

ABSTRAK

Dalam merangkum suatu teks, dibutuhkan waktu yang cukup banyak oleh karena itu peringkasan teks otomatis diperlukan. Berbagai macam metode untuk melakukan peringkasan teks otomatis sudah tersedia, oleh sebab itu eksplorasi dilakukan untuk mengetahui lebih dalam mengenai peringkasan teks otomatis. 2 metode yang telah difokuskan yaitu TF - IDF dan Intersection, perbedaan besar dari kedua metode ini adalah, TF – IDF menggunakan perhitungan secara global sedangkan Intersection menggunakan perhitungan secara lokal. Tujuan dari eksplorasi ini adalah untuk mengetahui pembobotan lokal atau global yang lebih baik dalam meringkas dokumen, untuk mengetahui jawaban dari pertanyaan ini beberapa pengujian telah dilakukan, seperti melakukan survei dengan memperlihatkan hasil ringkasan dari kedua metode dan membandingkannya dengan abstrak asli dari dokumen yang di uji dan penilaian berdasarkan kata unik yang berada pada hasil ringkasan dan abstrak asli, menggunakan algoritma khusus yang telah dibuat sebelumnya untuk melakukan pengujian. Hasil pengujian memperlihatkan perbedaan yang tidak terlalu besar ,dan metode intersection bekerja dengan baik pada makna yang berfokus pada beberapa kalimat saja dikarenakan perhitungan lokal, sedangkan TF-IDF bekerja dengan baik pada makna yang tersebar luas pada beberapa paragraf, dan hasil kedua pengujian memperlihatkan hal yang sama yaitu semakin besar dokumen maka penilaian baik dari survei maupun kata unik menurun.

Kata kunci: intersection, kata, Peringkasan, teks, tf-idf

ABSTRACT

Summarizing a text is time-consuming, therefore automatic summarization is needed. A variety method to perform automatic summarization already exists, so the exploration has done to know more about automatic text summarization. 2 methods have been focused which is TF - IDF and Intersection, major differences of these two methods are TF - IDF using the calculation globally while Intersection calculation using the calculation locally. The purpose of this exploration is to determine the weighting locally or globally better in summarizing the document, to know the answers to these question some testing has been done, such as a survey that showing the summary results of both methods and compare them with abstract of the document and a calculation based on the unique words from the two results of the methods and abstract, using special algorithms that have been made previously to perform the testing. The test results show that the difference is not too great, and the method of intersection is well performed for focused points of certain sentences because of the local calculation, while TF-IDF is well performed for wider points of certain paragraphs because of the global calculation. And the results of the both tests show the same thing that the larger the document both survey as well as the unique word decreases in the value of rating.

Keywords: Intersection, summarization, term, text, TF – IDF

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITASI LAPORAN PENILITIAN.....	ii
PERNYATAAN PUBLIKASI LAPORAN PENILITIAN.....	iii
PRAKATA.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiv
DAFTAR NOTASI/ LAMBANG.....	xv
DAFTAR SINGKATAN	xviii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Pembahasan	2
1.4 Ruang Lingkup.....	2
1.5 Sumber Data.....	3
1.6 Sistematika Penyajian	3
BAB 2 KAJIAN TEORI	4
2.1 Ringkasan.....	4
2.2 Peringkasan teks Otomatis	4
2.2.1 Jenis-jenis Ringkasan	4
2.3 Kata	5
2.4 Kalimat.....	5

2.5 Paragraf	5
2.6 Term Frequency Inverse Document Frequency (TF-IDF)	6
2.7 Cosine Similarity.....	7
2.8 Intersection.....	7
2.9 PageRank	8
2.10 Contoh Program	8
BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	12
3.1 Analisis Masalah	12
3.1.1 Analisis Data Pengguna	13
3.2 Analisis Sistem.....	14
3.2.1 Tokenizing.....	15
3.2.2 Stopword	16
3.3 Perancangan Sistem	16
3.3.1 UI Design	17
3.3.2 Use Case Diagram.....	18
3.3.2.1 Penjelasan Use Case Diagram.....	19
3.3.3 Activity Diagram.....	19
3.3.3.1 Activity Diagram Memasukan Data	19
3.3.3.2 Activity Diagram Eksekusi Program	20
3.4 Contoh Algoritma	21
3.4.1 Contoh TF-IDF	21
3.4.2 Contoh Intersection	27
BAB 4 IMPLEMENTASI SISTEM	29
4.1 Arsitektur Sistem.....	29
4.1.1 Arsitektur Intersection.....	30
4.1.2 Arsitektur TF-IDF & Cosine Similarity.....	32

4.2 Penjelasan Algoritma	35
4.2.1 Algoritma Pemisahan Teks	37
4.2.2 Algoritma Pembuatan Grafik	37
4.2.2.1 Algoritma Intersection	39
4.2.2.2 Algoritma TF-IDF dan Cosine Similarity	41
4.2.2.2.1 Algoritma TF-IDF	45
4.2.2.2.2 Algoritma Cosine Similarity	46
4.2.3 Algoritma Perhitungan PageRank	48
4.2.3.1 Algoritma PageRank	52
4.2.4 Algoritma Penyusunan Kalimat	55
4.2.4.1 Penyusunan Kalimat	56
4.2.5 Algoritma Penilaian Kata Unik	58
4.2.5.1 Algoritma pencari kata unik	60
BAB 5 PENGUJIAN	61
5.1 Hasil Survei	61
5.1.1 Survei I	61
5.1.2 Survei II	68
5.1.3 Survei III	74
5.1.4 Kesimpulan Survei	80
5.2 <i>Benchmarking</i>	81
BAB 6 SIMPULAN DAN SARAN	84
6.1 Simpulan	84
6.2 Saran	84
DAFTAR PUSTAKA	1
RIWAYAT HIDUP PENULIS	3

DAFTAR GAMBAR

Gambar 2.1 Rumus Cosine Similarity	7
Gambar 2.2 Rumus Intersection.....	7
Gambar 2.3 Contoh program autosummarizer.....	9
Gambar 2.3 Contoh Program smmry	10
Gambar 2.5 Contoh Program 3 tools4noobs	11
Gambar 3.1 Hasil Survey Pembaca.....	13
Gambar 3.2 Hasil Survey Penulis	14
Gambar 3.3 Tokenizing.....	15
Gambar 3.4 Gambaran Rancangan Sistem.....	17
Gambar 3.5 Rancangan UI awal	18
Gambar 3.6 Usecase pengguna	18
Gambar 3.7 Activity Diagram pemasukan data	20
Gambar 3.8 Activity Diagram Eksekusi program.....	21
Gambar 3.9 Contoh Algoritma TF-IDF Cosine	23
Gambar 4.1 Intersection Diagram	30
Gambar 4.2 TF - IDF Diagram	33
Gambar 4.3 Config Variable	35
Gambar 4.4 Algoritma pengambilan teks	36
Gambar 4.5 Susunan variable awal	36
Gambar 4.6 Function getSentences.....	37
Gambar 4.7 Function makeGraph	38
Gambar 4.8 Function IntersectionAlgorithm	41
Gambar 4.10 Funtion TfIdf And Cosine Similarity Algorithm	42
Gambar 4.11 Funtion TfIdf And Cosine Similarity Algorithm 2	43
Gambar 4.12 function TF.....	45
Gambar 4.13 Function IDF	45
Gambar 4.14 Function TFforQuery	46
Gambar 4.15 Function Cosine Similarity.....	47
Gambar 4.16 Function calculatePageRank 1	48
Gambar 4.17 Function calculatePageRank 2	49

Gambar 4.18 Function calculatePageRank 3	50
Gambar 4.19 Function calculatePageRank 4	50
Gambar 4.20 Keseluruhan function calculatePageRank	51
Gambar 4.21 fucntion runPageRankOnce 1	52
Gambar 4.22 function runPageRankOnce 2	54
Gambar 4.23 Final Code	55
Gambar 4.24 Function constructSummary 1	56
Gambar 4.25 Function constructSummary 2	57
Gambar 4.26 Algoritma Penilaian.....	59
Gambar 4.27 Function uniqueWord.....	60
Gambar 5.1 Contoh Survey I	62
Gambar 5.2 Grafik Survey I Intersection	63
Gambar 5.3 Table Survey I Intersection	64
Gambar 5.4 Grafik Survey I TF - IDF	65
Gambar 5.5 Table Survey I TF - IDF.....	66
Gambar 5.6 Grafik Survey Final	67
Gambar 5.7 Table Survey I Final.....	68
Gambar 5.8 Contoh Survey II	68
Gambar 5.9 Grafik Survey II Intersection.....	69
Gambar 5.10 Table Survey II Intersection	70
Gambar 5.11 Grafik Survey II TF - IDF	71
Gambar 5.12 Tabel Survey II TF - IDF	72
Gambar 5.13 Grafik Survey II Final	73
Gambar 5.14 Table Survey II Final	74
Gambar 5.15 Contoh Survey III.....	74
Gambar 5.16 Grafik Survey III Intersection	75
Gambar 5.17 Table Survey III Intersection.....	76
Gambar 5.18 Grafik Survey III TF – IDF	77
Gambar 5.19 Table Survey III TF – IDF	78
Gambar 5.20 Grafik Survey III Final	79
Gambar 5.21 Table Survey III	80
Gambar 5.22 Kesimpulan Survey	80

Gambar 5.22 Tabel Testing Algorithm	81
Gambar 5.24 Tabel Testing Algorithm	82

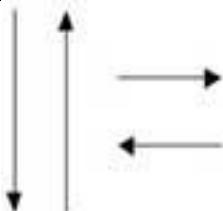


DAFTAR TABEL

Table 3.1 Penjelasan <i>usecase</i> memasukan data	19
Table 3.2 Penjelasan <i>usecase</i> mengeksekusi program.....	19



DAFTAR NOTASI/ LAMBANG

Gambar	Nama	Keterangan
	<i>Flow Direction symbol</i>	Simbol ini digunakan untuk menghubungkan satu simbol dengan lainnya
	<i>Terminator Symbol</i>	Simbol ini menandakan dimulainya proses atau berakhirnya proses terakhir
	<i>Initial State</i>	Kondisi awal sebuah objek sebelum ada perubahan keadaan
	<i>Processing Symbol</i>	Simbol yang menunjukkan pengolahan yang dilakukan oleh computer
	<i>Association line</i>	Fungsi yang hanya menggambarkan keterlibatan <i>actor</i> dengan <i>use case</i>
	<i>Manual Operation Symbol</i>	Simbol yang menunjukkan pengolahan suatu data yang dilakukan secara manual

	<i>System Boundary</i>	Batas gambaran antara sistem dan <i>actor</i>
	<i>Manual Input Symbol</i>	Simbol untuk pemasukan data secara manual on-line keyboard
	<i>Document Symbol</i>	Simbol ini mengambarkan sebuah dokumen input maupun output
	<i>Stored Data Symbol</i>	Mewakili suatu media penyimpanan atau biasa disebut <i>database</i>
	<i>Actor</i>	Orang/sistem yang berkaitan dan berinteraksi dengan program
	<i>Final State</i>	Kondisi akhir ketika objek berhenti memberikan respon
	<i>State</i>	Kondisi sebuah entitas
	<i>Transition</i>	Perubahan kondisi suatu objek yang disebabkan oleh suatu event

◊	<i>Decision</i>	Kondisi percabangan yang diharuskan memilih satu kondisi
---	-----------------	--



DAFTAR SINGKATAN

TF-IDF	Term Frequency Inverse Document Frequency
--------	---

