

ABSTRAK

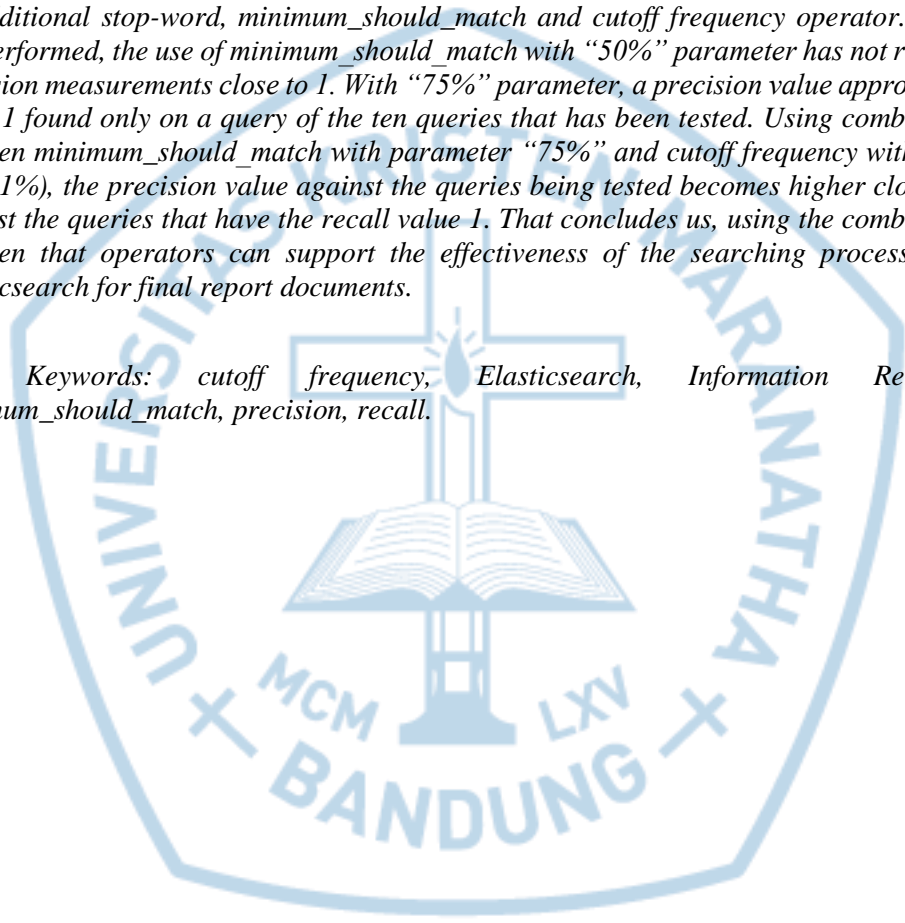
Informasi adalah kebutuhan primer di era teknologi seiring dengan kemajuan pesat perkembangan teknologi saat ini. Pada saat ini telah banyak dikembangkan *framework-framework* yang dapat digunakan untuk mengelola dokumen dan melakukan pencarian teks untuk mendapatkan informasi. Tujuan temu – balik informasi (*Information Retrieval*) adalah bagaimana mendapatkan dokumen relevan dalam koleksi dokumen yang dikembalikan. Salah satu *framework* temu-balik informasi yang populer adalah *Elasticsearch* dan tujuan penelitian adalah untuk mengimplementasi *framework Elasticsearch* sebagai fitur pencarian pada pembuatan sistem pencarian dokumen laporan tugas akhir. Penerapan *Elasticsearch* di dalam sistem yang dikembangkan memerlukan konfigurasi-konfigurasi tambahan untuk meningkatkan nilai efektifitas pengukuran *precision* dan *recall* dalam pengujian pencarian. Konfigurasi yang dimaksud adalah dengan penggunaan *stop-word* tambahan, operator *minimum_should_match* dan *cutoff frequency*. Pada pengujian yang dilakukan, penggunaan operator *minimum_should_match* dengan parameter “50%” belum menghasilkan pengukuran *precision* mendekati nilai 1 (satu). Dengan parameter “75%” nilai *precision* yang mendekati 1 hanya ditemukan pada satu dari sepuluh *query* yang diujikan. Dengan menggunakan kombinasi antara *minimum_should_match* parameter “75%” dan *cutoff frequency* sebesar 0.01 (1%) nilai *precision* terhadap *query-query* yang diujikan menjadi lebih tinggi mendekati 1 terhadap *query-query* yang memiliki nilai *recall* 1. Hal tersebut menunjukkan kombinasi 2 (dua) operator tersebut dapat mendukung efektifitas *Elasticsearch* melakukan proses pencarian terhadap dokumen laporan tugas akhir.

Kata kunci: *cutoff frequency*, *Elasticsearch*, *minimum_should_match*, *precision*, *recall*, temu – balik informasi.

ABSTRACT

Information is a primary requirement in the technological era along with the rapid advances in current technological developments. Nowadays there are many library that can be used to manage documents and performing full – text search operation for obtaining information. The purpose of information retrieval is how to obtain relevant documents in the set of the documents returned. One of the most popular full – text search library is Elasticsearch and the purpose of this research is to implementing Elasticsearch as a search feature in a system for final report IR system. The application of Elasticsearch in the developed system requires additional configuration to improve the effectiveness of precision and recall measurements in search testing. That configuration is with the use of an additional stop-word, minimum_should_match and cutoff frequency operator. In the test performed, the use of minimum_should_match with “50%” parameter has not resulted precision measurements close to 1. With “75%” parameter, a precision value approaching value 1 found only on a query of the ten queries that has been tested. Using combination between minimum_should_match with parameter “75%” and cutoff frequency with value 0.01 (1%), the precision value against the queries being tested becomes higher close to 1 against the queries that have the recall value 1. That concludes us, using the combination between that operators can support the effectiveness of the searching process using Elasticsearch for final report documents.

Keywords: cutoff frequency, Elasticsearch, Information Retrieval, minimum_should_match, precision, recall.



DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
DAFTAR KODE PROGRAM	xv
DAFTAR SINGKATAN	xvi
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	5
2.1 Information Retrieval (IR)	5
2.1.1 Indexing Process	7
2.1.1.1 Stopword Removal	7
2.1.1.2 Stemming & Lemmatization	8
2.1.2 Information Retrieval Model	8

2.1.3 Vector Space Model.....	9
2.2 Evaluasi Information Retrieval	13
2.2.1 Precision & Recall.....	14
2.3 Apache Lucene.....	16
2.4 Apache Tika	17
2.5 Elasticsearch.....	17
BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	19
3.1 Analisis-Ruang Audio Visual Perpustakaan UKM.....	19
3.2 Gambaran Keseluruhan.....	20
3.2.1 Persyaratan Antarmuka Eksternal	20
3.2.2 Antarmuka Dengan Pengguna.....	20
3.2.3 Antarmuka Perangkat Keras	20
3.2.4 Antarmuka Perangkat Lunak	20
3.3 Arsitektur Pengembangan Sistem IR yang dibuat	20
3.4 Preprocessing Dokumen Laporan Tugas Akhir.	22
3.5 Pemodelan Perangkat Lunak.....	22
3.5.1 Aktivitas Cari Dokumen	23
3.5.2 Aktivitas Impor Dokumen.....	24
3.6 Desain Antarmuka.....	24
3.6.1 Manage Document	25
BAB 4 IMPLEMENTASI.....	27
4.1 Implementasi Struktur Mapping Data pada Elasticsearch	27
4.1.1 Analyzer	28
4.2 Implementasi Class	30
4.2.1 Class MainFrame	30
4.2.2 Class Diagram CollectionDocumentPanel.....	31

4.2.3 Class PDFConvertDialog	31
4.2.4 Class DataMahasiswaDialog.....	32
4.2.5 Class ImportDocumentDialog.....	32
4.2.6 Class TableModelDataMahasiswa	33
4.2.7 Class TableModelQueryResult	33
4.2.8 Class SearchElastic	34
4.2.8.1 Operator minimum_should_match	37
4.2.8.2 Operator Cutoff Frequency	38
4.2.9 Class IndexElastic	39
4.2.10 Class PDFFileReader	45
4.2.11 Class CSVFileReader.....	45
4.2.12 Class TextFileReader	47
4.2.13 Class DataKonfigurasi	47
4.2.14 Class DataMahasiswa	49
4.3 Sequence Diagram	49
4.3.1 Sequence Diagram Pencarian Dokumen.....	49
4.3.2 Sequence Diagram Indexing & Penghapusan Dokumen	50
4.3.3 Sequence Diagram Indexing Multi Dokumen.....	51
4.3.4 Sequence Diagram Konfigurasi Pencarian.....	52
4.4 Implementasi Tampilan.....	52
4.4.1 Tampilan Muka	53
4.4.2 Form Manage Document.....	56
4.4.3 Form Input Data Mahasiswa	57
4.4.4 Form Impor Multi Dokumen.....	58
4.4.5 Form Pengaturan Pencarian	60
4.4.6 Form Konverter PDF ke TXT	61

BAB 5 PENGUJIAN	62
5.1 Cara Pengujian	62
5.2 Evaluasi Pengujian	63
5.2.1 Evaluasi Standar	63
5.2.2 Evaluasi dengan Konfigurasi Minimum Should Match	64
5.2.3 Evaluasi dengan Konfigurasi Minimum Should Match dan Cutoff Frequency	66
5.3 Pengujian Black Box	68
5.3.1 Black Box Testing Form Muka (Pencarian)	68
5.3.2 Black Box Testing Form Manage Document	69
5.3.3 Black Box Testing Form Pengaturan Pencarian	70
5.3.4 Black Box Testing Form Konverter PDF ke TXT	71
5.3.5 Black Box Testing Form Input Data Mahasiswa	71
5.3.6 Black Box Testing Form Impor Multi Dokumen	72
BAB 6 SIMPULAN DAN SARAN	74
6.1 Simpulan	74
6.2 Saran	74
DAFTAR PUSTAKA	76
LAMPIRAN A Dokumen Pengujian Acak & Klasifikasi	A-1

DAFTAR GAMBAR

Gambar 1.1 Chart Pengunjung Ruang Audio Visual dari tahun 2011 s.d. tahun 2016	2
Gambar 2.1 Komponen Sistem IR [2].....	6
Gambar 2.2 Text processing phases in an IR system [3]	7
Gambar 2.3 Contoh Model Ruang Vektor [5]	11
Gambar 2.4 Komponen - komponen Aplikasi Pencarian.....	16
Gambar 2.5 Arsitektur Tika [1].....	17
Gambar 3.1 Flowchart Ruang Audio Visual.....	19
Gambar 3.2 Penggunaan Framework dengan Pembuatan Aplikasi	20
Gambar 3.3 Penggunaan Framework untuk Pengembangan Aplikasi.....	21
Gambar 3.4 Use Case Aplikasi Pencarian.....	23
Gambar 3.5 Activity Pencarian Dokumen	23
Gambar 3.6 Activity Impor Dokumen	24
Gambar 3.7 Tampilan Awal Sistem	25
Gambar 3.8 Form Manage Document.....	25
Gambar 3.9 Form Import Document.....	26
Gambar 3.10 Form Document Data	26
Gambar 4.1 Class Diagram Sistem Pencarian Tugas Akhir	30
Gambar 4.2 Class MainFrame	31
Gambar 4.3 Class CollectionDocumentPanel	31
Gambar 4.4 Class PDFConvertDialog	32
Gambar 4.5 Class DataMhsDialog.....	32
Gambar 4.6 Class ImportDocumentDialog.....	33
Gambar 4.7 Class TableModelDataMahasiswa	33
Gambar 4.8 Class TableModelQueryResult	34
Gambar 4.9 Class SearchElastic	34
Gambar 4.10 Class PDFFileReader	45
Gambar 4.11 Class CSVFileReader.....	46
Gambar 4.12 Class TextFileReader	47
Gambar 4.13 Class DataKonfigurasi	47

Gambar 4.14 Class DataMahasiswa.....	49
Gambar 4.15 Sequence Diagram – Cari Dokumen.....	50
Gambar 4.16 Sequence Diagram – Indeks Dokumen.....	51
Gambar 4.17 Sequence Diagram – Indeks Multi Dokumen.....	52
Gambar 4.18 Sequence Diagram – Atur Konfigurasi.....	52
Gambar 4.19 Form Tampilan Awal.....	53
Gambar 4.20 Form Hasil Pencarian – Tanpa Hasil.....	54
Gambar 4.21 Form Hasil Pencarian – Dengan Hasil.....	55
Gambar 4.22 Form Detail Dokumen.....	56
Gambar 4.23 Windows Explorer – Lokasi Dokumen dalam Bentuk pdf.....	56
Gambar 4.24 Form Manage Document.....	57
Gambar 4.25 Form Input Data Mahasiswa.....	58
Gambar 4.26 Form Impor Multi Dokumen.....	59
Gambar 4.27 JFileChooser untuk memilih dokumen CSV.....	59
Gambar 4.28 Form Impor Multi Dokumen – TXT.....	60
Gambar 4.29 Form Konfigurasi Pencarian.....	60
Gambar 4.30 Form Konverter PDF.....	61
Gambar 5.1 Chart Nilai Rata – Rata Precision Pengujian.....	68

DAFTAR TABEL

Tabel 2.1 Term Frequency dari Dokumen dan Keyword	12
Tabel 2.2 Hasil Perhitungan TF/IDF.....	12
Tabel 2.3 Tabel Hasil Penghitungan	13
Tabel 2.4 Tabel Kemungkinan Relevansi Dokumen [2].....	14
Tabel 4.1 Tabel Parameter Operator <code>minimum_should_match</code>	38
Tabel 5.1 Hasil Pencarian Standar	63
Tabel 5.2 Hasil Pencarian Standar dengan Limit Pencarian	64
Tabel 5.3 Hasil Pencarian menggunakan Operator Minimum Should Match	64
Tabel 5.4 Hasil Pencarian menggunakan Operator Minimum Should Match dengan Limit.....	65
Tabel 5.5 Hasil Pencarian menggunakan Operator Minimum Should Match (75%)	65
Tabel 5.6 Hasil Pencarian menggunakan Operator Minimum Should Match (75%) dengan limit.....	66
Tabel 5.7 Hasil Pencarian menggunakan Operator Minimum Should Match dan Cutoff Frequency	67
Tabel 5.8 Hasil Pencarian menggunakan Operator Minimum Should Match dan Cutoff Frequency dengan limit	67
Tabel 5.9 Tabel Pengujian Form Muka.....	68
Tabel 5.10 Tabel Pengujian Form Manage Document	70
Tabel 5.11 Tabel Pengujian Form Pengaturan Pencarian	70
Tabel 5.12 Tabel Pengujian Form Konverter PDF ke TXT	71
Tabel 5.13 Tabel Pengujian Form Input Data Mahasiswa.....	72
Tabel 5.14 Tabel Pengujian Form Impor Multi Dokumen	72

DAFTAR KODE PROGRAM

Kode Program 4.1 Struktur Mapping Laporan Tugas Akhir Mahasiswa	28
Kode Program 4.2 Konfigurasi Analyzer Laporan Tugas Akhir Mahasiswa pada Elasticsearch.....	29
Kode Program 4.3 Method search & searchAsList.....	35
Kode Program 4.4 Method searchDocument	36
Kode Program 4.5 method searchMatchQuery	37
Kode Program 4.6 Contoh Penggunaan Operator minimum_should_match.....	37
Kode Program 4.7 Contoh Penggunaan Operator Cutoff Frequency	39
Kode Program 4.8 Class IndexElastic.....	39
Kode Program 4.9 Method insertDataMahasiswa	40
Kode Program 4.10 Method insertMultipleDataTXT	42
Kode Program 4.11 Method insertMultipleDataPDF	43
Kode Program 4.12 Method executeBulk & copyPDFData	44
Kode Program 4.13 Method readPDF.....	45
Kode Program 4.14 Method searchData & getData.....	47
Kode Program 4.15 Method getTextContent.....	47
Kode Program 4.16 Method readFile.....	48
Kode Program 4.17 Method saveConfig.....	49

DAFTAR SINGKATAN

FP	Fakultas Psikologi
IR	<i>Information Retrieval</i>
RPL	Rekayasa Perangkat Lunak
UML	<i>Unified Modelling Language</i>
PDF	<i>Portable Document Format</i>
CSV	<i>Comma Separated Values</i>
UKM	Universitas Kristen Maranatha
TA	Tugas Akhir

