

Perangkingan Fatwa Islam dengan Metode Latent Semantic Indexing Menggunakan Teknik Singular Value Decomposition (SVD)

Fajar Arifin¹, Eko Darwiyanto², Moch Arif Bijaksana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹fajarifin@students.telkomuniversity.ac.id, ²ekodarwiyanto@telkomuniversity.ac.id,

³arifbijaksana@telkomuniversity.ac.id

Abstrak

Fatwa merupakan bagian penting yang ada di dalam agama Islam dimana tujuan dari fatwa yaitu untuk menjawab permasalahan yang ada di kehidupan sehari-hari umat Islam. Fatwa dapat di artikan sebagai hasil dari perundingan antara kelompok atau organisasi yang khusus mendalami permasalahan agama Islam, dimana fatwa merujuk kepada Al-Qur'an maupun hadits. Permasalahan yang ingin di selesaikan dalam tugas akhir ini yaitu ingin mengetahui kemiripan yang ada di dalam fatwa, dimana fatwa yang diambil berasal dari situs resmi MUI, NU, MUHAMMADIYAH dan PERSIS. Dataset yang diambil dari situs tersebut lalu dibandingkan dengan dataset yang lainnya namun dengan permasalahan/topik yang sama. Dengan menggunakan metode Latent Semantic Indexing serta menggunakan Teknik singular value decomposition diharapkan bisa mendapatkan hasil kemiripan antara fatwa-fatwa yang dibandingkan. Dataset yang digunakan merupakan data dengan Bahasa Indonesia. Dalam pencarian kemiripan fatwa penelitian ini menggunakan perhitungan cosine similarity. Untuk menganalisis hasil yang di dapatkan dari cosine similarity akan dibandingkan dengan jaccard similarity.

Kata kunci : Fatwa, Latent Semantic Indexing, Cosine Similarity, Jaccard Similarity

Abstract

Fatwa is an important part of the Islamic religion where the purpose of the fatwa is to answer the problems that exist in the daily lives of Muslims. Fatwa can be interpreted as the result of negotiations between groups or organizations that specifically explore the problem of Islam, where the fatwa refers to the Qur'an and the hadith. The problem to be resolved in this thesis is to find out the similarities in the fatwa, where the fatwa taken comes from the official website of MUI, NU, MUHAMMADIYAH and PERSIS. The dataset taken from the site is then compared with other datasets but with the same problem / topic. By using the Latent Semantic Indexing method and using the singular value decomposition technique, it is hoped that the similarity between the fatwas can be compared. The dataset used is Indonesian data. In searching for the similarity of fatwa this study uses cosine similarity calculations. To analyze the results obtained from cosine similarity will be compared with jaccard similarity.

Keywords: Fatwa, Latent Semantic Indexing, Cosine Similarity, Jaccard Similarity

1. Pendahuluan

1.1 Latar Belakang

Fatwa adalah suatu hasil yang diambil dari Al-Qur'an ataupun hadits mengenai masalah yang dihadapi oleh umat islam dimana hasil tersebut di rundingkan secara berkelompok atau bersama-sama (dalam organisasi/lembaga) dan menghasilkan kesepakatan bersama. Penggunaannya dalam kehidupan beragama di Indonesia, fatwa dikeluarkan oleh berbagai lembaga yang memiliki kemampuan khusus atau ahli di bidang agama seperti Majelis Ulama Indonesia, Muhammadiyah, Nahdlatul 'Ulama dan Persatuan Islam.

Ketika muncul sebuah masalah keagamaan, sering muncul fatwa yang berbeda dari para ulama dari berbagai organisasi. Bagi orang awam hal tersebut membuat bingung mereka, fatwa mana yang akan di ikuti . Dalam rangka untuk mempermudah masyarakat untuk memilih fatwa mana yang ingin di ikuti maka di perlukan aplikasi yang mengumpulkan dokumen-dokumen fatwa dan membandingkannya, penelitian ini dilakukan untuk maksud tersebut.

Latent Semantic Indexing (LSI) adalah suatu metode yang sudah ada dan sudah digunakan dalam teknologi guna mempermudah manusia untuk mencari data atau dokumen yang memiliki kemiripan satu sama lainnya. Pada penelitian ini digunakan metode perhitungan Singular Value Decomposition (SVD) untuk mendapatkan matrik dekomposisi.

Pencarian adalah proses melacak nilai pada term yang ada didalam dokumen dan dalam kasus ini akan digunakan metode LSI. Dalam metode LSI terdapat penggunaan TF.IDF yang digunakan untuk menghitung bobot/nilai pada kata di dalam dokumen. Untuk meningkatkan hasil perbandingan kemiripan antar fatwa diharapkan penggunaan TF.IDF menghasilkan nilai similaritas yang relevan.

1.2 Topik dan Batasannya

Sistem membuka semua teks di dalam folder yang tersedia dalam bentuk/format .txt. di dalam fatwa tersebut terdapat kata/kalimat yang dicurigai mengandung kemiripan semantik yang akan dibandingkan dengan fatwa lainnya. Adapun *output* dari sistem adalah nilai kemiripan semantik dari masing-masing fatwa serta presentase peringkat yang mengandung kemiripan semantik yang tinggi. Keluaran dari sistem tersebut adalah nilai kemiripan semantic dari tertinggi sampai terendah.

Dalam menyelesaikan masalah pada tugas akhir ini, terdapat beberapa batasan masalah diantaranya:

1. *Dataset* yang digunakan adalah fatwa-fatwa berbahasa Indonesia yang didapatkan dalam situs resmi MUI, MUHAMMADIYAH, NU dan PERSIS. Karakteristik kemiripan semantik dari bagian fatwa yang di bandingkan dengan fatwa lainnya.
2. Fatwa yang akan dibandingkan telah ditentukan sebelumnya, adapun minimal jumlah fatwa yang dibandingkan adalah dua, dimana ada pasangan fatwa yang memiliki kemiripan semantik dan ada pasangan fatwa yang tidak memiliki kemiripan semantik..
3. Kemiripan semantik ditandai dengan *rangking*, dimana pasangan yang memiliki kemiripan semantik yang tinggi menunjukkan nilai *similarity* mendekati angka satu, sedangkan pasangan yang tidak memiliki kemiripan semantik menunjukkan nilai *similarity* mendekati angka nol.
4. Pada penelitian ini baru mengumpulkan fatwa tentang masalah hukum mengenai bunga bank.

1.3 Tujuan

Adapun tujuan dari tugas akhir ini adalah sebagai berikut:

1. Mengetahui nilai kemiripan semantik dengan menggunakan metode *latent semantic Indexing* pada Fatwa yang di dapatkan dari situs MUI, NU, MUHAMMADIYAH DAN PERSIS.
2. Mengetahui pengaruh jumlah dimensi pada pencarian nilai kemiripan.
3. Mengetahui efektivitas *latent semantic Indexing* dalam mendeteksi kemiripan semantik pada bagian-bagian Fatwa.

2. Studi Terkait

2.1 Fatwa

Fatwa adalah menerangkan hukum syara' dalam persoalan sebagai jawaban atas pertanyaan yang diajukan oleh peminta fatwa (mustafti) baik secara perorangan maupun kolektif[8]. Dimana permasalahan tersebut merupakan permasalahan yang umum atau sering di alami oleh manusia. Fatwa dikeluarkan oleh berbagai lembaga yang memiliki kemampuan khusus atau ahli di bidang agama seperti Majelis Ulama Indonesia, Muhammadiyah, Nahdlatul Ulama dan Persatuan Islam.

2.2 Latent Semantic Indexing

Latent Semantic Indexing (LSI) adalah suatu metode yang sudah ada dan sudah digunakan dalam teknologi guna mempermudah manusia untuk mencari data atau dokumen yang memiliki kemiripan satu sama lainnya. Adapun langkah-langkah detil dari proses LSI adalah sebagai berikut:

1. *Text preprocessing*, pada tugas akhir ini yang di gunakan dalam proses preprocessing yaitu tokenisasi, *stopwords removal* dan *stemming*. Untuk penjelasan lebih detailnya akan di bahas pada bab 3 mengenai rancangan sistem.
2. Setelah dilakukan preprocessing tahap selanjutnya yaitu pemberian bobot pada tiap kata. Untuk menghitung bobot kata TF (Term Frequency) dan IDF (Inverse Document Frequency). Dibawah ini rumus untuk mengitung TF-IDF:

$$Wdt = TF \times IDF$$

$$Wdt = TF \times \log \frac{n}{df}$$

Keterangan:

Wdt : bobot *term* (kata) pada fatwa

TF : jumlah kemunculan *term* dalam pada fatwa

n : jumlah semua fatwa yang ada dalam *dataset*

df : jumlah *term* yang terkandung dalam semua fatwa

3. Pada tahapan ke tiga dalam LSI yaitu penggunaan SVD yang digunakan untuk mendekomposisi matriks dari matriks A menjadi matriks U, S dan V^T sebelum dilakukannya proses *cosine similarity*. Dapat dilihat pada gambar dibawah ini merupakan rumus dari SVD :

$$A_{td} \approx U_{tm} S_{mm} V_{md}^T$$

Keterangan:

A: matriks asal

U: matriks eigenvector dari AA^T

S : matriks diagonal

V^T : transpose dari matriks V

t : jumlah baris matriks

d : jumlah kolom matriks

m: rank, dimana $\text{rank} (< \min(t,d))$

2.3 Cosine Similarity

Cosine Similarity adalah suatu rumus perhitungan matematika yang digunakan untuk menghitung nilai persamaan yang terdapat diantara dua vector dengan mencari kosinus. Dalam tugas akhir ini di gunakanlah *cosine similarity* untuk perbandingan antara fatwa satu dengan yang lainnya. Dibawah ini merupakan rumus dari *cosine similarity*:

$$\cos \theta = \frac{A \cdot B}{|A| |B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan:

A = vektor A

B = vektor B

|A| = panjang vektor A

|B| = panjang vektor V

2.4 Jaccard Similarity

Jaccard Similarity di gunakan untuk membandingkan apakah hasil dari cosine similarity dan jaccard similarity memiliki perbandingan yang berbanding lurus atau malah hasilnya berbeda. Di bawah ini rumus yang digunakan untuk menghitung jaccard similarity:

$$Jaccard(A, B) = \frac{A \cap B}{A \cup B}$$

Keterangan:

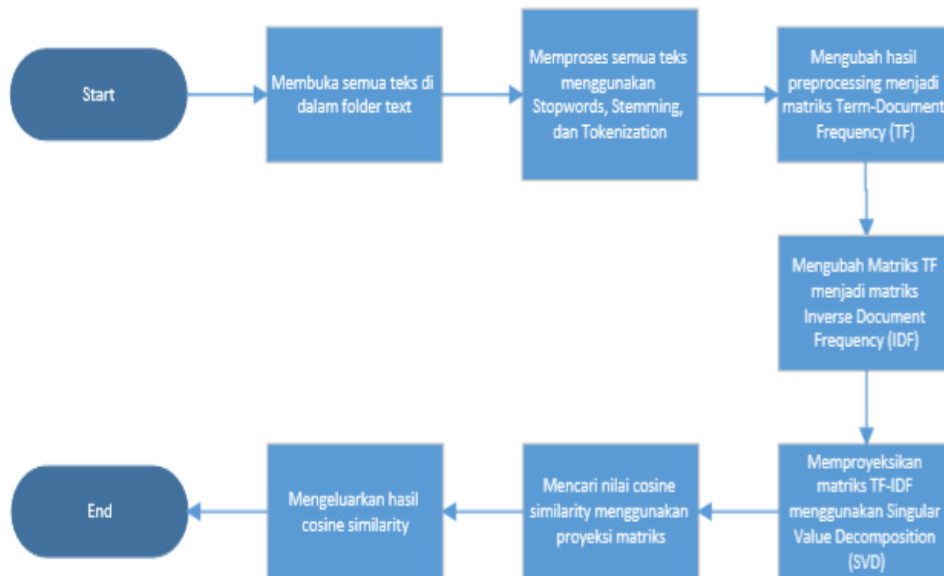
A =Query pada dokumen 1

B =Query pada dokumen 2

3. Sistem yang Dibangun

3.1 Gambaran Umum Sistem

Gambaran sistem secara umum dapat dilihat pada Gambar 1.



Gambar 1 Diagram alir

Berikut penjelasan dari gambaran umum sistem:

1. Sistem membuka semua teks di dalam folder yang tersedia dalam format .txt.
2. Sistem melakukan *preprocessing* terhadap data, yaitu dengan melakukan tokenisasi, *stopwords removal*, serta stemming.
3. Setelah *preprocessing* didapatkan data bersih yang kemudian akan diterapkan metode *latent semantic indexing*.
4. Setelah menerapkan metode, sistem melakukan perhitungan nilai kemiripan semantik dari dokumen menggunakan *cosine similarity*.

3.2 Rancangan Sistem

3.2.1 Preprocessing

Dataset pada halaman yang diambil dari situs Muhammadiyah, MUI, NU dan PERSIS dimana data belum siap digunakan, oleh karena itu dibutuhkan *preprocessing* data terlebih dahulu. Berikut ini merupakan penjelasan dari Gambar 1:

- a. Tokenisasi, merupakan suatu proses pada preprocessing dimana sebuah kalimat di ubah/ dipisahkan berdasarkan kata, pada proses tokenisasi ini sangatlah penting untuk ke tahap selanjutnya, ditandai dengan spasi dan tanda “|”, untuk lebih jelasnya dapat dilihat pada Tabel 1, kalimat diambil dari [9].

Tabel 1 Contoh Tokenisasi

Sebelum	Sesudah
ini permasalahan umat islam di berbagai negara, kalimat diambil dari [9].	ini permasalahan umat islam di berbagai negara

- b. Pada tahap berikutnya yaitu *Stopwords Removal*, dimana dalam tahap ini kata yang sudah di dapatkan dari tokenisasi akan diproses kembali. Proses yang dilakukan yaitu menghilangkan kata umum atau kata yang sering muncul namun tidak memiliki makna khusus sehingga tidak mempengaruhi arti sebenarnya yang terdapat di dalam fatwa tersebut. contohnya dapat dilihat pada Tabel 2.

Tabel 2 Contoh Stopwords Removal

Sebelum	Sesudah
ini permasalahan umat islam di berbagai negara	permasalahan umat islam berbagai negara

- c. Stemming diperlukan dimana pada proses ini akan membentuk kata menjadi kata dasar. Kata yang di proses pada tahap stemming di gunakan untuk menghindari ketidakcocokan kata yang di dapat. Berikut ini hasil dari stemming :

Tabel 3 Contoh Stemming

Sebelum	Sesudah
permasalahan umat islam berbagai negara	masalah umat islam bagai negara

3.2.2 Latent Semantic Indexing

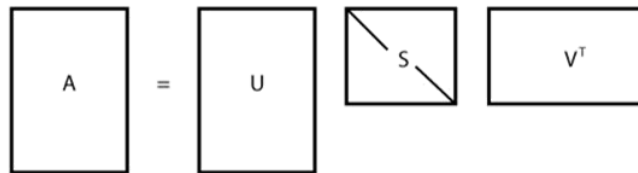
Berikut merupakan penjelasan *latent semantic Indexing* :

1. Pembobotan *Term*

Di dalam tahapan ini di kenal sebagai pemberian bobot pada tiap kata di dalam dokumen, dimana dokumen yang dipakai dalam tugas akhir ini berupa fatwa. Untuk menghitung bobot kata TF (Term Frequency) dan IDF (Inverse Document Frequency).

2. *Singular Value Decomposition* (SVD)

Untuk tahap selanjutnya setelah diperoleh hasil dari TF-IDF barulah ke tahap selanjutnya yaitu perhitungan *Singular Value Decomposition* (SVD). SVD sendiri merupakan sebuah matriks A yang di ubah menjadi tiga matriks yaitu U,S, dan V^T disebut sebagai dekomposisi matriks. Sebagai ilustrasinya dapat dilihat pada Gambar 2.



Singular Value Decomposition

Gambar 2 Ilustrasi SVD

	Q	D1	D2	D3	df	n/df	idf=log(n/df)
Masalah	1	1	1	1	3	1	0
Umat	1	1	1		2	1,5	0,17609
Islam	1		1	1	2	1,5	0,17609
negara				1	1	3	0,47712

Itu merupakan contoh sederhana dari perhitungan TF dimana terdapat 3 dokumen

Term	idf=log(n/df)	W			
		Q	D1	D2	D3
Masalah	0	0	0	0	0
Umat	0,17609	0,17609	0,17609	0,17609	0
Islam	0,17609	0,17609		0,17609	0,17609
negara	0,47712	0	0	0	0,47712

Merupakan contoh perhitungan pada table TF-IDF

	Q	D1	D2	D3
Jumlah kuadrat bobot	0.062015	0.031008	0.062015	0.258651
Panjang vektor	0.249029	0.17609	0.249029	0.508578
Jumlah perkalian bobot query dengan dokumen		0.031008	0.062015	0.031008
Cosine similarity		0.707107	1	0.244829

Perhitungan cosine similarity

3.2.3 Cosine Similarity

Cosine similarity digunakan untuk mengetahui nilai kemiripan antara fatwa yang satu dengan fatwa yang lainnya. Dimana pada dasarnya nilai pada cosine similarity hanyalah dari nilai nol sampai nilai satu, dengan penjelasannya yaitu apabila hasil perbandingan antara fatwa mendekati nilai satu maka fatwa tersebut memiliki nilai kemiripan yang tinggi sedangkan apabila nilai yang dihasilkan mendekati nilai nol memiliki arti bawa kemiripan fatwa tersebut rendah.

4. Evaluasi

4.1. Hasil Pengujian

Pengujian dalam penelitian ini, merupakan pengujian terhadap program mesin perankingan dokumen dalam pencocokan dengan dokumen yang lainnya dari hasil *dataset* yang diambil dari MUI, NU, MUHAMMADIYAH dan PERSIS (PERSATUAN ISLAM) yang semuanya mengacu kepada Al-Qur'an dan hadits. *dataset* berupa teks berbahasa Indonesia terjemahan dari Al-Qur'an dan hadits sedangkan *Output* sistem adalah berupa nilai kemiripan dari seluruh dokumen dimana yang memiliki nilai kemiripan tertinggi berada di peringkat teratas.

Pengujian ini dilakukan untuk mengetahui nilai *similarity* dari sistem berdasarkan dataset yang telah dibuat. Hasil dari skenario ini dapat membuktikan bahwa sistem memiliki penanganan terhadap beberapa bentuk variasi penulisan. Adapun tahap-tahap yang dilakukan sebagai berikut:

1. *Dataset* dengan berbagai macam bentuk variasi penulisan berdasarkan terjemahan ayat Al-Quran dan hadits yang diambil dari MUI, NU, MUHAMMADIYAH dan PERSIS (PERSATUAN ISLAM).
2. Sistem mengurutkan hasil *output* berdasarkan kemiripan dengan *query* secara *descending*, dimana posisi paling atas adalah hasil *output* yang memiliki nilai *similarity* paling baik.

Setelah di lakukan proses preprocessing hal selanjutnya yaitu dengan melakukan perhitungan TF di mana apabila terdapat kata yang sama di dalam dua dokumen yang dibandingkan maka akan bernilai 1 untuk lebih jelasnya dapat di lihat pada gambar di bawah ini:

```
Word frequency matrix:
      10  11  12  130  134  14  ...  yang  yusuf  zahrah  zakat  zalim  zina
Doc 1  0  0  0  0  0  0  ...   3    0    0    0    1    0
Doc 2  0  0  0  1  0  0  ...   4    0    0    1    0    1
Doc 3  1  1  1  0  1  1  ...   1    2    1    0    0    0
Doc 4  1  0  0  0  0  0  ...   1    0    0    0    0    0

[4 rows x 765 columns]
```

Gambar 1. Hasil dari perhitungan TF

Setelah perhitungan TF selesai hal selanjutnya yang dilakukan yaitu menghitung TF IDF pada dokumen, berikut ini gambar yang dihasilkan dari perhitungan TF IDF:

```
Tfidf matrix:
      10  11  12  ...  zakat  zalim  zina
Doc 1  0.000000  0.000000  0.000000  ...  0.000000  0.035258  0.000000
Doc 2  0.000000  0.000000  0.000000  ...  0.020873  0.000000  0.020873
Doc 3  0.015435  0.019578  0.019578  ...  0.000000  0.000000  0.000000
Doc 4  0.031869  0.000000  0.000000  ...  0.000000  0.000000  0.000000

[4 rows x 765 columns]
```

Gambar 2. Hasil dari perhitungan TF IDF

Hasil dari perhitungan cosine similarity yang di gunakan untuk menghitung nilai kemiripan dokumen yang satu dengan dokumen yang lainnya, di mana dokumen yang nilainya mendekati nilai 1,000 merupakan dokumen dengan nilai kemiripan paling tinggi. Hasil dari perhitungan cosine similarity dapat dilihat di bawah ini:

```
Cosine Similarity:
      Doc 1  Doc 2  Doc 3  Doc 4
Doc 1  1.000  0.434  0.245  0.191
Doc 2  0.434  1.000  0.536  0.358
Doc 3  0.245  0.536  1.000  0.332
Doc 4  0.191  0.358  0.332  1.000
```

Gambar 3. Hasil dari perhitungan Cosine Similarity

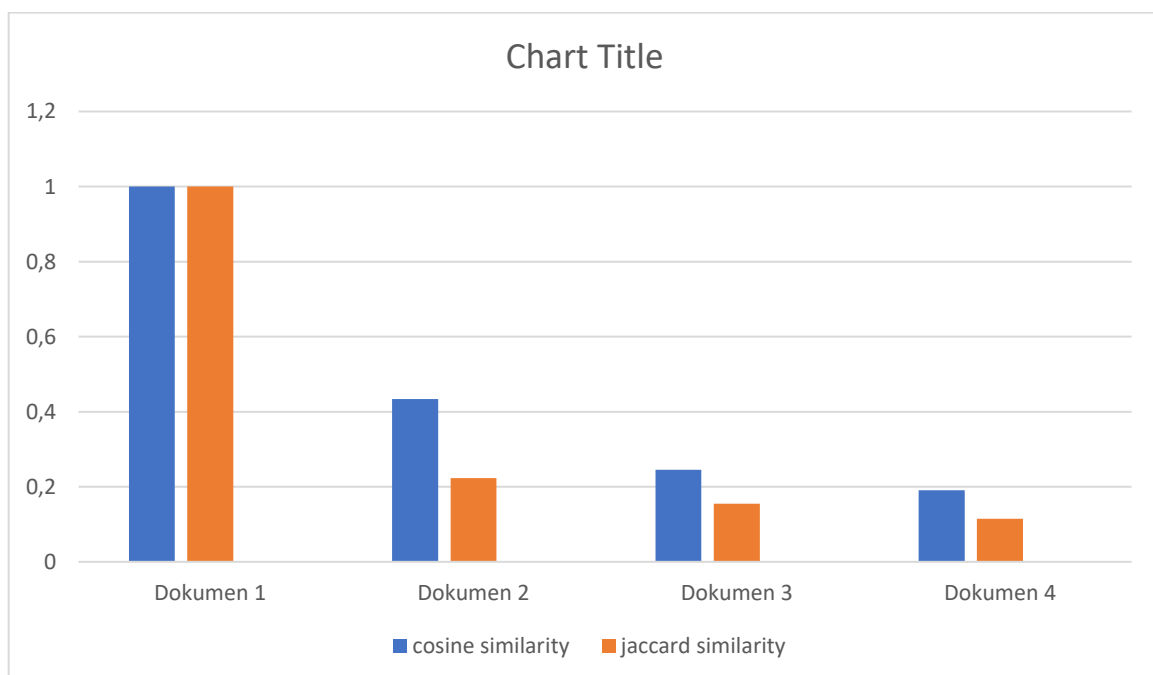
Setelah hasil dari perhitungan cosine similarity barulah dapat di buat perankingan dokumen, dimana dokumen yang nilainya mendekati nilai 1,000 maka dokumen tersebut yang memiliki peringkat/ranking tertinggi atau dapat disebut juga peringkat pertama. Gambaran pada program dapat di lihat di bawah ini:

```
Cosine Similarity Doc 1:
      Doc
Doc 1  1.000
Doc 2  0.434
Doc 3  0.245
Doc 4  0.191

Jaccard Similarity Doc 1:
      Doc
Doc 1  1.000
Doc 2  0.223
Doc 3  0.155
Doc 4  0.115
```

Gambar 4. Hasil dari perankingan dokumen

Untuk mempermudah pembacaan perankingan dapat di buat menggunakan diagram, gambaran umumnya seperti di bawah ini :



Gambar 5. Hasil perankingan dokumen ke 2

4.2. Analisa Hasil Pengujian

Dari hasil pengujian yang telah dilakukan dapat di analisis bahwa di dapatkan hasil rata-rata terlebih dahulu. Setelah di peroleh hasil rata-rata dari kedua metode tersebut, barulah menghitung hasil presentasi dari setiap metode. Dibawah ini merupakan hasil dari setiap metode yang di dapatkan yaitu:

	Cosine Similarity	Jaccard Similarity
Rata-rata	0,467	0,373
Persentase	55%	44%

5. Kesimpulan

Dari hasil pengujian yang telah dilakukan, berikut ini kesimpulan yang di dapat yaitu:

1. Dari perbandingan antara kedua metode tersebut dapat di lihat bahwa kemiripan antar dokumen yang di bandingkan memiliki nilai perbandingan yang hampir serupa antara kedua metode.
2. Keakuratan dokumen teks yang dijadikan sebagai term sangat terpengaruh oleh kelengkapan daftar stopword dan hasil root kata dari stemming untuk teks bahasa Indonesia.

Daftar Pustaka

- [1] Giacomo Domeniconi, Gianluca Moro, Roberto Pasolini, Claudio Sartori, "A Study on Term Weighting for Text Categorization: A Novel Supervised Variant of tf.idf," in DATA 2015 - Proceedings of 4th International Conference on Data Management Technologies and Applications., At Colmar, Alsace, France, 20-22 July, 2015., Alsace, 2015.
- [2] T. M. Akbar, "Analisis Perbandingan Metode Pembobotan Kata TF.IDF dan TF.RF Terhadap Performansi Kategorisasi Teks," Universitas Telkom, Bandung, 2012.
- [3] N. K. E. Putri, "Analisis Perbandingan Metode Pembobotan TF.CHI2 dan TF.RF Terhadap Kategorisasi Teks Berbahasa Indonesia," Universitas Telkom, Bandung, 2014.
- [4] Kurniawati, "Term Weighting Berbasis Indeks Kelas Menggunakan Metode TF.IDF.ICSD δ F," Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 2016.
- [5] G. A. Pratama, "Pencarian Multi Word pada Al Quran dan Hadits berdasarkan Berita Media menggunakan Metode TF-IDF," Universitas Telkom, Bandung, 2016.
- [6] C. Supriyanto and A. Affandy, "Kombinasi Teknik Chi Square Dan Singular Value Decomposition Untuk Reduksi Fitur Pada Pengelompokan Dokumen," Semantik, vol. 2011, Semantik, 2011.
- [7] L. Muflikhah and B. Baharudin, "Document Clustering Using Concept Space and Cosine Similarity Measurement," 2009 Int. Conf. Comput. Technol. Dev., pp. 58–62, 2009.
- [8] Yusuf Qardawi, Fiqh Prioritas, "Mansyurat Kuliah Da'wah Islamiyah", 1990.
- [9] kementerian luar negeri. "ini permasalahan umat islam di berbagai negara". https://kemlu.go.id/portal/id/read/129/halaman_list_lainnya/organisasi-kerja-sama-islam-oki.2019