

**USULAN
PENELITIAN PENGEMBANGAN DOSEN**



**STUDI PENERAPAN TALA *STEMMING* DALAM
SISTEM REKOMENDASI DOKUMEN**

TIM PENGUSUL:

WAYAN GEDE SUKA PARWITA (0822088901)

NI PUTU DIAN INDRA PRATIWI (0814099002)

PROGRAM STUDI TEKNIK INFORMATIKA

STMIK STIKOM INDONESIA

DENPASAR

JUNI 2018

HALAMAN PENGESAHAN

1. Judul Penelitian : Studi Penerapan Tala *Stemming* Dalam Sistem Rekomendasi Dokumen
2. Bidang Penelitian : Teknik Informatika
3. Ketua Peneliti :
 - a. Nama Lengkap : Wayan Gede Suka Parwita, M.Cs
 - b. Jenis Kelamin : Laki-Laki
 - c. Disiplin Ilmu : Ilmu Komputer
 - d. Pangkat/Golongan : III/b
 - e. Jabatan Fungsional : Asisten Ahli
 - f. Program Studi : Teknik Informatika
4. Anggota Peneliti :
 - a. Nama Lengkap : Ni Putu Dian Indra Pratiwi, M.Pd.
 - b. Jenis Kelamin : Perempuan
 - c. Disiplin Ilmu :
 - d. Pangkat/Golongan :
 - e. Jabatan Fungsional :
 - f. Program Studi : Teknik Informatika
5. Jumlah Biaya yang Diusulkan : Rp 4.975.000,-

Denpasar, 21 Juni 2018
Mengetahui
Kepala Program Studi TI

Ketua Peneliti

I Putu Gede Budayasa, M.T.I.
NIDN : 0820068402

Wayan Gede Suka Parwita, M.Cs.
NIDN: 0822088901

Menyetujui
Kepala Lembaga Penelitian dan Pengabdian Masyarakat

Ida Bagus Ary Indra Iswara, S.Kom., M.Kom
NIDN: 0824048801

DAFTAR ISI

HALAMAN COVER	i
HALAMAN PENGESAHAN.....	2
DAFTAR ISI.....	iii
DAFTAR GAMBAR.....	v
DAFTAR TABEL	vi
DAFTAR LAMPIRAN	vii
RINGKASAN	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
1.5 Luaran Penelitian.....	2
BAB II TINJAUAN PUSTAKA.....	4
2.1 Ekstraksi <i>Keyword</i>	4
2.1.1 Tokenisasi.....	4
2.1.2 <i>Stopword Removal</i>	5
2.1.3 <i>Stemming</i>	5
2.1.4 Tala <i>Stemmer</i>	6
2.1.5 Pembobotan.....	7
2.2 Cosine Similarity	7
2.3 Evaluasi Recommendation System	8
2.3.1 Precision	8
2.3.2 Recall.....	9
2.3.3 F-Measure.....	9
BAB IIIMETODE PENELITIAN.....	11
3.1 Data Pengujian	11

3.1.1 Data Rekomendasi.....	11
3.1.2 Dokumen Penelitian	11
3.1.3 <i>Stopword</i>	11
3.2 Tahapan Penelitian	11
3.3 Pengujian Sistem	12
BAB IV BIAYA DAN JADWAL PENELITIAN.....	13
4.1 Anggaran Biaya.....	13
4.2 Jadwal Penelitian.....	13
DAFTAR PUSTAKA	14

DAFTAR GAMBAR

Gambar 3.1 Tahap Penelitian.....	11
----------------------------------	----

DAFTAR TABEL

Tabel 3.1 Pembagian kondisi hasil yang memungkinkan.....	9
Tabel 4.1 Anggaran penelitian	13
Tabel 4.2 Jadwal penelitian.....	13

DAFTAR LAMPIRAN

Lampiran 1. Justifikasi Anggaran Penelitian	14
Lampiran 2. Susunan Organisasi Tim Peneliti/Pelaksana dan Pembagian Tugas	15
Lampiran 3. Biodata Ketua Tim Peneliti	16
Lampiran 4. Biodata Anggota Tim Peneliti	19

RINGKASAN

Pengembangan sistem rekomendasi berbasis ekstraksi keyword dapat menggunakan Tala Stemming. Hal ini mengingat ekstraksi keyword dalam dokumen membutuhkan sumber daya yang besar. Penggunaan Tala Stemming merupakan salah satu opsi yang lebih ramah sumber daya karena tidak perlu membandingkan setiap eliminasi imbuhan dengan kamus kata. Dengan mempertimbangkan kelemahan Tala Stemming, perlu dilakukan studi pada pengaruh penggunaan Tala Stemming dalam sistem rekomendasi berbasis ekstraksi keyword, dimana akan dilakukan perbandingan untuk menentukan kualitas rekomendasi saat penggunaan Tala stemming dan saat tidak menggunakan Tala stemming.

Penelitian dimulai dengan analisis sistem yang telah dibangun pada penelitian Parwita, dkk (2017). Analisis ini memetakan komponen yang akan digunakan dalam pengujian sistem rekomendasi. Setelah analisis sistem, selanjutnya dilakukan pengumpulan data seperti stopword yang diambil dari penelitian Tala (2003), rekomendasi dosen, dan dokumen penelitian dosen. Implementasi pengujian dilakukan dengan membangun algoritma pengujian memanfaatkan komponen sistem rekomendasi yang telah dibangun. Algoritma yang dibangun yaitu untuk 2 pengujian yang dilakukan. Pengujian pertama adalah pengujian sistem informasi memanfaatkan proses stemming, dan pengujian kedua yaitu pengujian sistem informasi tanpa proses stemming. Hasil pengujian merupakan nilai precision, recall, dan f-measure yang digunakan dalam melakukan analisis perbandingan hasil pengujian.

Keyword : Stemming, Tala Stemming, Sistem Rekomendasi.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stemming merupakan salah satu proses dalam ekstraksi *keyword* yang mengubah kata menjadi kata dasar (stem). Ada 2 pendekatan dalam pengembangan proses *stemming* yaitu dengan pendekatan berbasis aturan dan pendekatan berbasis kamus kata. Pendekatan berbasis aturan banyak digunakan karena menggunakan sumber daya yang lebih kecil dibandingkan dengan pendekatan berbasis kamus. Proses *stemming* berbasis aturan mulai dikembangkan oleh Porter (1980) (Porter *Stemmer*) untuk bahasa Inggris. Porter *stemmer* lalu dikembangkan untuk bahasa Indonesia oleh Tala yang saat ini sering disebut dengan algoritma Tala *Stemming*.

Tala *Stemming* menggunakan aturan dalam imbuhan bahasa Indonesia. Penggunaan Tala *Stemming* merupakan salah satu opsi dalam pengembangan sistem ekstraksi teks dalam *information retrival*. Akan tetapi, Tala *stemming* juga belum sepenuhnya akurat untuk mencari kata dasar. Ada beberapa kata yang gagal diubah ke dalam kata dasar oleh Algoritma Tala. Selain itu, walaupun lebih ramah sumber daya dibandingkan dengan pendekatan berbasis kamus kata, Tala *Stemming* tetap membutuhkan sumber daya yang besar dalam melakukan ekstraksi setiap kata yang ada dalam dokumen. Hal ini berdampak pada meningkatnya waktu yang dibutuhkan dalam ekstraksi dokumen.

Pengembangan sistem rekomendasi berbasis ekstraksi *keyword* dapat menggunakan Tala *Stemming*. Hal ini mengingat ekstraksi *keyword* dalam dokumen membutuhkan sumber daya yang besar. Penggunaan Tala *Stemming* merupakan salah satu opsi yang lebih ramah sumber daya karena tidak perlu membandingkan setiap eliminasi imbuhan dengan kamus kata. Dengan mempertimbangkan kelemahan Tala *Stemming*, perlu dilakukan studi pada pengaruh penggunaan Tala *Stemming* dalam sistem rekomendasi berbasis ekstraksi *keyword*, dimana akan dilakukan perbandingan untuk menentukan kualitas rekomendasi saat penggunaan Tala *stemming* dan saat tidak menggunakan Tala *stemming*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan dapat dirumuskan masalah yang akan dikaji dalam penelitian ini yaitu bagaimana penerapan tala *stemming* dan pengaruhnya dalam sistem rekomendasi berbasis ekstraksi *keyword*.

1.3 Tujuan Penelitian

Tujuan penelitian ini yaitu melakukan penerapan tala *stemming* dan mengetahui pengaruh tala *stemming* pada akurasi sistem rekomendasi.

1.4 Manfaat Penelitian

Manfaat yang didapat dari penelitian ini yaitu dapat mengetahui bagaimana penerapan tala *stemming* dalam sistem rekomendasi berbasis ekstraksi *keyword* dan seberapa besar pengaruh tala *stemming* terhadap akurasi sistem rekomendasi.

1.5 Luaran Penelitian

Hasil penelitian ini akan dipublikasikan dalam bentuk publikasi ilmiah hasil penelitian yaitu pada Jurnal Ilmiah Nasional Sistem Informasi ber-ISSN.

No	Jenis Luaran		Indikator Capaian		
			TS0	TS+1	TS+2
1	Publikasi Ilmiah ²⁾	Internasional			
		Nasional terakreditasi			
		Nasional tidak terakreditasi	draf	published	
2	Pemakalah dalam temu ilmiah ³⁾	Internasional			
		Nasional			
3	<i>Invited speaker</i> dalam temu ilmiah ⁴⁾	Internasional			
		Nasional			
4	<i>Visiting Lecturer</i> ⁵⁾	Internasional			
5	Hak Kekayaan Intelektual (HAKI) ⁶⁾	Paten			
		Paten Sederhana			
		Hak Cipta			
		Merek Dagang			
		Rahasia Dagang			
		Desain Produk Industri			
		Indikasi Geografis			

		Perlindungan Varietas Tanaman			
		Perlindungan topografi sirkuit terpadu			
6	Teknologi Tepat Guna ⁷⁾				
7	Model/Purwarupa/Desain/Karya seni/ Rekayasa Sosial ⁸⁾				
8	Buku Ajar (ISBN) ⁹⁾				
9	Tingkat Kesiapan Teknologi (TKT) ¹⁰⁾				

BAB II

TINJAUAN PUSTAKA

2.1 Ekstraksi *Keyword*

Dalam dokumen ilmiah, *keyword* adalah kata pokok yang merepresentasikan masalah yang diteliti atau istilah-istilah yang merupakan dasar pemikiran dan dapat berupa kata tunggal atau gabungan kata. Similaritas *keyword* dokumen dapat digunakan untuk menentukan relevansi dokumen terhadap dokumen lain (Weiss, dkk., 2005). *Automatic keyword extraction system* memiliki tugas untuk mengidentifikasi kumpulan kata, frase kunci, *keyword*, atau segmen kunci dari sebuah dokumen yang dapat menggambarkan arti dari dokumen (Hulth, 2003). Tujuan dari ekstraksi otomatis adalah menekan kelemahan pada ekstraksi manual yang dilakukan manusia yaitu pada kecepatan, ketahanan, cakupan, dan juga biaya yang dikeluarkan.

Salah satu pendekatan yang dapat digunakan dalam *automatic keyword extraction* yaitu pendekatan tata bahasa. Pendekatan ini menggunakan fitur tata bahasa dari kata-kata, kalimat, dan dokumen. Metode ini memperhatikan fitur tata bahasa seperti bagian kalimat, struktur sintaksis, dan makna yang dapat menambah bobot. Fitur tata bahasa tersebut dapat digunakan sebagai penyaring untuk *keyword* yang buruk. Dalam ekstraksi *keyword* dengan pendekatan tata bahasa berbasis struktur sintaksis, ada beberapa tahap yang dilakukan yaitu tokenisasi, *stopword removal*, *stemming*, dan pembobotan kata (Oelze, 2009).

2.1.1 Tokenisasi

Teks elektronik adalah urutan linear simbol (karakter, kata-kata atau frase). Sebelum dilakukan pengolahan, teks perlu disegmentasi ke dalam unit-unit linguistik seperti kata-kata, tanda baca, angka, *alpha-numeric*, dan lain-lain. Proses ini disebut tokenisasi. Tokenisasi sederhana (*white space tokenization*) merupakan tokenisasi yang memisahkan kata berdasarkan karakter spasi, tab, dan baris baru (Weiss, dkk., 2005). Namun, tidak setiap bahasa melakukan hal ini (misalnya bahasa Cina, Jepang, Thailand). Dalam bahasa Indonesia, selain tokenisasi sederhana diperlukan juga tokenisasi yang memisahkan kata-kata berdasarkan karakter lain seperti “/” dan “-”.

2.1.2 *Stopword Removal*

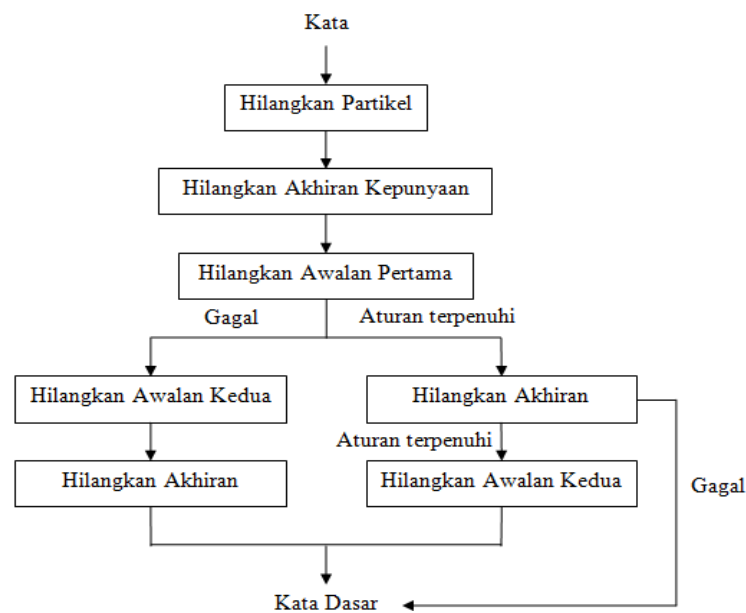
Stopword removal adalah pendekatan mendasar dalam *preprocessing* yang menghilangkan kata-kata yang sering muncul (*stopword*). Fungsi utamanya adalah untuk mencegah hasil proses selanjutnya terpengaruh oleh *stopword* tersebut. Banyak diantara *stopword* tersebut tidak berguna dalam *Information Retrival* (IR) dan *text mining* karena kata-kata tersebut tidak membawa informasi (seperti ke, dari, dan, atau). Cara biasa untuk menentukan apa yang dianggap sebagai *stopword* adalah menggunakan *stoplist*. *Stoplist* merupakan kumpulan kata atau kamus yang berisi daftar *stopword*. Langkah penghilangan *stopword* ini adalah langkah yang sangat penting dan berguna (Srividhya dan Anitha, 2010).

2.1.3 *Stemming*

Algoritma *stemming* adalah proses yang melakukan pemetaan varian morfologi yang berbeda dari kata-kata ke dalam kata dasar/kata umum (*stem*). *Stemming* berguna pada banyak bidang komputasi linguistik dan *information retrieval* (Lovins, 1968). Dalam kasus bahasa Indonesia, sejauh ini hanya ada dua algoritma untuk melakukan proses *stemming* yaitu algoritma yang dikembangkan oleh Nazief dan Adriani serta algoritma yang dikembangkan oleh Tala. Algoritma Nazief dan Adriani dikembangkan dengan menggunakan pendekatan *confix stripping* dengan disertai pemindaian pada kamus. Sedangkan *stemming* yang dikembangkan Tala menggunakan pendekatan yang berbasis aturan (*rule-based*).

Pengembangan Algoritma Tala didasarkan pada kenyataan bahwa sumber daya seperti kamus besar digital untuk bahasa mahal karena kurangnya penelitian komputasi di bidang linguistik. Maka, ada kebutuhan untuk algoritma *stemming* tanpa keterlibatan kamus. Algoritma Tala sendiri dikembangkan dari algoritma Porter *stemmer* yang dimodifikasi untuk bahasa Indonesia. Algoritma Tala menghasilkan banyak kata yang tidak dipahami. Ini disebabkan oleh ambiguitas dalam aturan morfologi Bahasa Indonesia. Dalam beberapa kasus kesalahan tidak mempegaruhi kinerja, tetapi dalam kasus lain menurunkan kinerja (Tala, 2003).

2.1.4 Tala Stemmer



Gambar 2.1 Skema Tala Stemmer (Tala, 2003)

Algoritma Tala memproses awalan, akhiran, dan kombinasi keduanya dalam kata turunan. Walaupun dalam bahasa Indonesia terdapat sisipan, jumlah kata yang diturunkan menggunakan sisipan sangat sedikit. Karena hal tersebut dan juga demi penyederhanaan, sisipan akan diabaikan.

Algoritma Porter *stemmer* dibangun berdasarkan ide tentang akhiran pada bahasa Inggris yaitu kebanyakan merupakan kombinasi dari akhiran yang lebih sederhana dan lebih kecil. Beberapa perubahan dilakukan pada algoritma Porter *stemmer* agar sesuai dengan Bahasa Indonesia. Perubahan dilakukan pada bagian kumpulan aturan dan penilaian kondisi. Karena algoritma Porter *stemmer* hanya dapat menangani akhiran, maka perlu penambahan agar dapat menangani awalan, akhiran, dan juga penyesuaian penulisan dalam kasus dimana terjadi perubahan karakter pertama kata dasar. Gambar 2.1 menunjukkan langkah-langkah proses pada algoritma Tala.

Dalam Bahasa Indonesia, unit terkecil dari suatu kata adalah suku kata. Suku kata paling sedikit terdiri dari satu huruf vokal. Desain implementasi algoritma Tala belum dapat mengenali seluruh suku kata. Ini disebabkan karena adanya dua huruf vokal yang dianggap satu suku kata yaitu ai, au, dan oi. Kombinasi dua huruf vokal (terutama ai, oi) tersebut dapat menjadi masalah, apalagi jika berada pada akhir sebuah kata. Ini disebabkan oleh sulitnya membedakannya dengan kata yang mengandung akhiran *-i*. Hal ini

menyebabkan kombinasi huruf vokal ai/oi akan diperlakukan seperti kata turunan. Huruf terakhir (-i) akan dihapus pada hasil proses *stemming*. Kebanyakan kata dasar terdiri dari minimal dua suku kata. Inilah alasan kenapa kata yang akan diproses memiliki minimal dua suku kata.

2.1.5 Pembobotan

Tahapan ini dilakukan dengan tujuan untuk memberikan suatu bobot pada *term* yang terdapat pada suatu dokumen. *Term* adalah satu kata atau lebih yang dipilih langsung dari *corpus* dokumen asli dengan menggunakan metode *term-extraction*. Fitur tingkat *term*, hanya terdiri dari kata-kata tertentu dan ekspresi yang ditemukan dalam dokumen asli (Feldman dan Sanger, 2007).

Dalam pengkategorian teks dan aplikasi lain di *information retrieval* maupun *machine learning*, pembobotan *term* biasanya ditangani melalui metode yang diambil dari metode pencarian teks, yaitu yang tidak melibatkan tahap belajar (Debole dan Sebastiani, 2003). Ada tiga asumsi monoton yang muncul di hampir semua metode pembobotan dapat dalam satu atau bentuk lain yaitu (Zobel dan Moffat, 1998):

- a. *Term* yang langka tidak kalah penting daripada *term* yang sering muncul (asumsi IDF).
- b. Kemunculan berkali-kali dari *term* pada dokumen tidak kalah penting daripada kemunculan tunggal (asumsi TF).
- c. Untuk pencocokan *term* dengan jumlah pencocokan yang sama, dokumen panjang tidak lebih penting daripada dokumen pendek (asumsi normalisasi).

Bobot diperlukan untuk menentukan apakah *term* tersebut penting atau tidak. Bobot yang diberikan terhadap sebuah *term* bergantung kepada metode yang digunakan untuk membobotinya.

2.2 Cosine Similarity

Pendekatan *cosine similarity* sering digunakan untuk mengetahui kedekatan antara dokumen teks. Perhitungan *cosine similarity* dimulai dengan menghitung *dot product*. *Dot product* merupakan perhitungan sederhana untuk setiap komponen dari kedua vektor. Vektor merupakan representasi dari masing-masing dokumen dengan jumlah *term* pada masing-masing dokumen sebagai

dimensi dari vektor (Manning, dkk., 2009). Vektor ditunjukkan oleh notasi (2.1) dan (2.2). Hasil *dot product* bukan berupa vektor tetapi berupa skalar. Persamaan (2.3) merupakan perhitungan *dot product* dimana n merupakan dimensi dari vector (Axler, 1997).

$$\vec{a} = (a_1, a_2, a_3, \dots, a_n) \quad (2.1)$$

$$\vec{b} = (b_1, b_2, b_3, \dots, b_n) \quad (2.2)$$

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \quad (2.3)$$

a_n dan b_n merupakan komponen dari vektor (bobot *term* masing-masing dokumen) dan n merupakan dimensi dari vektor. *Cosine similarity* merupakan perhitungan yang mengukur nilai *cosine* dari sudut antara dua vektor (atau dua dokumen dalam *vector space*). *Cosine similarity* dapat dilihat sebagai perbandingan antara dokumen karena tidak hanya mempertimbangkan besarnya masing-masing jumlah kata (bobot) dari setiap dokumen, tetapi sudut antara dokumen. Persamaan (2.4) dan (2.5) adalah notasi dari metode *cosine similarity* dimana $\|\vec{a}\|$ merupakan *Euclidean norm* dari vektor a dan $\|\vec{b}\|$ merupakan *Euclidean norm* vektor b (Han, dkk., 2011).

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta \quad (2.4)$$

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (2.5)$$

Dari notasi (2.5) dapat dibentuk persamaan matematika yang ditunjukkan oleh persamaan (2.6) (Lops, dkk., 2011).

$$Similarity(x, y) = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2}} \quad (2.6)$$

Dimana:

a_i : term ke- i yang terdapat pada dokumen a .

b_i : term ke- i yang terdapat pada dokumen b .

2.3 Evaluasi Recommendation System

2.3.1 Precision

Precision bersama *recall* merupakan salah satu pengujian dasar dan paling sering digunakan dalam penentuan efektifitas *information retrieval system* maupun *recommendation system*. *True positive* (tp) pada *information retrieval*

merupakan *item* relevan yang dihasilkan oleh sistem. Sedangkan *false positive* (fp) merupakan semua *item* yang dihasilkan oleh sistem. Sehingga dalam *information retrieval*, *precision* dihitung dengan persamaan (2.7) (Manning, dkk., 2009).

$$Precision = \frac{tp}{tp + fp} = \frac{\text{relevant item retrieved}}{\text{retrieved item}} \quad (2.7)$$

Istilah *positive* dan *negative* mengacu pada prediksi yang dilakukan oleh sistem. Sedangkan istilah *true* dan *false* mengacu pada prediksi yang dilakukan oleh pihak luar atau pihak yang melakukan observasi. Pembagian kondisi tersebut dapat dilihat pada Tabel 3.11 (Manning, dkk., 2009).

Tabel 3.1 Pembagian kondisi hasil yang memungkinkan

	<i>Relevant</i>	<i>Nonrelevant</i>
<i>Retrieved</i>	<i>True positive (tp)</i>	<i>False positive (fp)</i>
<i>Not retrieved</i>	<i>False negative (fn)</i>	<i>True negative (tn)</i>

2.3.2 Recall

Recall digunakan sebagai ukuran dokumen yang relevan yang dihasilkan oleh sistem. *False negative* (fn) merupakan semua *item* relevan yang tidak dihasilkan oleh sistem. Dalam evaluasi *information retrieval system*, *recall* dihitung dengan persamaan (2.8) (Manning, dkk., 2009).

$$Recall = \frac{tp}{tp + fn} = \frac{\text{relevant item retrieved}}{\text{relevant item}} \quad (2.8)$$

2.3.3 F-Measure

F-measure merupakan nilai tunggal hasil kombinasi antara nilai *precision* dan nilai *recall*. F-measure dapat digunakan untuk mengukur kinerja dari *recommendation system* ataupun *information retrieval system*. Karena merupakan rata-rata harmonis dari *precision* dan *recall*, F-measure dapat memberikan penilaian kinerja yang lebih seimbang. Persamaan (2.9) merupakan persamaan untuk menghitung F-measure (Jannach, dkk., 2010).

$$F_{measure} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (2.9)$$

BAB III

METODE PENELITIAN

3.1 Data Pengujian

Data yang dikumpulkan untuk penelitian yaitu data kesesuaian penelitian dosen dengan pengajuan UPP mahasiswa, dokumen penelitian dosen, dan *stopword*.

3.1.1 Data Rekomendasi

Pengumpulan data rekomendasi untuk pengujian dilakukan dengan metode kuesioner ke masing-masing dosen untuk mengetahui UPP yang sesuai dengan bidang penelitian dosen. Dalam lembar kuesioner berisi nim dan judul UPP yang diajukan oleh mahasiswa. UPP yang digunakan diambil dari basis data Sintesys tanpa dilakukan perubahan. Contoh lembar kuesioner dapat dilihat pada Lampiran 1.

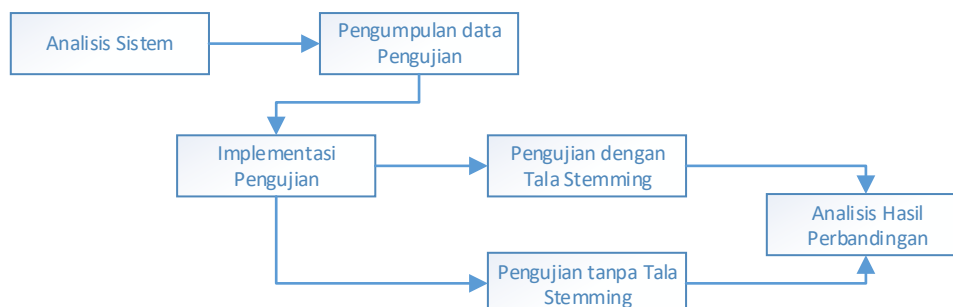
3.1.2 Dokumen Penelitian

Dokumen untuk set data merupakan 54 dokumen penelitian yang dimiliki oleh 22 orang dosen STMIK STIKOM Indonesia. Sedangkan set data UPP menggunakan 185 pengajuan yang diambil dari basis data Sintesys.

3.1.3 Stopword

Stopword yang digunakan merupakan *stopword* yang diusulkan oleh Tala dalam penelitian pengembangan *stemmer* untuk bahasa Indonesia.

3.2 Tahapan Penelitian



Gambar 3.1 Tahap Penelitian

Penelitian dimulai dengan analisis sistem yang telah dibangun pada penelitian Parwita, dkk (2017). Analisis ini memetakan komponen yang akan

digunakan dalam pengujian sistem rekomendasi. Setelah analisis sistem, selanjutnya dilakukan pengumpulan data seperti *stopword* yang diambil dari penelitian Tala (2003), rekomendasi dosen, dan dokumen penelitian dosen. Implementasi pengujian dilakukan dengan membangun algoritma pengujian memanfaatkan komponen sistem rekomendasi yang telah dibangun. Algoritma yang dibangun yaitu untuk 2 pengujian yang dilakukan. Pengujian pertama adalah pengujian sistem informasi memanfaatkan proses *stemming*, dan pengujian kedua yaitu pengujian sistem informasi tanpa proses *stemming*. Hasil pengujian merupakan nilai *precision*, *recall*, dan *f-measure* yang digunakan dalam melakukan analisis perbandingan hasil pengujian. Gambar 3.1 merupakan gambaran mengenai tahapan penelitian yang dilakukan.

3.3 Pengujian Sistem

Pengujian dilakukan 2 kali yaitu melakukan rekomendasi dengan menggunakan proses *stemming* dan pengujian tanpa menggunakan *stemming*. Pengujian dilakukan dengan batasan minimum similarity dari 5% hingga 40%. Hasil pengujian ini kemudian dibandingkan sehingga menemukan pengaruh penggunaan proses *stemming*. Adapun nilai yang dibandingkan adalah nilai *precision*, *recall*, dan *f-measure*.

BAB IV

BIAYA DAN JADWAL PENELITIAN

4.1 Anggaran Biaya

Adapun rincian anggaran biaya yang diajukan pada penelitian ini dapat dilihat pada Tabel 4.1.

Tabel 4.1 Anggaran penelitian

No.	Jenis Pengeluaran	Biaya yang Diusulkan (Rp)
1	Gaji dan Upah	3.000.000
2	Bahan habis pakai dan peralatan	1.075.000
3	Perjalanan	150.000
4	Lain-lain (publikasi, seminar, laporan, lainnya)	750.000
	Jumlah	4.975.000

4.2 Jadwal Penelitian

Penelitian ini akan dilaksanakan dalam jangka waktu tujuh bulan dengan jadwal kegiatan seperti yang ditunjukkan pada tabel 4.2.

Tabel 4.2 Jadwal penelitian

No.	Kegiatan	Bulan ke-					
		1	2	3	4	5	6
1.	Persiapan dan Pengumpulan Data						
2.	Studi Literatur						
3.	Analisis Kebutuhan Sistem						
4.	Perancangan Sistem						
5.	Implementasi dan Pengujian Sistem						
6.	Pengukuran Akurasi						
7.	Penyusunan Laporan Penelitian						
8.	Publikasi Ilmiah Hasil Penelitian						

DAFTAR PUSTAKA

- Hulth, A., 2003, Improved Automatic *Keyword* Extraction Given More Linguistic Knowledge, Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, 216-223.
- Lovins, J. B., 1968, Development of a *Stemming* Algorithm, Mechanical Translation and Computational Linguistics, 11, Massachusetts Institute of Technology, Cambridge, Massachusetts, Maret dan Juni 1968, 22-31.
- Manning, C. D., Raghavan, P., dan Schütze, H., 2009, An Introduction to Information Retrieval, Cambridge University Press, Cambridge, England.
- Oelze, I., 2009, Automatic *Keyword* Extraction for Database Search, Ph.D. Thesis, University of Hannover, Hannover.
- Parwita, W. G. S., Swari, M. H. P., Welda, 2018, Perancangan Sistem Rekomendasi Dokumen Dengan Pendekatan Content-Based Filtering, Journal of Computer Engineering, System and Science, vol 3, issue 1, hal 65-75, Medan, Indonesia.
- Tala, F. Z., 2003, A Study of *Stemming* Effects on Information Retrieval in Bahasa Indonesia, Master Thesis, Universiteit van Amsterdam.
- Weiss, S. M., Indurkha, N., Zhang, T., dan Damearau, F. J., 2005, Text Mining : Predictive Method for Analizing Unstructured Information, Springer, New York, USA.

Lampiran 1. Justifikasi Anggaran Penelitian

1. Gaji dan Upah

- | | |
|---|---------------|
| a. Biaya Programmer sebesar Rp7.353 / jam
(alokasi waktu 17 jam/minggu selama 24 minggu) | Rp. 3.000.000 |
|---|---------------|

2. Bahan Habis Pakai dan Peralatan

- | | |
|---|-------------|
| a. Biaya Pengolahan Data (Listrik, Komputer) | Rp. 50.000 |
| b. Foto Copy (FC) | Rp. 175.000 |
| c. Alat Tulis Kantor (Tinta Printer, Kertas, dll.) | Rp. 150.000 |
| d. Penelusuran pustaka dan pembelian buku referensi | Rp. 700.000 |

3. Biaya Perjalanan Penelitian

- | | |
|---|-------------|
| a. Biaya Survey Pendahuluan (Rapat, Konsumsi Rapat) | Rp. 100.000 |
| b. Biaya Pengumpulan Data | Rp. 50.000 |

4. Lain-lain

- | | |
|-----------------------------|-------------|
| a. Biaya Dokumentasi | Rp. 50.000 |
| b. Biaya Publikasi | Rp. 600.000 |
| c. Biaya Penyusunan Laporan | Rp. 100.000 |
-

Total Biaya

Rp 4.975.000

(Empat Juta Sembilan Ratus Tujuh Puluh Lima Ribu Rupiah)

Lampiran 2. Susunan Organisasi Tim Peneliti/Pelaksana dan Pembagian Tugas

No.	Nama Lengkap / NIDN	Instansi Asal	Bidang Ilmu	Alokasi Waktu (Jam/Minggu)	Uraian Tugas
1.	Wayan Gede Suka Parwita /0822088901	STMIK STIKOM Indonesia	Ilmu Komputer, Sistem Informasi	10 Jam / minggu	Menganalisis permasalahan, merancang dan membangun sistem.
2.	Ni Putu Dian Indra Pratiwi/	STMIK STIKOM Indonesia		7 Jam / minggu	Menganalisis permasalahan, mengumpulkan data, menyusun laporan.

Lampiran 3. Biodata Ketua Tim Peneliti

A. Identitas Diri

1.	Nama Lengkap	Wayan Gede Suka Parwita, M.Cs.
2.	Jenis Kelamin	Laki-laki
3.	Jabatan Fungsional	Tenaga Pengajar
4.	NIK	1507236
5.	NIDN	0822088901
6.	Tempat dan Tanggal Lahir	Denpasar, 22 Agustus 1989
7.	E-Mail	gede.suka@gmail.com
8.	Nomor HP	081337841111
9.	Alamat Kantor	Jl. Tukad Pakerisan 97 Denpasar, Bali
10.	Nomor Telepon/Faks	0361 - 256 995/ 0361 - 246 875
11.	Lulusan yang Telah Dihilangkan	-
12.	Mata Kuliah yg Diampu	a. Object Oriented Analysis and Design
		b. Struktur Data
		c. Sistem Basis Data
		d. Bahasa Basis Data
		e. Riset Teknologi Informasi

B. Riwayat Pendidikan

	S-1	S-2
Nama Perguruan Tinggi	Universitas Udayana	Universitas Gadjah Mada
Bidang Ilmu	Ilmu Komputer	Ilmu Komputer
Tahun Masuk-Lulus	2007-2011	2011-2015
Judul Tugas Akhir/Tesis	Perancangan dan Implementasi Sistem Pencarian Buku Menggunakan Algoritma Pemetaan Transaksi	<i>Hybrid Recommendation System</i> Memanfaatkan Penggalan <i>Frequent Itemset</i> dan Perbandingan <i>Keyword</i>
Nama Pembimbing	Ngurah Agus Sanjaya Er, M.Kom., Luh Gede Astuti, M.Kom.	Edi Winarko, M.Sc., Ph.D.

C. Pengalaman Penelitian Dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jml (juta Rp)
1.	2016	Pengembangan Sistem Monitoring Tugas Akhir dan Kerja Praktek STMIK STIKOM Indonesia	Penelitian Pengembangan Dosen STIKI (PPDS)	3.325.000
2.	2016	Sistem Rekomendasi Dosen Pembimbing Tugas	Penelitian Dosen Pemula	20.000.000

		Akhir di STMIK STIKOM Indonesia Menggunakan Content-Based Filtering	(PDP) Ristek Dikti	
3.	2017	Pengaruh Stopword Terhadap Akurasi Sistem Rekomendasi Dokumen	Penelitian Pengembangan Dosen STIKI (PPDS)	3.000.000

D. Pengalaman Pengabdian Kepada Masyarakat dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jml (juta Rp)
1.	2015	Pelatihan 40 Wanita Pelaku UKM se-Kota Denpasar	BKBPP Kota Denpasar	27.023.500
3.	2015	Input Data Keluarga Menggunakan Microsoft Excel Pada Kegiatan PKM 2015 Bekerja Sama Dengan BKKBN Provinsi Bali	BKKBN Provinsi Bali	910.000.000
4.	2016	Edukasi Program 1 Juta Domain Bekerjasama Dengan Kementerian Komunikasi dan Informatika	Kementrian Kominfo RI	50.000.000
5.	2016	Pendataan Penduduk Kabupaten Badung dalam Program "Krama Badung Sehat" Menggunakan Aplikasi Terintegrasi	Kabupaten Badung	250.500.000

E. Publikasi Artikel Ilmiah Dalam Jurnal dalam 5 Tahun Terakhir

No.	Judul Artikel Ilmiah	Nama Jurnal	Volume/Nomor/Tahun
1.	<i>Hybrid Recommendation System</i> Memanfaatkan Penggalan <i>Frequent Itemset</i> dan Perbandingan <i>Keyword</i>	Indonesian Journal of Computing and Cybernetics Systems (IJCCS)	Vol 9, No 2 (2015)
2.	Perencanaan Strategi SI/TI pada STMIK STIKOM Indonesia	S@cies (STIKI Applied Sciences)	Vol 6, No 1 (2016)

	Dengan Metode Ward and Peppard		
3.	Pengembangan Sistem Monitoring Tugas Akhir dan Kerja Praktek STMIK STIKOM Indonesia	S@cies (STIKI Applied Sciences)	Vol 7, No 1 (2016)
4.	Perancangan Sistem Rekomendasi Dokumen Dengan Pendekatan Content-Based Filtering	CESS (Journal of Computer Engineering, System and Science)	Vol 3, No 1 (2018)

F. Pemakalah Seminar Ilmiah (*Oral Presentation*) dalam 5 Tahun Terakhir

No.	Nama Pertemuan Ilmiah/Seminar	Judul Artikel Ilmiah	Waktu dan Tempat

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidak-sesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Hibah Penelitian Pengembangan Dosen STIKI.

Denpasar, 21 Juni 2018
Pembuat,

(Wayan Gede Suka Parwita)

Lampiran 4. Biodata Anggota Tim Peneliti

A. Identitas Diri

1.	Nama Lengkap	
2.	Jenis Kelamin	
3.	Jabatan Fungsional	
4.	NIK	
5.	NIDN	
6.	Tempat dan Tanggal Lahir	
7.	E-Mail	
8.	Nomor HP	
9.	Alamat Kantor	
10.	Nomor Telepon/Faks	
11.	Lulusan yang Telah Dihasilkan	
12.	Mata Kuliah yg Diampu	1.
		2.
		3.
		4.
		5.

B. Riwayat Pendidikan

	S-1	S-2
Nama Perguruan Tinggi		
Bidang Ilmu		
Tahun Masuk-Lulus		
Judul Tugas Akhir/Tesis		
Nama Pembimbing		

C. Pengalaman Penelitian Dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jml (juta Rp)
1				
2				

D. Pengalaman Pengabdian Kepada Masyarakat dalam 5 Tahun Terakhir

No.	Tahun	Judul Penelitian	Pendanaan	
			Sumber	Jml (juta Rp)
1				

E. Publikasi Artikel Ilmiah Dalam Jurnal dalam 5 Tahun Terakhir

No.	Judul Artikel Ilmiah	Nama Jurnal	Volume/Nomor/Tahun

F. Pemakalah Seminar Ilmiah (*Oral Presentation*) dalam 5 Tahun Terakhir

No.	Nama Pertemuan Ilmiah/Seminar	Judul Artikel Ilmiah	Waktu dan Tempat

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidak-sesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Hibah Penelitian Pengembangan Dosen STIKI.

Denpasar, 15 April 2016
Pengusul,

(_____)

SURAT PERNYATAAN KETUA PENELITI/PELAKSANA

Yang bertanda tangan dibawah ini:

Nama : Wayan Gede Suka Parwita, M.Cs.
NIDN : 0822088901
Pangkat / Golongan : -
Jabatan Fungsional : Tenaga Pengajar

Dengan ini menyatakan bahwa proposal penelitian saya yang dengan judul : *Studi Penerapan Tala Stemming Dalam Sistem Rekomendasi Dokumen*, yang diusulkan dalam Hibah Penelitian Pengembangan Dosen STIKI untuk tahun anggaran 2018 **bersifat original dan belum pernah dibiayai oleh lembaga / sumber dana lain.**

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku dan mengembalikan seluruh biaya penelitian yang sudah diterima ke STMIK STIKOM Indonesia (STIKI).

Demikian pernyataan ini dibuat dengan sesungguhnya dan dengan sebenar-benarnya.

Mengetahui, Kepala LPPM STIKOM Indonesia,	Denpasar, 21 Juni 2018 Yang menyatakan,
--	--

Ida Bagus Ary Indra Iswara, M.Kom.
NIP/NIK: 1403210

Wayan Gede Suka Parwita, M.Cs.
NIP/NIK: 1507236