halaman: 1)
- automatic search in a single document the core sentences indonesian language 1syamsul arifin (halaman: 1)
- the method used in this study is the term frequency (halaman: 1)
- in roomates to calculate the weight of each word (halaman: 1)
- a fakultas teknik universitas muhammadiyah jember email (halaman: 1)
- by applying text mining and term frequency (halaman: 1)
- core sentence (halaman: 1)
- 0910651057), 2lutfi ali muharom s (halaman: 1, 1)
- fakultas teknik universitas muhammadiyah jember email (halaman: 1)
- inverse document frequency (halaman: 1, 1)
- com pencarian kalimat inti (halaman: 1)
- tetx mining (halaman: 1)
- automatically (halaman: 1)
- abstrak 1syamsul arifin (halaman: 1)
- menghasilkan pencarian otomatis kalimat inti dokumen (halaman: 1)
- search (halaman: 1)
- kom jurusan teknik informatika (halaman: 1)
- kom jurusan teknik informatik (halaman: 1)
- proses pencarian kalimat inti (halaman: 1)
- menerapkan text mining (halaman: 1)
- text mining (halaman: 1, 1, 3, 3, 3, 3)
- menghasilkan summarization dokumen (halaman: 1, 5)
- internet (halaman: 1)
- 3 ratih ayuninghem (halaman: 1, 1)
- menemukan kalimat inti (halaman: 1)
- frekuensi dokumen invers (halaman: 1, 1)
- term frekuensi (halaman: 1, 1)
- kalimat inti (halaman: 1, 1, 2, 2)
- idf (halaman: 1, 1, 1, 1, 1, 4, 4, 5, 5)
- pembobotan dokumen (halaman: 1)
- dokumen tunggal (halaman: 1)
- dokumen berita (halaman: 1, 5)
- situs olahraga (halaman: 1, 2)
- menghitung bobot (halaman: 1, 2, 4)
- bahasa indonesia (halaman: 1, 2, 5)
- pencarian (halaman: 1, 4, 5)
- s (halaman: 1, 1)
- dokumen (halaman: 1, 1, 1, 4, 4, 4, 4, 4, 4, 5)
- otomatis (halaman: 1, 1, 1)
- kalimat (halaman: 1, 1, 1, 1, 1, 1, 3, 3, 4, 4)
- yahoo (halaman: 1, 1)
- tf (halaman: 1, 1, 1, 1, 1, 1, 4, 5)
- syamsul_arifin19 (halaman: 1, 1)
- st (halaman: 1, 1)
- si (halaman: 1, 1)
- pengembangan (halaman: 1)
- penelitian (halaman: 1, 1, 1, 1, 3, 4)
- metode (halaman: 1, 2, 3, 3, 5)
- m (halaman: 1, 1)
- kunci (halaman: 1)
- kumpulan (halaman: 1)
- keyword (halaman: 1)
- hasil (halaman: 1, 4, 4, 5)

- diperoleh (halaman: 1)
- bentuk (halaman: 1)
- aplikasi (halaman: 1, 1, 4)
- metode term frequency – invers document frequency (halaman: 2)
- tinjauan pustaka kalimat inti kalimat inti (halaman: 2)
- tahap penulisan artikel meliputi latar (halaman: 2)
- pencarian otomatis kalimat inti (halaman: 2, 2, 4, 5)
- kalimat topik bertugas memberitahukan (halaman: 2)
- dokumen tunggal berbahasa indonesia (halaman: 2)
- menangani kesalahan penulisan (halaman: 2)
- mencari kalimat inti (halaman: 2)
- mudah memahami makna (halaman: 2)
- jenis sumber apapun (halaman: 2)
- merupakan pandangan mendalam (halaman: 2)
- kamus bahasa indonesia (halaman: 2)
- data set mengambil (halaman: 2)
- ide pokok penulis (halaman: 2)
- menghindari pembacaan teks (halaman: 2, 2)
- tulisan kategori views (halaman: 2)
- melayani list dokumen (halaman: 2)
- tujuan penelitian a (halaman: 2)
- manfaat penelitian a (halaman: 2)
- pembaca gagasan pokok (halaman: 2)
- penulis artikel bertujuan (halaman: 2)
- mengambil sumber informasi (halaman: 2, 2)
- berfungsi menghubungkan alinea (halaman: 2)
- inti kalimat (halaman: 2)
- kalimat topik (halaman: 2)
- sumber tertulis (halaman: 2)
- relevan mengambil (halaman: 2)
- pendapat penulis (halaman: 2)
- pendahuluan latar (halaman: 2)
- laporan penelitian (halaman: 2)
- kalimat utama (halaman: 2, 2)
- pandangan), (halaman: 2)
- metode tf (halaman: 2, 2)
- data keuangan (halaman: 2)
- ide utama (halaman: 2)
- berisi ide (halaman: 2)
- bahasa latin (halaman: 2)
- situs goal (halaman: 2)
- sengaja dihilangkan (halaman: 2)
- salah satunya (halaman: 2, 3)
- ringkas sesuai (halaman: 2, 2, 5)
- proses pembuktian (halaman: 2)
- penilaian penulisnya (halaman: 2)
- penentu bobot (halaman: 2)
- mengurutkan bobot (halaman: 2)
- menawarkan pemecahan (halaman: 2)
- memiliki kesatuan (halaman: 2)
- media online (halaman: 2)
- media cetak (halaman: 2)
- bidang komputasi (halaman: 2)
- teks mining (halaman: 2, 2)

- menyerap informasi (halaman: 2, 2)
- menyebabkan informasi (halaman: 2)
- menghasilkan informasi (halaman: 2)
- ketersediaan informasi (halaman: 2)
- informasi tersedia (halaman: 2)
- informasi sejarah (halaman: 2)
- bersifat tulisan (halaman: 2)
- mengolah dokumen (halaman: 2, 2)
- dokumen berasal (halaman: 2)
- artikel artikel (halaman: 2)
- surat resmi (halaman: 2)
- surat perjanjian (halaman: 2)
- surat negara (halaman: 2)
- surat kabar (halaman: 2)
- perkembangan teknologi (halaman: 2)
- peninggalan terlukis (halaman: 2)
- membantu pembaca (halaman: 2, 2)
- louis gottschalk (halaman: 2)
- kesaksian lisan (halaman: 2)
- dibutuhkan pembobotan (halaman: 2)
- dokumen artikel (halaman: 2, 5, 5)
- isi dokumen (halaman: 2, 2)
- menjadikan ringkasan (halaman: 2)
- peringkasan dokumen (halaman: 2)
- kebutuhan pembaca (halaman: 2, 2, 5)
- term (halaman: 2, 2, 4)
- tahap (halaman: 2, 5, 5)
- petilasan arkeologis (halaman: 2)
- diperoleh rumusan (halaman: 2)
- isi paragraf (halaman: 2)
- bahasa inggris memiliki karakteristik tersendiri (halaman: 3)
- buku the text mining handbook (halaman: 3)
- tahap pemotongan string input berdasarkan (halaman: 3)
- langkah pencarian otomatis kalimat inti (halaman: 3)
- metode penelitian desain sistem model (halaman: 3)
- dicari kalimat intinya 2 (halaman: 3)
- inversedocument frequency metodeterm frequency (halaman: 3)
- dokumen tunggaltiap kalimat dianggap (halaman: 3)
- text mining text mining (halaman: 3)
- pencarian kalimat inti (halaman: 3, 5)
- dokumen menjalani proses tokenizing (halaman: 3)
- gambar desain sistem (halaman: 3)
- pengaruh tata bahasanya (halaman: 3)
- natural language processing (halaman: 3)
- menggali informasi dimana (halaman: 3)
- dianggap subyek riset (halaman: 3)
- banyaknya klaimdimana suatukata (halaman: 3)
- tokenizing penggunaan istilah (halaman: 3)
- text mining mengadopsi (halaman: 3)
- menganalisa unstructured teks (halaman: 3)
- memecah string teks (halaman: 3)
- proses pemotongan dokumen (halaman: 3)
- adalahcarapemberian bobot hubungan (halaman: 3)
- menjalani proses stemming (halaman: 3)

- mengalikan nilai tf (halaman: 3)
- proses stemming bertujuan (halaman: 3)
- proses merubah huruf (halaman: 3)
- pengumpulan data statistik (halaman: 3)
- proses pemecahan dokumen (halaman: 3)
- tahap tokenizing (halaman: 3)
- inversedocument frequency (halaman: 3)
- berdasarkan frekuensi (halaman: 3)
- proses tokenizing (halaman: 3)
- string alphanumerik (halaman: 3, 3)
- data mining (halaman: 3, 3, 3)
- pemilahan kalimat (halaman: 3)
- document frequency (halaman: 3)
- kegiatan riset (halaman: 3)
- proses stemming (halaman: 3, 3)
- term frequency (halaman: 3, 3)
- proses penyaringan (halaman: 3, 3)
- proses casefolding (halaman: 3)
- penyimpanan teks (halaman: 3)
- konten teks (halaman: 3)
- user berinteraksi (halaman: 3)
- unit terkecil (halaman: 3, 3)
- tool analisis (halaman: 3)
- sepenuhnya sempurna (halaman: 3)
- perhitungan bobot (halaman: 3)
- mengidentifikasikan isi (halaman: 3)
- machine learning (halaman: 3)
- information retrieval (halaman: 3)
- frekuensi kemunculan (halaman: 3, 3)
- d). (halaman: 3)
- bersifat suffix (halaman: 3)
- bentuk dasarnya (halaman: 3)
- analisa konten (halaman: 3)
- 2006]. (halaman: 3)
- 2005). (halaman: 3)
- 2001], (halaman: 3)
- sekumpulan dokumen (halaman: 3)
- pendefinisian dokumen (halaman: 3)
- menginput dokumen (halaman: 3)
- memilah dokumen (halaman: 3)
- frekuensi dokumen (halaman: 3)
- dokumen dansemakin (halaman: 3)
- dokumen 4 (halaman: 3)
- dokumen 3 (halaman: 3)
- menghilangkan imbuhan (halaman: 3)
- bidang khusus (halaman: 3)
- berisi kumpulan (halaman: 3)
- didapatkan jenis (halaman: 3)
- mengandung imbuhan (halaman: 3)
- langkah (halaman: 3)
- proses (halaman: 3, 3, 3, 4, 4)
- teks (halaman: 3)
- stemming (halaman: 3, 4)
- statistik (halaman: 3)

- huruf (halaman: 3)
- gambar (halaman: 3)
- tampilan sistem pencarian otomatis kalimat inti dokumen tunggal berbahasa indonesia (halaman: 4)
- gambar sistem pencarian otomatis kalimat inti gambar diatas (halaman: 4)
- $= \log$ (halaman: 4)
- sistem pencarian otomatis kalimat inti (halaman: 4)
- bahasa pemrograman microsoft visual basic (halaman: 4)
- pembahasan pencarian otomatis kalimat inti (halaman: 4)
- kalimat menggunakan fungsi split () (halaman: 4)
- gambar proses text mining proses tf (halaman: 4)
- pengembangan aplikasi peringkasan teks otomatis (halaman: 4)
- sistem pencarian kalimat inti (halaman: 4)
- tanda titik ".", tanda (halaman: 4)
- df pembobotan diperoleh berdasarkan (halaman: 4)
- memotong string dokumen 5 (halaman: 4)
- metode tf – idf (halaman: 4)
- nilai w terbesar dijadikan (halaman: 4)
- gambar pembobotan tf (halaman: 4)
- proses text mining (halaman: 4, 4, 5)
- tanda seru "!" (halaman: 4)
- dimana dokumen dipecah (halaman: 4)
- dihitung nilai kemunculannya (halaman: 4)
- kalimat yang berisi term (halaman: 4)
- pembobotan tf (halaman: 4, 5)
- inilah kalimat (halaman: 4)
- kalimat dalam dokumen (halaman: 4)
- kalimat 9 (halaman: 4)
- proses stoplist (halaman: 4)
- proses pengurutan (halaman: 4)
- proses mengembalikan (halaman: 4)
- nilai kumulatif (halaman: 4)
- nilai idf (halaman: 4, 4)
- term dihitung (halaman: 4)
- single dokumen (halaman: 4)
- idf). (halaman: 4)
- d 7 (halaman: 4)
- bentuk dasar (halaman: 4)
- kemunculan term (halaman: 4, 4)
- pembahasan (halaman: 4)
- w (halaman: 4, 4)
- kemunculan (halaman: 4)
- bobot (halaman: 4)
- "?" (halaman: 4)
- tokenizing (halaman: 4)
- terbentuk (halaman: 4)
- t (halaman: 4)
- sorting (halaman: 4)
- ringkasan (halaman: 4)
- relevan (halaman: 4)
- persamaan (halaman: 4, 4)
- output (halaman: 4)
- n (halaman: 4)
- mustaqfiri (halaman: 4)
- muncul (halaman: 4, 4)

- membuang (halaman: 4)
- kalaimat (halaman: 4)
- istilah (halaman: 4, 4, 4)
- grossman (halaman: 4)
- ditemukan (halaman: 4)
- dinamis (halaman: 4)
- dilanjutkan (halaman: 4)
- dikembangkan (halaman: 4)
- dfi (halaman: 4)
- delimiter (halaman: 4)
- berbeda (halaman: 4)
- 8 (halaman: 4)
- 6 (halaman: 4)
- 2011 (halaman: 4)
- 1998 (halaman: 4)
- on theoretical arguments for idf ", journal of documentation (halaman: 5)
- essays in honour of karen spärck jones (halaman: 5)
- " bahasa indonesia mata kuliah pengembangan kepribadian (halaman: 5)
- 2002), penggunaan digital tree hibrida (halaman: 5)
- 2006), text mining application programming (halaman: 5)
- natural language processing and information retrieval (halaman: 5)
- 1998), charting a new course (halaman: 5)
- a primer of historical method (halaman: 5)
- m ., abidin z ., kusumawati (halaman: 5)
- " understanding inverse document frequency (halaman: 5)
- peringkasan teks otomatis berbahasa indonesia (halaman: 5)
- 1998), information retrieval (halaman: 5)
- 1986), understanding history (halaman: 5)
- institut teknologi sepuluh nopember (halaman: 5)
- terjemahan nugroho notosusanto). (halaman: 5)
- r .(2011). (halaman: 5)
- metode maximum marginal relevance (halaman: 5)
- dimbali 5 januarii 2012 (halaman: 5)
- berhasil mengambil sumber informasi (halaman: 5)
- a z ., (halaman: 5)
- aplikasi information retrieval (halaman: 5)
- gambar hasil pencarian kesimpulan (halaman: 5)
- otomatis kalimat inti (halaman: 5)
- algorithm and heuristics (halaman: 5)
- dokumen berbahasa indonesia (halaman: 5)
- mengubah produk teks (halaman: 5)
- berhasil mengolah data (halaman: 5)
- thomson learning inc (halaman: 5)
- kluwer academic publisher (halaman: 5)
- jurusan teknik informatika (halaman: 5)
- daftar pustaka arifin (halaman: 5)
- d ., ophir (halaman: 5)
- basis data saintek (halaman: 5)
- gambar pengurutan bobot (halaman: 5)
- hasil pencarian tahap (halaman: 5)
- m ., (halaman: 5)
- pengembangan tugas (halaman: 5)
- jones k (halaman: 5)
- new york (halaman: 5, 5)

- s ., (halaman: 5, 5, 5)
- penulis mengambil (halaman: 5)
- pengurutan bobot (halaman: 5)
- l ., (halaman: 5)
- f ., (halaman: 5)
- 2007). (halaman: 5)
- 2004). (halaman: 5)
- tata bahasa (halaman: 5)
- bahasa inggris (halaman: 5)
- menentukan hasil (halaman: 5)
- hasil percobaan (halaman: 5)
- uji coba (halaman: 5)
- ui press (halaman: 5)
- saran kesimpulan (halaman: 5)
- pt ". (halaman: 5)
- penulis menyarankan (halaman: 5)
- menghitung akurasi (halaman: 5)
- ejournal matics (halaman: 5)
- dapat membantu (halaman: 5)
- artikel asli (halaman: 5)
- menampilkan hasil (halaman: 5)
- menghasilkan program (halaman: 5)
- saran penulis (halaman: 5)
- pengembang diharapkan (halaman: 5)
- diharapkan pengembang (halaman: 5)
- isbn 1 (halaman: 5)
- bahasa (halaman: 5)
- 5 (halaman: 5)
- kesimpulan (halaman: 5)
- saran (halaman: 5)
- pengembang (halaman: 5)
- menampilkan (halaman: 5)
- program (halaman: 5, 5)
- pembaca (halaman: 5, 5)
- 1 (halaman: 5, 5)

3. Metode Word2Vec:

- Kata kunci: dokumen
 - > hasil (similarity: 0.42, halaman: 1, 2, 3, 4, 5)
 - > cari (similarity: 0.38, halaman: 1, 2, 3, 4, 5)
 - > kalimat (similarity: 0.38, halaman: 1, 2, 3, 4, 5)
 - > text (similarity: 0.35, halaman: 1, 3, 4, 5)
 - > mining (similarity: 0.34, halaman: 1, 2, 3, 4, 5)
 - > baca (similarity: 0.33, halaman: 2, 5)
 - > inti (similarity: 0.33, halaman: 1, 2, 3, 4, 5)
 - > bahasa (similarity: 0.33, halaman: 1, 2, 3, 4, 5)
 - > proses (similarity: 0.33, halaman: 1, 2, 3, 4)
 - > term (similarity: 0.32, halaman: 1, 2, 3, 4)
- Kata kunci: metode
 - > is (similarity: 0.20, halaman: 1)
 - > frequency (similarity: 0.19, halaman: 1, 2, 3, 5)

- > teknik (similarity: 0.18, halaman: 1, 3, 5)
- > term (similarity: 0.18, halaman: 1, 2, 3, 4)
- > inti (similarity: 0.17, halaman: 1, 2, 3, 4, 5)
- > hitung (similarity: 0.17, halaman: 1, 2, 3, 4, 5)
- > aplikasi (similarity: 0.17, halaman: 1, 4, 5)
- > indonesia (similarity: 0.16, halaman: 1, 2, 4, 5)
- > isi (similarity: 0.15, halaman: 2, 3, 4, 5)
- > sumber (similarity: 0.15, halaman: 2, 5)

- Kata kunci: frekuensi
Tidak ditemukan dalam model.