

Metode Pembobotan Kata Berbasis *Cluster* Untuk Perangkingan Dokumen Berbahasa Arab

Cluster-Based Term Weighting Method for Ranking Arabic Documents

Amelia Devi P A¹, Lutfiyatul A², Agus Z A³, Maryamah⁴, Rizka W S⁵, Rarasmaya I⁶
^{1,2,3,4}Jurusan Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut
Teknologi Sepuluh Nopember, Surabaya
⁵Jurusan Teknologi Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut
Teknologi Sepuluh Nopember, Surabaya
⁶Jurusan Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut
Teknologi Sepuluh Nopember, Surabaya
Email: ¹ameliadev26@gmail.com, ²luthfiaizah.if@gmail.com, ³agusza@cs.its.ac.id,
⁴maryamahfaisol02@gmail.com, ⁵wakhidatus@its.ac.id, ⁶raras@its.ac.id

Abstrak

Perangkingan dokumen telah menjadi topik yang banyak dibahas pada sistem temu kembali informasi untuk memberikan urutan dokumen paling relevan berdasarkan kueri yang diberikan oleh pengguna. Namun, penelitian tentang perangkingan dokumen dalam bahasa Arab masih belum banyak dilakukan karena memiliki morfologi yang unik dan literatur dalam bahasa Arab yang masih sedikit. Selain itu, didalam proses perangkingan juga diperlukan perhitungan pembobotan kata yang optimal supaya dapat memberikan hasil yang sesuai. Pembobotan kata yang paling umum digunakan adalah *term frequency-inverse document frequency* (TF.IDF) yang hanya menghitung pembobotan setiap kata berdasarkan pengelompokan dokumen saja sehingga dapat menyebabkan relevansi antar dokumen menjadi rendah karena memiliki tingkat kemiripan antar dokumen yang berbeda. Oleh karena itu, diperlukan suatu pembobotan kata berdasarkan pengelompokan dokumen dalam suatu *cluster* untuk meningkatkan relevansi antar dokumen. Tujuan penelitian ini membuat metode perangkingan dokumen berbahasa Arab dengan menggunakan pembobotan berbasis *cluster* yang menggabungkan TF.IDF dan *inverse cluster frequency* (ICF) sehingga diperoleh TF.IDF.ICF. Evaluasi dilakukan menggunakan dokumen *e-book* berbahasa Arab yang telah dikelompokkan menjadi tiga *cluster*. Hasil penelitian membuktikan pembobotan TF.IDF.ICF mampu menemukan dokumen paling relevan terhadap kueri yang dimasukkan oleh pengguna serta memperoleh nilai rata-rata *precision* dan *F1-Measure* yang lebih tinggi daripada hanya menggunakan pembobotan TF.IDF yaitu sebesar 68% dan 78%.

Kata Kunci: Perangkingan Dokumen Arab, TF.IDF.ICF, Pembobotan Kata Berbasis *Cluster*

Abstract

Document ranking has become a much-discussed topic of information retrieval systems to rank documents based on their relevance based on a user-supplied query. However, research on document ranking in Arabic has not been widely carried out because it has a unique morphology, and literature in Arabic is still small. Besides, in the ranking process, it is also necessary to calculate the optimal term weighting so that it can provide appropriate results. The most commonly used term weighting is term frequency-inverse document frequency (TF.IDF) which only calculates the weighting of each term based on document grouping so that it can cause the relevance between documents to be low because they have a similarity level between different documents. Therefore, we need term weighting based on the grouping of documents in a cluster to increase the relevance between documents. This study proposes a document ranking method in Arabic using cluster-based weighting that combines TF.IDF and inverse cluster frequency (ICF) to obtain TF.IDF.ICF. The evaluation was carried out using an Arabic e-book document which has been grouped into three clusters. The results of the research prove that

TF.IDF.ICF weighting can find the documents most relevant to the query entered by the user and obtain higher average values for precision and F1-Measure than using TF.IDF weighting, namely 68% and 78%.

Keywords: Arabic Document Ranking, TF.IDF.ICF, Cluster-Based Term Weighting

1. PENDAHULUAN

Sistem temu kembali informasi atau *information retrieval* merupakan sistem untuk menemukan kembali informasi yang paling relevan berdasarkan kebutuhan dari pencari informasi atau pengguna [1]. Salah satu konsep dari sistem temu kembali informasi adalah proses pencarian dokumen [2]. Sebuah permintaan berupa kueri yang diberikan oleh pengguna akan sangat menentukan hasil yang diperoleh dari proses penemuan kembali informasi [3]. Dokumen-dokumen yang dihasilkan dari proses pencarian yang dilakukan oleh pengguna akan diurutkan sesuai dengan tingkat relevansinya atau biasa disebut dengan perangkingan dokumen. Hasil teratas dari proses perangkingan dokumen tersebut akan menjadi dokumen yang paling relevan terhadap kueri yang telah dimasukan oleh pengguna [4]. Perangkingan dokumen ini merupakan pembahasan yang umum diteliti dalam sistem temu kembali informasi [5].

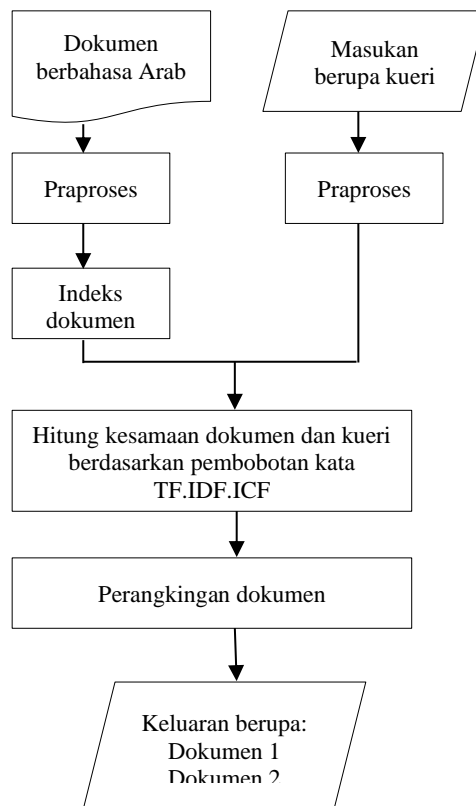
Namun, penelitian tentang perangkingan dokumen dalam bahasa Arab masih belum banyak dilakukan [1]. Hal tersebut dikarenakan ada beberapa perbedaan antara bahasa Arab dengan bahasa lainnya, diantaranya bahasa Arab dituliskan dari kanan ke kiri serta di dalam bahasa Arab, satu kata dapat memiliki arti lebih dari satu [3]. Selain itu, bahasa Arab juga memiliki morfologi yang unik dan literatur dalam bahasa Arab yang masih sedikit sehingga belum banyak yang melakukan penelitian dengan menggunakan dokumen berbahasa Arab [6].

Pada umumnya, perangkingan dilakukan dengan menghitung bobot dari setiap kata dalam dokumen. Pembobotan kata yang banyak digunakan adalah TF.IDF (*term frequency inverse document frequency*) [1]. Pembobotan kata TF (*Term frequency*) merupakan banyaknya kemunculan kata dalam dokumen, sedangkan IDF (*inverse document frequency*) merupakan banyaknya kemunculan kata dalam dokumen lain [1]. Pada pembobotan TF.IDF ini, kata yang jarang muncul memiliki nilai yang lebih tinggi daripada kata yang sering muncul [5]. Penelitian yang dilakukan oleh Maryamah dkk [7] mengusulkan sebuah metode pembobotan berbasis kelas dan topik untuk berita berbahasa Indonesia. Metode tersebut merupakan gabungan dari TF.IDF.ICF dan ITF sehingga menjadi TF.IDF.ICF.ITF. *Inverse topic frequency* (ITF) pada penelitian tersebut memperhatikan kemunculan kata dalam dokumen yang sebelumnya sudah dibagi menjadi beberapa topik dan bertujuan untuk mengelompokkan dokumen secara otomatis. Namun hasil yang diperoleh dari metode pembobotan TF.IDF.ICF.ITF masih kurang optimal dan masih rendah dibandingkan metode TF.IDF. Penelitian Ren dkk [8] mengusulkan sebuah metode yaitu *inverse class frequency* (ICF) dan *inverse class space density Frequency* (ICS_dF) untuk proses pembobotan kata. Penambahan metode ICS_dF pada penelitian tersebut memberikan nilai diskriminatif positif pada kata yang sering muncul maupun yang jarang muncul. Penelitian yang dilakukan oleh Holle dkk [5] mengusulkan sebuah metode pembobotan kata dengan mempertimbangkan preferensi pengguna pada sistem perangkingan dokumen berbahasa Arab yaitu dengan menghitung preferensi kata menggunakan IPF. Kemudian nilai preferensi dari kata pada dokumen yang memiliki kesamaan dengan kueri dikalikan dengan α sebagai penguat. Nilai α adalah variabel bernilai antara 0 sampai sampai 1. Selanjutnya menggabungkan dengan metode pembobotan yang sudah ada menjadi TF.IDF.IBF.IPF α . Penelitian Fauzi dkk [1] melakukan perangkingan dokumen buku berbahasa Arab dengan menghitung kemiripan kueri yang dimasukan oleh pengguna menggunakan pembobotan kata TF.IDF.ICF.IBF. Metode tersebut menggabungkan pembobotan kata TF.IDF dengan ICF (*inverse class frequency*) dan IBF (*inverse book frequency*). Pembobotan kata ICF mempertimbangkan kemunculan kata pada kumpulan kelas, sedangkan pembobotan kata IBF mempertimbangkan kemunculan kata pada kumpulan buku atau kitab.

Pada perhitungan pembobotan IBF tersebut didasarkan pada pengelompokan dokumen dalam sebuah buku, di mana dalam sebuah buku terdapat beberapa bab atau tema yang berbeda, sehingga menyebabkan relevansi antar dokumen tersebut bisa jadi rendah karena memiliki tingkat kemiripan antar dokumen yang berbeda dan membahas topik yang juga berbeda sama sekali. Oleh karena itu, diperlukan suatu pembobotan kata berdasarkan pengelompokan dokumen dalam suatu *cluster* pada sistem perangkingan dokumen berbahasa Arab supaya mendapatkan tingkat relevansi antar dokumen yang tinggi karena mampu menemukan dokumen paling relevan terhadap kueri yang dimasukkan oleh pengguna dan dapat memberikan dokumen lain dalam satu *cluster* yang memiliki tingkat kemiripan yang sama dengan dokumen paling relevan tersebut. Dengan demikian, penelitian ini bertujuan membuat suatu sistem perangkingan dokumen berbahasa Arab dengan menggunakan pembobotan kata berbasis *cluster* yaitu TF.IDF.ICF untuk menemukan dokumen paling relevan terhadap kueri yang dimasukkan oleh pengguna dengan menggunakan *inverse cluster frequency* yang mengelompokkan dokumen berbahasa Arab berdasarkan masing-masing *cluster* dan menggabungkannya dengan TF.IDF.

2. METODE PENELITIAN

Terdapat beberapa tahapan metode perangkingan yang dilakukan dalam penelitian ini. Tahap pertama adalah melakukan praproses pada dokumen berbahasa Arab. Tahap praproses ini juga dilakukan pada kueri yang dimasukkan oleh pengguna. Kemudian dari tahap praproses akan didapatkan beberapa kata-kata. Kata-kata tersebut kemudian dilakukan pengindeksan tiap kata berdasarkan dokumen dan *cluster*. Tahap selanjutnya adalah perhitungan kesamaan dokumen dan kueri yang dimasukkan oleh pengguna berdasarkan perhitungan pembobotan kata TF.IDF.ICF. Perangkingan dokumen menghasilkan dokumen yang terurut sesuai dengan tingkat relevansinya.



Gambar 1 Tahapan-tahapan metode usulan

2.1 Pengumpulan Data dan Praproses

Pada penelitian ini digunakan dataset berbahasa Arab yang sebelumnya telah terbagi menjadi beberapa *cluster*. Total ada 150 dokumen *e-book* berbahasa Arab yang terdiri dari tiga *cluster* seperti yang ditunjukkan pada Tabel 1. Setiap *cluster* berisi 50 dokumen.

Tabel 1 Dataset yang digunakan dalam penelitian

<i>Cluster</i>	Total
1	50
2	50
3	50
Total	150

Kemudian, dari dataset tersebut dilakukan praproses. Praproses adalah proses perubahan data terstruktur dari data tidak terstruktur yang disesuaikan dengan kebutuhan [9]. Semua dokumen dan kueri yang dimasukkan oleh pengguna akan melalui tahapan praproses terlebih dahulu. Pada tahapan ini, terdiri dari beberapa tahap yaitu tokenisasi, pemfilteran, normalisasi, menghapus *stopword* dan *stemming* [5]. Tokenisasi adalah proses mengubah dokumen menjadi kata per kata dengan menghilangkan spasi. Pemfilteran adalah proses menghapus *diartic* (harokat), angka dan tanda baca. Normalisasi merupakan proses yang sangat penting teks Arab memiliki banyak macam cara penulisan pada kata yang sama. Normalisasi dilakukan dengan misalnya mengubah (ء), (إ), (أ), (ؤ), (ئ) menjadi alif (ا), mengubah ta marbutoh (ة) menjadi ha (ه), mengubah ya (ي) menjadi ya (ي) [4]. Tahap selanjutnya adalah penghapusan *stopword Arabic* [5]. Kemudian pada tahap terakhir yaitu *stemming*, yang mengubah bentuk kata menjadi kata dasar dengan melakukan penghapusan imbuhan. Misalnya menghapus kata sambung wa (و), beberapa prefix (ال, ب, ل, ك, ف, ل, قال, كال, بال, وال) dan beberapa suffix (ا, ه, ها, و, ان, ين, وا, ون, ها, ه, ا) [4].

2.2 Pembobotan Kata dan Indeks Dokumen

Pembobotan dilakukan untuk menyatakan kepentingan kata terhadap dokumen atau sekumpulan dokumen. Pada tahap pembobotan ini dilakukan menggunakan *term frequency* (TF), *inverse document frequency* (IDF) dan *inverse cluster frequency* (ICF).

2.2.1 Term Frequency (TF)

Term frequency (TF) merupakan perhitungan banyaknya kemunculan kata dalam dokumen, seperti dirumuskan sebagai berikut:

$$TF(d, k) = f(d, k) \quad (1)$$

Persamaan 1 diatas merupakan rumus dari pembobotan TF dimana $f(d, k)$ adalah frekuensi kemunculan kata k pada dokumen d .

2.2.2 Inverse Document Frequency (IDF)

Pembobotan kata IDF adalah perhitungan kemunculan kata pada sekumpulan dokumen yang mana kata yang jarang muncul memiliki nilai yang lebih tinggi dibandingkan dengan kata yang sering muncul. Dengan demikian, dapat dikatakan bahwa kepentingan kata akan berkebalikan dengan frekuensi kemunculannya dalam sekumpulan dokumen, seperti persamaan dibawah ini:

$$IDF(k) = 1 + \log (Nd/df(k)) \quad (2)$$

Persamaan 2 menjelaskan bahwa Nd adalah jumlah dari seluruh dokumen dalam dataset dan $df(k)$ adalah jumlah dokumen yang mengandung kata k .

2.2.3 Inverse Cluster Frequency (ICF)

Selain menghitung TF.IDF, penelitian ini juga menambahkan perhitungan *inverse cluster frequency* (ICF) untuk menghitung pembobotan kata berdasarkan *cluster*. Apabila IDF memperhatikan munculnya kata dalam sekumpulan dokumen, maka pembobotan ICF disini lebih memperhatikan kemunculan kata pada suatu dataset yang sebelumnya telah terbagi menjadi beberapa *cluster*. Kata yang jarang muncul dalam suatu *cluster* merupakan kata yang memiliki nilai lebih tinggi. Kepentingan kata dalam pembobotan ICF akan berkebalikan dengan frekuensi kemunculannya dalam suatu *cluster* tertentu. Untuk menghitung ICF dapat dilihat persamaan di bawah ini:

$$ICF(k) = 1 + \log (Nc/cf(k)) \quad (3)$$

Persamaan 3 merupakan rumus dari perhitungan *inverse cluster frequency* dimana Nc adalah jumlah seluruh *cluster* yang ada dalam dataset dan $cf(k)$ adalah jumlah *cluster* yang mengandung kata k .

$$TF.IDF.ICF = TF(d, k) \times IDF(k) \times ICF(k) \quad (4)$$

Kemudian persamaan 1,2 dan 3 dikalikan untuk mendapatkan rumus pembobotan kata TF.IDF.ICF seperti yang terlihat pada persamaan 4 diatas. Setelah dilakukan proses pembobotan kata, tahap menghitung kesamaan antara indeks kata pada dokumen dengan kueri yang telah dimasukan oleh pengguna dilakukan berdasarkan perhitungan TF.IDF.ICF yang sudah didapatkan tadi. Dari perhitungan kesamaan tersebut akan dilakukan proses perangkaian dokumen terlebih dahulu hingga didapatkan hasil dokumen yang sudah terurut.

3. HASIL DAN PEMBAHASAN

Perhitungan pembobotan kata TF.IDF yang dirumuskan dengan mengalikan antara persamaan 1 dan 2 dilakukan setelah proses pengumpulan data dan praproses. Hasil perhitungan pembobotan kata TF.IDF ditunjukkan pada Tabel 2, dan hasil perhitungan pembobotan kata TF.IDF.ICF terlihat pada Tabel 3 di bawah ini. Perhitungan pembobotan kata TF.IDF.ICF tersebut didapatkan dengan menggunakan persamaan 4.

Tabel 2 Hasil perhitungan pembobotan kata TF.IDF

No	Term	Arti Term	TF.IDF
1	الصلاة	Sholat	4.2468
2	القواعد	Aturan	3.7405
3	الفرق	Perbedaan	3.1761
4	كريمة	Ramah	3.1761
5	زمن	Waktu	3.1761

6	العدل	Keadilan	3.1761
7	الخلق	Penciptaan	3.1761
8	قدر	Takdir	2.8751
9	الفضل	Kebaikan	2.3979
10	قول	Perkataan	2.273

Tabel 3 Hasil perhitungan pembobotan kata TF.IDF.ICF

No	Term	Arti Term	TF.IDF.ICF
1	القواعد	Aturan	5.5252
2	الصلاة	Sholat	4.9946
3	الفرق	Perbedaan	4.6015
4	كريمة	Ramah	4.6015
5	زمن	Waktu	4.6015
6	العدل	Keadilan	4.6015
7	الخلق	Penciptaan	4.6015
8	قدر	Takdir	4.2468
9	الفضل	Kebaikan	2.8202
10	قول	Perkataan	2.6733

Pada penelitian ini juga dilakukan perbandingan antara pembobotan kata TF.IDF dan TF.IDF.ICF. Perbandingan tersebut didapatkan dengan menghitung nilai *precision* (P), *recall* (R), dan *F1-Measure*. *Precision* adalah kesesuaian atau kecocokan dokumen antara dokumen yang diinginkan pengguna dengan jawaban yang dihasilkan dari proses pencarian dokumen, sedangkan *recall* menjelaskan keberhasilan proses pencarian dokumen dalam menemukan kembali informasi yang diminta oleh pengguna [10]. *F1-Measure* adalah pengaruh relatif antara *precision* dan *recall* [11]. Perhitungan *precision*, *recall* dan *F1-Measure* dirumuskan pada persamaan 5, 6, dan 7 di bawah ini.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F1 - Measure = \frac{2 \times P \times R}{P+R} \quad (7)$$

Dari persamaan di atas menjelaskan bahwa TP (*true positif*) adalah banyaknya dokumen yang berhasil ditemukan dan relevan. FP (*false positif*) menunjukkan banyaknya dokumen yang ditemukan tapi tidak relevan. FN (*false negative*) berarti banyaknya dokumen yang relevan tapi tidak ditemukan. Perhitungan *F1-Measure* dilakukan dengan mengalikan *precision* dan *recall* dikali dengan 2 kemudian dibagi dengan jumlah dari *precision* dan *recall*.

Tabel 4 Perbandingan antara TF.IDF.ICF dan TF.IDF

Kueri	TF.IDF		TF.IDF.ICF	
	P	R	P	R
Q1	0.50	0.67	0.75	0.67
Q2	0.50	1.00	0.50	1.00
Q3	0.67	0.75	0.67	0.75
Q4	0.50	1.00	0.50	1.00
Q5	1.00	1.00	1.00	1.00
Rata-rata	63%	88%	68%	88%
F1	76%		78%	

Tabel 4 di atas merupakan hasil perbandingan antara pembobotan kata TF.IDF dengan TF.IDF.ICF yang dihitung berdasarkan nilai *precision*, *recall* dan *F1-Measure* menggunakan persamaan 5, 6 dan 7. Berdasarkan hasil perbandingan tersebut menunjukkan bahwa pembobotan TF.IDF.ICF memiliki nilai rata-rata *precision* dan *F1-Measure* yang lebih tinggi dibandingkan dengan pembobotan kata TF.IDF. Jika nilai *precision* yang didapatkan dalam penelitian ini dibandingkan dengan hasil penelitian yang dilakukan oleh Amril Mutoi [12] yang hanya menggunakan pembobotan kata TF.IDF pada sistem perankingan dokumen, maka nilai *precision* menggunakan pembobotan kata TF.IDF.ICF yang diusulkan dalam penelitian ini mendapatkan nilai yang lebih tinggi daripada nilai *precision* menggunakan pembobotan kata TF.IDF yang diusulkan oleh Amril Mutoi yang hanya sebesar 3,9% saja. Perolehan nilai rata-rata *precision* yang meningkat itu menandakan bahwa dengan menggunakan pembobotan yang diusulkan yaitu TF.IDF.ICF dapat meningkatkan keberhasilan proses pencarian dokumen dengan memberikan dokumen yang paling relevan menurut pengguna. Selain itu, jika nilai *recall* yang didapatkan dalam penelitian ini dibandingkan dengan penelitian Fatkhul Amin dkk [13] yang menggunakan pembobotan kata TF.IDF pada sistem perankingan dokumen, maka nilai *recall* dengan menggunakan pembobotan kata TF.IDF.ICF mendapatkan nilai yang lebih tinggi daripada nilai *recall* menggunakan pembobotan kata TF.IDF yang diusulkan oleh Fatkhul Amin dkk yang hanya sebesar 3%. Perolehan nilai rata-rata *recall* yang meningkat menandakan bahwa dengan menggunakan pembobotan kata TF.IDF.ICF dapat meningkatkan keberhasilan proses pencarian dokumen dalam menemukan dokumen yang relevan. Perolehan nilai rata-rata *recall* dan *precision* yang meningkat, maka meningkatkan pula nilai *F1-Measure* pada pembobotan TF.IDF.ICF karena nilai *F1-Measure* sangat terkait dengan nilai *recall* dan *precision*. Terjadinya peningkatan nilai *recall*, *precision* dan *F1-Measure* dengan menggunakan pembobotan TF.IDF.ICF dikarenakan dalam perhitungan ICF memperhatikan perhitungan kata pada masing-masing *cluster* yang mana dalam suatu *cluster* memiliki tingkat kemiripan antar dokumen yang tinggi sehingga dapat meningkatkan relevansi antar dokumen dan dapat menemukan dokumen paling relevan terhadap kueri yang dimasukkan oleh pengguna. Namun, pembobotan kata dengan menggunakan TF.IDF.ICF, dimana ICF disini berarti *inverse cluster frequency*, akan memberi nilai rendah pada kata-kata yang muncul di beberapa *cluster* sehingga

masih perlu memperhatikan jumlah kemunculan suatu kata dalam sekumpulan dokumen yang menjadi anggota pada suatu *cluster* tertentu.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil membuat sistem perangkingan dokumen berbahasa Arab dengan menggunakan pembobotan kata berbasis *cluster* yaitu TF.IDF.ICF yang mampu menemukan dokumen yang paling relevan berdasarkan kueri yang dimasukkan oleh pengguna dengan memberikan dokumen lain dalam satu *cluster* yang memiliki tingkat kemiripan yang sama dengan dokumen paling relevan tersebut. Nilai rata-rata *precision* dan *F1-Measure* dengan menggunakan pembobotan kata TF.IDF.ICF lebih tinggi daripada menggunakan pembobotan TF.IDF yaitu sebesar 68% dan 78%. Namun, di dalam penelitian ini masih membutuhkan pengembangan yaitu pembobotan kata ICF (*inverse cluster frequency*) yang lebih memperhatikan jumlah kemunculan kata yang terdapat dalam sekumpulan dokumen yang menjadi anggota dari suatu *cluster* tertentu.

DAFTAR PUSTAKA

- [1] M. A. Fauzi, A. Z. Arifin, and A. Yuniarti, 2017, Arabic Book Retrieval Using Class and Book Index Based Term Weighting, *International Journal of Electrical and Computer Engineering (IJECE)*, No. 6, Vol. 7, pp. 3705–3710.
- [2] A. Aulia, D. Khairani, and N. Hakiem, 2017, *Development of A Retrieval System for Al Hadith in Bahasa (Case Study: Hadith Bukhari)*, UIN Syarif Hidayatullah, Jakarta.
- [3] E. Elabd, E. Alshari, and H. Abdulkader, 2015, Semantic Boolean Arabic Information Retrieval, *International Arab Journal of Information Technology*, No. 3, Vol. 12, pp. 311-316.
- [4] R. Sholikah, D. Kartika, A. Z. Arifin, and D. Purwitasari, 2017, Term Weighting Based on Positive Impact Factor Query for Arabic Fiqh Document Ranking, *Jurnal Ilmu Komputer dan Informasi*, No. 1, Vol. 10, pp. 29-36.
- [5] K. F. H. Holle, A. Z. Arifin, and D. Purwitasari, 2015, Preference Based Term Weighting For Arabic Fiqh Document Ranking, *Jurnal Ilmu Komputer dan Informasi*, No. 1, Vol. 8, pp. 45-52.
- [6] M. Alhanjouri, 2017, Pre Processing Techniques for Arabic Documents Clustering, *International Journal of Engineering and Management Research*, No. 2, Vol. 7, pp. 70-79.
- [7] Maryamah, M. A. P. Subali, L. S. Qolby, A. Z. Arifin, and M. A. Fauzi, 2018, Metode Pembobotan Berbasis Topik dan Kelas untuk Berita Online Berbahasa Indonesia, *Jurnal Linguistik Komputasional*, No. 1, Vol. 1, pp. 11–16.
- [8] F. Ren and M. G. Sohrab, 2013, Class-Indexing-Based Term Weighting for Automatic Text Classification, *Information Science*, pp. 109-125.
- [9] A. T. Ni'mah and A. Z. Arifin, 2020, Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis, *Rekayasa*, No. 2, Vol.13, pp. 172-180.

- [10] N. P. Lestari, 2016, Uji Recall and Precision Sistem Temu Kembali Informasi OPAC Perpustakaan ITS Surabaya, *Skripsi*, Program Studi Sarjana Ilmu Informasi dan Perpustakaan, Univ. Airlangga, Surabaya.
- [11] N. Muslimah, Indriati, and R. . Wihandika, 2019, Klasifikasi Film Berdasarkan Sinopsis dengan Menggunakan Improved K-Nearest Neighbor (K-NN), *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, No. 1, Vol. 3, pp. 196–204.
- [12] A. M. Siregar, 2017, Perbandingan Pembobotan Kata Dalam Sistem Temu Balik Informasi, *Techno Xplore Jurnal Teknologi dan Ilmu Komputer*, No. 2, Vol. 2, pp. 1-8.
- [13] F. Amin and J. A. Razaq, 2018, Pemeringkatan Hasil Pencarian Dokumen Teks Pada Sistem Temu Kembali Informasi Berbahasa Jawa Menggunakan Metode Dice Similarity, *INFOKAM*, No. 2, pp. 120–129.