

1. Menjelaskan secara detail metode kueri pencarian dan similarity search pada proyek pengembangan STBI, yaitu

Metode kueri pencarian menggunakan *Information Retrieval* merupakan proses pencarian yang didasarkan pada teks lengkap atau indeks dengan berbasiskan konten lainnya.

Kemudian, VSM merupakan merupakan model aljabar untuk merepresentasikan objek apapun sebagai vektor untuk mengukur tingkat kemiripan. VSM juga mampu melakukan perankingan terhadap dokumen-dokumen yang relevan terhadap kata kunci yang dimasukkan. VSM juga digunakan untuk mengetahui tingkat kesamaan (similarity) term dengan cara pembobotan term. VSM juga merepresentasikan suatu dokumen dan *query* dalam bentuk vektor yang memiliki arah dan juga jarak.

Metode Similarity Search menggunakan *Cosine Similarity* merupakan metode yang digunakan untuk menganalisa kemiripan yang memiliki fungsi untuk mendapatkan kata atau term yang diinginkan berdasarkan *query*, serta jarak untuk mengurutkannya. Vektor pada dokumen merupakan vektor yang berisi frekuensi jumlah kemunculan kata pada suatu dokumen. Perhitungan untuk nilai *term* pada masing-masing dokumen adalah kemiripan pada dua buah vektor dalam ruang dimensi yang didapatkan dari *cosinus* pada sebuah sudut dari perkalian antara kedua buah vektor yang dibandingkan karena *cosinus* pada nilai 0 merupakan 1 dan kurang dari 1 untuk sudut yang lainnya.

Cosine score yang bernilai 0 mempunyai makna bahwa kedua buah vektor berada pada sumbu 90° secara ortogonal dan tidak memiliki kecocokan, sedangkan *cosine score* yang semakin dekat dengan nilai 1 maka semakin kecil sudut dan semakin besar kecocokan antara 2 vektor [20]. Dua buah vektor dikatakan mirip ketika nilai pada *cosine similarity*nya adalah 1.

2. Menjelaskan secara detail metode evaluasi yang digunakan untuk evaluasi prototype STBI, kemudian mendiskusikan hasil evaluasi yang diperoleh
Penjelasan secara detail mengenai metode evaluasi yang digunakan yaitu:

Metode evaluasi menggunakan *Precision* dan *Recall*.

Precision merupakan ukuran ketepatan dari jumlah dokumen yang dapat ditemukan dan dianggap relevan oleh proses pencarian untuk keperluan pencarian dokumen.

Recall merupakan rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen dalam koleksi yang dianggap relevan.

Rumus persamaan yang dipergunakan yaitu:

$$Precision = \frac{\text{Jumlah dokumen Relevan yang terambil}}{\text{Jumlah dokumen yang terambil}}$$

$$Recall = \frac{Jumlah\ dokumen\ Relevan\ yang\ terambil}{Jumlah\ dokumen\ Relevan}$$

Berikut salah satu contoh hasil *retrieval* untuk *query* “dermatitis” dapat dilihat pada Gambar 1. Hasil *retrieval* memberikan informasi kelima kolom dari data rekam medis SOAP.

[] Search query >> dermatitis

Rank	Scores	Document number	ANAMNESIA	TINDAKAN	TERAPI	CATATAN FISIK	DIAGNOSA
	Scores	Document Number	FS_ANAMNESIA	FS_TINDAKAN	FS_TERAPI	FS_CATATAN_FISIK	FS_DIAGNOSA
Ranking							
1	0.436	98	<p>GATAL DI PAHA makin banyak </p>	NaN	<p>R/ Hexilon 8 mg tab no X</p><p>S 1-0-0 </p>	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
2	0.436	97	<p>KONTROL sempat membaik </p>	NaN	<p>R/ Hexilon 8 mg tab no XV</p><p>S 1-0-0 </p>	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
3	0.436	806	<p>gatal di kaki</p>	NaN	<p>R/ </p>IKADERM CREAM gr 10 </p> BA...	<p>Suhu : - C,Nadi : - x/ menit,Respirasi : - x...	<p>Dermatitis Numular </p>
4	0.436	805	<p>gatal di kaki kambuh </p>	NaN	<p>R/ </p>IKADERM CREAM gr 10 </p> BA...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
5	0.436	804	<p>gatal2 dikaki sdh 1 bln</p>	NaN	<p>R/ </p>IKADERM CREAM gr 10 </p> BA...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
6	0.436	143	<p>gatal di tangan dan kaki kambuh </p>	NaN	<p>R/ HEXILON 8MG </p> tab no X</p><p>...</p>	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
7	0.436	301	<p>gatal2 di kaki </p>	NaN	<p>R/ Desoxymetason gr 15 </p> <p>Fuladi...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
8	0.410	207	<p>GATAL DI KAKI</p>	NaN	<p>R/ </p>TERMISIL CR </p> gr 10 </p> ...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik </p><p> </p>Derma...
9	0.399	246	<p>KONTROL, membaik</p>	NaN	<p>R/ Metylprednisolon 8 mg tab no X </p> ...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik </p>
10	0.399	247	<p>gatal di </p>badan 2 th terakhir, kambuh2a...	NaN	<p>R/ Metylprednisolon 8 mg tab no X </p> ...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik </p>

Gambar 1. Contoh Hasil *Retrieval query* “dermatitis”

Hasil *screenshot coding* yang berkaitan dengan *Precision* dapat dilihat pada Gambar 2 dan 3, kemudian jika *Recall* dapat dilihat pada Gambar 4 dan 5.

Precision

```
[ ] 1 def prec(ret):
2     rel_items_retrieved = 0
3     ret_items = 0
4
5     for a in range(1, len(ret)+1):
6         ret_items += 1
7         if(ret['Scores'].loc[ret.index == a] > 0).all() :
8             rel_items_retrieved += 1
9         elif(ret['Scores'].loc[ret.index == a] == 0).all() :
10            continue
11
12    precision = rel_items_retrieved / ret_items
13    return precision
```

```
[ ] 1 con_precision = []
2
3 def precision(ret):
4     data = prec(ret)
5     con_precision.append(data)
```

Gambar 2. Coding *Precision*

```
[ ] 1 def avg_precision(data):
    2     jumlah = sum(data)
    3     panjang = len(data)
    4
    5     hasil = jumlah / panjang
    6     return hasil
```

```
[ ] 1 avg_precision(con_precision)
```

0.5739130434782608

Gambar 3. Coding Average Precision

Recall

```
▶ 1 def rec(ret, document):
    2     rel_items_retrieved = 0
    3     rel_items = 0
    4
    5     for a in range(1, len(ret)+1):
    6         if(ret['Scores'].loc[ret.index == a] > 0).all() :
    7             rel_items_retrieved += 1
    8         elif(ret['Scores'].loc[ret.index == a] == 0).all() :
    9             continue
    10
    11     for b in range(len(document)):
    12         if(document[b] != 0):
    13             rel_items += 1
    14         elif(document[b] == 0):
    15             break
    16
    17     recall = rel_items_retrieved / rel_items
    18     return recall
```

```
▶ 1 con_recall = []
    2
    3 def recall(ret, document):
    4     data = rec(ret, document)
    5     con_recall.append(data)
```

Gambar 4. Coding Recall

```
[ ] 1 def avg_recall(data):
    2     jumlah = sum(data)
    3     panjang = len(data)
    4
    5     hasil = jumlah / panjang
    6     return hasil
```

```
[ ] 1 avg_recall(con_recall)
```

0.7809222305533547

Gambar 5. Coding Average Recall

mengetahui kinerja dari *information retrieval* yang dikembangkan, dilakukan *retrieve data* untuk kedua kolom dengan menginputkan masing-masing 10 *query*. Berikut *list query* yang digunakan untuk menguji model dapat dilihat pada Tabel 1.

Tabel 1. List query pada Pengujian Model

No	FS_ANAMNESA	Jumlah Dokumen	FS_DIAGNOSA	Jumlah Dokumen
1.	kontrol	10	hamil	10
2.	gigi sakit	10	bronkitis akut	10
3.	ca paru	10	neurosa	10
4.	nyeri bahu kanan	6	dermatitis	10
5.	nyeri	4	et	10
6.	telinga nyeri	4	bph	10
7.	lutut kanan sakit	4	abses gigi	10
8.	lutut kiri sakit	4	meningioma	10
9.	batuk pilek	2	Kb	4
10.	batuk	2	Rhinitis	4

Dari hasil pengujian yang dilakukan, didapatkan hasil seperti yang terdapat pada Tabel 2 di bawah ini.

Tabel 2. Nilai Hasil Pengujian dengan *Precision* dan *Recall*

	<i>Average Precision</i>	<i>Average Recall</i>
FS_ANAMNESA	0.522	0.816
FS_DIAGNOSA	0.574	0.781
<i>Average</i>	0.548	0.796

Dari proses *retrieve* yang dilakukan didapatkan hasil yang cukup baik, yakni dengan *precision* sebesar 54,8 % dan *recall* sebesar 79,6 %. Jadi didapatkan bahwa *low precision* dan *high recall* menunjukkan bahwa model dapat memprediksi lebih banyak *true positive*.

Hasil *retrieval* dan nilai evaluasi sangat **bergantung pada banyak dokumen** di dalam korpus dan *query* yang dimasukkan, sehingga sangat mungkin banyak dokumen yang nantinya di-*retrieve* dan nilai *precision* dan *recall* bisa saja berubah.

3. Perbedaan antara sistem rekomendasi dan sistem tanya jawab

Menurut saya proyek STBI saya termasuk pada kategori mana?

Sistem Rekomendasi yaitu sistem yang memiliki tujuan untuk memperkirakan informasi yang dapat menarik pengguna dan juga membantu pengguna (*user*) dalam menentukan pilihannya.

Sedangkan, Sistem tanya jawab merupakan sistem yang memiliki tujuan untuk mencari jawaban yang paling mirip diantara beberapa jawaban dari pertanyaan yang terpilih.

Menurut saya, proyek STBI yang dikerjakan termasuk dalam kategori Sistem tanya jawab. Dikarenakan sistem kami melakukan *retrieve* untuk semua kolom dan menampilkan 10 rangking yang secara signifikan memunculkan tergantung pada kata yang *user* atau pengguna masukkan pada sistem.

4. a) Membuat rangkuman lengkap proyek STBI

Rangkuman terdiri dari Introduction, Methods, Results, Discussion dan Referencess.

Introduction

Indikator pelayanan fasilitas kesehatan adalah kecepatan dan ketepatan dalam menyediakan medical record. Menurut Peraturan Kementerian Kesehatan Indonesia nomor 129/Menkes/SK/II/2008 menjelaskan bahwa salah satu standar minimum dalam sebuah pelayanan di fasilitas kesehatan adalah dapat menyediakan rekam medis dalam waktu 10 menit dan mempunyai kelengkapan informasi sebesar 100%.

Menurut UU Praktik Kedokteran dalam penjelasan pasal 46 ayat (1) yang dimaksud dengan Rekam Medis adalah berkas yang berisi catatan dan dokumen tentang identitas pasien, pemeriksaan, pengobatan, tindakan dan pelayanan lain yang telah diberikan kepada pasien.

Information Retrieval (IR) atau Temu Balik Informasi dalam ilmu komputer adalah proses mendapatkan informasi yang relevan, dalam hal ini berkaitan dengan merepresentasikan, mencari, dan memanipulasi data besar kumpulan teks elektronik dan data bahasa manusia lainnya.

Vector Space Model (VSM) adalah model aljabar untuk merepresentasikan objek apa pun sebagai vektor untuk mengukur tingkat kemiripan. Baik dikatakan mirip secara ukuran, maupun nilainya.

Perangkingan termasuk tahapan dengan menggunakan bantuan *Cosine Similarity*. *Cosine Similarity* merupakan metode yang menganalisa kemiripan yang memiliki fungsi untuk mendapatkan kata atau *term* yang diinginkan berdasarkan *query* serta jarak untuk mengurutkannya.

Evaluasi yang digunakan adalah *Precision* dan *Recall*. *Precision* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan. Sedangkan, *Recall* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen dalam koleksi yang dianggap relevan.

Penelitian ini melakukan proses *retrieve* kumpulan dokumen rekam medis dari *dataset* yang berisi 1000 baris data. *Dataset* tersebut merupakan data Rekam Medis SOAP yang berasal dari REKMED, aplikasi pencatatan rekam medis pasien berbasis *cloud*. *Dataset* ini mempunyai 5 jumlah kolom, yang terdiri dari FS_ANAMNESA, FS_TINDAKAN, FS_TERAPI, FS_CATATAN_FISIK, dan FS_DIAGNOSA. Metode

yang digunakan adalah melakukan *data preprocessing*, penghitungan kemiripan dengan *cosine similarity*, perangkingan berdasarkan *cosine similarity*, dan menggunakan beberapa *evaluation matrix*. Diharapkan hasil dari penelitian ini mampu menawarkan model pencarian dokumen rekam medis yang sesuai dengan *query* yang dimasukkan oleh pengguna.

Methods

A. Preprocessing

Preprocessing adalah tahap pertama yang dilakukan untuk membersihkan dan melakukan normalisasi data rekam medis yang sudah ada. Berikut adalah rincian dari tiap tahap *preprocessing* yang dilakukan dalam penelitian ini:

1. Menghapus baris kosong, di mana baris yang tidak memiliki data baik di kolom FS_ANAMNESA atau FS_DIAGNOSA dihapus untuk tahap berikutnya.
2. Lowercase, dimana setiap huruf diubah menjadi huruf kecil.
3. Menghapus tanda baca dan tag HTML. Data rekam medis memiliki banyak tanda baca dan tag HTML seperti <p>,
, dan lain sebagainya sehingga perlu dihapus agar proses selanjutnya menjadi lebih efisien.
4. Menghapus stopwords. Kata-kata yang tidak memiliki makna signifikan seperti kata sambung, kata depan, atau kata-kata lainnya dihapus agar tidak terlalu banyak noise.
5. Menghapus spasi berlebih antar kata atau karakter.
6. Normalisasi dengan mengganti kata-kata typo, seperti 'peerut' menjadi 'perut', 'jln' menjadi 'jalan', 'riwy' menjadi 'riwayat', dan lain sebagainya. Di dalam penelitian ini dihimpun sebanyak 1290 pasang kata yang dijadikan acuan untuk mengganti kata dalam teks rekam medis.

B. Pembentukan Matriks TF-IDF

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan salah satu cara untuk menghitung bobot kata berdasarkan frekuensi kemunculannya di dalam sebuah korpus. TF mewakili frekuensi kata, menunjukkan berapa kali mereka muncul dalam korpus persamaan (1).

$$tf(t, d) = \frac{f_d(t)}{\max f_d(w)} \quad (1)$$

$$idf(t, D) = \ln\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right) \quad (2)$$

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (3)$$

Di mana: $f_d(t) :=$ frekuensi kemunculan kata t dalam dokumen D

D : = korpus dokumen

IDF menunjukkan tingkat kepentingan sebuah kata dalam korpus secara keseluruhan. Nilai IDF didapatkan dengan menghitung logaritma dari nilai *inverse* proporsi dokumen pada persamaan (2). Bobot TF-IDF dihitung dengan mengalikan dua nilai tersebut sebagaimana terlihat pada persamaan (3). Semakin besar nilai TF-IDF, menunjukkan semakin signifikan kata yang bersangkutan di dalam korpus.

C. Penghitungan Kemiripan

Cosine Similarity merupakan sebuah ukuran untuk menghitung jarak yang digunakan pada data berupa vektor dari dokumen. Dokumen berisikan sebuah data yang dapat terdiri dari ratusan bahkan ribuan atribut. Setiap atribut mewakili *term* atau kata yang berisi nilai berarti frekuensi kemunculan dalam dokumen tertentu. Vektor pada dokumen merupakan vektor yang berisi frekuensi jumlah kemunculan kata pada suatu dokumen. Perhitungan untuk nilai *term* pada masing-masing dokumen adalah kemiripan pada dua buah vektor dalam ruang dimensi yang didapatkan dari *cosinus* pada sebuah sudut dari perkalian antara kedua buah vektor yang dibandingkan karena *cosinus* pada nilai 0 merupakan 1 dan kurang dari 1 untuk sudut yang lainnya. *Cosine score* yang bernilai 0 mempunyai makna bahwa kedua buah vektor berada pada sumbu 90° secara ortogonal dan tidak memiliki kecocokan, sedangkan *cosine score* yang semakin dekat dengan nilai 1 maka semakin kecil sudut dan semakin besar kecocokan antara 2 vektor. Penghitungan *cosine similarity* dijelaskan pada persamaan (4), yaitu:

$$Sim(\alpha) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \cdot B_i}{\sqrt{\sum_{i=1}^n (A_i)^2 \cdot \sum_{i=1}^n (B_i)^2}} \quad (4)$$

Dimana:

A = Vektor Dokumen

B = Vektor *Query*

$A \cdot B$ = Perkalian antara vektor A dan vektor B

$|A|$ = Panjang vektor A

$|B|$ = Panjang vektor B

$|A||B|$ = *Cross Product* antara $|A|$ dan $|B|$

α = Sudut yang terbentuk antara vektor A dan B

D. Evaluasi

Precision merupakan ukuran ketepatan dari jumlah dokumen yang dapat ditemukan dan dianggap relevan oleh proses pencarian untuk keperluan pencarian dokumen. Sedangkan, *recall* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen dalam koleksi yang dianggap relevan. Rumus Persamaan (5) menjelaskan mengenai *Precision* dan Persamaan (6) menjelaskan mengenai *Recall*.

$$Precision = \frac{\text{Jumlah dokumen Relevan yang terambil}}{\text{Jumlah dokumen yang terambil}} \quad (5)$$

$$Recall = \frac{Jumlah\ dokumen\ Relevan\ yang\ terambil}{Jumlah\ dokumen\ Relevan} \quad (6)$$

Results

Berikut adalah *list 10 query* yang digunakan untuk menguji model dapat dilihat pada Tabel I dan Tabel II.

TABEL I

LIST QUERY KOLOM FS_ANAMNESA PADA PENGUJIAN MODEL

No	FS_ANAMNESA	Jumlah Dokumen
1.	kontrol	10
2	gigi sakit	10
3.	ca paru	10
4.	nyeri bahu kanan	6
5.	nyeri	4
6.	telinga nyeri	4
7.	lutut kanan sakit	4
8.	lutut kiri sakit	4
9.	batuk pilek	2
10.	batuk	2

TABEL II

LIST QUERY UNTUK KOLOM FS_DIAGNOSA PADA PENGUJIAN MODEL

No	FS_DIAGNOSA	Jumlah Dokumen
1.	hamil	10
2.	bronkitis akut	10
3.	neurosa	10
4.	dermatitis	10
5.	et	10
6.	bph	10
7.	abses gigi	10
8.	meningioma	10
9.	Kb	4
10.	Rhinitis	4

Gambar 3 menunjukkan semua kolom yang sudah berhasil di-*retrieve*. Sedangkan gambar 4 mejelaska mengenai kolom *ranking*, *scores*, *document number* dan FS_ ANAMNESA.

[] Search query >> dermatitis

Rank	Scores	Document number	ANAMNESA	TINDAKAN	TERAPI	CATATAN FISIK	DIAGNOSA
	Scores	Document Number	FS_ANAMNESA	FS_TINDAKAN	FS_TERAPI	FS_CATATAN_FISIK	FS_DIAGNOSA
Ranking							
1	0.436	98	<p>GATAL DI PAHA makin banyak 1th <...</p>	NaN	<p>R/ Hexilon 8 mg tab no X</p><p>S 1-0-0&nb...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
2	0.436	97	<p>KONTROL sempat membaik </p>	NaN	<p>R/ Hexilon 8 mg tab no XV</p><p>S 1-0-0&n...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
3	0.436	806	<p>-gatal di kaki</p>	NaN	<p>R/ IKADERM CREAM gr 10 BA...	<p>Suhu : - C,Nadi : - x/ menit,Respirasi : - X...	<p>Dermatitis Numular </p>
4	0.436	805	<p>gatal di kaki kambuh </p>	NaN	<p>R/ IKADERM CREAM gr 10 BA...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
5	0.436	804	<p>gatal2 dikaki sdh 1 bln</p>	NaN	<p>R/ IKADERM CREAM gr 10 BA...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
6	0.436	143	<p>gatal di tangan dan kaki kambuh </p>	NaN	<p>R/ HEXILON 8MG tab no X</p> <p>...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
7	0.436	301	<p>gatal2 di kaki </p>	NaN	<p>R/ Desoxymetason gr 15 <p>Fuladi...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Numular </p>
8	0.410	207	<p>GATAL DI KAKI</p>	NaN	<p>R/ TERMISIL CR gr 10 <...</p>	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik <p> Derma...
9	0.399	246	<p>KONTROL, membaik</p>	NaN	<p>R/ Metylprednisolon 8 mg tab no X <p>...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik </p>
10	0.399	247	<p>gatal di badan 2 th terakhir, kambuh2a...	NaN	<p>R/ Metylprednisolon 8 mg tab no X <p>...	<p>Suhu : C,Nadi : x/ menit,Respirasi : x/ menit...	<p>Dermatitis Atopik </p>

Gambar 3. Semua kolom yang berhasil di *retrieve*

Search query >> dermatitis

Rank	Scores	Document number	ANAMNESA	TINDAKAN	TERAPI	CATATAI
	Scores	Document Number	FS_ANAMNESA			
Ranking						
1	0.436	98	<p>GATAL DI PAHA makin banyak 1th <...</p>			
2	0.436	97	<p>KONTROL sempat membaik </p>			
3	0.436	806	<p>-gatal di kaki</p>			
4	0.436	805	<p>gatal di kaki kambuh </p>			
5	0.436	804	<p>gatal2 dikaki sdh 1 bln</p>			
6	0.436	143	<p>gatal di tangan dan kaki kambuh </p>			
7	0.436	301	<p>gatal2 di kaki </p>			
8	0.410	207	<p>GATAL DI KAKI</p>			
9	0.399	246	<p>KONTROL, membaik</p>			
10	0.399	247	<p>gatal di badan 2 th terakhir, kambuh2a...			

Gambar 4. Kolom *ranking*, *scores*, *document number*, dan FS_ ANAMNESA yang berhasil di *retrieve*.

Gambar 4 terlihat bahwa setiap dokumen yang berhasil di *retrieve* mempunyai bobot yang didapatkan dari penghitungan *cosine similarity*.

Discussion

Bobot tersebut digunakan untuk menentukan ranking dari dokumen yang berhasil di *retrieve*. Semakin tinggi nilai bobot pada suatu dokumen maka semakin tinggi ranking dokumen tersebut. Semakin tinggi ranking suatu dokumen menunjukkan bahwa dokumen tersebut semakin relevan terhadap *query* yang dimasukkan.

Average precision dan *average recall* merupakan *evaluation matrix* yang digunakan pada penelitian ini. *Average precision* digunakan untuk mengetahui kemampuan model dalam mendapatkan hasil yang sesuai dengan *query* yang dimasukkan, sedangkan *average recall* untuk mendapatkan apa saja dokumen yang mengandung *query* yang dimasukkan. Dari *query* yang telah dimasukkan, didapatkan *average precision* dan *average recall* seperti yang terdapat pada Tabel III.

TABEL III
NILAI HASIL PENGUJIAN DENGAN PRECISION DAN RECALL

	<i>Average Precision</i>	<i>Average Recall</i>
FS_ANAMNESA	0.522	0.816
FS_DIAGNOSA	0.574	0.781
<i>Average</i>	0.548	0.796

Dari proses *retrieve* yang dilakukan didapatkan hasil yang cukup baik, yakni dengan *average precision* sebesar 54.8 % dan *average recall* sebesar 79.6 %. Jadi didapatkan *low precision* dan *high recall* menunjukkan bahwa model dapat memprediksi lebih banyak *true positive*.

Hasil evaluasi model berdasarkan 1000 dokumen rekam medis dan pengujian dengan 20 query pencarian dalam penelitian ini memberikan nilai rata-rata *precision* sebesar 0.548 dan nilai rata-rata *recall* sebesar 0.796. Nilai *low precision* dan *high recall* menunjukkan bahwa model dapat memprediksi lebih banyak *true positive*. Hasil *retrieval* dan nilai evaluasi bergantung pada banyak dokumen dalam korpus dan *query* yang dimasukkan, sehingga sangat mungkin banyak dokumen yang nantinya di-*retrieve* dan nilai *precision* dan *recall* bisa saja berubah jika ada penyesuaian jumlah dokumen dan penggunaan *query* pencarian lainnya.

References

- [1] I. Rosadi and M. I. Purnama, "Analysis Of Time Analysis Of Outstanding Medical Records To Improve The Quality Of Services At Dustira Hospital, Cimahi Imam," *Int. J. Heal. Sci.*, vol. 1, no. 1, pp. 1–5, 2021.
- [2] Menteri Kesehatan Republik Indonesia, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 24 Tahun 2022*, 2022.
- [3] Menteri Kesehatan Republik Indonesia, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 129/Menkes/SK/II/2008*, 2008.
- [4] E. W. R. P, R. Indrawati, and L. Widjaja, "A Service Quality Review of Medical Record Department In Private Hospital, South Jakarta," *J. Multidiscip. Acad.*, vol. 05, no. 02, pp. 101–105, 2021, [Online]. Available: <http://www.kemalapublisher.com/index.php/JoMA/article/view/554>.
- [5] Indonesia. *Undang-Undang Nomor 29 Tahun 2004 tentang Praktik Kedokteran*. Lembaran Negara RI Tahun 2004 Nomor 116, Tambahan Lembaran RI Nomor 4431. Jakarta: Sekretariat Negara, 2004.

- [6] Menteri Kesehatan Republik Indonesia, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 269 Tahun 2008*, 2008.
- [7] S. Buttcher, C. L. A. Clarke, and G. V. Cormack, *Information Retrieval Implementing and Evaluating Search Engines*. Cambridge, Massachusetts: The MIT Press, 2010.
- [8] N. Hasanah, *Sistem Pencarian Skripsi Berbasis Information Retrieval di FASTIKOM UNSIQ*. Jurnal PPKM, 1(1), 105-113, 2017.
- [9] F. Faridah, K. Munadi, and F. Arnia, "Aplikasi Histogram Discrete Cosine Transform (DCT) untuk Sistem Temu Kembali Citra Termal Berbasis Konten," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 2, no. 1, 2019.
- [10] Y. Kang, J. Li, J. Yang, Q. Wang, and Z. Sun, "Semantic analysis for enhanced medical retrieval," *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017.
- [11] R. Shaptala, A. Kyselova, and G. Kyselov, "Exploring the vector space model for online courses," *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2017.
- [12] E. Wahyudi, S. Sfenrianto, M. J. Hakim, R. Subandi, O. R. Sulaeman, and R. Setiyawan, "Information retrieval system for searching JSON files with vector space model method," *2019 International Conference of Artificial Intelligence and Information Technology (ICAIT)*, 2019.
- [13] P. E. Mas'udia, M. D. Atmadja, and L. D. Mustafa, "Information Retrieval Tugas Akhir dan Perhitungan Kemiripan Dokumen Mengacu pada Abstrak Menggunakan Vector Space Model," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 8, no. 1, pp. 355–362, 2017.
- [14] A. T. Adiyanto and D. H. UN, "Information Retrieval Sistem Kearsipan Pencarian Dokumen Di Dinas Pemberdayaan Perempuan Dan Perlindungan Anak Kota Semarang Menggunakan Metode Vector Space Model." *JURNAL MAHAJANA INFORMASI*, 7(1), pp.9-23, <https://doi.org/10.51544/jurnalmi.v7i1.2538>, 2022.
- [15] R. C. N. Santi, S. Eniyati, R. Retnowati, and H. Yulianton, "PENGUNAAN SISTEM TEMU KEMBALI DALAM PENCARIAN KATA UNTUK TERJEMAHAN AL QURAN", 2019.
- [16] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online Dengan Menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 1, p. 306-312, ISSN 2548-964X, Agu. 2017.
- [17] M. Widiyasri, E. Tjandra and L. M. Chandra, "Peningkatan Kinerja Pencarian Dokumen Tugas Akhir menggunakan Porter Stemmer Bahasa Indonesia dan Fungsi Peringkat Okapi BM25," *TEKNIKA*, vol. 6, no. 1, pp. 54-60, 2017.
- [18] E. S. Munif, and Y. Kristian, "PENCARIAN TEMA SEJENIS SINOPSIS NOVEL BAHASA INDONESIA DENGAN MENGGUNAKAN GVSM." *Joutica: Journal of Informatic Unisla*, 6(2), 492-501, 2021.
- [19] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, "Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms," *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning*, 2021.
- [20] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining Concepts and Techniques*, 3rd edition. 225 Wyman Street, Waltham, MA 02451, USA: Morgan Kaufmann, 2012.

b) Proyek STBI Anda merupakan *Multimedia Information Retrieval Systems* (MIRS), Argumentasi anda bagaimana?

Penjelasan mengenai *Multimedia Information Retrieval Systems* merupakan Sistem Informasi dengan konsep masukan dan juga keluaran dalam bentuk data multimedia. Adanya perubahan pada blok masukan, basis data dan blok keluaran yang menjadi ciri perlunya modifikasi yang perlu ada pada setiap *Multimedia Information Retrieval Systems*. Umumnya data masukan yang kerap digunakan adalah teks, audio, image dan video (Setyanto, 2005).

Jadi menurut argumentasi saya, proyek STBI kelompok saya tidak termasuk dalam MIRS, dikarenakan tidak memerlukan perubahan yang signifikan pada saat input dan juga menghasilkan data keluarannya. Karena data inputannya berupa kata dan juga outputnya berupa nilai hasil evaluasi dengan menggunakan metode Cosine Similarity dengan Vector Similarity Model (VSM).