

**PENERAPAN *VECTOR SPACE MODEL* DALAM PENCARIAN  
DOKUMEN JURNAL BERBAHASA INDONESIA DENGAN *QUERY*  
BERUPA UCAPAN**



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Komputer  
Pada Departemen Ilmu Komputer/Informatika**

**Disusun Oleh:**

**ERZAN MIFTAH FARIDI**

**24010311120004**

**DEPARTEMEN ILMU KOMPUTER/INFORMATIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO  
2017**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini,

Nama : Erzan Miftah Faridi

NIM : 24010311120004

Judul : Penerapan *Vector Space Model* dalam Pencarian Dokumen Jurnal Berbahasa Indonesia dengan *Query* Berupa Ucapan

Dengan ini saya menyatakan bahwa dalam tugas akhir atau skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.



## HALAMAN PENGESAHAN

Judul : Penerapan *Vector Space Model* dalam Pencarian Dokumen Jurnal Berbahasa Indonesia dengan *Query* Berupa Ucapan

Nama : Erzan Miftah Faridi

NIM : 24010311120004

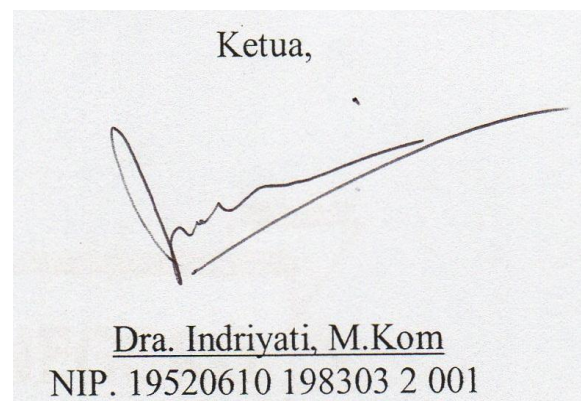
Telah diujikan pada sidang tugas akhir pada tanggal 18 Januari 2017 dan dinyatakan lulus pada tanggal 18 Januari 2017

Semarang, 18 Januari 2017

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika

Panitia Penguji Tugas  
Akhir



## HALAMAN PENGESAHAN

Judul : Penerapan *Vector Space Model* dalam Pencarian Dokumen Jurnal Berbahasa Indonesia dengan *Query* Berupa Ucapan


Nama : Erzan Miftah Faridi

NIM : 24010311120004

Telah diujikan pada sidang tugas akhir pada tanggal 18 Januari 2017

Semarang, 18 Januari 2017

Dosen Pembimbing,



Sukmawati Nur Endah, S.Si, M.Kom  
NIP.197805022005012002

## ABSTRAK

Pencarian informasi dengan keragaman dan banyaknya dokumen yang ada dapat ditangani dengan kajian – kajian yang ada pada *information retrieval*, khususnya penerapan pada mesin pencari. Mesin pencari yang berkembang sekarang kebanyakan masih menggunakan *query* berupa teks dan masih jarang yang menggunakan masukan *query* berupa ucapan bahasa Indonesia. Penelitian ini menerapkan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan. Sebelum melakukan pencarian dokumen, *query* ucapan dikenali dalam bentuk teks menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC) sebagai metode ekstraksi ciri dan *Hidden Markov Model* (HMM) untuk pengenalan ucapannya. Pengujian yang dilakukan meliputi pengujian pengenalan ucapan dan pengujian keakuratan mesin pencari dengan *query* berupa ucapan. Berdasarkan pengujian *10-fold cross validation* dengan 1.000 data, ucapan dapat dikenali sebagai teks dengan akurasi sebesar 89,4%. Sedangkan untuk pengujian keakuratan mesin pencari dilakukan dengan menghitung *precision recall* dari 5 dokumen teratas dengan jumlah dokumen 50 dan 8 *user*, didapatkan nilai *Mean Average Precision* (MAP) sebesar 0,904. Hasil pengujian tersebut mengindikasikan bahwa pengenalan ucapan sudah mampu digunakan sebagai masukan *query* untuk mesin pencari dan mesin pencari dapat menghasilkan dokumen ter-*retrieve* yang cukup relevan.

**Kata kunci** : mesin pencari, pengenalan ucapan, *vector space model*, pencarian dokumen jurnal.

## ABSTRACT

The increasing number and variety on information can be handled by the study of information retrieval which is being implemented in search engine. Nowadays most of developing search engines use text based query and its rare to see search engine using voice based query specially in Indonesian language. This research applied vector space model to search Indonesian journals with voice based query, this research used Mel Frequency Cepstral Coefficients (MFCC) as feature extraction process and Hidden Markov Model (HMM) as the learning method. This research used two types of test they were speech recognition test and the accuracy of search engine test. By using the 10-fold cross validation with 1,000 data, the accuracy of speech recognition was 89,4%. While to get the accuracy of search engine, the precision recall of top five document (from 50 document and 8 users) was counted and the result was Mean Average Precision (MAP) value 0,904. From the result of the test indicates that speech recognition was capable to be used as the query for search engine and the search engine can produce the relevant retrieved document.

**Keyword :** search engine, speech recognition, vector space model, journal speech recognition.

## KATA PENGANTAR

Puji syukur penulis haturkan kehadiran Allah SWT atas rahmat, anugerah, dan kesempatan yang diberikan-Nya penulis dapat menyelesaikan laporan tugas akhir.

Tugas akhir yang berjudul “**Penerapan *Vector Space Model* dalam Pencarian Dokumen Jurnal Berbahasa Indonesia dengan *Query* Berupa Ucapan**” disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Jurusan Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan tugas akhir ini penulis banyak mendapat bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terima kasih kepada:

1. Ragil Saputra, S.Si, M.Cs, selaku Ketua Departemen Ilmu Komputer / Informatika FSM UNDIP
2. Helmie Arif Wibawa, S.Si, M.Cs, selaku Koordinator Tugas Akhir Departemen Ilmu Komputer / Informatika
3. Sukmawati Nur Endah, S.Si, M.Kom, selaku dosen pembimbing
4. Semua pihak yang telah membantu hingga selesainya tugas akhir ini, yang tidak dapat penulis sebutkan satu persatu. Semoga Allah membalas segala kebaikan yang telah diberikan kepada penulis

Penulis menyadari bahwa dalam laporan ini masih banyak terdapat kekurangan baik dari penyampaian materi maupun isi dari materi itu sendiri. Hal ini dikarenakan keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan.

Semoga laporan tugas akhir ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, Januari 2017

Penulis

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiv
DAFTAR LAMPIRAN .....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan dan Manfaat .....	3
1.4 Ruang Lingkup .....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1 Penelitian Terdahulu Tentang Penerapan VSM untuk Pencarian Dokumen. ....	4
2.2 Sinyal Ucapan.....	4
2.2.1. Representasi Sinyal Ucapan .....	5
2.2.2. Karakteristik Sinyal Ucapan.....	7
2.2.2.1. FonemVokal.....	8
2.2.2.2. Fonem Diftong .....	8
2.2.2.3. Fonem Konsonan Nasal (Sengau) .....	9
2.2.2.4. Fonem Konsonan Frikatif (Geseran) .....	9
2.2.2.5. Fonem Konsonan Stop.....	10
2.3 Pengenalan Ucapan.....	10



2.3.1.	Pengertian Pengenalan Ucapan .....	10
2.3.2.	Parameter Pengenalan Ucapan .....	11
2.4	Pra Pemrosesan.....	13
2.5	<i>Mel-Frequency Cepstral Coefficient</i> (MFCC) .....	13
2.5.1.	Proses <i>Pre – emphasize Filtering</i> .....	14
2.5.2.	Proses <i>Frame blocking</i> .....	15
2.5.3.	Proses <i>Windowing</i> .....	16
2.5.4.	<i>Fast Fourier Transfer</i> (FFT).....	18
2.5.5.	Proses <i>Mel Frequency Wrapping</i> .....	19
2.5.6.	Proses <i>Discrete Cosine Transform</i> (DCT).....	21
2.5.7.	Proses <i>Cepstral Liftering</i> .....	22
2.6	<i>Hidden Markov Model</i> (HMM).....	23
2.7	Arsitektur Information Retrieval .....	25
2.7.1.	Tokenisasi .....	27
2.7.2.	Penghapusan <i>Stopword</i> .....	27
2.7.3.	Pembobotan.....	27
2.8	<i>Vector Space Model</i> .....	28
2.9	<i>Precision</i> dan <i>Recall</i> .....	32
2.10	<i>Average Precision</i> .....	33
2.11	<i>Mean Average Precision</i> .....	34
2.12.	<i>Unified Modelling Language</i> (UML).....	35
2.12.1	<i>Use case diagram</i> .....	35
2.12.2	<i>Class Diagram</i> .....	37
2.12.3	<i>Sequence Diagram</i> .....	40
2.12.4	<i>Activity Diagram</i> .....	41
2.13.	Model Proses Pengembangan Perangkat Lunak.....	43
2.14.	Metode <i>Unified Process</i> .....	44

BAB III METODE PENELITIAN .....	48
3.1    Arsitektur Sistem .....	48
3.2    Tahapan Penelitian.....	49
3.2.1.    Pengenalan Ucapan.....	51
3.2.2.    Pembuatan Mesin Pencari .....	66
3.3    Contoh Proses.....	70
BAB IV PENGEMBANGAN PERANGKAT LUNAK.....	91
4.1.    Fase <i>Inception</i> .....	91
4.1.1.    Kebutuhan Fungsional .....	91
4.1.2.    Model <i>Use Case</i> .....	91
4.2.    Fase <i>Elaboration</i> .....	94
4.2.1. <i>Design Model</i> .....	94
4.2.1.1. <i>Sequence Diagram</i> .....	94
4.2.1.2. <i>Class Diagram</i> .....	96
4.2.2.    Data Model.....	97
4.2.3.    Menyusun <i>Prototype</i> Antarmuka .....	98
4.3.    Fase <i>Construction</i> .....	100
4.3.1.    Spesifikasi Perangkat.....	100
4.3.2.    Implementasi Kelas .....	100
4.3.3.    Implementasi Antarmuka .....	101
4.3.4.    Menyusun Rencana Pengujian Fungsional .....	102
BAB V HASIL DAN ANALISA .....	104
5.1    Data Penelitian.....	104
5.1.1.    Data Penelitian Pengenalan Ucapan .....	104
5.1.2.    Data Penelitian Pencarian Dokumen .....	104
5.2    Skenario Pengujian .....	107
5.2.1    Pengujian Pengenalan Ucapan .....	107

5.2.2	Pengujian Pencarian Dokumen .....	107
5.3	Hasil Penelitian.....	108
5.3.1.	Hasil Penelitian Pengenalan Ucapan .....	108
5.3.2.	Hasil Penelitian Pencarian Dokumen .....	108
5.4	Analisa Hasil .....	110
BAB VI PENUTUP .....		113
6.1.	Kesimpulan.....	113
6.2.	Saran .....	113
DAFTAR PUSTAKA .....		115
LAMPIRAN .....		118

## DAFTAR GAMBAR

Gambar 2. 1	Contoh Sinyal Ucapan “It’s time” (Rabiner & Juang, 1993) .....	5
Gambar 2. 2	Spektogram Pita Lebar, Spektogram Pita Sempit dan Amplituda Ucapan dari kalimat “Every Salt Breeze Comes from Sea” (Rabiner & Juang, 1993).....	7
Gambar 2. 3	Bentuk Sinyal Ucapan Vokal Bahasa Indonesia pada suara pria (Mutiara, 2009).....	9
Gambar 2. 4	Bentuk Sinyal Ucapan Vokal Bahasa Indonesia pada suara wanita (Mutiara, 2009) .....	9
Gambar 2. 5	Block Diagram Untuk MFCC.....	14
Gambar 2. 6	Contoh dari pre-emphasize pada sebuah frame .....	15
Gambar 2. 7	Diagram Block Proses Penentuan Frame (Andriana, 2011).....	16
Gambar 2. 8	Sinyal Ucapan Domain Waktu menjadi Domain Frekuensi.....	19
Gambar 2. 9	Magnitude Dari Rectangular Dan Triangular Filterbank .....	19
Gambar 2. 10	Prinsip Frekuensi Wrapping .....	20
Gambar 2. 11	Triangular Filterbank dengan mel scale .....	21
Gambar 2. 12	Perbandingan Spektrum dengan dan tanpa Cepstral Liftering .....	22
Gambar 2. 13	Contoh Markov Chain .....	23
Gambar 2. 14	Evolusi temporal dari Hidden Markov Model.....	24
Gambar 2. 15	Proses Temu Balik Informasi Dokumen Teks.....	25
Gambar 2. 16	Proses pengindeksan pada suatu kata (Yates R, 1999) .....	26
Gambar 2. 17	Cosine dari $\theta$ didapatkan dari $d_j$ dan $q$ (Yates R, 1999).....	29
Gambar 2. 18	Contoh Use Case Diagram .....	37
Gambar 2. 19	Contoh Class Diagram.....	39
Gambar 2. 20	Contoh Sequence Diagram .....	41
Gambar 2. 21	Activity Diagram Surat Menyurat.....	43
Gambar 2. 22	Alur Kerja Unified Process.....	45
Gambar 2. 23	Siklus hidup unified process .....	45
Gambar 3. 1	Arsitektur Sistem Aplikasi.....	49
Gambar 3. 2	Diagram Blok Penelitian .....	50
Gambar 3. 3	<i>Flowchart</i> Proses <i>Input</i> Ucapan Untuk Pengenalan Kata. ....	52
Gambar 3. 4	<i>Flowchart</i> Proses <i>Input</i> Ucapan Untuk Data Pelatihan.....	52

Gambar 3. 5	<i>Flowchart MFCC</i> .....	54
Gambar 3. 6	<i>Flowchart DC-Removal</i> .....	55
Gambar 3. 7	<i>Flowchart Pre-Emphasize</i> .....	56
Gambar 3. 8	<i>Flowchart Frame Blocking</i> .....	57
Gambar 3. 9	<i>Flowchart Windowing</i> .....	58
Gambar 3. 10	<i>Flowchart FFT</i> .....	59
Gambar 3. 11	<i>Flowchart Filterbank</i> .....	60
Gambar 3. 12	<i>Flowchart DCT</i> .....	61
Gambar 3. 13	<i>Flowchart Cepstral Liftering</i> .....	62
Gambar 3. 14	<i>Flowchart Pelatihan HMM</i> .....	63
Gambar 3. 15	<i>Flowchart proses Pelatihan Parameter HMM</i> .....	65
Gambar 3. 16	<i>Flowchart Pengenalan Ucapan</i> .....	66
Gambar 3. 17	<i>Flowchart Pra Pemrosesan</i> .....	67
Gambar 3. 18	<i>Flowchart Pembobotan Kata</i> .....	68
Gambar 3. 19	<i>Flowchart Proses VSM</i> .....	69
Gambar 4. 1	Use Case Diagram penerapan vector space model dalam pencarian dokumen jurnal berbahasa Indonesia dengan query berupa ucapan .....	92
Gambar 4. 2	Sequence Diagram Melakukan Pelatihan .....	95
Gambar 4. 3	Sequence DiagramMenambah Data Ucapan .....	95
Gambar 4. 4	Sequence Diagram Melakukan Pencarian Dokumen .....	95
Gambar 4. 5	Sequence DiagramMenambah Data Dokumen .....	96
Gambar 4. 6	Class Diagram penerapan vector space model dalam pencarian dokumen jurnal berbahasa Indonesia dengan query berupa ucapan .....	96
Gambar 4. 7	Desain Antarmuka Halaman Home .....	98
Gambar 4. 8	Desain Antarmuka Halaman Pelatihan.....	99
Gambar 4. 9	Desain Antarmuka Halaman Upload Dokumen dan Pembobotan.....	99
Gambar 4. 10	Tampilan Halaman Home.....	101
Gambar 4. 11	Tampilan Halaman Pelatihan.....	102
Gambar 4. 12	Tampilan Halaman Upload Dokumen dan Pembobotan.....	102
Gambar 5. 1	Grafik <i>Average Precision</i> .....	110
Gambar 5. 2	Grafik <i>Mean Average Precision</i> Setiap User .....	110
Gambar 5. 3	(a) Sinyal Suara “algoritma genetika” tanpa noise. (b) Sinyal Suara “algoritma genetika” dengan noise. ....	111

## DAFTAR TABEL

Tabel 2. 1	Perkembangan Penelitian Tentang Penerapan VSM untuk Pencarian Dokumen.....	4
Tabel 2. 2	Fonem-fonem Bahasa Indonesia (Departemen Pendidikan Nasional, 2009). ..	7
Tabel 2. 3	Tabel Parameter Pengenalan Suara. ....	12
Tabel 2. 4	Nilai bobot kata dari contoh <i>vector space model</i> .....	30
Tabel 2. 5	Perhitungan <i>Recall</i> dan <i>Precision</i> .....	32
Tabel 2. 6	<i>User Judgement</i> .....	34
Tabel 2. 7	Komponen <i>Use Case Diagram</i> .....	36
Tabel 2. 8	Komponen <i>Class Diagram</i> .....	38
Tabel 2. 9	Komponen <i>Sequence Diagram</i> .....	40
Tabel 2. 10	Komponen <i>ActivityDiagram</i> .....	42
Tabel 3. 1	Hasil ekstraksi ciri data latih untuk sinyal ucapan “pengenalan” .....	75
Tabel 3. 2	Hasil ekstraksi ciri data latih untuk sinyal ucapan “suara” .....	75
Tabel 3. 3	Nilai Parameter $\Sigma$ untuk State 1 .....	76
Tabel 3. 4	Nilai Parameter B .....	76
Tabel 3. 5	Nilai $\alpha$ .....	78
Tabel 3. 6	Nilai $\beta$ .....	79
Tabel 3. 7	Nilai $\gamma$ .....	79
Tabel 3. 8	Nilai $\gamma_{\text{observasi}}$ pada State 1 .....	80
Tabel 3. 9	Nilai $\gamma_{\text{observasi}}$ pada State 2 .....	81
Tabel 3. 10	Nilai $\gamma_{\text{observasi}}$ pada State 3 .....	81
Tabel 3. 11	Nilai $\mu$ baru .....	82
Tabel 3. 12	Nilai Parameter $\Sigma$ baru pada State 1 .....	83
Tabel 3. 13	Nilai Parameter $\Sigma$ baru pada State 2 .....	83
Tabel 3. 14	Nilai Parameter $\Sigma$ baru pada State 3 .....	83
Tabel 3. 15	Nilai Parameter A setelah Pelatihan .....	85
Tabel 3. 16	Nilai Parameter $\Pi$ setelah Pelatihan .....	85
Tabel 3. 17	Nilai Parameter $\mu$ setelah Pelatihan .....	86
Tabel 3. 18	Nilai Parameter Sigma pada State 1 setelah Pelatihan .....	86
Tabel 3. 19	Nilai Parameter Sigma pada State 2 setelah Pelatihan .....	86
Tabel 3. 20	Nilai Parameter Sigma pada State 3 setelah Pelatihan .....	86

Tabel 3. 21	Contoh sinyal ucapan yang akan diuji .....	86
Tabel 3. 22	Nilai Parameter B berdasarkan Parameter $\mu$ dan $\Sigma$ .....	87
Tabel 3. 23	Nilai $\alpha$ sinyal ucapan terhadap Model .....	88
Tabel 3. 24	Nilai <i>likelihood</i> sinyal ucapan terhadap Model.....	88
Tabel 3. 25	Pembobotan TF-IDF.....	89
Tabel 4. 1	Kebutuhan Fungsional Aplikasi .....	91
Tabel 4. 2	Tabel Karakteristik Pengguna .....	92
Tabel 4. 3	Daftar Use Case.....	92
Tabel 4. 4	Use Case Detail untuk Melakukan Pelatihan .....	93
Tabel 4. 5	Use Case Detail untuk Menambah Data Ucapan .....	93
Tabel 4. 6	Use Case Detail untuk Melakukan Pencarian Dokumen .....	94
Tabel 4. 7	Use Case Detail untuk Menambah Data Dokumen .....	94
Tabel 4. 8	Struktur Folder data_pelatihan. ....	97
Tabel 4. 9	Struktur Folder database_pelatihan. ....	97
Tabel 4. 10	Struktur Folder database_jurnal. ....	97
Tabel 4. 11	Implementasi Class.....	100
Tabel 4. 12	Rencana Pengujian penerapan vector space model dalam pencarian dokumen jurnal berbahasa Indonesia dengan query berupa ucapan .....	103
Tabel 5. 1	Daftar Ucapan .....	104
Tabel 5. 2	Daftar <i>Query</i> .....	105
Tabel 5. 3	Daftar Dokumen yang Digunakan dalam Penelitian .....	105
Tabel 5. 4	Hasil Pengujian Pengenalan Ucapan .....	108
Tabel 5. 5	Hasil Perhitungan Nilai <i>Average Precision</i> .....	108
Tabel 5. 6	Hasil Nilai <i>Mean Average Precision</i> setiap <i>user</i> .....	109
Tabel 7. 1.	Deskripsi dan Hasil Uji Melakukan Pelatihan .....	119
Tabel 7. 2.	Deskripsi dan Hasil Uji Menambah Data Ucapan .....	119
Tabel 7. 3.	Deskripsi dan Hasil Uji Pencarian Dokumen .....	120
Tabel 7. 4.	Deskripsi dan Hasil Uji Menambah Dokumen .....	120
Tabel 7. 5.	Hasil Perhitungan Precision dan Recall.....	121

## DAFTAR LAMPIRAN

Lampiran 1. Pengujian dan Hasil Uji Fungsional Sistem .....	119
Lampiran 2. Hasil Evaluasi Vector Space Model.....	121
Lampiran 3. Source Code Implementasi.....	124



# BAB I

## PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, tujuan dan manfaat, dan ruang lingkup penelitian tugas akhir mengenai penerapan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan.

### 1.1 Latar Belakang

Peningkatan jumlah dan keragaman informasi yang beredar di internet, membuat pengguna semakin sulit mendapatkan informasi yang sesuai dengan yang dikehendaki. Kebutuhan penggunapun mulai bergeser dari yang dulunya mencari informasi secara kuantitatif menjadi kualitatif. Informasi yang berkualitas dipengaruhi oleh relevansi, keakuratan dan tepat waktu. Informasi biasanya disampaikan dalam bentuk teks, citra, suara, dan lain – lain. Banyaknya informasi tersebut dapat diselesaikan dengan kajian – kajian yang ada pada *information retrieval*, agar pengguna dapat menemukan informasi yang sesuai keinginan. *Information retrieval* (IR) sendiri adalah salah satu bidang dalam ilmu komputer yang membahas tentang pencarian informasi. Salah satu kajian dari IR yang dapat digunakan adalah aplikasi mesin pencari (*search engine*).

Kebanyakan mesin pencari yang berkembang sekarang adalah mesin pencari yang menggunakan *query* berupa teks. Masih terdapat batasan pada penggunaan mesin pencari tersebut, yaitu jika perangkat atau pengguna perangkat tersebut tidak dapat memasukkan *query* berupa teks. Walaupun sudah ada mesin pencari yang menggunakan masukan *query* berupa ucapan, tetapi yang berfokus pada pencarian jurnal berbahasa Indonesia belum ada. Jurnal sendiri merupakan karangan ilmu pengetahuan yang menyajikan fakta dan ditulis menurut metodologi penulisan yang baik dan benar (Arifin, 2008). Para pelajar khususnya mahasiswa sekarang membutuhkan jurnal untuk digunakan sebagai referensi dalam pengerjaan tugas, tetapi sering terkendala pada penggunaan bahasa asing. Untuk itu dalam penelitian kali ini akan dibuat sebuah mesin pencari yang fokus pada penggunaan *query* berupa ucapan yang digunakan untuk mencari jurnal berbahasa Indonesia.

Pengenalan ucapan atau *speech recognition* sendiri memungkinkan suatu perangkat untuk mengenali dan memahami kata – kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata – kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi matrik angka yang kemudian disesuaikan dengan kode – kode tertentu untuk mengidentifikasi kata – kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan. Hasil dari pengenalan ucapan inilah yang akan digunakan untuk *query* dari mesin pencari yang dibuat.

Pengenalan ucapan yang diterapkan pada mesin pencari ini menggunakan metode *Mel Frequency Cepstral Coefficients* (MFCC) untuk ekstraksi cirinya. Metode MFCC memiliki beberapa kelebihan diantaranya adalah mampu menangkap informasi penting dalam sinyal suara, menghasilkan data seminimal mungkin tanpa menghilangkan informasi–informasi yang ada dan mereplikasikan organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara (Andriana, 2011). Untuk metode pengenalan ucapan digunakan metode *Hidden Markov Model* (HMM), karena HMM mempunyai tingkat akurasi yang lebih tinggi dibanding metode lain (Rabiner & Juang, 1993).

Ada beberapa metode IR yang telah diketahui, diantaranya adalah *Probabilistic Model*, *Set-theoretic Models*, dan *Algebraic Model*. *Probabilistic Model* contohnya adalah penerapan Teorema Bayes, sedangkan *Set-theoretic Models* contohnya seperti *Standard Boolean* dan *Extended*, dan yang *Algebraic Model* contohnya adalah *Vector Space Model*. Untuk metode dari IR yang diterapkan pada mesin pencari ini adalah *Vector Space Model* (VSM). Di Indonesia penelitian yang menggunakan metode ini adalah penelitian tentang ”Implementasi *Search Engine* (Mesin Pencari) Menggunakan Metode *Vector Space Model*” (Fatkhul Amin, 2011). Penelitian ini berisi tentang penggunaan metode *Vector Space Model* untuk diimplementasikan pada *search engine*. Hasil dari penelitian tersebut dapat menemukan dokumen dengan tepat dan tingkat akurasi yang tinggi.

Oleh karena itu dalam penelitian ini dibuat penerapan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, dapat dirumuskan permasalahan yang dihadapi yaitu bagaimana menerapkan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan.

## 1.3 Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam penelitian tugas akhir ini adalah menghasilkan penerapan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan.

Manfaat dari penerapan mesin pencari ini adalah dapat mempermudah pencarian jurnal menggunakan *query* ucapan.

## 1.4 Ruang Lingkup

Adapun ruang lingkup dari penerapan *vector space model* dalam pencarian dokumen jurnal berbahasa Indonesia dengan *query* berupa ucapan adalah:

1. Suara yang di-*input*-kan berupa hasil rekaman langsung
2. *Output* berupa dokumen dengan format .pdf
3. Objek perekam menggunakan dialek yang umum (dalam hal ini adalah pulau jawa)
4. Objek perekam berusia antara 19 tahun - 25 tahun
5. Objek perekaman untuk data pelatihan menggunakan 10 orang *speaker*
6. Perekaman inputan suara menggunakan headset Logitech Stereo H150 dengan jarak  $\pm 3$  cm dari sekitaran mulut seseorang
7. Inputan suara dilakukan pada speaker dengan kondisi normal
8. Inputan maksimal 2 kata
9. Sistem ini akan diimplementasikan berbasis desktop dengan bahasa pemrograman MATLAB.