

***Document Retrieval Menggunakan Vector Space Model Untuk
Indonesian Question Answering System***

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Srata-1 Pada
Jurusan Teknik Informatika



Oleh:

Defita Auli Ramadhia
NIM: 09021381823078

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2023**

LEMBAR PENGESAHAN SKRIPSI

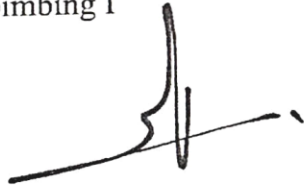
Document Retrieval Menggunakan Vector Space Model Untuk Indonesian Question Answering System

Oleh:

Defita Auli Ramadhia
NIM: 09021381823078

Palembang, 10 Januari 2023

Pembimbing I



Dr. Abdiansah., S.Kom., M.Cs
NIP. 198410012009121005

Pembimbing II,



Kanda Januar Miraswan, M.T.
NIP. 199001092019031012

Mengetahui,
Ketua Jurusan Teknik Informatika,



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari **Selasa** tanggal **20 Desember 2022** telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Defita Auli Ramadhia

NIM : 09021381823078

Judul : *Document Retrieval Menggunakan Vector Space Model Untuk Indonesian Question Answering System*

dan dinyatakan **LULUS**.

1. Ketua Penguji

Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003



2. Penguji I

Novi Yusliani, M.T.

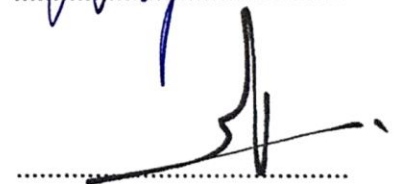
NIP. 19821108201212001



3. Pembimbing I

Dr. Abdiansah., S.Kom., M.Cs

NIP. 198410012009121005



4. Pembimbing II

Kanda Januar Miraswan, M.T.

NIP. 199001092019031012



Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.

NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini :

Nama : Defita Auli Ramadhia
NIM : 09021381823078
Program Studi : Teknik Informatika Bilingual
Judul Skripsi : *Document Retrieval Menggunakan Vector Space Model Untuk Indonesian Question Answering System*

Hasil pengecekan Software *iThenticate/Turnitin* : 18%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 10 Januari 2023



Defita Auli Ramadhia
NIM. 09021381823078

MOTTO DAN PERSEMBAHAN

“Your life is as good as your mindset”

“Allah tidak membebani seseorang melainkan sesuai dengan kemampuannya”

(Qs. Al-Baqarah: 286)

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan”

(Qs. Al-Insyirah ayat 5-6)

“Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu, Allah mengetahui sedang kamu tidak mengetahui ”

(Qs. Al-Baqarah: 216)

Kupersembahkan Karya Tulis ini Kepada:

- Orang tua dan Kakakku tercinta
- Dosen Pembimbing
- Sahabat dan Teman Seperjuangan
- Fakultas Ilmu Komputer Universitas Sriwijaya
- Diri sendiri

ABSTRACT

Document retrieval is a system search purposeful document for return relevant documents with query user. Component *document retrieval* have same function with machine seeker , that is gather desired information. *Document retrieval* is part from *Question Answering System* that uses algorithm tf-idf and calculations *cosine similarity*. One from system meet return information is *Indonesian Question Answering System (IQAS)* . System ask answer is something system that allows user state needs the information in question , then return quote text short as answer. Method *Vector Space Model* applied in study this method for represent system meet return information to in vector and account similarity function in the matching process a number of vector. Evaluation results on research this produce same value that is the accuracy value is 85% and the value precision 85%. Study this show that VSM can used for count level the resemblance that will sorted based on specified value .

Keyword : Document Retrieval, Vector Space Model, Indonesian Question Answering System

ABSTRAK

Document retrieval merupakan sebuah sistem pencarian dokumen yang bertujuan untuk mengembalikan dokumen yang relevan dengan kueri pengguna. Komponen *document retrieval* memiliki fungsi yang sama dengan mesin pencari, yaitu mengumpulkan informasi yang diinginkan. *Document retrieval* merupakan bagian dari *Question Answering System* yang menggunakan algoritma tf-idf dan perhitungan *cosine similarity*. Salah satu dari sistem temu kembali informasi adalah *Indonesian Question Answering System (IQAS)* . Sistem tanya jawab merupakan suatu sistem yang memungkinkan pengguna menyatakan kebutuhan informasinya dalam pertanyaan, kemudian mengembalikan kutipan teks singkat sebagai jawaban. Metode *Vector Space Model* diterapkan dalam penelitian ini metode untuk merepresentasikan sistem temu kembali informasi ke dalam vektor dan memperhitungkan fungsi similarity dalam proses pencocokan beberapa vektor. Hasil evaluasi pada penelitian ini menghasilkan nilai yang sama yaitu nilai accuracy 85% dan nilai preccision 85%. Penelitian ini menunjukkan bahwa VSM dapat digunakan untuk menghitung tingkat kemiripan yang akan diurutkan berdasarkan nilai yang ditentukan.

Kata Kunci: *Document Retrieval, Vector Space Model, Indonesian Question Answering System*

KATA PENGANTAR

Puji dan syukur atas kehadiran Allah Subhanahu Wa Ta'ala, atas segala nikmat, rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul “**Document Retrieval Menggunakan Vector Space Model Untuk Indonesian Question Answering System**” ini disusun untuk memenuhi salah satu syarat kelulusan tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari bahwa dalam pengerjaan Skripsi ini penulis banyak mendapatkan dukungan serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan banyak terimakasih yang tak terhingga kepada pihak yang telah banyak membantu dan mendukung, yaitu kepada:

1. Allah Subhanahu Wa Ta'ala telah memberikan penulis kesehatan, kekuatan, kesabaran, kemudahan, kelancaran dan kebaikan yang tak bisa dihitung dengan angka sekalipun.
2. Kedua orang tuaku tercinta Papa dan Mama, yang senantiasa tak henti-hentinya mendoakan penulis agar selalu diberi kemudahan dan terus memberikan penulis semangat di setiap harinya, serta dukungan yang begitu besar bagi penulis dalam menempuh pendidikan.
3. Kakakku Riski Septika Saputri yang senantiasa memberikan doa dan semangat bagi penulis.
4. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Alvi Syahrini Utami, M.Kom. selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Dr. Abdiansah, S.Kom., M.Cs dan Bapak Kanda Januar Miraswan, M.T. selaku Dosen Pembimbing Skripsi di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membimbing,

memberikan motivasi dan banyak membantu penulis dengan sabar.

7. Mba Wiwin selaku admin Jurusan Teknik Informatika yang telah membantu mengurus seluruh berkas.
8. Seluruh dosen dan staff Fakultas Ilmu Komputer Universitas Sriwijaya
9. Raisha Fathiya dan Kgs. M. Rusdiansyah Muharrom yang selalu memberikan motivasi dan banyak memberi bantuan dalam penulisan Skripsi ini.
10. Suna, Della, Prita, Kurnia, Tiansyah, Cindy Wijaya, Pretty, Arya Pradata dan Alredho yang telah bersedia meluangkan waktunya untuk menjawab pertanyaan dari penulis ketika sedang bingung dalam mengerjakan skripsi.
11. Kak Abdul Aziz Wicaksono, Nurmas, Gilang, Zora, Hana, Dinda, Mala yang membuat penulis selalu menguatkan mental ketika menghadapi masalah serta memberi ketenangan disaat penulis merasa sendiri.
12. Sahabatku Deska Srinadilla sekaligus bu psikolog yang menenangkan kondisi penulis disaat penulis berada dititik terendah dalam menulis skripsi.
13. Teman-teman seperjuangan Ti Bil B 2018 Roaina, Putri Pebreisnaini, Uswatun Khasana, Salsabila Nabrima Putri, Salsabela Maulina (Detektif Team), Syecky, Farid, Rifki, Ubaidillah, Ihtiar Alfath, Wahyu Dwi Saputra, Deny Septiawan dan yang tidak bisa disebutkan satu persatu.
14. Teman-teman BPH HMIF 2020 Kabinet Simfoni Fakultas Ilmu Komputer Universitas Sriwijaya.
15. Teman kost warna warni Novi, Sarah, Mifta, Nurul, Azhari, dan Arfi yang telah menghibur dan memberikan semangat dalam menulis skripsi.

Penulis juga mengucapkan terima kasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan Skripsi ini dan tidak dapat disebutkan satuper satu. Penulis berharap semoga tulisan ini dapat menjadi bahan pembelajaran dimasa yang akan datang. Penulis mohon maaf apabila terdapat kesalahan dan kekhilafan dalam penulisan laporan ini. Penulis

sangat mengharapkan kritik dan saran yang sifatnya membangun demi kesempurnaan tulisan ini.

Palembang, 10 Januari 2023

Defita Auli Ramadhia

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN TANDA LULUS	iii
HALAMAN PERNYATAAN	iv
HALAMAN MOTTO DAN PERSEMBAHAN.....	v
ABSTRACK	vi
ABSTRAK	xi
KATA PENGANTAR	xi
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvi
DAFTAR LAMPIRAN.....	xviii
BAB I PENDAHULUAN	1-1
1.1 Pendahuluan	1-1
1.2 Latar Belakang Masalah	1-1
1.3 Rumusan Masalah	1-2
1.4 Tujuan Penelitian.....	1-3
1.5 Manfaat Penelitian.....	1-3
1.6 Batasan Masalah.....	1-3
1.7 Sistematika Penulisan.....	1-4
1.8 Kesimpulan.....	1-5
BAB II KAJIAN LITERATUR	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Document Retrieval	II-1
2.2.2 Indonesian Question Answering System (IQAS)	II-1
2.2.2.1 Arsitektur Question Answering System.....	II-2
2.2.3 Vector Space Model.....	II-2

2.2.4	Information Retrieval (IR)	II-6
2.2.5	Text Preprocessing	II-6
2.2.6	Pemrograman Bahasa Python	II-7
2.2.7	Rational Unified Process (RUP)	II-8
2.3	Confusion Matrix	II-10
2.4	Penelitian Lain yang Relevan.....	II-11
2.5	Kesimpulan.....	II-13
BAB III METODE PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data	III-1
3.2.1	Jenis dan Sumber Data.....	III-1
3.3.2	Metode Pengumpulan Data.....	III-1
3.3	Tahapan Penelitian	III-2
3.3.1	Menentukan Kerangka Kerja	III-2
3.3.2	Menentukan Kriteria Pengujian	III-7
3.3.3	Menentukan Format Data Pengujian	III-7
3.3.4	Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian	III-7
3.3.5	Melakukan Pengujian Penelitian.....	III-8
3.3.6	Analisis Hasil Pengujian dan Membuat Kesimpulan.....	III-8
3.4	Metode Pengembangan Perangkat Lunak	III-9
3.4.1	Fase Insepsi.....	III-9
3.4.2	Fase Elaborasi	III-9
3.4.3	Fase Konstruksi.....	III-10
3.4.4	Fase Transisi	III-10
3.5	Kesimpulan.....	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK		IV-1
4.1	Pendahuluan	IV-1
4.2	Fase Insepsi	IV-1
4.2.1	Pemodelan Bisnis.....	IV-1

4.2.2	Kebutuhan Sistem	IV-1
4.2.3	Analisis dan Desain.....	IV-2
4.2.3.1	Analisis Kebutuhan Perangkat Lunak.....	IV-3
4.2.3.2	Analisis Data.....	IV-3
4.2.3.3	Analisis Text Preprocessing.....	IV-3
4.2.3.4	Analisis TF-IDF (Term Frequency-Inverse Document Frequency)	IV-8
4.2.3.5	Analisis VSM (Vector Space Model)	IV-11
4.2.3.6	Analisis Confusion Matrix.....	IV-17
4.2.3.7	Desain Perangkat Lunak	IV-17
4.3	Fase Elaborasi	IV-27
4.3.1	Pemodelan Bisnis.....	IV-27
4.3.2	Perancangan Data.....	IV-27
4.3.3	Perancangan Antarmuka	IV-27
4.3.4	Kebutuhan Sistem	IV-29
4.3.5	Diagram Aktivitas.....	IV-29
4.3.6	Diagram <i>Sequence</i>	IV-33
4.4	Fase Konstruksi.....	IV-38
4.4.1	Kebutuhan Sistem	IV-38
4.4.2	Diagram Kelas	IV-38
4.4.3	Implementasi.....	IV-40
4.4.3.1	Implementasi Kelas.....	IV-40
4.4.3.2	Implementasi Antarmuka.....	IV-42
4.5	Fase Transisi.....	IV-43
4.5.1	Pemodelan Bisnis.....	IV-43
4.5.2	Rencana Pengujian.....	IV-44
4.5.2	Implementasi.....	IV-44
4.6	Kesimpulan.....	IV-46
BAB V HASIL DAN ANALISIS PENELITIAN.....		V-1
5.1	Pendahuluan	V-1

5.2	Hasil Program.....	V-1
5.3	Hasil Percobaan.....	V-4
5.3.1	Skenario Percobaan.....	V-4
5.3.2	Hasil Pengujian.....	V-4
5.4	Analisis Hasil Pengujian.....	V-7
5.4.1	Analisis kinerja sistem <i>document retrieval</i> dengan menggunakan <i>Vector Spacel Model</i>	V-7
5.5	Kesimpulan.....	V-9
BAB VI KESIMPULAN DAN SARAN.....		VI-1
6.1	Pendahuluan.....	VI-1
6.2	Kesimpulan.....	VI-1
6.2	Saran.....	VI-1
DAFTAR PUSTAKA.....		xix

DAFTAR TABEL

Tabel II-1. Model <i>Confusion Matrix</i>	II-10
Tabel III-1. Format Data Pengujian.....	III-7
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Kebutuhan Non Fungsional.....	IV-2
Tabel IV-3. Contoh Data/Tokenisasi Kalimat.....	IV-4
Tabel IV-4. Proses <i>Case Folding</i>	IV-5
Tabel IV-5. Proses <i>Stopword Removal</i>	IV-6
Tabel IV-6. Hasil Proses <i>Stemming</i>	IV-8
Tabel IV-7. Hasil Perhitungan TF-IDF.....	IV-9
Tabel IV-8. Pembobotan Kata TF-IDF.....	IV-10
Tabel IV-9. Hasil Pembobotan TF dan IDF Seluruh Data.....	IV-12
Tabel IV-10. Hasil Pembobotan Seluruh Data.....	IV-13
Tabel IV-11. Perkalian Skalar antara Data dan Query.....	IV-14
Tabel IV-12. Contoh Perhitungan Vector.....	IV-15
Tabel IV-13. Hasil <i>Cosine Similarity</i>	IV-16
Tabel IV-14. Urutan Hasil <i>Cosine Similarity</i>	IV-17
Tabel IV-15. Definisi <i>Actor</i>	IV-19
Tabel IV-16. Definisi <i>Use Case</i>	IV-19
Tabel IV-17. Skenario <i>Use Case</i> Melakukan Pencarian Jawaban.....	IV-20
Tabel IV-18. Skenario <i>Use Case</i> Melakukan Pencarian Informasi diInternet.....	IV-21
Tabel IV-19. Skenario <i>Use Case</i> Melakukan Reset.....	IV-22
Tabel IV-20. Skenario <i>Use Case</i> Melakukan Manual.....	IV-23
Tabel IV-21. Skenario <i>Use Case</i> Melakukan Keluar.....	IV-24
Tabel IV-22. Skenario <i>Use Case</i> Melakukan Pengambilan Dataset.....	IV-25
Tabel IV-23. Skenario <i>Use Case</i> Melakukan Proses.....	IV-26
Tabel IV-24. Implementasi Kelas.....	IV-40
Tabel IV-25. Rencana Pengujian.....	IV-44
Tabel IV-26. Implementasi dan Rencana Pengujian.....	IV-45
Tabel V-1. Uji Coba Pertanyaan <i>Open Domain</i>	V-5
Tabel V-2. Uji Coba Pertanyaan <i>Close Domain</i>	V-6
Tabel V-3. Perhitungan nilai akurasi dan presisi.....	V-7

DAFTAR GAMBAR

Gambar II-1. Arsitektur Umum <i>Question Answering System</i>	II-2
Gambar II-2. Ilustrasi Dokumen dan <i>Query</i> dalam ruang <i>vector</i>	II-3
Gambar II-3. Matriks <i>term-document</i>	II-4
Gambar II-4. Rational Unified Process (RUP)	II-8
Gambar III-1. Tahapan Penelitian.....	III-2
Gambar III-2. Kerangka Kerja Perangkat Lunak.....	III-3
Gambar III-3. Diagram Tahapan <i>Text Preprocessing</i>	III-4
Gambar III-4. Diagram Pembobotan TF-IDF	III-5
Gambar III-5. Diagram <i>Similarity Measure</i>	III-6
Gambar IV-1. Diagram <i>Use Case Open Domain</i> Pencarian Jawaban untuk Pertanyaan Siapa.....	IV-19
Gambar IV-2. Diagram <i>Use Case Close Domain</i> Pencarian Jawaban untuk Pertanyaan Siapa.....	IV-19
Gambar IV-3. Antarmuka Tampilan 1 <i>Open Domain</i>	IV-29
Gambar IV-4. Antarmuka Tampilan 2 <i>Close Domain</i>	IV-30
Gambar IV-5. Diagram Aktivitas Melakukan Pencarian Jawaban	IV-31
Gambar IV-6. Diagram Aktivitas Proses Pencarian Informasi di Internet.....	IV-32
Gambar IV-7. Diagram Aktivitas Melakukan Reset	IV-32
Gambar IV-8. Diagram Aktivitas Membuka Manual	IV-33
Gambar IV-9. Diagram Aktivitas Melakukan Keluar	IV-33
Gambar IV-10. Diagram Aktivitas Melakukan Tombol Ambil.....	IV-34
Gambar IV-11. Diagram Aktivitas Melakukan Tombol Proses.....	IV-34
Gambar IV-12. Diagram <i>Sequence</i> Melakukan Pencarian Jawaban.....	IV-35
Gambar IV-13. Diagram <i>Sequence</i> Melakukan Proses Pencarian Informasi di Internet	IV-36
Gambar IV-14. Diagram <i>Sequence</i> Melakukan Reset.....	IV-37
Gambar IV-15. Diagram <i>Sequence</i> Membuka Manual	IV-37
Gambar IV-16. Diagram <i>Sequence</i> Melakukan Keluar.....	IV-38
Gambar IV-17. Diagram <i>Sequence</i> Melakukan Pengambilan Dataset.....	IV-38
Gambar IV-18. Diagram <i>Sequence</i> Melakukan Proses	IV-39
Gambar IV-19. Diagram Kelas	IV-41
Gambar IV-20. Implementasi Antarmuka Tampilan 1 <i>Open Domain</i>	IV-44
Gambar IV-21. Implementasi Antarmuka Tampilan 2 <i>Close Domain</i>	IV-45

Gambar V-1. Hasil Pengujian.....	V-2
Gambar V-2. Hasil Pengujian 2.....	V-3
Gambar V-3. Grafik Hasil Klasifikasi pada Jawaban Pertanyaan “Siapa”	V-6

DAFTAR LAMPIRAN

Lampiran 1. Log-Book Penelitian

Lampiran 2. Seluruh Data Pertanyaan Dan Jawaban

Lampiran 3. Seluruh Data Hasil Penelitian

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab pendahuluan akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan serta kesimpulan secara umum dalam penelitian. Bab ini juga berisikan penjelasan mengenai gambaran umum dari keseluruhan kegiatan yang dilakukan dalam penelitian.

1.2 Latar Belakang Masalah

Document retrieval adalah sebuah sistem pencarian dokumen yang bertujuan untuk mengembalikan dokumen yang relevan dengan kueri pengguna. Kueri adalah sekumpulan kata kunci (Monz, 2003). Komponen *document retrieval* memiliki fungsi yang sama dengan mesin pencari, yaitu mengumpulkan informasi yang diinginkan. *Document retrieval* merupakan bagian dari *Question Answering System* yang menggunakan algoritma tf-idf dan perhitungan *cosine similarity* (Isnurthina, 2016).

Sistem temu kembali informasi (*Information Retrieval*) saat ini banyak memiliki fitur *query* berupa pertanyaan sehingga informasi atau jawaban yang dikembalikan kepada pengguna akan lebih cepat dalam proses penemuan masalah. Salah satu dari sistem temu kembali informasi adalah *Question Answering System* (QAS) atau Sistem Tanya Jawab (Setiarini et al., 2020). Sistem tanya jawab merupakan suatu sistem yang memungkinkan pengguna menyatakan kebutuhan

informasinya dalam pertanyaan, kemudian mengembalikan kutipan teks singkat sebagai jawaban (Syauqi & Nurwahdah, 2015). *Indonesian Question Answering System* (IQAS), di sisi lain bertujuan untuk mengembalikan (tepat) menjawab pertanyaan (Monz, 2003).

Pada *Indonesian Question Answering System* diperlukan metode yang cocok diterapkan pada system *document retrieval*, salah satunya adalah dengan metode VSM (*Vector Space Model*). VSM adalah suatu metode untuk merepresentasikan sistem temu kembali informasi ke dalam vektor dan memperhitungkan fungsi similarity dalam proses pencocokan beberapa vektor. Suatu sistem temu kembali terdiri atas dua bagian, yaitu penyimpanan dokumen dan pemrosesan *query*, baik *query* maupun dokumen-dokumen yang disimpan dinyatakan dalam bentuk vector (Wicaksono et al., 2016). Hal ini juga yang menjadikan penulis menjadikan VSM sebagai metode penelitian ini karena cara kerja model ini efisien, mudah dalam representasi, dan dapat diimplementasikan pada *document-matching*. Dengan menggunakan *document retrieval* dengan metode VSM untuk pengambilan dokumen diharapkan dapat menangani pencarian data yang mana data yang dihasilkan cepat dan banyak.

1.3 Rumusan Masalah

Berdasarkan latar belakang tersebut dapat disimpulkan masalah penelitian ini adalah

1. Bagaimana membangun sistem *document retrieval* pada *Indonesian Question System Answering* ?

2. Bagaimana kinerja dari sistem *document retrieval* dengan menggunakan *Vector Spacel Model* ?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan sistem *document retrieval* untuk *Indonesian Question System Answering*.
2. Mengetahui kinerja dari sistem *document retrieval* dengan menggunakan *Vector Spacel Model*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Sistem ini dapat digunakan sebagai bagian dari sistem *Indonesian Question Answering System* pada komponen *document retrieval*.
2. Hasil penelitian dapat dijadikan sebagai rujukan untuk penelitian terkait di masa mendatang.

1.6 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset pengujian diambil menggunakan teknik web *scrapping*.
2. Data diambil dari hasil pencarian mesin pencari Google menggunakan beberapa kata kunci yang berbentuk pertanyaan.
3. Data yang digunakan merupakan open domain dan close domain.

4. Sistem hanya menerima kalimat tanya dalam bentuk pertanyaan “siapa”.
5. Data yang digunakan dalam penelitian ini disimpan dalam file berekstensi .csv.
6. Data yang digunakan terdiri dari dua kelas yaitu kelas positif dan kelas negatif.
7. Pengujian dilakukan menggunakan *confusion matrix* dan secara manual.
8. Metode pembobotan yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF).
9. Penelitian dikembangkan dengan bahasa pemrograman python.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini mengikuti panduan jurusan Teknik Informatika Universitas Sriwijaya sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat penelitian dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas dasar-dasar teori dari penelitian seperti definisi *Document Retrieval* penjelasan *Indonesian Question Answering System*, *Vector Space Model*, *Information Retrieval (IR)*, *text preprocessing*, Algoritma TF-IDF, bahasa pemrograman *python* dan *Rational Unified Process (RUP)*.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan membahas tahapan yang akan dilaksanakan pada dalam penelitian seperti pengumpulan data, analisis data dan perancangan sistem. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini berisi pembahasan mengenai setiap tahapan pengembangan perangkat lunak yang dilakukan. Metode yang digunakan adalah RUP (*Rational Unified Process*) yang terdiri atas proses Insepsi, Elaborasi, Konstruksi, dan Transisi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini berisi hasil pengujian pada perangkat lunak yang telah dikembangkan dan bab ini juga akan memaparkan pembahasan mengenai analisis dari hasil pengujian yang telah dilakukan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang dapat digunakan untuk penelitian selanjutnya.

1.8 Kesimpulan

Pada bab ini telah dibahas mengenai latar belakang penelitian serta acuan penting dalam penelitian seperti latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

DAFTAR PUSTAKA

- Abdiansah, & Hartati, S. (2015). Analisis Ekstraksi Pengetahuan Eksternal untuk Question Answering System. *Seminar Nasional Aplikasi Teknologi Informasi, June*, 29–38.
- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. *Konferensi Nasional Sistem Dan Informatika 2009, KNS&I09-036*, 196–201.
- Amin, F. (2013). Sistem Temu Kembali Informasi dengan Pemeringkatan Metode Vector Space Model. *Jurnal Teknologi Informasi DINAMIK, 18(2)*, 122–129.
<https://doi.org/10.22441/fifo.v9i1.1444>
- Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. (2009). *Safe-Level-SMOTE : Safe-Level-Synthetic Minority Over-Sampling TEchnique*. 475–476.
- Cao, Y., Xu, J., Liu, T. Y., Li, H., Huang, Y., & Hon, H. W. (2006). Adapting ranking SVM to document retrieval. *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006(49)*, 186–193.
<https://doi.org/10.1145/1148170.1148205>
- Curea, E. (2017). *Document retrieval and question answering in medical documents. A large-scale corpus challenge. September*, 1–7.
https://doi.org/10.26615/978-954-452-044-1_001
- Gumilar, M. D., Sembiring, F., & Erfina, A. (2021). Implementasi Progressive Web App Pada Sistem Informasi E-Learning Untuk. *Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi, Vol.10, No*, 187–196.

- Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. (2014). Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. *Proceedings - 2014 6th International Conference on Information Technology and Electrical Engineering: Leveraging Research and Technology Through University-Industry Collaboration, ICITEE 2014*, 0–3. <https://doi.org/10.1109/ICITEED.2014.7007894>
- Isnurthina, D. (2016). *Semantic Suffix Tree Clustering untuk Peningkatan Hasil Document Retriever pada Sistem Tanya Jawab Bahasa Indonesia*. 2(1), 392–398.
- Kadarina, T. M., & Hajar, M. H. I. (2019). Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara. *Jurnal Abdi Masyarakat (JAM)*, 5(1), 11. <https://doi.org/10.22441/jam.2019.v5.i1.003>
- Kurniawan, A., Solihin, F., & Hastarita, F. (2014). Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan Menggunakan Cosine Similarity. *Jurnal SimanteC*, 4(2 Desember), 115–124. <http://download.portalgaruda.org/article.php?article=433438&val=6202&title=PERANCANGAN DAN PEMBUATAN APLIKASI PENCARIAN INFORMASI BEASISWA DENGAN MENGGUNAKAN COSINE SIMILARITY>
- Mansouri, A., Affendey, L. S., Mamat, A., & Kadir, R. A. (2008). Semantically factoid question answering using fuzzy SVM named entity recognition.

Proceedings - International Symposium on Information Technology 2008, ITSim, 2. <https://doi.org/10.1109/ITSIM.2008.4631684>

Monz, C. (2003). Document retrieval in the context of question answering. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2633, 571–579. https://doi.org/10.1007/3-540-36618-0_44

Nurjannah, M., Hamdani, H., & Astuti, I. F. (2013). PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. *Jurnal Informatika Mulawarman*, 8(3), 110–113.

Sanjaya, F. (2017). Pemanfaatan Sistem Temu Kembali Informasi dalam Pencarian Dokumen Menggunakan Metode Vector Space Model. *J-INTECH (Journal Information and Technology)*, 05(02), 1689–1699.

Sati, B. A. B., Ali, M. A. S., & Abdou, S. M. (2016). Arabic Text Question Answering from an Answer Retrieval Point of View: a survey. *International Journal of Advanced Computer Science and Applications*, 7(7), 478–484. <https://doi.org/10.14569/ijacsa.2016.070766>

Setiarini, D., Gusmita, R. H., & Agustin, F. (2020). Sistem Tanya Jawab Menggunakan Kata Tanya Non-Factoid pada Terjemah Bahasa Indonesia Surat Al Baqarah. *Jurnal Linguistik Komputasional (JLK)*, 3(1), 25. <https://doi.org/10.26418/jlk.v3i1.31>

- Syahrudin, A. N., & Kurniawan, T. (2018). Input Dan Output Pada Bahasa Pemrograman Python. *Jurnal Dasar Pemrograman Python STMIK, January*, 1–7.
- Syauqi, A., & Nurwahdah, A. (2015). Sistem Tanya Jawab dengan Web Semantik. *Seminar Nasional Aplikasi Teknologi Informasi*, 43–51.
- Tia, T. K., & Kusuma, W. A. (2018). Model Simulasi Pengembangan Perangkat Lunak Menggunakan Rational Unified Process (Rup). *Teknika: Engineering and Sains Journal*, 2(1), 33. <https://doi.org/10.51804/tesj.v2i1.226.33-40>
- Turney, P. D., & Pantel, P. (2010). *From Frequency to Meaning: Vector Space Models of Semantics*. 37, 141–188.
- Wardani, P. K. (2017). Penerapan Metode Rational Unified Process pada Aplikasi Monitoring Periodic Service Alat Berat. *Indonesian Journal of Applied Informatics*, 1(2), 1. <https://doi.org/10.20961/ijai.v1i2.11002>
- Wicaksono, V. B., Saptono, R., & Sihwi, S. W. (2016). Analisis Perbandingan Metode Vector Space Model dan Weighted Tree Similarity dengan Cosine Similarity pada kasus Pencarian Informasi Pedoman Pengobatan Dasar di Puskesmas. *Jurnal Teknologi & Informasi ITSmart*, 4(2), 73. <https://doi.org/10.20961/its.v4i2.1768>