

# PROSIDING SEMMAU 2017

SEMINAR NASIONAL & KONFERENSI  
SISTEM INFORMASI,  
INFORMATIKA & KOMUNIKASI

TEMA:  
E - BUSINESS SEBAGAI DAYA DUKUNG  
INDUSTRI KREATIF

Kupang, 25 November 2017

BUKU 3

ISBN: 978-602-73628-0-2



**STIKOM UYELINDO KUPANG**



# PROSIDING SEMMAU 2017

---

**Penulis,**  
Pemakalah SEMMAU 2017

**Penerbit,**  
STIKOM UYELINDO KUPANG

# PROSIDING SEMMAU 2017

---

## KOMITE

### Penulis:

Pemakalah Seminar Nasional & Konferensi Sistem Informasi, Informatika & Komunikasi (SEMMAU 2017)

ISBN : 978-602-73628-0-4

### Komite Program:

Prof. Daniel Herman Fredy Manongga, M.Sc., Ph.D. (UKSW)  
Prof. Dr. Ir. Eko Sedyono, M. Kom (UKSW)  
Prof. Mustafid (UNDIP)  
Prof. Dr.Ir. Kuswara Setiawan, M.T. (UPH)  
Dr. Ir. Rila Mandala, M.Eng. (ITB)  
Dr. Achmad Nizar, S. Kom., M.Kom. (UI)  
Ir. Dana Indra Sensuse, M.Lis., Ph.D. (UI)

### Penyunting:

Max ABR. Soleman Lenggu. S. Kom., M.T.  
Skolastika Siba Igon, S. Kom., M.T  
Reza Hardi Nugroho  
Henry Max Matchless Ratmo  
Fransiskus Xaverius Pey Tae  
Eko D. Rihibiha  
Yohana Stefania Pipa Wea  
Immanuel M. Laka  
Ritwan Banu  
Maissy P. Babar  
Frialdhy S. Ketty  
Maestryn A. Taeko  
Muhammad Bdariyadi  
Barnabas Sarbunan

### Desain Sampul:

Max Lenggu

### Redaksi :

#### Dapur Semmau

Lembaga Penelitian, Publikasi dan Pengembangan pada Masyarakat  
Jl. Perintis Kemerdekaan 1, Kayu Putih, Kupang, NTT, Indonesia.  
Telp.(0380)8554501, Fax (0380) 8554501  
Email : [semmau@uyelindo.ac.id](mailto:semmau@uyelindo.ac.id)  
<http://www.semmau.uyelindo.ac.id>.

### Penerbit :

Sekolah Tinggi Manajemen Informatika & Komputer (STIKOM) Uyelindo Kupang.  
Jl. Perintis Kemerdekaan 1, Kayu Putih, Kupang, NTT, Indonesia.  
Telp.(0380)8554501, Fax (0380) 8554501  
Email : [stikom@uyelindo.ac.id](mailto:stikom@uyelindo.ac.id)  
<http://www.uyelindo.ac.id>.

*Cetakan ketiga November 2017*

*Hak Cipta di Lindungi Undang-undang*

*Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun tanpa ijin tertulis dari penerbit.*

# PROSIDING SEMMAU 2017

---

## KATA PENGANTAR

Segala puji dan syukur selayaknya tercurah kehadirat Allah Yang Maha Kuasa yang tanpa henti mengucurkan rahmat dan karuniaNya, baik kurunia sehat, rejeki, kecerdasan, kemauan, dan bahkan juga karunia dalam bentuk kesadaran dan kemampuan bersyukur kepadaNya, dan dengan ijinnya Prosiding Seminar Nasional dan Konferensi Sistem Informasi, Teknik Informatika, dan Komunikasi (SEMMAU) tahun 2017 dengan Tema “*E-BUSSINES* SEBAGAI DAYA DUKUNG INDUSTRI KREATIF” dapat kami terbitkan.

Buku Prosiding ini berisi sekumpulan *Paper* dari hasil penelitian ilmiah yang telah diseleksi, untuk dipresentasikan dalam kegiatan Seminar Nasional dan Konferensi Sistem Informasi, Teknik Informatika, dan Komunikasi (SEMMAU) tahun 2017 dan bertempat di *Ballroom* Swiss Belinn Kristal Hotel Kupang Nusa Tenggara Timur pada tanggal 25 November 2017, kegiatan ini diikuti oleh peserta pemakalah yang berasal dari berbagai perguruan tinggi yang tersebar di kawasan Nusa Tenggara Timur (NTT), maupun di luar NTT, yang terdiri dari 26 makalah dari para peserta pemakalah.

Seminar Nasional yang bertemakan “*E-BUSSINES* SEBAGAI DAYA DUKUNG INDUSTRI KREATIF” ini menghadirkan pembicara utama berkelas nasional yakni Prof. Dr. Ir. Eko Sedyono, M. Kom.

Ucapan terima kasih kami sampaikan kepada Reviewer Paper dan pihak-pihak yang telah membantu penyelenggaraan Seminar Nasional dan Konferensi Sistem Informasi, Teknik Informatika, dan Komunikasi (SEMMAU) tahun 2017 ini. Semoga prosiding ini dapat bermanfaat dan dapat digunakan dengan sebaik-baiknya.

Akhir kata, jika ada yang kurang berkenan selama penyelenggaraan kegiatan seminar maupun dalam penerbitan buku prosiding ini mohon dimaafkan. Semoga apa yang telah kita lakukan ini bermanfaat bagi kemajuan bangsa dan negara dimasa depan. Amin.

Kupang, November 2017  
Panitia,

Sumarlin

# PROSIDING SEMMAU 2017

---

## DAFTAR ISI

	Halaman
<b>OPTIMALISASI PEMAHAMAN MATERI TEKNOLOGI INFORMASI DAN KOMUNIKASI DENGAN KONTEN BERBASIS MULTIMEDIA (STUDI KASUS PADA SMP SATU ATAP MORO – ENDE).</b> <i>Agustinus Lambertus Suban.</i>	394 - 399
<b>APLIKASI PENILAIAN DOSEN DAN STAF PADA STIKOM ARTHA BUANA KUPANG BERBASIS ANDROID.</b> <i>Ahmad Haidaroh, Fajar Riski Maulidan.</i>	400 - 408
<b>PETA RUTE ANGKUTAN UMUM KOTA KUPANG BERBASIS <i>MOBILE GEOGRAPHIC INFORMATION SYSTEM</i></b> <i>Benyamin Jago Belalawe, Benediktus Yoseph Bhae, Petrus Katemba.</i>	409 - 413
<b>ANALISIS POPULARITAS WEBSITE PEMERINTAH KOTA DI SUMATERA SELATAN.</b> <i>Dien Novita, Lisa Amelia Fransen</i>	414 -419
<b>E-COMMERCE SEBAGAI UPAYA PENGEMBANGAN USAHA KECIL DAN MENENGAH DALAM PEMASARAN KERIPIK HASIL INOVASI DI KOTA BOYOLALI.</b> <i>Donna Setiawati.</i>	420 - 427
<b>SISTEM INFORMASI PUBLIK RUMAH TANGGA MISKIN PADA KABUPATEN SUMBA TIMUR.</b> <i>Edwin Ariesto Umbu Malahina, Emanuel Safirman Bata</i>	428 - 435
<b>CARA PEMBERANTASAN HAMA TANAMAN KAKAO DENGAN METODE <i>MULTIMEDIA DEVELOP LIFE CYCLE</i> GUNA MENINGKATKAN PRODUKSI TANAMAN KAKAO BERBASIS ANDROID.</b> <i>Febriyanti Alwisye Wara, Imelda Dua Reja.</i>	436 - 439
<b>PERANCANGAN SISTEM INFORMASI PELINTAS BATAS WILAYAH NEGARA RI DAN RDTL.</b> <i>Fransiskus M.H. Tjiptabudi, Skolastika S. Igon.</i>	440 - 446
<b>SISTEM PENDUKUNG KEPUTUSAN UNTUK PENENTUAN MINAT DENGAN METODE <i>ANALYTICAL HIERARCHY PROCESS</i>.</b> <i>Gregorius William Meno, Dony M. Sihotang, Tiwuk Widiastuti.</i>	447 - 452
<b>APLIKASI ANDROID SEBAGAI MEDIA PEMBELAJARAN KARDIOVASKULER PADA PEREDARAN DARAH MANUSIA.</b> <i>Imelda Dua Reja, Febriyanti Alwisye Wara, Bastian Texaniwin Nakoda.</i>	453 - 458
<b>PEMANFAATAN APLIKASI PEMBELAJARAN BUDAYA TIMOR TENGAH SELATAN SEBAGAI UPAYA MENINGKATKAN RASA CINTA TANAH AIR.</b> <i>Mardhalia Saitakela, Skolastika S. Igon.</i>	459 – 462

## PROSIDING SEMMAU 2017

---

<b>PENGARUH PENERAPAN SISTEM INFORMASI AKUNTANSI TERHADAP KINERJA INDIVIDU PEGAWAI DI KOPERASI PEGAWAI NEGERI SIPIL BHAKTI HUSADA.</b>	<b>463 – 467</b>
<i>Maria Florentina Rumba, Yosafat Pati Koten</i>	
<b>SISTEM TEMU BALIK INFORMASI DOKUMEN TEKS MENGGUNAKAN VECTOR SPACE MODEL.</b>	<b>468 – 473</b>
<i>Mariam Fatima Somu, Paulina Aliandu, Paskalis Andi Nani.</i>	
<b>PENENTUAN LARVA TIRAM MUTIARA TERBAIK SEBAGAI PENGHASIL MUTIARA UNGGUL DENGAN LOGIKA FUZZY.</b>	<b>474 – 481</b>
<i>Marselina Dorce Tlaan, Marinus I.J. Lamabelawa.</i>	
<b>PEMETAAN TEMPAT WISATA KABUPATEN ROTE NDAO BERBASIS WEB GIS.</b>	<b>482 – 486</b>
<i>Menhya Snae, Max ABR S Lenggu, Benediktus Y. Bhae.</i>	
<b>DETEKSI CALON KREDITUR MOTOR DENGAN NAÏVE BAYES CLASSIFIER (STUDI KASUS: PT. FIF CABANG KUPANG).</b>	<b>487 – 494</b>
<i>Miransyah Koroh, Marlinda Vasty Overbeek.</i>	
<b>PENERAPAN GOOGLE MAPS DALAM MENENTUKAN LOKASI-LOKASI WIFI CORNER DI KOTA KUPANG BERBASIS ANDROID.</b>	<b>495 – 498</b>
<i>Muhammad Harits Ardiyanto, Yohanes Suban Belutowe.</i>	
<b>SISTEM TEMU BALIK DOKUMEN TEKS MENGGUNAKAN METODE BOOLEAN DAN TERM WEIGHT TF.IDF.</b>	<b>499 – 504</b>
<i>Ni Putu Anggi Yuliani, Paulina Aliandu, Paskalis Andi Nani.</i>	
<b>PENERAPAN METODE SMART (SIMPLE MULTI ATTRIBUTE RATING TECHNIQUE) DAN ALGORITMA K-NN (K-NEAREST NEIGHBOR) DALAM PENENTUAN STATUS KESEHATAN BAYI BARU LAHIR DI RUMAH SAKIT BHAYANGKARA KUPANG</b>	<b>505 – 511</b>
<i>Omar Bilham Tamonob, Kornelis Letelay, Sebastianus Mola.</i>	
<b>SISTEM PENDUKUNG KEPUTUSAN PENENTUAN CALON PENERIMA BEASISWA PADA PROGRAM STUDI TEKNIK INFORMATIKA UNIVERSITAS NUSA NIPA MENGGUNAKAN METODE AHP.</b>	<b>512 – 515</b>
<i>Petrus Wolo, Stefania Memen Tupen, Yosep P. Minggu.</i>	
<b>FUTURE COMPUTING PROFESSION.</b>	<b>516 – 522</b>
<i>Raul Bernardino, Hasibun Asikin</i>	
<b>INFORMASI LOKASI DAERAH PENGHASIL KOMODITI UNGGULAN DI KABUPATEN KUPANG BERBASIS ANDROID.</b>	<b>523 – 527</b>
<i>Remerta Noni Naatonis, Skolastika S. Igon.</i>	
<b>PENERAPAN CUSTOMER RELATIONSHIP MANAGEMENT PADA APLIKASI PEMASARAN KAIN TENUN IKAT KHAS FLORES BERBASIS ANDROID.</b>	<b>528 – 534</b>
<i>Sumarlin, Dewi Angraini.</i>	

## PROSIDING SEMMAU 2017

---

- DESAIN MODEL INFORMASI DANA DESA BERBASIS WEB (STUDI KASUS DESA TANINI KECAMATAN TAKARI KABUPATEN KUPANG).** 535 – 543  
*Yohanes Payong, Antonio Soares, Venansius A. K. Ga'a*
- ANALISA CITRA UNTUK KLASIFIKASI KAYU BANGUNAN.** 544 – 547  
*Yohanes Suban Belutowe*
- IMPLEMENTASI METODE *SIMPLE ADDITIVE WEIGHTING* UNTUK SELEKSI PENERIMAAN BEASISWA BIDIK MISI DI POLITEKNIK NEGERI SAMARINDA.** 548 – 554  
*Yusni Nyura, Damar Nurcahyono.*

## SISTEM TEMU BALIK INFORMASI DOKUMEN TEKS MENGGUNAKAN *VECTOR SPACE MODEL*

Mariam Fatima Somu<sup>1</sup>, Paulina Aliandu<sup>2</sup>, Paskalis Andi Nani<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Informatika, Unika Widya Mandira Kupang  
<sup>1</sup>adjafaty2@gmail.com, <sup>2</sup>paulinaaliandu@gmail.com, <sup>3</sup>paskalisnani@gmail.com

### Abstrak

Buku merupakan media informasi yang memiliki peran penting dalam perkembangan ilmu pengetahuan. Saat ini telah banyak toko buku yang menyediakan berbagai jenis buku, salah satunya adalah Toko Buku Suci. Di toko buku suci, terdapat masalah yang sering dialami oleh karyawan dalam menemukan buku cetak yang dicari atau di inginkan oleh pembeli. Atas dasar permasalahan tersebut maka dibangun sebuah Sistem Temu Balik Informasi untuk mempermudah pencarian buku. Metode *Vector Space Model* merupakan salah satu alternatif yang dapat diimplementasikan untuk memecahkan masalah tersebut. Dengan metode *Vector Space Model* kita dapat melihat tingkat kesamaan dengan cara pembobotan *term* (kata) dan perhitungan similaritas. Ini adalah suatu metode pembobotan dengan menghitung jarak antar dokumen dan perhitungan kemiripan dokumen dan *query*. Sistem Temu Kembali Informasi dokumen teks ini dapat memberi kemudahan dalam mencari dan menemukan dokumen teks resensi buku yang diinginkan dengan lebih efisien dengan merekomendasikan buku sesuai dengan kebutuhan dengan rata-rata waktu pencarian informasi yang berlangsung cepat yaitu 0,01454 detik.

*Kata kunci: Sistem Temu Balik Informasi, Vector Space Model, Teks, Tf.Idf.*

### 1. PENDAHULUAN

Kemajuan zaman dan teknologi saat ini telah merubah pola dalam perusahaan ataupun organisasi-organisasi. Perubahan dan perkembangan teknologi yang demikian pesat seperti otomatisasi dan komputerisasi dalam segala bidang usaha menuntut kalangan teknologi informasi dan para praktisi yang berkecimpung di dalamnya untuk lebih siap menghadapi kemajuan yang ada. Seiring dengan berkembangnya organisasi dalam suatu perusahaan, kebutuhan informasi yang cepat dan efisien menjadi suatu hal yang sangat penting dalam menentukan kemajuan perusahaan.

Toko buku merupakan tempat menjual berbagai macam buku. Keberadaan Toko buku sangat membantu masyarakat sekitar dalam mendapatkan informasi dan wawasan melalui buku atau media cetak lainnya ditambah dengan fasilitas-fasilitas penunjang yang tersedia, sehingga diharapkan mampu menampung kegiatan-kegiatan yang berhubungan dengan dunia perbukuan secara langsung maupun tidak langsung. Toko Buku Suci merupakan salah satu dari toko buku yang ada di Kupang. Toko Buku Suci juga merupakan salah satu alternatif tempat untuk mencari buku, jika buku-buku yang diinginkan pembeli tidak ada di toko buku lain atau ketika stok buku telah habis. Selain itu, pencarian buku di Toko Buku Suci memakan waktu yang lama, karena belum adanya

sistem yang mampu mempercepat pencarian judul dan resensi buku yang diinginkan pengguna. Pengunjung yang datang membeli buku tidak hanya dari kalangan orang dewasa namun juga anak-anak, pekerja, siswa atau siswi dan mahasiswa atau mahasiswi yang sering mengalami kesulitan dalam mencari buku karena saat pembeli menanyakan buku yang ingin dibeli, pelayan memberikan buku yang tidak sesuai dengan keinginan pembeli walaupun pengarang atau judul buku sama. Jumlah buku yang ada di Toko Buku Suci kurang lebih seribu buku, dengan berbagai jenis buku untuk pelajaran TK, Paud, SD, SMP, SMA, Perguruan tinggi (perkuliahan) dan referensi populer umum. Sedangkan tataletak penyusunan buku belum diletakan secara teratur. Hal ini juga merupakan suatu masalah yang harus dipecahkan, karena kurangnya pengorganisasian letak buku, serta pencarian buku yang akan memakan waktu lama, jika karyawan kurang paham dengan sistem kerja di toko buku yang masih manual. Untuk mengatasi hal tersebut, diperlukan suatu *information retrieval system* yang dapat melakukan temu balik informasi dengan cepat dan lebih baik.

*Information retrieval system* merupakan suatu sistem yang berfungsi untuk menemukan kembali informasi yang relevan dengan kebutuhan pemakai. Dalam hal ini, temu kembali informasi berkaitan dengan representasi, penyimpanan dan akses

terhadap representasi dokumen. Untuk menyediakan solusi pada mesin pencarian agar mampu menyediakan informasi dokumen teks pada *database* yang tepat dan *query* tertentu. Hasil dari pencarian akan direpresentasikan dengan urutan (*ranking*) kemiripan dokumen dengan *query*.

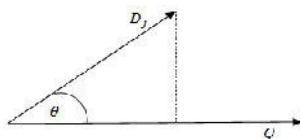
Sistem yang dibangun membantu pembeli untuk mengenali buku yang diinginkan dan teks bebas untuk menjelaskan resensi buku dan gambar kover buku. Untuk menyelesaikan persoalan di atas, perlu dibuat sebuah Sistem Temu Balik Informasi Dokumen Teks Menggunakan *Vector Space Model*, untuk pencarian buku cetak di Toko Buku Suci Kupang. Sistem yang dibuat bertujuan membantu pelayan Toko Buku Suci dalam pencarian buku cetak berdasarkan *query* yang diinginkan sehingga lebih efisien.

## 2. KAJIAN LITERATUR

*Information Retrieval* (IR) merupakan sistem untuk merepresentasikan, menyimpan, mengorganisasikan dan memperoleh informasi [1]. Sistem temu kembali informasi bertujuan untuk menjawab kebutuhan informasi *user* dengan sumber informasi yang tersedia [2]. IR teks membutuhkan *preprocessing* dalam persiapan dokumennya. Pada tahapan *preprocessing*, *query* yang dimasukkan pengguna dikonversi sesuai aturan tertentu untuk mengekstrak *term-term* penting yang sejalan dengan *term-term* yang sebelumnya telah diekstrak dari dokumen dan menghitung relevansi antara *query* dan dokumen berdasarkan *term-term* tersebut. Sebagai hasilnya sistem mengembalikan suatu daftar dokumen terurut *descending* (rangking) sesuai nilai kemiripan dengan *query* pengguna [3].

### 2.1 Vector Space Model

*Vector Space Model* (VSM) adalah metode untuk melihat tingkat kedekatan atau kesamaan (*similarity*) *term* dengan cara pembobotan *term*. Dokumen dipandang sebagai sebuah vektor yang memiliki *magnitude* (jarak) dan *direction* (arah). Pada *vector space model*, sebuah istilah direpresentasikan dengan sebuah dimensi dari ruang vektor. Relevansi sebuah dokumen ke sebuah *query* didasarkan pada similaritas diantara vektor dokumen dan vektor *query* [1].



Gambar 1 Representasi dokumen dan *query* pada ruang vektor [1]

#### 2.1.1 Menghitung bobot dokumen dengan *term-frequency and inverse document frequency* (*tf-idf*)

Formula yang digunakan pada *term frequency* (*tf*) diberikan berdasarkan jumlah kemunculan suatu kata di dalam dokumen. *Inverse document frequency* (*idf*) dihitung dengan menggunakan formula sebagai berikut [4]:

$$idf_i = \log\left(\frac{N}{df_j}\right) \quad (1)$$

Dimana:

$N$  = total jumlah dokumen dalam koleksi

$df_j$  = jumlah dokumen yang berisi *term*  $j$

$idf_i$  = frekuensi kemunculan suatu istilah dalam dokumen  $i$ .

Dengan demikian rumus untuk *Tf-Idf* adalah mengalikan nilai *term frequency* (*Tf*) dengan nilai *Inverse document frequency* (*Idf*) menggunakan formula sebagai berikut [5]:

$$W_{ij} = tf_{ij} * idf_{ij} \quad (2)$$

Dimana:

$tf_{ij}$  = frekuensi *term*  $j$  dalam dokumen  $i$

$W_{ij}$  = bobot *term*  $j$  dalam dokumen  $i$

$idf_i$  = frekuensi kemunculan suatu istilah dalam dokumen.

Setelah bobot ( $W$ ) masing-masing dokumen diketahui, maka dilakukan proses *sorting* atau pengurutan dimana semakin besar nilai  $W$ , semakin besar tingkat similaritas dokumen tersebut terhadap kata kunci, demikian sebaliknya.

#### 2.1.2 Menghitung Jarak *Query* dan Dokumen

Perhitungan jarak dokumen dan jarak *query* dilakukan setelah perhitungan bobot. Jika,  $Q$  adalah vektor *query* dan  $D_j$  adalah vektor dokumen, merupakan 2 buah vektor dalam ruang berdimensi- $n$ , dan  $\theta$  adalah sudut yang dibentuk oleh 2 vektor tersebut. Maka, [1] :

$$Q \bullet D_j = |Q| |D_j| \cos \theta \quad (3)$$

Dimana,  $Q \bullet D_j$  adalah hasil perkalian dalam (*inner product*) kedua vektor, sedangkan perhitungan jarak *query* menggunakan persamaan (4) dan dokumen menggunakan persamaan (5) sebagai berikut :

$$|Q| = \sqrt{\sum_{j=1}^n (W_{i,q})^2} \quad (4)$$

Dengan  $|Q|$  adalah jarak *query*, dan  $W_{iq}$  adalah bobot *query* dokumen ke- $i$ , maka jarak *query* ( $|Q|$ ) dihitung untuk didapatkan jarak *query* dari bobot *query* dokumen ( $W_{i,q}$ ) yang terambil oleh sistem. Jarak *query* bisa

dihitung dengan persamaan akar jumlah kuadrat dari *query*.

$$|D_j| = \sqrt{\sum_{i=1}^n (W_{i,j})^2} \quad (5)$$

Dengan  $|D_j|$  adalah jarak dokumen, dan  $W_{i,j}$  adalah bobot dokumen ke-i, maka jarak dokumen ( $|D_j|$ ) dihitung untuk didapatkan jarak dokumen dari bobot dokumen ( $W_{i,j}$ ) yang terambil oleh sistem. Jarak dokumen bisa dihitung dengan persamaan akar jumlah kuadrat dari dokumen. Persamaan di atas merupakan panjang vektor atau jarak *euclidean* suatu vektor dengan titik nol.

### 2.1.3 Menghitung Similaritas *Query* dan Dokumen

Setelah jarak dari tiap dokumen dan *query* didapatkan, dilakukan perhitungan similaritas antara *query* dengan dokumen. Perhitungan kesamaan kedua vektor adalah sebagai berikut:

$$Sim(Q, D_j) = \sum_{i=1}^n W_{iq} \cdot W_{ij} \quad (6)$$

Dengan  $W_{ij}$  adalah bobot term dalam dokumen,  $W_{iq}$  adalah bobot *query*, dan  $Sim(Q, D_j)$  adalah Similaritas antara *query* dan dokumen. Similaritas antara *query* dan dokumen (*inner product*) atau  $Sim(Q, D_j)$  digunakan untuk mendapatkan bobot dengan didasarkan pada bobot *term* dalam dokumen  $W_{ij}$  dan bobot *query*  $W_{iq}$  atau dengan cara menjumlah bobot Q dikalikan dengan bobot dokumen.

### 2.1.4 Menghitung *Cosine Similarity* (*cosine coefficient*)

Pengukuran *cosine similarity* (menghitung nilai cosine sudut antara dua *vector*) menggunakan persamaan (7) dan (8)

$$Sim(Q, D_j) = \cos(Q, D_j) = \frac{Q \cdot D_j}{|Q||D_j|} = \frac{1}{|Q||D_j|} \sum_{i=1}^n W_{iq} \cdot W_{ij} \quad (7)$$

Dari formula diatas dapat dinormalisasi menjadi [5]:

$$Sim(Q, D_j) = \frac{\sum_{i=1}^n W_{iq} \cdot W_{ij}}{\sqrt{\sum_{j=1}^n (W_{iq})^2 \cdot \sum_{i=1}^n (W_{ij})^2}} \quad (8)$$

Similaritas antara *query* dan dokumen atau  $Sim(Q, D_j)$  berbanding lurus terhadap jumlah bobot *query* (Q) dikali bobot dokumen ( $D_j$ ) dan berbanding terbalik terhadap akar jumlah kuadrat ( $|Q|$ ) dikali akar jumlah kuadrat dokumen ( $|D_j|$ ). Perhitungan similaritas menghasilkan bobot dokumen yang mendekati nilai 1 atau menghasilkan bobot dokumen yang lebih besar dibandingkan

dengan nilai yang dihasilkan dari perhitungan *inner product*.

## 2.2 Precision and Recall

Parameter yang digunakan dalam mengukur performansi sistem temu kembali adalah:

### (i) Precision

Dalam bidang pencarian informasi, *precision* (disebut juga *positive prediction value*) merupakan metrik untuk mengukur kinerja sistem dalam mendapatkan data yang relevan [6]. *Precision* ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen yang terambil oleh sistem baik relevan maupun tidak relevan.

$$precision = \frac{TP}{TP+FP} \quad (9)$$

Keterangan :

TP=True Positif (jumlah data benar yang bernilai positif)

FP=False Positif (data salah yang bernilai positif)

### (ii) Recall

*Recall* (disebut juga sensitivitas) merupakan metrik untuk mengukur kinerja sistem dalam mendapatkan data relevan yang terbaca [8]. *Recall* ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen relevan yang ada dalam koleksi dokumen (terambil ataupun tak terambil sistem).

$$recall = \frac{TP}{TP+FN} \quad (10)$$

Keterangan :

TP=True Positif (jumlah data benar yang bernilai positif)

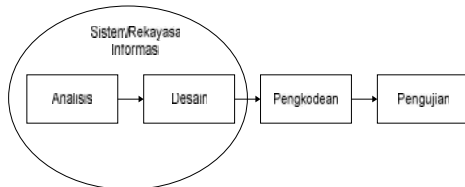
FN=False Negatif (data salah bernilai negatif)

## 2.3 Penelitian Terkait

Beberapa penelitian yang berkaitan dengan sistem temu balik informasi pernah dilakukan [7] yang menghasilkan sebuah sistem yang dapat mencari informasi dari isi *file* dokumen yang disimpan di dalam sistem. Penelitian ini dikembangkan dengan metode *waterfall*. Metode pembobotan yang digunakan dalam penelitian ini adalah tf-idf. Selain itu penelitian [8] menghasikan sistem temu balik informasi pencarian skripsi menggunakan pembobotan tf.idf dan *vector space model*. Sistem temu balik informasi dengan menggunakan metode VSM dirancang untuk mendapatkan hasil pencarian dengan nilai similaritas tinggi. Metode VSM dipilih karena cara kerja model ini efisien, mudah dalam representasi dan dapat diimplementasikan pada dokumen teks.

### 3. METODE PENELITIAN

Metode yang digunakan pada penelitian sistem temu balik melalui beberapa tahapan dan kebutuhan Hardware maupun Software. Sistem temu balik didesain metode pengembangan *Software Development Life Cycle* (SDLC) air terjun (*Waterfall*) atau disebut sekuensial linier (*Sequential Linear*). Proses model *waterfall* terlihat seperti Gambar



Gambar 2 Model *waterfall* [9]

#### 3.1. Analisis sistem

Analisis sistem merupakan suatu analisis yang terdiri dari suatu sistem yang utuh kedalam suatu komponen dengan maksud untuk mengevaluasi permasalahan-permasalahan dan kebutuhan yang diharapkan sehingga dapat diusulkan teori perbaikannya.

Adapun analisis juga dilakukan terhadap peran sistem dan peran pengguna yaitu sebagai berikut:

- (i) Analisis kebutuhan sistem  
Analisis kebutuhan sistem dilakukan untuk mengetahui fasilitas yang harus disediakan oleh sistem. Fungsi utama dari sistem temu balik ini adalah untuk membantu pencarian pencarian informasi agar dapat dilakukan secara cepat, efisien dan sesuai dengan *query* yang dimasukan oleh *user*. Untuk menghasilkan *output* yang baik, maka sistem harus didukung oleh sistem perangkat keras (*hardware*) dan perangkat lunak (*software*).
- (ii) Analisis peran sistem  
Dalam perancangan sistem ini, sistem yang akan dibangun memiliki peranan sebagai berikut :
  - a. Sistem yang dibuat dapat mengolah, menginput, menyimpan, melihat, mengubah dan menghapus dokumen teks.
  - b. Sistem dapat memberikan *output* berupa hasil pencarian dokumen teks.
- (iii) Analisis peran pengguna  
Dalam sistem ini terdapat dua (2) kategori pengguna dalam sistem ini yaitu *admin* dan *user*, dimana *admin* tersebut merupakan *entity* yang terlibat langsung di dalam sistem ini.
  - a. *Admin* yang akan menginput, melihat, menambah, menghapus, mengedit dokumen teks. Sistem juga dapat melakukan pencarian (*search*) dokumen teks.

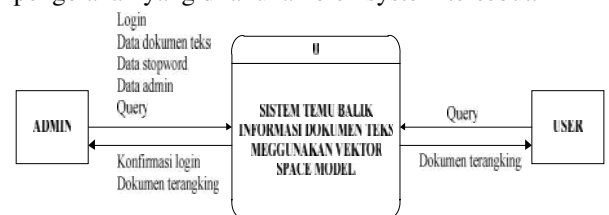
- b. *User* yang menggunakan sistem ini dapat melakukan pencarian (*search*) dokumen teks berdasarkan *query* yang dimasukan.

#### 3.2. Perancangan Perangkat Lunak

Tahap perancangan adalah proses penggunaan berbagai teknik dan prinsip untuk tujuan mendefinisikan proses atau sistem secara detail. Adapun perancangan perangkat lunak adalah sebagai berikut:

##### 3.2.1 Pemodelan proses

Pemodelan proses pada penelitian ini dilakukan dengan menggunakan tools Data Flow Diagram (DFD) adalah alat yang menggambarkan aliran data melalui sistem dan kerja/tugas atau pengolahan yang dilakukan oleh sistem tersebut.



Gambar 3 Diagram konteks

Diagram konteks, adalah sebuah teknik grafik yang menggambarkan aliran data dari keseluruhan sistem. Dengan melihat gambar 3 menjelaskan bahwa sistem yang akan dibangun memiliki dua aktor yaitu user dan admin.

##### 3.2.2 Perancangan basis data dan antar muka

Adapun Basis data dan struktur table yang akan dibuat pada sistem temu kembali informasi adalah table *tb\_dokumen*, *tb\_indeks*, *tb\_bobot*, *tb\_vektor*, dan *tb\_similarity*, *tb\_stopword*, *tb\_admin*.

Pada tahap ini antarmuka dari perangkat lunak akan dibuat dalam bentuk storyboard untuk visualisasi ide dari aplikasi yang akan dibangun, sehingga dapat memberikan gambaran dari aplikasi yang akan dihasilkan.

##### 3.2.3 Implementasi dan pengujian

Implementasi sistem menggunakan PHP dan komputasi dengan bantuan *store procedure*. Pengujian performansi sistem menggunakan *precision and recall* dan unjuk kerja sistem dengan metoda *black box*.

## 4 HASIL DAN PEMBAHASAN

### 4.1 Implementasi sistem

Tahap implementasi menerjemakan hasil analisa dan perancangan sehingga menghasilkan suatu sistem temu kembali informasi yang dapat digunakan pengguna untuk mencari informasi dokumen teks.

# PROSIDING SEMMAU 2017

## 4.1.1 Proses *retrieval*

Implementasi dari proses *retrieval* adalah proses untuk mencari dokumen teks yang relevan berdasarkan kata kunci yang di masukan pengguna. Hasil dari implementasi proses *retrieval* dapat dilihat pada tabel berikut.

Tabel 1. Hasil Perhitungan Indeks

No	Kata kunci	Id_Dok	Kemiripan
1	Website	1	1.46992
		2	1.0481
		6	0.301328
2	Pemrograman Dekstop	1	0.337332
		3	0.949616
		4	1.67086
		5	1.89707
3	HTML, CSS, PHP & Mysql	2	1.16929
		4	0.910098
		5	0.768527
		6	0.301328
4	Dekstop Database Python	2	0.698732
		3	0.288526
		4	0.796822
		6	0.903983
5	PC Menjadi Server Web Dan Database	1	0.337332
		2	0.458823
		3	0.237404
		4	1.21346
		5	0.384263
		6	0.903983

Hasil pencarian dari sistem yang ada dalam Tabel 1 hasil pencarian adalah beberapa data implementasi dari hasil pencarian dokumen berdasarkan kata kunci yang dimasukkan pengguna dengan mencari nilai *cosine similarity* berdasarkan formula 4. Dapat dilihat untuk kata kunci "Website", sistem menemukan 3 dokumen yang relevan yaitu id\_dok 1 dengan kemiripan 1.46992 dan id\_dok 2 dengan kemiripan 1.0481 dan id\_dok 6 dengan kemiripan 0.301328.

## 4.1.2 Pengujian *Indexing*

Pengujian ini dilakukan untuk mengetahui waktu pembuatan index pada sistem, waktu pembuatan index akan diuji berdasarkan banyaknya dokumen yang di index dan lamanya proses index.

Sample data yang dimasukkan sebanyak 10 dokumen dan akan dihitung waktu proses index setiap dokumen. Pengujian indexing sebanyak 10 dokumen lama proses indexing yang dibutuhkan sistem yaitu 3.19 detik dengan jumlah term sebanyak 2156 term yang di berikan bobot. Untuk melihat hasil keseluruhan pengujian waktu indexing yang telah dilakukan dapat dilihat pada tabel 2 pengujian waktu indexing.

Tabel 2 Pengujian Waktu Indexing

No	Jumlah dokumen	Waktu indeks	Jumlah token
1	1	0.432	144
2	2	0.483	294
3	3	1.13	532
4	4	1.24	752
5	5	1.81	901
6	6	2.3	1194
7	7	2.39	1480
8	8	2.66	1741
9	9	3.03	1939
10	10	3.19	2156

## 4.1.3 Pengujian *Retrieval*

Hasil pengujian *recall* dan *precision* pada sistem dengan menggunakan data sampel sebanyak 10 dokumen. Diperlihatkan pada tabel 3 pengujian *recall* dan *precision*. Pengujian dilakukan sebanyak 5 kali, dengan menggunakan kata kunci yang berbeda di setiap pengujian.

Tabel 3 Deskripsi Dokumen

Pengujian	Query	T1	Hasil Pencarian			Recall	Precision
			T2	T3	T4		
<b>K01</b>	Website	2	2	0	2	1	0.667
<b>K02</b>	Pemrograman Dekstop	1	1	4	5	1	0.2
<b>K03</b>	HTML, CSS, PHP & Mysql	1	1	3	4	1	0.25
<b>K04</b>	Dekstop Database Python	1	1	3	4	1	0.25
<b>K05</b>	PC Menjadi Server Web Dan Database	1	1	5	6	1	0.1667

Keterangan :

T1 = Dokumen relevan dalam koleksi

T2 = Dokumen relevan

T3 = Tidak relevan

T4 = Jumlah dokumen ditemukan

Nilai dari *precision* memiliki nilai yang berbeda hampir di setiap pengujian. Nilai dari

# PROSIDING SEMMAU 2017

---

*precision* tergantung dari keunikan kata kunci yang diberikan, semakin unik kata kunci yang diberikan maka semakin tinggi nilai *precision* yang diperoleh sebaliknya semakin umum kata kunci yang diberikan maka akan semakin kecil nilai dari *precision* tersebut.

## 5 KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan sistem temu kembali informasi berhasil dibangun dan dapat mengimplementasikan pengindeksan otomatis dengan menggunakan metode pembobotan TF-IDF dan dapat menemukan dokumen yang relevan terhadap kata kunci dengan menghitung nilai *cosine similarity*. Waktu proses *indexing* sistem berdasarkan pengujian waktu *indexing*, waktu proses *indexing* terus meningkat dengan banyaknya dokumen yang di index dengan waktu pengujian *indexing* pada 10 dokumen teks 3,19 detik dengan banyaknya kata yang terindeks adalah 2156.

## REFERENSI

- [1] Baeza-Yates, R., Riberio-Neto, B., 1999, *Modern Information Retrieval, Addison Wesley-Pearson international edition*, Boston. USA.
- [2] Salton, G., 1989, *Automatic Text Process The Transformation, Analysis, and Retri of Information by Computer*, Addison Wesley Publishing Company, Inc. Un States of America
- [3] Cios, K. J., Pedrycz, W., Swiniarski, R. 2007, *Data Mining A Knowledge Disco Approach*, Springer, New York
- [4] Manning, C., Raghavan, P. dan Schutze, H., 2009, *Introduction to Information Retrieval*, Cambridge University Press, England.
- [5] Lu, G., 1999, *Multimedia Database Management Systems*, Artech House Inc. Canton Street.
- [6] Prasetyo, 2014, *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*, Andi, Yogyakarta
- [7] Nadirman, F., 2006. *Sistem Temu Kembali Informasi Dengan Metode Vector Space Model Pada Pencarian File Dokumen Berbasis Teks*, Tugas Akhir, Unpublished, Universitas Gajah Mada Yogyakarta
- [8] Lake, L.K., 2015, *Information Retrieval System Dokumen Teks Abstrak Skripsi Dengan Metode Pembobotan Tf-Idf Dan Pengukuran Similarity Menggunakan Vector Space Model*, Tugas Akhir, Unpublished, Universitas Katolik Widya Mandira Kupang.
- [9] Rossa, A., S., Shalahuddin, M., 2015, *Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Ojek*, Informatika, Bandung.

# STIKOM UYELINDO KUPANG

PROGRAM STUDI :  
SISTEM INFORMASI (S1) TERAKREDITASI  
TEKNIK INFORMATIKA (S1) TERAKREDITASI  
TEKNIK INFORMATIKA (D3) TERAKREDITASI

Jalan Perintis Kemerdekaan I -Kayu Putih Kupang-NTT

Telp; 0380-8554500, 85554499, Fax.0380-8554502

Website: <http://www.uyelindo.ac.id>

Website: <http://www.semmanu.uyelindo.ac.id>

Email: [stikom@uyelindo.ac.id](mailto:stikom@uyelindo.ac.id), [semmanu@uyelindo.ac.id](mailto:semmanu@uyelindo.ac.id)



ISBN



978-602-73628-0-2