Jurnal Mahajana Informasi. Vol. 7, No. 1, Juni 2022
e-ISSN : 2527-8290

Information Retrieval Sistem Kearsipan Pencarian Dokumen Di Dinas Pemberdayaan Perempuan Dan Perlindungan Anak Kota Semarang Menggunakan Metode Vector Space Model

Anggoro Teguh Adiyanto*[1], Dewi Handayani UN[2]
[1,2]Program Studi Teknik Informatika, Universitas Stikubank Semarang
e-mail: *[1]anggoroteguh33@gmail.com, [2] dewi_h@edu.unisbank.ac.id

Abstrak

Pengelolaan arsip secara tertib dan terpadu dengan memanfaatkan teknologi merupakan salah satu indikator keberhasilan reformasi birokrasi untuk membangun. Saat ini Kementerian Pemberdayaan Perempuan dan Perlindungan Anak sedang mengintegrasikan pengelolaan kearsipan ke dalam sistem e-office yang merupakan bagian dari pelaksanaan Sistem Pemerintahan Berbasis Elektronik (SPBE). Sementara itu, kondisi pengelolaan kearsipan di Kemen PPPA pada 2018 hingga kini masih memprihatinkan karena masih berada pada katagori "tidak baik". Hal tersebut berakibat pada pencarian data yang dibutuhkan menyulitkan petugas terkait. Untuk dapat melakukan pencarian berdasar substansi yang paling mirip, terdapat teknologi yang disebut *information Text Retrieval user*. *Information text retrieval* adalah salah satu metode yang digunakan untuk menyimpan data dengan cara memprosesnya (menghilangkan *stop word*) dan menyimpan tiap kata beserta informasi dari kata tersebut (letak kata, jumlah bobot). Information retrieval berfokus pada proses yang terlibat di dalam representasi, media penyimpanan, mencari dan menemukan informasi yang relevan dari informasi yang diinginkan oleh user. Metode *Vector Space Model* merupakan salah satu alternatif yang dapat diimplementasikan untuk memecahkan masalah ini. *Vector space model* (VSM) atau model ruang vektor adalah suatu metode untuk merepresentasikan dokumen dan query dalam bentuk vektor pada ruang multidimensional . Dengan metode *Vector Space Model* dapat dilihat tingkat kedekatan atau kesamaan dengan cara pembobotan term. Metode ini dapat di terapkan dalam pengembangan sistem *information retrieval* pada pengelolaan kearsipan di Kemen PPPA, sehingga dapat memberikan perbaikan dalam kearsipan di terutama dalam pencarian data. Pada percobaan menggunakan sampel dalam sistem dan analisis manual, metode VSM dapat memberikan solusi pencarian dokumen dengan tepat.

Kata Kunci: Information retrieval, vector space model, data kearsipan.

Abstract

Management of archives in an orderly and integrated manner by utilizing technology is one indicator of the success of bureaucratic reform to build. Currently, the Ministry of Women's Empowerment and Child Protection is integrating archive management into the e-office system which is part of the implementation of the Electronic-Based Government System (SPBE). Meanwhile, the condition of archive management at the Ministry of PPPA in 2018 is still concerning because it is still in the "not good" category. This results in the search for the required data making it difficult for the relevant officers. To be able to search based on the most similar substance, there is a technology called information Text Retrieval user. Information text retrieval is one of the methods used to store data by processing it (removing stop words) and storing each word along with information from the word

(word location, number of weights). Information retrieval focuses on the processes involved in the representation, storage media, searching and finding relevant information from the information desired by the user. The Vector Space Model method is an alternative that can be implemented to solve this problem. Vector space model (VSM) or vector space model is a method for representing documents and queries in vector form in multidimensional space. With the Vector Space Model method, it can be seen the level of closeness or similarity by means of term weighting. This method can be applied in the development of an information retrieval system in archive management at the PPPA Ministry, so that it can provide improvements in archives, especially in data retrieval. In experiments using samples in the system and manual analysis, the VSM method can provide an appropriate document search solution.

Keywords: Information retrieval, vector space model, archival data.

PENDAHULUAN

Saat ini Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (Kemen PPPA) sedang mengintegrasikan pengelolaan kearsipan ke dalam sistem e-office yang merupakan bagian dari pelaksanaan Sistem Pemerintahan Berbasis Elektronik (SPBE). Sementara itu, kondisi pengelolaan kearsipan di Kemen PPPA pada 2018 hingga kini masih memprihatinkan karena masih berada pada katagori "tidak baik" (PPPA, 2021). Hal tersebut berakibat pada pencarian data yang dibutuhkan menyulitkan petugas terkait.

Dengan berkembangnya era informasi pada saat ini DPPPA Kota Semarang timbul sebuah gagasan untuk membangun sebuah sistem mesin pencari yang dapat memberikan pembobotan dari masing masing nama data yang sesuai dengan kata kunci yang dicari, sehingga akan dapat memilih yang mana data yang menjadi paling diprioritaskan. Pencarian pada database yang dilakukan hanya mampu mencari judul yang sesuai berdasarkan kata kunci yang diinputkan, misalnya, jika kata kunci yang dimasukkan adalah "sistem cerdas" maka akan ditampilkan semua dokumen yang mengandung kata "sistem cerdas" namun sistem tidak bisa mengukur mana dokumen yang paling diprioritaskan.

Untuk dapat melakukan pencarian berdasar substansi yang paling mirip, terdapat teknologi yang disebut information Text Retrieval user [9]. Information text retrieval adalah salah satu metode yang digunakan untuk menyimpan data dengan cara memprosesnya (menghilangkan stop word) dan menyimpan tiap kata beserta informasi dari kata tersebut (letak kata, jumlah bobot, dll). Information retrieval berfokus pada proses yang terlibat di dalam representasi, media penyimpanan, mencari dan menemukan informasi yang relevan dari informasi yang diinginkan oleh user [9].

Metode Vector Space Model merupakan salah satu alternatif yang dapat diimplementasikan untuk memecahkan masalah ini. Vector space model (VSM) atau model ruang vektor adalah suatu metode untuk merepresentasikan dokumen dan query dalam bentuk vektor pada ruang multidimensional [8]. Dengan metode Vector Space Model dapat dilihat tingkat kedekatan atau kesamaan dengan cara pembobotan term.

Pada proses stemming atau mencari kata dasar pada kata, sistem akan menggunakan algoritma tala. Penelitian yang telah dilakukan oleh Mardi Siswo Utomo dalam [8], menyatakan bahwa Stemmer tala merupakan adopsi dari algoritma stemmer bahasa inggris terkenal porter stemmer. Stemmer ini menggunakan rule base analisis untuk mencari root sebuah kata. Pada sistem pencarian buku akan menggunakan stemming algoritma tala karna algoritma tala merupanya pengembangan dari algoritma porter [8].

Banyak algoritma dan teknik telah dikembangkan di bidang olah data dan pengambilan informasi namun mengambil data dari basis data besar terus menjadi masalah. Dalam penelitian sebelumnya [5], menggunakan model ruang vektor untuk melakukan Proses pencarian yang sebelumnya harus menyertakan judul dokumen secara lengkap, setelah menerapkan konsep information retrival system, pencarian dapat dilakukan dengan lebih cepat tepat dan akurat, tanpa perlu melakukan pencarian judul dokumen secara terperinci, sistem akan menyamakan keyword yang di masukan dengan dokumen yang tersimpan pada aplikasi dengan mengambil informasi dari database sebagai bahan uji coba.

LANDASAN TEORI

Tinjauan pustaka

Pencarian dokumen teks pada mesin pencari dengan metode Vector Space Model (VSM) mampu merepresentasikan dokumen dan query dalam bentuk vektor dimensional. Metode VSM untuk pencarian kata untuk mencari kata "Tuhan" pada Al Quran melalui tahapan pertama yaitu proses text preproccessing yang menggunakan 4 tahapan yaitu *case folding, tokenizing, filtering, stemming (algoritma porter stemmer) dan term weighting* (metode TF-IDF) yang berfungsi untuk memaksimalkan hasil pencarian. Dan proses similaritas (cosine similarity) yang berfungsi untuk mendapatkan kata yang dimaksud dengan query serta jaraknya untuk pengurutan [13].

Selama ini perolehan informasi terhadap banyaknya data yang terambil pada saat pencarian dianggap tidak efektif, yang berdampak pada rendahnya akurasi hasil pencarian sehingga diperlukan aplikasi mesin pencari yang memiliki tingkat presisi yang tinggi. Amin dkk,2021 membangun sistem temu kembali informasi (STKI) menggunakan metode vector space model (VSM) yang bisa melakukan pencarian informasi dengan tingkat presisi tinggi dalam bentuk mesin pencari. Aplikasi perolehan informasi diimplementasikan terhadap banyaknya jumlah data cerita rakyat Nusantara untuk tujuan menemukan cerita yang relevan dengan kebutuhan pengguna.

Vector space model atau VSM memiliki efektifitas dalam pencarian kata karena hasil pencariannya berdasarkan kemiripan vectorquery dan vector dokumen. Implememtasi Algoritma VSM dengan tahapan : preproccessing text menggunakan 4 tahapan, pembobotan term menggunakan metode TF-IDF, dan perangkingan menggunakan metode Cosine Similarity dalam penelitian yang dilakukan oleh Hadiono dkk, 2018 membahas tentang banyaknya informasi berbasis teks yang dapat disimpan, memunculkan potensi kesulitan untuk mendapatkan informasi yang diperlukan serta dampaknya bila membutuhkan waktu lama jika pencarian dokumen dilakukan satu persatu.

Metode *Vector Space Mode*l merupakan salah satu alternatif yang dapat diimplementasikan untuk masalah perolehan informasi berdasar tingkat kedekatan atau kesamaan dengan cara pembobotan term. Metode pembobotan data dengan menghitung jarak antar dokumen. Implementasi metode vector space model pada perolehan informasi dengan memberikan rekomendasi pada pencarian buku sesuai dengan prioritas pengguna berdasarkan pembobotan pada judul buku [8].

Sistem temu kembali judul tugas akhir dan perhitungan kemiripan dokumen menggunakan vector space model dengan melakukan pencarian berdasar substansi yang paling mirip dalam teknologi *information text reitrieval*, dimana sistem ini secara otomatis melakukan *indexing* secara *offline* dan temu kembali (*retrieval*) secara real time. Dokumen akan ditampilkan diurutkan berdasarkan dokumen yang paling mirip [9]. Tahapan yang dilakukan unutk perolehan kembali informasi tanpa perlu melakukan pencarian judul dokumen secara terinci, dimana sistem akan menyamakan keyword yang dimasukkan dengan dokumen

yang tersimpan. Pertama, menghitung skor kemiripan menggunakan rata-rata tertimbang dari setiap item ukuran kosinus kemudian menghitung ukuran kemiripan dan untuk menentukan sudut antara vektor dokumen dan vektor query karena VSM didasarkan pada geometri di mana setiap istilah memiliki dimensi sendiri dalam ruang multi-dimensi, pertanyaan dan dokumen adalah titik atau vektor dalam ruang ini [5].

(Sanjaya, 2017) menyatakan perlu diterapkannya metode ilmu pencarian yang dikenal dengan temu kembali informasi (Information Retrieval). Salah satunya metode dalam sistem temu kembali adalah.

Sistem temu kembali informasi dengan metode Vector Space Model (VSM) sebelum melakukan pencarian dokumen akan dilakukan indexing dengan memecah isi teks dari dokumen-dokumen tersebut menjadi index term. Index term ini yang akan digunakan untuk proses pencarian. Proses pembentukan index term dari teks yang terdapat di dalam dokumen akan melalui beberapa tahapan yaitu *parsing*, *text preprocessing* , penghitungan bobot , dan juga pengukuran kesamaan [11].

Salah satu solusi untuk mengatasi masalah pencarian dokumen skripsi menggunakan sistem temu kembali information dapat digambarkan sebagai sebuah proses untuk mendapatkan dokumen yang relevan dari sekumpulan dokumen yang ada melalui pencarian *query* yang diinputkan pengguna. *Information Retrieval* (IR) merupakan suatu metode untuk menemukan kembali data tidak terstruktur yang tersimpan pada sekumpulan dokumen, kemudian menyediakan informasi mengenai subyek yang dibutuhkan [7].

Sistem temu kembali informasi untuk perolehan kembali informasi di perpustakaan dengan tingkat presisi yang tinggi dan hasil pengukuran pada jeda waktu pencarian yang cepat dan efektif dengan waktu rata-rata 0,48 detik telah mampu mencakup dan memproses banyak dan beragam koleksi dari berbagai subjek yang digunakan oleh pengguna perpustakaan. Tampilan katalog juga memudahkan pengguna perpustakaan untuk menggunakan dan mencari informasi [16].

Text Mining

Text mining adalah teknik yang membantu pengguna untuk menemukan informasi yang berguna dari sejumlah besar dokumen teks digital di web atau database. Oleh karena itu penting bahwa model penambangan teks yang baik harus mengambil informasi yang memenuhi kebutuhan pengguna dalam kerangka waktu yang relatif efisien [14].

Text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seseorang user berinteraksi dengan sekumpulan dokumen menggunakan tool analisis yang merupakan komponen-komponen dalam data mining. Text mining digunakan untuk mengolah dokumen sebelum dilakukan proses similarity. Didalam proses text mining terdapat proses preprocessing. Preprocessing text merupakan tindakan menghilangkan karakter-karakter tertentu yang terkandung dalam dokumen, seperti koma, tanda petik dan lain-lain serta mengubah semua huruf kapital menjadi huruf kecil [12].

Berikut tahapan-tahapan proses didalam text mining/preprocessing[17]:

1. Casefolding

Casefolding adalah tahap mengubah huruf besar menjadi huruf kecil dalam dokumen maka penghapusan tanda baca selain huruf "a" to "z" yang dianggap sebagai karakter pembatas.

2. Tokenizing

Tokenizing adalah tahap memecah kalimat menjadi kata-kata. Dengan kata terbelah pertama, string yang sudah ada masukan akan lebih sederhana karena ditampilkan dalam setiap kata menurut ruang yang membaginya, sehingga dengan bentuk itu, akan mempermudah proses perubahan menjadi batang kata.

3. Filtering

Filtering adalah tahap penghapusan kata-kata tidak dianggap mengandung arti atau pemikiran yang seharusnya ada ada (stopwords). Stopword berisi kata-kata umum yang sering muncul dalam sebuah dokumen dalam jumlah banyak namun tidak memiliki kaitan dengan tema tertentu.

4. Stemming

Stemming adalah mengembalikan kata-kata yang diperoleh dari hasil filtering ke bentuk dasarnya, menghilangkan imbuhan awal (*prefix*) dan imbuhan akhir (*sufix*) sehingga didapat kata dasar.

TF-IDF (Term Frequency-Inverse Document Frequency)

Basis pembobotan TF-IDF merupakan jenis pembobotan yang melibatkan pengukuran statistik untuk mengukur seberapa penting sebuah kata dalam kumpulan dokumen. Tingkat kepentingan meningkat ketika sebuah kata muncul beberapa kali dalam sebuah dokumen tetapi diimbangi dengan frekuensi kemunculan kata tersebut dalam kumpulan dokumen (Wisnu & Hetami, 2015). TF merupakan pembobotan yang sederhana dimana penting tidaknya sebuah kata diasumsikan sebanding dengan jumlah kemunculan kata tersebut dalam dokumen, sementara IDF merupakan pembobotan yang mengukur seberapa penting sebuah kata dalam dokumen apabila dilihat secara global pada seluruh dokumen [3].

Perhitungan IDF menggunakan persamaan 1

$$(t) = \log(D / df(t)) \text{ ..... (1)}$$

Dimana:

df(t) = Jumlah dokumen yang mengandung kata ke-t dari kata kunci

D = Jumlah semua dokumen yang ada di dalam database

IDF = Rasio frekuensi dokumen pada kata ke-t dari kata kunci

Perhitungan TF-IDF menggunakan persamaan 2

$$TF\text{-}IDF(d,t) = TF(d,t) * IDF(t) \text{ ..... (2)}$$

Dimana:

d = dokumenke-d

t = kata ke-t dari kata kunci

tf = frekuensi banyaknya kata ke-t dari kata kunci pada dokumen ke-d

TF-IDF = bobot dokumen ke-d terhadapkata kunci ke-t

IDF = rasio frekunsi dokumen pada kata ke-t dari kata kunci.

Vector Space Model

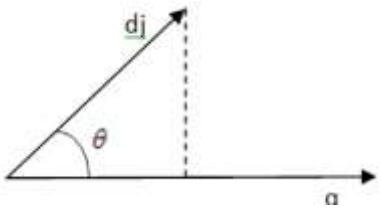
Menurut (Singh dan Singh, 2016) menjelaskan VSM adalah model aljabar yang digunakan untuk Informasi Pengambilan kembali. Ini merupakan dokumen bahasa alami dalam cara formal dengan menggunakan vektor dalam multidimensional ruang. Vector Space Model (VSM) adalah cara mewakili dokumen melalui kata-kata itu berisi. Konsep di balik vector space model adalah dengan menempatkan istilah, dokumen, dan pertanyaan dalam ruang term-document adalah mungkin untuk menghitung persamaan antara kueri dan istilah atau dokumen, dan memungkinkan hasil perhitungan untuk diberi peringkat sesuai dengan ukuran kemiripan diantara mereka. VSM memungkinkan keputusan dibuat tentang dokumen mana yang mirip satu sama lain dan untuk pertanyaan.

Dimensi sesuai dengan jumlah term dalam dokumen yang terlibat. Pada model ini:

1. Vocabulary merupakan kumpulan semua term berbeda yang tersisa dari dokumen setelah preprocessing dan mengandung t term index. Term-term ini membentuk suatu ruang vektor.
2. Setiap term i di dalam dokumen atau query j, diberikan suatu bobot (weight) bernilai real wij.
3. Dokumen dan query diekspresikan sebagai vektor t dimensi dj = (w1, w2, ..., wtj) dan

terdapat n dokumen di dalam koleksi, yaitu j = 1, 2, ..., n.

Sebuah dokumen dj dan sebuah query q direpresentasikan sebagai vektor dimensi seperti pada gambar.



Gambar 1 Representasi Dokumen dan Query pada Ruang Vektor (Leman & Andesa, 2017)

Proses perhitungan Vector space model melalui tahapan perhitungan term frequency (tf) menggunakan persamaan (1) (Fatkhul Amin, 2012) dalam (Leman & Andesa, 2017).

tf = tfij ..... (3)

Dengan tf adalah term frequency, dan tfij adalah banyaknya kemunculan term ti dalam dokumen dj, Term frequency (tf) dihitung dengan menghitung banyaknya kemunculan term ti dalam dokumen dj.

Perhitungan Inverse Document Frequency (idf), menggunakan persamaan :

idfi=log N/df ..... (4)

Dengan idfi adalah inverse document frequency, N adalah jumlah dokumen yang terambil oleh sistem, dan dfi adalah banyaknya dokumen dalam koleksi dimana term ti muncul di dalamnya, maka Perhitungan idfi digunakan untuk mengetahui banyaknya term yang dicari (dfi) yang muncul dalam dokumen lain yang ada pada database.

Perhitungan term frequency Inverse Document Frequency (tfidf), menggunakan persamaan :

Wij=tfij.log N/df ..... (5)

Dengan Wij adalah bobot dokumen, N adalah Jumlah dokumen yang terambil oleh sistem, tfij adalah banyaknya kemunculan term ti pada dokumen dj, dan dfi adalah banyaknya dokumen dalam koleksi dimana term ti muncul di dalamnya. Bobot dokumen (Wij) dihitung untuk didapatkannya suatu bobot hasil perkalian atau kombinasi antara term frequency (tfij) dan Inverse Document Frequency (idf).

Perhitungan Jarak query, menggunakan persamaan :

|q|=√(∑_(i=1)^t▒ 〖(Wig)〗) ..... (6)

Dengan |q| adalah Jarak query, dan Wiq adalah bobot query dokumen ke-i, maka Jarak query (|q|) dihitung untuk didapatkan jarak query dari bobot query dokumen (Wiq) yang terambil oleh sistem. Jarak query bisa dihitung dengan persamaan akar jumlah kuadrat dari query.

Perhitungan Jarak Dokumen, menggunakan persamaan :

|dj|=√(∑_(i=1)^t▒ 〖(Wij)〗) ..... (7)

Dengan |dj| adalah jarak dokumen, dan Wij adalah bobot dokumen ke-i, maka Jarak dokumen (|dj|) dihitung untuk didapatkan jarak dokumen dari bobot dokumen dokumen (Wij) yang terambil oleh sistem. Jarak dokumen bisa dihitung dengan persamaan akar jumlah kuadrat dari dokumen.

Menghitung index terms dari dokumen dan query (q,dj). menggunakan persamaan :

q,dj=∑_(i=1)^t▒ 〖Wig.Wij〗 ..... (8)

Dengan Wij adalah bobot term dalam dokumen, Wiq adalah bobot query.

Pengukuran Cosine Similarity menghitung nilai kosinus sudut antara dua vector menggunakan persamaan :

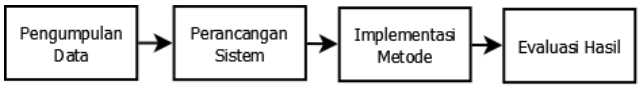
sim(q.dj)=(q .dj)/(|q|*|dj|) ..... (9)

Similaritas antara query dan dokumen atau Sim(q,dj) berbanding lurus terhadap jumlah bobot query (q) dikali bobot dokumen (dj) dan berbanding terbalik terhadap akar jumlah kuadrat q (|q|) dikali akar jumlah kuadrat dokumen (|dj|). 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.

METODE PENELITIAN

Skema Alur Penelitian

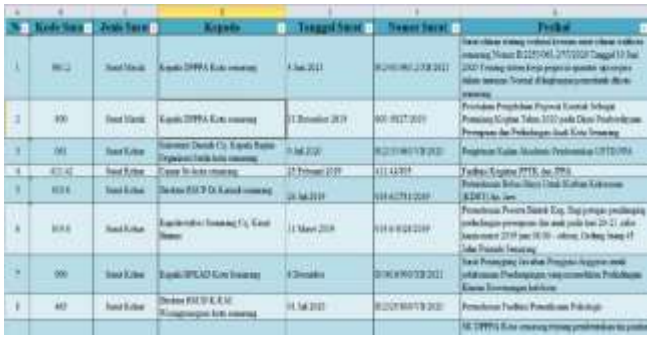
Dalam melakukan penelitian ini, untuk mempermudahnya maka dijabarkan langkah-langkah apa saja yang akan diambil dalam melakukan penelitian ini.



Gambar 2 Skema Alur Penelitian

Pengumpulan Data

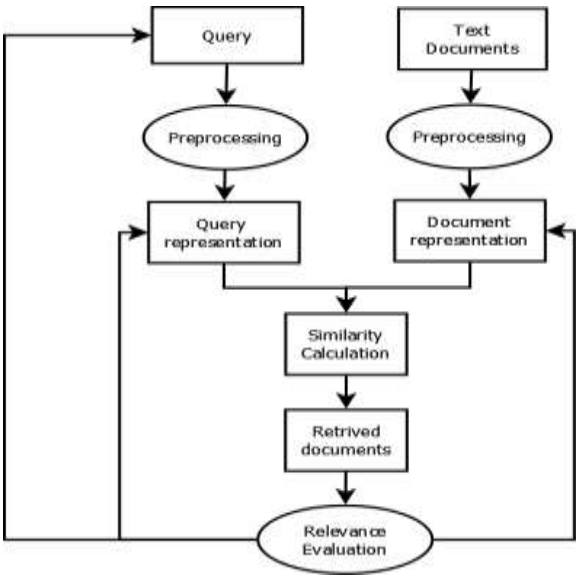
Objek penelitian dapat ditemukan dengan cara mengambil data sample untuk dijadikan data pengolahan pada sistem. Objek dalam penelitian ini yaitu Dokumen Dinas Pemberdayaan. Berikut ini tangkapan layar data yang diperoleh dan akan dijadikan sebagai sample dalam penelitian.



Gambar 3 Data Sample

Analisis Data

Penerapan metode pada sistem memiliki dua komponen yaitu, melakukan pre-processing terhadap database dan kemudian menerapkan metode tertentu untuk menghitung kedekatan relevansi atau similarity antara dokumen di dalam database yang telah dilakukan proses awal atau preprocessing dengan query pengguna. Query yang dimasukkan pengguna dikonversi sesuai aturan tertentu untuk membentuk term-term penting yang sejalan dengan term-term yang sebelumnya telah diekstrak dari dokumen dan menghitung relevansi antara query dan dokumen berdasarkan pada term-term tersebut. Sebagai hasilnya, sistem mengembalikan suatu daftar dokumen terurut sesuai nilai kemiripannya dengan query pengguna.



Gambar 4 Bagan ALur SIstem

Berdasarkan tahapan metode penelitian pada gambar 3.2 diatas, terdapat dua langkah utama yang dijelaskan sebagai berikut :

1) Preprocessing

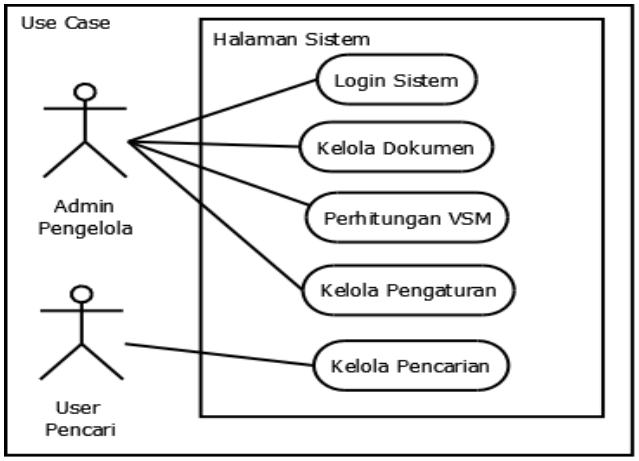
   Pada tahap ini terdiri dari beberapa tahapan yaitu, text mining, case folding, cleaning data, tokenizing, filtering, stemming dan indexing.

2) Similarity calculation

   Pada penelitian ini similarity akan menggunakan metode Vector Space Model (VSM). Relevansi sebuah dokumen ke sebuah query didasarkan pada similaritas diantara vektor dokumen dan vektor query.

Rancangan Model

Rancangan Model yang disampaikan dalam penelitian ini, yaitu use case diagram, diperlihatkan seperti gambar 5.



Gambar 5 Use Case Diagram

Rincian kebutuhan fungsional untuk mendukung sistem kearsipan pencarian dokumen menggunakan metode VSM yaitu :

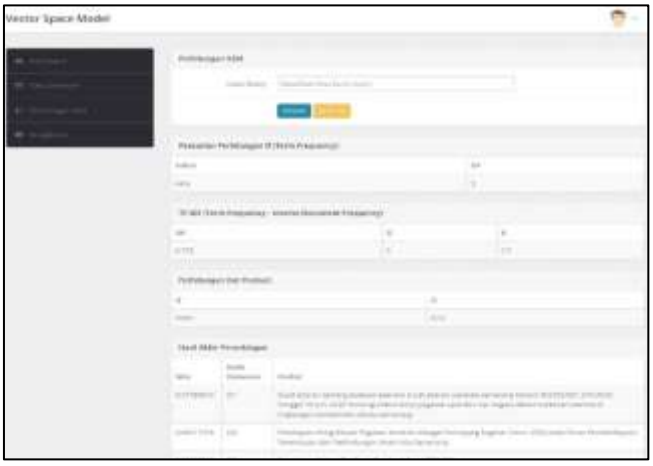
Kebutuhan Fungsional	Penjelasan
Kelola Dokumen	Admin dapat menambah dokumen baru, mengedit dokumen, dan menghapus dokumen
Perhitungan VSM	Admin dapat melihat alur perhitungan dari VSM
Pengaturan	Admin dapat mengelola data login (username dan password)
Halaman Pencarian	User dapat melakukan pencarian dokumen pada halaman ini, dengan hasil pencarian dari VSM.

HASIL DAN PEMBAHASAN

Hasil Implementasi Perancangan

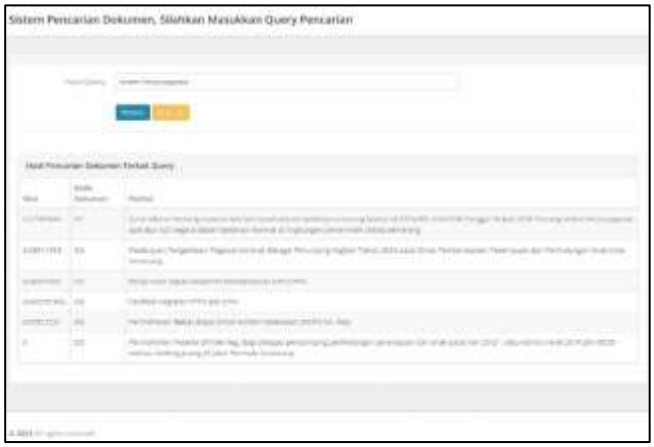
Hasil gambaran implementasi perancangan sistem yang dibuat disampaikan untuk mengetahui gambaran pada tampilan sistem. Terdapat 3 halaman utama yaitu halaman penambahan dokumen, halaman untuk melihat perhitungan vector space model dan halaman pencarian yang diasumsikan untuk user.

Gambar 6 Halaman kelola dokumen



Gambar 7 Halaman detail vector space model

Pada halaman detail vector space model ini, admin dapat mengetahui jalannya perhitungan dari metode vector space model, sampai hasil akhir.



Gambar 8 Halaman pencarian dokumen untuk user

Evaluasi Hasil

Proses *testing* menggunakan sampel data sebanyak 3 record, berikut ini diasumsikan seorang user memasukkan query "sistem kerja pegawai" pada halaman sistem pencarian dokumen seperti pada gambar 8.

Query (Q)	sistem kerja pegawai
D1	Surat edaran tentang evaluasi keenam surat edaran walikota semarang Nomor B/2253/061.2/VI/2020 Tanggal 10 Juni 2020 Tentang sistem kerja pegawai aparatur sipi negara dalam tantanan Normal di lingkungan pemerintah dikota semarang
D2	Pesetujuan Pengelolaan Pegawai Kontrak Sebagai Penunjang Kegitan Tahun 2020 pada Dinas Pemberdayaan Perempuan dan Perlindungan Anak Kota Semarang
D3	Pengiriman Kajian Akademis Pembentukan UPTD PPA

Pencarian Perhitungan *tf (Term Frequency)*:

Table 1 Hasil perhitungan tf (Term Frequency) (Fauzi, 2018)

Token	Q	Dokumen (tf)			DF
		1	2	3	
akademis	0	0	0	1	1
anak	0	0	1	0	1
aparatur	0	1	0	0	1
dinas	0	1	0	0	1
edaran	0	1	0	0	1
evaluasi	0	1	0	0	1
kajian	0	0	0	1	1
kegitan	0	0	1	0	1
kerja	1	1	0	0	2
kontrak	0	0	1	0	1
lingkungan	0	1	0	0	1
negara	0	1	0	0	1
Nomor	0	1	0	0	1
normal	0	1	0	0	1
pada	0	0	2	0	2
pegawai	1	1	1	0	3
pembentukan	0	0	0	1	1
pemberdayaan	0	0	1	0	1
pemerintah	0	1	0	0	1
pengelolaan	0	0	1	0	1
pengiriman	0	0	0	1	1
penunjang	0	0	1	0	1
perempuan	0	0	1	0	1
perlindungan	0	0	1	0	1
pesetujuan	0	0	1	0	1
sebagai	0	0	1	0	1
semarang	0	1	1	0	2
sipil	0	1	0	0	1
sistem	1	1	0	0	2
surat	0	1	0	0	1
tantanan	0	1	0	0	1
tentang	0	1	0	0	1
walikota	0	1	0	0	1

Keterangan :
D1, D2, D3, D4, D5, D6 = Dokumen
tf = banyak kata yang dicari pada sebuah dokumen.
D = total dokumen,
df = Banyak dokumen yang mengandung kata yang dicari.

Dari hasil Perhitungan tf , data sample dari jumlah dokumen yang ada dihasilkan 33 token dari 3 dokumen dan satu query, untuk mendapatkan jarak dokumen dan query, di perlukan perhitungan idf yang di hasilkan dari tokenasi hasil perhitungan berikut ini:

Table 2 Term Frequency - Inverse Document Frequency

Token	N/df	idf
akademis	3/1 = 3	Log(n/df) = 0,477121
anak	3	0,477121
aparatur	3	0,477121
dinas	3	0,477121
edaran	3	0,477121
evaluasi	3	0,477121
kajian	3	0,477121
kegitan	3	0,477121
kerja	1,5	0,176091
kontrak	3	0,477121
lingkungan	3	0,477121
negara	3	0,477121
Nomor	3	0,477121
normal	3	0,477121
pada	1,5	0,176091
pegawai	1	0
pembentukan	3	0,477121
pemberdayaan	3	0,477121
pemerintah	3	0,477121
pengelolaan	3	0,477121
pengiriman	3	0,477121
penunjang	3	0,477121
perempuan	3	0,477121
perlindungan	3	0,477121
pesetujuan	3	0,477121
sebagai	3	0,477121
semarang	1,5	0,176091
sipil	3	0,477121
sistem	1,5	0,176091
surat	3	0,477121
tantanan	3	0,477121
tentang	3	0,477121
walikota	3	0,477121

Table 3 Term Frequency - Inverse Document Frequency (lanjut)

	Tf * idf			
Token	**Q**	**D1**	**D2**	**D3**
akademis	0	0	0	0,477
anak	0	0	0,477	0
aparatur	0	0,477	0	0
dinas	0	0,477	0	0
edaran	0	0,477	0	0
evaluasi	0	0,477	0	0
kajian	0	0	0	0,477
kegitan	0	0	0,477	0
kerja	0,176	0,176	0	0
kontrak	0	0	0,477	0
lingkungan	0	0,477	0	0
negara	0	0,477	0	0
Nomor	0	0,477	0	0
normal	0	0,477	0	0
pada	0	0	0,352	0
pegawai	0	0	0	0
pembentukan	0	0	0	0,477
pemberdayaan	0	0	0,477	0
pemerintah	0	0,477	0	0
pengelolaan	0	0	0,477	0
pengiriman	0	0	0	0,477
penunjang	0	0	0,477	0
perempuan	0	0	0,477	0
perlindungan	0	0	0,477	0
pesetujuan	0	0	0,477	0
sebagai	0	0	0,477	0
semarang	0	0,176	0,176	0
sipil	0	0,477	0	0
sistem	0,176	0,176	0	0
surat	0	0,477	0	0
tantanan	0	0,477	0	0

tentang	0	0,477	0	0
walikota	0	0,477	0	0

TF.IDF (Term Frequency. Inverse Document Frequency) merupakan perhitungan statistik yang bertujuan untuk memberikan gambarkan seberapa penting sebuah kata terhadap sebuah koleksi dokumen yang tersedia. TF-IDF digunakan untuk pembobotan dalam Information Retrieval. Nilai TF.IDF akan meningkat sejalan dengan banyaknya jumlah kata yang sering muncul. Setelah itu kemudian menghitung perhitungan jarak dokumen sebagai berikut :

Table 4 Jarak Dokumen

Q	D1	D2	D3
0	0	0	0,477
0	0	0,477	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
0	0	0	0,477
0	0	0,477	0
0,176	0,176	0	0
0	0	0,477	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
0	0	0,352	0
0	0	0	0
0	0	0	0,477
0	0	0,477	0
0	0,477	0	0
0	0	0,477	0
0	0	0	0,477
0	0	0,477	0
0	0	0,477	0
0	0	0,477	0
0	0	0,477	0
0	0,176	0,176	0
0	0,477	0	0
0,176	0,176	0	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
0	0,477	0	0
Sqrt(Q)	**Sqrt(D)**		
0,839	1,811	1,559	0,954

Relevansi sebuah dokumen ke sebuah query didasarkan pada similaritas diantara vektor dokumen dan query, panjang dokumen cenderung memiliki frekuensi kemunculan kata yang besar. Setelah diketahui perhitungan jarak antara Q-D dengan menggunakan rumus Sqrt $D = sqrt\ (\sum_j^n = 1 \sum_j^2)$.

Tabel 3.1 Perhitungan Dot Product

Q	D1	D2	D3	Q * D
0	0	0	0,477	0
0	0	0,477	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0	0	0,477	0
0	0	0,477	0	0
0,176	0,176	0	0	0,062
0	0	0,477	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0	0,352	0	0
0	0	0	0	0
0	0	0	0,477	0

0	0	0,477	0	0
0	0,477	0	0	0
0	0	0,477	0	0
0	0	0	0,477	0
0	0	0,477	0	0
0	0	0,477	0	0
0	0	0,477	0	0
0	0	0,477	0	0
0	0	0,477	0	0
0	0,176	0,176	0	0
0	0,477	0	0	0
0,176	0,176	0	0	0,062
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0
0	0,477	0	0	0

Menghitung index terms dari dokumen dan query sebagai berikut :

Q * D1 = *0,176 * 0,176 + 0,176 * 0,176*
= 0,0620
Q * D2 = *0,1761 * 0 = 0*
Q * D3 = *0,1761 * 0 = 0*

Setelah mendapatkan nilai bobot TF.IDF dan jarak antara dokumen dengan query (Q-D), langkah selanjutnya menghitung similaritas dokumen atau menghitung nilai *Cosinus* sudut antara vector query dengan tiap dokumen dengan rumus :

$$\textbf{Cosine } \emptyset \textbf{ D}i = \frac{\boldsymbol{Q} * \boldsymbol{D}}{|\boldsymbol{Q}| * |\boldsymbol{D}|}$$

Diberikan contoh perhitungan untuk mencari nilai D1 sebagai berikut :

$$D1 = \frac{0{,}062}{0{,}839 * 1{,}811} = 0{,}0408048.$$

$$D2 = \frac{0}{0{,}839 * 1{,}559} = 0$$

$$D3 = \frac{0}{0{,}839 * 0{,}954} = 0$$

Sehingga dari analisa Vector Space Model diperoleh hasil untuk dokumen di atas adalah sebagai berikut :

Table 5 Hasil Akhir Perankingan

Dokumen	Nilai	Rangking
D1	0,0408048	1
D2	0	2
D3	0	3

Berdasarkan percobaan sistem dan analisa evaluasi yang dilakukan jika user memasukkan query "sistem kerja pegawai" maka pencarian vector space model akan menampilkan data dokumen ke 1 sebagai prioritas utamanya, dengan perihal "Surat edaran tentang evaluasi keenam surat edaran walikota semarang Nomor B/2253/061.2/VI/2020 Tanggal 10 Juni 2020 Tentang sistem kerja pegawai aparatur sipi negara dalam tantan an Normal di lingkungan pemerintah dikota semarang".

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan pengujian yang sudah dilakukan pada perolehan kembali informasi pada pengelolaan kearsipan di Kemen PPPA untuk sistem kearsipan pencarian dokumen di dinas pemberdayaan perempuan dan perlindungan anak kota semarang menggunakan metode vector space model diharapkan lebih dapat memberikan perbaikan secara sistematis dalam kearsipan terutama dalam pencarian data sehingga proses pencarian dokumen bisa dilakukan dengan cepat dan tepat.

Saran

Dalam pengembangan sistem information retrieval pencarian dokumen ini diperlukan perbaikan untuk mencapai hasil yang maksimal, perlu dilakukan pada pengembangan selanjutnya yaitu :

1. Menambah metode pencarian lain atau filter lainnya, sehingga mempercepat lagi waktu pencarian pada dokumen

2. Penelitian selanjutnya diharapkan dapat dikembangkan pencarian dokumen yang dapat menggunakan voice note atau query penangkapan suara user, sehingga user tidak perlu mengetik query.

DAFTAR PUSTAKA

[1] Alfa, F. A. P. J. (2018). *Pemeringkatan Hasil Pencarian Dokumen Teks Pada Mesin Pencari | Jurnal Dinamika Informatika*. https://unisbank.ac.id/ojs/index.php/fti2/article/view/8126

[2] Amin, E. R. P. N. F. (2021). Sistem Temu Kembali Cerita Rakyat Nusantara Menggunakan Metode Vector Space Model. *Jurnal Dinamika Informatika*. https://doi.org/10.35315/INFORMATIKA.V13I1.8554

[3] Asra, S. F. D. N. S. T. (2019). *View of Optimasi Algoritma Vector Space Model Dengan Algoritma K-Nearest Neighbour Pada Pencarian Judul Artikel Jurnal*. http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/27/22

[4] Dalimunthe, N. (2020). Focus Ilmu Administrasi Upmi Focus Ilmu Administrasi Upmi. *Jurnal Ilmu Administrasi UPMI*, *1*(2), 58–65.

[5] Fauzi, A. ; G. (2018). *View of Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model*. http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/61/55

[6] Hadiono, K. H. Y. D. A. D. (2018). Penggunaan Cosine Similarity Untuk Mencari Kesamaan Kandungan Obat. *Jurnal Dinamika Informatika*. https://doi.org/10.35315/INFORMATIKA.V10I1.8125

[7] Hasanah, N. (2017). *View Of Sistem Pencarian Skripsi Berbasis Information Retrieval Di Fastikom Unsiq*. https://ojs.unsiq.ac.id/index.php/ppkm/article/view/411/240

[8] Leman, D., & Andesa, K. (2017). Implementasi Vector Space Model Untuk Meningkatkan Kualitas Pada Sistem Pencarian Buku Perpustakaan. *Seminar Nasional Informatika (SNIf)*, *1*(1), 8–15. http://e-journal.potensi-utama.ac.id/ojs/index.php/SNIf/article/view/233/180

[9] Mas'udia, P. E., Atmadja, M. D., & Mustafa, L. D. (2017). Information Retrieval Tugas Akhir Dan Perhitungan Kemiripan Dokumen Mengacu Pada Abstrak Menggunakan Vector Space Model. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *8*(1), 355–362. https://jurnal.umk.ac.id/index.php/simet/article/view/1016

[10] PPPA, K. (2021). *Kementerian Pemberdayaan Perempuan Dan Perlindungan Anak*. https://www.kemenpppa.go.id/index.php/page/read/29/2509/kemen-pppa-tingkatkan-komitmen-dan-sinergitas-dalam-pengelolaan-arsip

[11] 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.

[12] Slamet, C., Atmadja, A. R., Maylawati, D. S., Lestari, R. S., Darmalaksana, W., & Ramdhani, M. A. (2018). *Recent citations Automated Text Summarization for Indonesian Article Using Vector Space Model*. https://doi.org/10.1088/1757-899X/288/1/012037

[13] Sri, Eniyati;Rina, Candra Noor Santi; Heribertus, Y. (2018). Penggunaan Sistem Temu Kembali Dalam Pencarian Kata Untuk Terjemahan Al Quran. *Proceeding SENDI_U*.

[14] Sugara, B. ; D. D. (2019). *View of Pemanfaatan Sistem Temu Kembali*

Informasi dalam Pencarian Dokumen Menggunakan Metode Vector Space Model. http://jurnal.stiki.ac.id/J-INTECH/article/view/189/162

[15] Susanto, Rani ; Andriana, A. D. (2016). *ResearchGate*. https://www.researchgate.net/publication/332968830_perbandingan_model_waterfall_dan_prototyping_untuk_pengembangan_sistem_informasi/link/5fb4bf8145851518fdb08eaf/download

[16] Yasin, A., & Rachman, M. A. (2019). *Information Retrieval System Evaluation: A Case Study At Badan Pengkajian Dan Penerapan Teknologi (BPPT) Library*. https://doi.org/10.24252/kah.v7i1a9

[17] Zeniarja, J., Salam, A., & Achsanu, I. (2020). Sistem Koreksi Jawaban Esai Otomatis (E-Valuation) dengan Vector Space Model pada Computer Based Test (CBT). *Seri Prosiding Seminar Nasional Dinamika Informatika*, *4*(1). http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/134