TERM WEIGHTING BERDASARKAN *CLASS DISTINGUISHING POWER* MENGGUNAKAN METODE TF.IGM UNTUK PERANGKINGAN DOKUMEN AL-QURAN

SKRIPSI

Oleh :

RURI NUR AINI
NIM. 12650022

JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2018

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

HALAMAN JUDUL

TERM WEIGHTING BERDASARKAN *CLASS DISTINGUISHING POWER* MENGGUNAKAN METODE TF.IGM UNTUK PERANGKINGAN DOKUMEN AL-QURAN

SKRIPSI

Diajukan kepada :
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
RURI NUR AINI
NIM.12650022

JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2018

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

LEMBAR PERSETUJUAN

***TERM WEIGHTING* BERDASARKAN *CLASS DISTINGUISHING POWER* MENGGUNAKAN METODE TF.IGM UNTUK PERANKINGAN DOKUMEN AL-QURAN**

SKRIPSI

Oleh :
RURI NUR AINI
NIM.12650022

Telah diperiksa dan disetujui untuk diuji

Tanggal : 01 November **2018**

Dosen Pembimbing I	Dosen Pembimbing II
Dr. Cahyo Crysdian	Khadijah F.H. Holle, M.Kom
NIP. 19740424 200901 1 008	NIDT. 19900626 20160801 2 077

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

ii

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

LEMBAR PENGESAHAN

TERM WEIGHTING BERDASARKAN _CLASS DISTINGUISHING POWER_ MENGGUNAKAN METODE TF.IGM UNTUK PERANGKINGAN DOKUMEN AL-QURAN

SKRIPSI

Oleh :
RURI NUR AINI
NIM.12650022

Telah Dipertahankan Didepan Dewan Penguji Skripsi
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal 19 Desember 2018

Susunan Dewan Penguji:		**Tanda tangan**
Penguji Utama	: Linda Salma Anggreani, M.T NIP. 19770803 200912 2 005	( )
Ketua Penguji	: Dr.M. Amin Hariyadi, M.T NIP. 19670118 200501 1 001	( )
Sekretaris Penguji	: Dr. Cahyo Crysdian NIP. 19740424 200901 1 008	( )
Anggota Penguji	: Khadijah F.H. Holle, M.Kom NIDT. 19900626 20160801 2 077	( )

Mengetahui,
Ketua jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

iii

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

HALAMAN ORISINALITAS PENELITIAN

Saya yang bertanda tangan dibawah ini :

Nama : Ruri Nur Aini

NIM : 12650022

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Judul skripsi: ***Term Weighting* Berdasarkan *Class Distinguishing Power* dengan Menggunakan Metode TF.IGM untuk Perangkingan Dokumen Al-Quran**

Menyatakan dengan sebenar-benarnya bahwa hasil penelitian saya ini tidak mengandung unsur-unsur penjiplakan karya penelitian atau karya ilmiah orang lain, kecuali yang secara tertulis saya kutip dalam laporan ini dan saya sebutkan sumber kutipan pada daftar pustaka. Apabila ternyata hasil penelitian ini terbukti terdapat unsur penjiplakan, maka saya bersedia untuk mempertanggung jawabkan, serta diproses sesuai aturan yang berlaku.

Malang, 01 - Nov - 2018
Yang membuat pernyataan,

Ruri Nur Aini
12650022

iv

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

MOTO

“ Jika melakukan sesuatu atas nama Allah,

Tidak akan ada yang namanya kecewa dan sakit hati”

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

HALAMAN PERSEMBAHAN

Tugas akhir ini saya persembahkan:

- Kepada **Ayah** dan **Ibu**, Terimakasih atas segala kasih sayang, do'a dan dukungannya meskipun tugas ini tidak dapat saya selesaikan dengan baik dan sedikit terlambat lulus.
- Kepada **Dosen Pembimbing Pak Cahyo** dan **Bu Khadijah**, Terimakasih yang telah sabar dalam membimbing saya yang sangat banyak kekurangan dalam menyelesaikan tugas ini.
- Kepada kakak-kakak (**Rahmat** dan **Nisa**) dan adik (**Rian**), terimakasih selalu mendukung dan menghibur saya selama proses.
- Kepada **Mas Faiq Faruqi** dan **Mas Alif Nur Iman**, terimakasih telah membantu dan sabar menemani belajar.
- Kepada teman-teman yang selalu menyemangati dan mendoakan, yang pernah berjuang bareng… **Niamah, Finda, Windi, Evi, Ariyani, Luvy, Laili, Laila, Rizqo.**
- Kepada adik-adik tersayang **Weni** dan **Neni** yang menemani sampai akhir perjuangan.

Terimakasih semuanya..
Terimakasih atas dukungan, motivasi, nasihat dan DOAnya…
Semoga Allah SWT melimpahkan berkah dan ridlo-Nya kepada kita semua..
Amin….

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

KATA PENGANTAR

Segala puji bagi Allah yang telah melimpahkan Rahmat dan Karunia-Nya kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini. Serta shalawat serta salam tercurahkan kepada Nabi Muhammad SAW yang membawa umatnya dari zaman gelap menuju zaman yang terang benderang yakni *addinul islam wal iman.*

Dan tidak lupa penulis sampaikan terimakasih yang sebesar-besarnya kepada pihak—pihak yang membantu dan mendukung penulis dalam menyelesaikan tugas akhir ini, baik secra langsung maupun tidak langsung.

1. Kepada Dosen pembimbing sekaligus sebagai Ketua Jurusan Teknik Informatika Bapak Dr. Cahyo Crysdian yang selalu mendukung, mengarahkan dan memberi masukan dalam pengerjaan tugas akhir ini.
2. Kepada Dosen pembimbing Ibu Khadijah Fahmi H. H, M.Kom yang telah banyak meluangkan waktu, memberi motivasi, mengarahkan dan sealu sabar membimbing penulis, sehingga penulis dapat menyelesaikan tugas akhir ini.
3. Kepada Bapak Zainal Abidin, M.Kom selaku Dosen Wali.
4. Kepada seluruh Dosen Teknik Informatika yang telah memberikan ilmu kepada penulis selama studi.
5. Dan kepada semua pihak yang tidak bisa penulis sebut satu persatu, terimakasih atas segala yang diberikan kepada penulis baik pengertian, perhatian dan dukungannya selama ini.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Akhirnya, penulis berharap bahwa tugas akhir ini dapat memberikan kontribusi keilmuan untuk melanjut dan mengembangkannya. Namun, penulis menyadari bahwa tugas ini masih banyak kekurangannya dan jauh dari sempurna. Semoga kekurangan dapat disempurnakan oleh penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Amin.

Malang, 2018

Penulis

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

DAFTAR ISI

HALAMAN JUDUL i
LEMBAR PERSETUJUAN ii
LEMBAR PENGESAHAN iii
HALAMAN ORISINALITAS PENELITIAN iv
MOTTO v
HALAMAN PERSEMBAHAN vi
KATA PENGANTAR vii
DAFTAR ISI ix
DAFTAR GAMBAR xi
DAFTAR TABEL xiii
ABSTRAK xiv
BAB 1 PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Identifikasi Masalah 4
1.3 Tujuan Penelitian 5
1.4 Manfaat Penelitian 5
1.5 Batasan Masalah 5
1.6 Sistematika Penulisan 6
BAB 2 TINJAUAN PUSTAKA 8
2.1 Penelitian Terkait 8
2.2 Sistem Temu Kembali Informasi 12
2.3 Preprocessing 14
2.4 Perangkingan Dokumen 16
2.4.1 *Term Weighting* 17
2.4.2 *Vector Space Model* (VSM) 20
2.4.3 *Cosine Similarity* 22
2.4.4 Perhitungan Mean Average *Precision* (MAP) 23
BAB 3 METODOLOGI PENELITIAN 24
3.1 Studi Literatur 24

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

3.2 Perancangan dan Imlementasi Sistem 25
3.2.1 Dataset 27
3.2.2 *Preprocessing* 29
3.2.3 *Term Weighting* 34
3.2.4 *Vector Space Model* dan *Cosine Similarity* 43
3.2.5 Output 46
3.3 Evaluasi Pengujian dan Anilisa Hasil 46
3.4 Lingkungan Pengembangan Sistem 47
BAB 4 PEMBAHASAN 49
4.1 Implementasi 49
4.2 Analisa Uji coba dan Hasil 50
4.3 Integrasi dengan Al Quran 65
BAB 5 PENUTUP 71
5.1 Kesimpulan 71
5.2 Saran 71

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

DAFTAR GAMBAR

Gambar 2.1 Kerangka sistem temu kembali informasi (Ingwersen, 1992)..... 13
Gambar 2. 2 Diagram Preprocessing 14
Gambar 2. 3 Representasi dokumen dan vektor pada ruang vektor (mandala, 2006). 21
Gambar 2.4 Representasi matriks kata-dokumen (Mandala. 2006)..... 21
Gambar 3. 1 Alur Metodologi Penelitian. 24
Gambar 3. 2 Blok Diagram Rancangan Sistem 26
Gambar 3. 3 Code Proses Tokenizing..... 30
Gambar 3. 4 Code Proses Normalizing dan Filtering 30
Gambar 3. 5 Code Proses Stopword Removal 32
Gambar 3. 6 Tabel Hasil Preprocessing pada database..... 33
Gambar 3. 7 Code Proses Penghapusan Suffix 34
Gambar 3. 8 Query SQL Pembobotan TF..... 35
Gambar 3. 9 Tabel hasil Pembobotan TF 35
Gambar 3. 10 SQL code pengindeksan term. 36
Gambar 3. 11 Code Proses sorting class 36
Gambar 3. 12 Tabel Hasil Pembobotan IGM..... 37
Gambar 3. 13SQL code perkalian bobot TF dan bobot IGM 38
Gambar 3. 14 Tabel hasil perkalian bobot TF dan bobot IGM..... 39
Gambar 3. 15 Code Proses Preprocessing query 42
Gambar 3. 16 Code Proses pembobotan query 42
Gambar 3. 17 SQL code perkalian bobot query dengan bobot term dokumen..... 43
Gambar 3. 18 SQL perhitungan vektor dokumen 43
Gambar 3. 19 SQL perhitungan vektor query..... 44
Gambar 3. 20 SQL code perhitungan similarity.....44
Gambar 4. 1 Contoh pencarian ayat 53
Gambar 4. 2 Tabel hasil q_term..... 53
Gambar 4. 3 Tabel hasil proses_bobot_input..... 54

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Gambar 4. 4 Tabel hasil dot_product 54
Gambar 4. 5 Tabel hasil similarity 55
Gambar 4. 6 Hasil proses pada program 56
Gambar 4. 7 Grafik Prosentase nilai MAP 58

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

DAFTAR TABEL

Tabel 3. 1 Daftar dokumen AL-Quran 27
Tabel 3. 2 Term hasil Stemming 32
Tabel 3. 3 Contoh Dokumen 40
Tabel 3. 4 Pembobotan TF dan IGM 41
Tabel 3. 5 Pembobotan TF.IGM 43
Tabel 3. 6 Hasil Pembobotan query 44
Tabel 3. 7 Perhitungan cosine similarity 45
Tabel 3. 8 Perangkingan dokumen 46
Tabel 3. 9 Skenario Uji coba 48
Tabel 4. 1 DaftarInput query pendek 52
Tabel 4. 2 Daftar Input query panjang 53
Tabel 4. 3 nilai prosentase MAP 58
Tabel 4. 4 Perhitungan MAP rank-k 10 query pendek 60
Tabel 4. 5 Perhitungan MAP rank-k 20 query pendek 61
Tabel 4. 6 Perhitungan MAP rank-k 10 query panjang 62
Tabel 4. 7 Perhitungan MAP rank-k 20 query panjang 63
Tabel 4. 8 Nilai perhitungan MAP query pendek dan panjang rank-k10 64
Tabel 4. 9 Nilai perhitungan MAP query pendek dan panjang rank-k 20 65

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

ABSTRAK

Ruri Nur Aini, 2018. ***Term Weighting* Berdasarkan *Class Distinguishing Power* Menggunakan Metode TF.IGM untuk Perangkingan Dokumen Al-Quran.** Pembimbing : (1) Dr. Cahyo Crysdian (2) Khadijah F.H Holle, M.Kom

Kata kunci : Perangkingan dokumen, indeks kelas, pembobotan kata, TF.IGM.

Besarnya jumlah data digital saat ini, menjadi tantangan dalam pencarian informasi. Adanya sistem temu kembali informasi akan meringankan para pencari untuk menemukan kembali data atau dokumen yang dibutuhkan. Salah satu pembahasannya yaitu perangkingan dokumen. Representasi perangkingan dokumen pada umumnya menggunakan *Vector Space Model* (VSM) berbasis pada pembobotan TF.IDF. Pembobotan TF.IDF tidak sepenuhnya dianggap efektif karena hanya memperhatikan pembobotan kata berbasis dokumen. Pada penelitian ini menggunakan pendekatan pembobotan kata berbasis indeks kelas yang diterapkan pada dataset Al-Quran berbahasa Arab yang telah terbagi dalam 14 indeks kelas sesuai pada kitab *Fahturrahman* oleh Syaikh Ilmi Zadeh Fu'ad Abd Al Baqiy. Pembobotan TF.IGM secara tepat mengukur *class distinguishing power* (kekuatan membedakan kelas) pada *term. Term* dengan kelas khusus kadang-kadang muncul dalam kelas lain, hal ini sangat sering terjadi dalam satu bahkan beberapa kelas. *Term* dalam kelas khusus ini dapat digunakan untuk membedakan teks antar kelas yang berbeda dan diberikan bobot yang lebih besar. Secara umum, *term* yang lebih terkonsentrasi distribusi kelas cenderung memiliki kekuatan membedakan kelas. Dataset akan melalui *preprocessing*, pembobotan, pengukuran kemiripan dengan *cosine similarity* dan dievaluasi dengan menggunakan MAP. Data uji terdiri dari masing-masing indeks kelas dan diketagorikan dalam *query* pendek dan *query* panjang. Hasil penelitian menunjukkan bahwa metode yang diusulkan menghasilkan nilai MAP tertinggi yaitu 84.24% dari *query* panjang.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

ABSTRACT

Ruri Nur Aini, 2018. **Term Weighting based on Class Distinguishing Power using TF.IGM method for ranking Quran Documents.** Pembimbing : (1) Dr. Cahyo Crysdian (2) Khadijah F.H Holle, M.Kom

Keywords: Document ranking, class index, term weighting, TF.IGM

Abstract - The large amount of digital data today is a challenge in information retrieval. The Information Retrieval System make it easy the users to rediscover the data or documents needed. One of the discussions is ranking documents. Document ranking representation generally uses Vector Space Model (VSM) based on TF.IDF weighting method. The TF.IDF is not fully considered effective because it only focused into the weighting of the document-based words. In this study, the class index based weighting approach was applied to the Arabic-language Al-Quran dataset which was divided into 14 classes index according to the Fahturrahman book by Shaykh Ilmi Zadeh Fu'ad Abd Al Baqiy. TF.IGM weighting precisely measures the class distinguishing power on the term. Term with special class sometimes appears in other classes, this often happens in one or even several classes. The terms in this special class can be used to distinguish text between different classes and given greater weight. In general, terms that are more concentrated in class distribution tend to have the power to distinguish classes. The dataset will be through preprocessing, weighting, measurement of similarity with cosine similarity and evaluated using MAP. The test data consists of each class index and is categorized in short queries and long queries. The results showed that the proposed method produced the highest MAP value of 84.24% of long queries.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

الملخص

روري نور عيني، ٢٠١٨. **ترجيح الأزمنة بأساس الصف مميز الطاقة باستخدام طريقة TF.IGM لترتيب وثيقة القرآن.** المشرف: (1)الدكتور جاهيو كريسديان (2) خديجة F.H هولي الماجيستر

كلمات البحث: ترتيب الوثائق، فهرس الصف، ترجيح الكلمات، TF.IGM.

يصير الكبير من البيانات الرقمية اليوم تحديًا في استرجاع المعلومات. سيخفف وجود نظام استرجاع المعلومات من الباحثين لإعادة اكتشاف البيانات أو الوثيقة المطلوبة. أحد المناقشات هي ترتيب الوثائق. يستخدم التمثيل لترقيم الوثائق عاميا نموذج الفضاء نافلات (VSM) مستنادًا إلى ترجيح TF.IDF. لا يعتبر ترجيح TF.IDF فعالا كاملا لأنه يهتم ترجيح الكلمات مستنادًا إلى الوثيقة فقط. استخدم في هذا البحث النهج ترجيح الكلمات مستنادًا إلى فهرس الصف إلى المطبق في مجموعة بيانات القرآن باللغة العربية مقسما إلى ١٤ فهرس الصف موافقا لكتاب فتح الرحمان للشيخ Iعلمي زاده فؤاد عبد الباقي. يقيس ترجيح TF.IGM بدقة الصف مميز الطاقة (القدرة على تمييز الطبقات) في المصطلح. يظهر مصطلح ذو فصل خاص أحيانًا في فصل آخر، غالبًا ما يحدث هذا في فئة واحدة حتى عدة الفصول. يمكن استخدام هذا المصطلح في فصل خاص لتمييز النص بين الصفوف المختلفة وأعطي وزن أكبر. بشكل العام، المصطلح الأكثر تركيزًا في توزيع الصفوف تميل إلى امتلاك القوة على التمييز بين الطبقات. ستمر مجموعة البيانات من خلال المعالجة المسبوقة، الترجيح، قياس التشابه بتشابه الجيب ويقيم باستخدام MAP. تتكون بيانات الاختبار من كل فهرس الصف ويصنف في سؤال قصير وسؤال طويل. أظهرت نتائج البحث أن الطريقة المقترحة أنتجت أعلى القيمة MAP هي ٨٤,٢٤٪ من سؤال طويل.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Besarnya jumlah data digital saat ini, menjadi tantangan dalam pencarian informasi. Adanya sistem temu kembali informasi atau juga yang disebut *information retrieval system* akan meringankan para pencari untuk menemukan kembali data atau dokumen yang dibutuhkan. Data atau dokumen tersebut dapat berupa teks, suara, gambar maupun video.

Salah satu pembahasan dari sistem temu kembali adalah perangkingan dokumen. Perangkingan dokumen dilakukan untuk menyediakan informasi dokumen yang sesuai dengan data yang diinginkan pengguna dari *query* pengguna (Esraa, 2010). Secara garis besar, pokok pembahasan IR meliputi 2 hal yakni pembobotan dan perangkingan (Holle, Arifin, & Purwitasari, 2015). Perangkingan digunakan untuk mendapatkan dokumen-dokumen yang relevan dengan *query* pengguna diurutkan dari tingkat relevansinya. Perangkingan ini diperlukan karena mungkin dokumen yang sesuai dengan *query* pencarian lebih dari satu.

Representasi perangkingan dokumen salah satunya menggunakan *Vector Space Model* (VSM) dari kumpulan dataset. Dokumen dalam *vector space model* direpresentasikan dalam bentuk matriks yang berisi bobot *term* pada dokumen. Bobot tersebut menyatakan kepentingan/kontribusi terhadap suatu dokumen atau kumpulan dokumen. Kepentingan suatu dokumen dapat dilihat dari frekuensi kemunculan bobot terhadap dokumen (Indra, 2015).

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Perangkingan dokumen juga dapat diimplementasikan pada pemilihan dan pembobotan kata dari *query* untuk meningkatkan kesesuaian konteks *query* dengan pencarian dokumen. Metode pembobotan kata yang sering digunakan adalah *Term Frequency & Inverse Document Frequency* (TF.IDF). Dimana dari setiap kata yang diasumsikan memiliki kepentingan yang proposional terhadap jumlah kemunculan kata pada dokumen yaitu *term frequency* (TF) akan dikalikan dengan kemunculan *term* pada kumpulan dokumen yaitu *inverse document frequency* (IDF).

Adapun penelitian tentang perangkingan dokumen berbahasa Arab yang dilakukan oleh Harrag dkk (2008) menggunakan *vector space model* (SVM) berbasis *term weighting* TF.IDF. Pada metode tersebut dokumen direpresentasikan sebagai sebuah vektor yang dibentuk dari nilai-nilai *term* yang menjadi indeksnya dan mengukur kemiripannya dengan *cosine similarity* (Harrag et al, 2008). Metode TF.IDF hanya berbasis pada *term* dalam satu dokumen, dan itu tidak cukup untuk menentukan indeks dokumen. Penentuan indeks yang akurat juga bergantung pada kepentingan *term* terhadap kelas (kelangkaan *term* pada keseluruhan kelas). Sehingga dibutuhkan *term weighting* berbasis kelas yang dinamakan *inverse class frequency* (ICF) (Wang dan Zang, 2013). Pada metode pembobotan ICF hanya memperhatikan *term* yang ada pada kelas tanpa memperhatikan jumlah *term* dalam dokumen yang menjadi anggota kelas. Oleh karena itu fungsi ICF direvisi dan menerapkan *inverse class space density frequency* ($ICS_{\delta}F$), dan dihasilkan metode TF.IDF.$ICS_{\delta}F$ yang menyediakan

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

diskriminasi positif pada *term* yang jarang dan sering muncul (Ren dan Sohrab, 2013).

Pembobotan yang dilakukan Fauzi (2013) berdasarkan *term* berbasis buku untuk perangkingan dokumen berbahasa Arab. Pembobotan yang dilakukan menggunakan metode TF.IDF.ICF yang akan dikalikan dengan *inverse book frequency* sehingga menjadi metode TF.IDF.ICF.IBF untuk meningkatkan performa perangkingan dokumen yang memiliki hirarki berupa buku-buku yang memiliki banyak halaman (Fauzi, 2013). Begitu pula pada penelitian yang dilakukan oleh Khadijah (2014) pembobotan kata untuk perangkingan dokumen fiqih berbahasa Arab berdasarkan tingkat preferensi dari pengguna yang menggunakan metode TF.IDF.IBF.IPF$\propto$ (Khadijah, 2014). Revisian metode ICF telah diimplementasikan juga pada perangkingan dokumen Al-Quran yang diajukan oleh Kurniawati (2016) berbasis indeks kelas dengan metode TF.IDF.$ICS_{\delta}F$. Meskipun pada penelitiannya tidak ditunjukkan *recall, presision* dan *f-measure* untuk evaluasinya, telah disebutkan bahwa nilai akurasinya lebih tinggi dibandingkan dengan pencarian menggunakan metode ICF saja (Kurniawati, 2016).

Sejak metode TF.IDF tidak sepenuhnya dianggap efektif dalam pengklasifikasian, Kewen dkk (2016) mengusulkan sebuah metode baru tentang pembobotan dimana pembobotan lokal TF dikalikan *inverse gravity moment* (IGM). TF.IGM menggabungkan model statistik baru untuk secara tepat mengukur *class distinguishing power* (kekuatan membedakan kelas) pada *term. Term* dengan kelas khusus kadang-kadang muncul dalam kelas lain, hal ini sangat

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

sering terjadi dalam satu bahkan beberapa kelas. *Term* dalam kelas khusus ini dapat digunakan untuk membedakan teks antar kelas yang berbeda dan diberikan bobot yang lebih besar. Secara umum, *term* yang lebih terkonsentrasi distribusi kelas cenderung memiliki kekuatan membedakan kelas. Ketidakseragaman distribusi sampel umumnya diukur dengan entropi. Namun dalam penelitian Kewen dkk (2016) mengukur ketidakseragaman atau tingkat konsentrasi distribusi diusulkan dengan model statistik baru yang disebut *inverse gravity moment* (IGM) yang mencerminkan daya pembela kelas pada *term.*

Oleh karena itu dalam penelitian ini, penulis mengusulkan implementasi metode *term weighting* berdasarkan *class distinguishing power* (kekuatan membedakan kelas) untuk perangkingan dokumen Al-Quran. Perhitungan *class distinguising power* menggunakan pembobotan IGM. Bobot IGM digabungkan dengan pembobotan lokal yang sudah ada yaitu TF sehingga menjadi pembobotan TF.IGM. Implementasi pembobotan TF.IGM pada perangkingan dokumen Al-Quran diharapkan memberikan *output* dokumen ter-*retrieve* yang memiliki urutan (rangking) baik sesuai *query* pengguna berdasarkan tema (indeks).

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya maka terdapat permasalahan yang diangkat dalam penelitian ini adalah seberapa tinggi tingkat relevansi yang ditunjukkan dengan nilai *mean average precision* (MAP) untuk perangkingan dokumen Al-Quran dengan metode TF.IGM?

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

1.3 Tujuan Penelitian

Adapun maksud dan tujuan penelitian ini untuk mengukur tingkat relevansi perangkingan dokumen Al-Quran dengan metode TF.IGM dari nilai *mean average precision* (MAP).

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut:

1. Dalam segi peneliti, dapat memahami dan implementasikan konsep/teori tentang *term weighting* dan perangkingan dokumen.
2. Dalam segi keilmuan, turut serta dalam mengintegrasikan ilmu pengetahuan teknologi dengan islam yang berupa penerapan metode *term weighting* TF.IGM terhadap perangkingan dokumen Al-Quran.
3. Dalam segi pembaca, metode TF.IGM dapat dijadikan sebagai acuan untuk melakukan *term weighting*.

1.5 Batasan Masalah

Pembahasan dalam penelitian ini dibatasi beberapa hal sebagai berikut:

1. Metode pembobotan TF.IGM diterapkan pada perangkingan dokumen berbahasa Arab.
2. Dataset bahasa Arab Al-Quran yang digunakan terdiri dari 14 kelas.
3. Tahap stemming bahasa arab menggunakan algoritma *light stemmer*.
4. Ukuran kemiripan menggunakan algoritma *cosine similarity*.
5. Evaluasi relevansi perangkingan menggunakan *Mean Average Precision* (MAP).

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

1.6 Sistematika Penulisan

Dalam penulisan skripsi ini, secara keseluruhan terdiri dari lima bab yang masing-masing bab disusun dengan sistematika sebagai berikut :

BAB I PENDAHULUAN

Bab ini merupakan bagian awal, dalam bab ini berisi latar belakang masalah, pernyataan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan laporan.

BAB II STUDI PUSTAKA

Bab ini berisi tentang teori-teori yang berhubungan dengan permasalahan yang diangkat dari penelitian ini, antara lain penjadwalan, penjadwalan mata pelajaran sekolah, algoritma *Artificial Bee Colony* dan penelitian-penelitian terkait.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang desain sistem sistem dan implementasi sistem yang akan dibuat sekaligus batasan-batasan sistem serta di dalamnya juga terdapat beberapa diagram-diagram sistem.

BAB IV UJI COBA DAN PEMBAHASAN

Bab ini berisi mengenai pengujian dan analisis dari hasil pengujian dari sistem yang telah dibangun berdasarkan hasil perancangan pada bab 3 sebelumnya.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran seluruh penelitian yang telah dilakukan.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB 2

TINJAUAN PUSTAKA

Pada bagian ini membahas tentang penelitian yang terkait dan konsep tentang teori yang digunakan dalam melakukan penelitian ini.

2.1 Penelitian Terkait

Penelitian yang dilakukan oleh Kewen dkk (2016) mengusulkan sebuah metode baru tentang pembobotan dimana pembobotan lokal TF dikalikan *inverse gravity moment* (IGM). Metode pembobotan TF.IGM inilah yang mennjadi metode pada penelitian ini. TF.IGM menggabungkan model statistik baru untuk secara tepat mengukur *class distinguishing power* (kekuatan membedakan kelas) pada *term. Term* dengan kelas khusus kadang-kadang muncul dalam kelas lain, hal ini sangat sering terjadi dalam satu bahkan beberapa kelas. *Term* dalam kelas khusus ini dapat digunakan untuk membedakan teks antar kelas yang berbeda dan diberikan bobot yang lebih besar. Pembeda pada dtribusi sampel umumnya diukur dengan entropi. Namun dalam penelitian Kewen dkk (2016) mengukur pembeda atau tingkat konsentrasi distribusi diusulkan dengan model statistik baru yang disebut *inverse gravity moment* (IGM) yang mencerminkan daya pembela kelas pada *term*.

Penelitian perangkingan dokumen Al-Quran sebelumnya pernah dilakukan oleh Kurniawati (2016), menggunakan pendekatan pembobotan berbasis pengindeksan kelas dan membandingkan pengaruhnya pada dua metode pembobotan lain yang dijadikan pendekatan dasar. Metode yang diterapkan pada

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

dataset Al-Quran yang dijadikan tolak ukur koleksi. Al-Quran memiliki banyak ayat, masing-masing ayat tersebut adalah dokumen yang akan dirangking berdasarkan *query* pengguna. TF.IDF melakukan pembobotan hanya berbasis dokumen tanpa memperhatikan kelas yang merupakan induk dokumen. Sedangkan TF.IDF.ICF memperhatikan indeks kelas dan menghiraukan anggota dari setiap kelas yang menyebabkan sulit untuk membedakan frekuensi *term* yang jarang muncul. Oleh karenanya, pada penelitiannya diimplementasikan fungsi *inverse class space density frequency* ($ICS_\delta F$) dengan menggabungkan fungsi pembobotan yang sudah ada sehingga menjadi TF.IDF.$ICS_\delta F$. Metode tersebut dapat memberikan nilai diskriminasi positif pada *term* yang jarang muncul.

Penelitian perangkingan dokumen berbahasa Arab juga pernah dilakukan oleh Khadijah (2014). Namun perangkingan yang dilakukannya adalah dokumen/buku fiqih bahasa Arab. Metode yang diajukan adalah metode pembobotan kata berbasis preferensi untuk merangkingkan dokumen sesuai dengan preferensi pengguna. Metode yang diajukan digabungkan dengan pembobotan kata berbasis indeks dokumen dan buku sehingga mampu memperhatikan aspek kesesuaian (*relevance*) dan keutamaan (*preference*). Metode pembobotan yang diusulkan disebut dengan *Inverse Preference Frequency with α value* (IPFα). Langkah pembobotan yang diusulkan yaitu dengan perhitungan nilai preferensi *term* dengan pembobotan IPF. Kemudian nilai preferensi dari *term* dokumen yang sama dengan *term query* dikalikan dengan $\boldsymbol{\alpha}$ sebagai penguat. IPFα digabungkan dengan metode pembobotan yang telah ada menjadi TF.IDF.IBF.IPFα. Pengujian metode yang diusulkan menggunakan

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

dataset dari beberapa dokumen fiqih berbahasa Arab. Evaluasi menggunakan perhitungan *recall*, *precision*, dan *f-measure*. Hasil uji coba menunjukkan bahwa dengan pembobotan TF.IDF.IBF.IPFα diperoleh perangkingan dokumen dengan urutan yang tepat dan sesuai dengan preferensi pengguna.

Pada penelitian yang dilakukan oleh Fauzi (2013) juga untuk dokumen fiqih berbahasa Arab, diusulkan metode baru *term weighting* yang berbasis pada indeks buku dan kelas. Metode ini memperhatikan frekuensi kemunculan *term* pada keseluruhan buku dan kelas. Metode ICF dan *inverse book frequency* (IBF) ini digabungkan dengan metode sebelumnya sehingga menjadi TF.IDF.ICF.IBF. Pengujian metode ini menggunakan dataset dari beberapa *e-book* berbahasa arab. Hasil penelitian menunjukkan bahwa metode yang diajukan terbukti dapat diaplikasikan pada perangkingan dokumen berbahasa arab dan memiliki performa yang lebih bagus dibanding metode sebelumnya. Karena dianggap terdapat beberapa buku berbahasa Arab yang memiliki puluhan bahkan ratusan halaman. Masing-masing halaman dari buku tersebut adalah sebuah dokumen yang akan dirangking berdasarkan *query* dari pengguna. TF.IDF hanya melakukan pembobotan berbasis pada dokumen tanpa memperhatikan indeks buku dan kelas yang merupakan induk dokumen tersebut sehingga kinerjanya kurang maksimal jika diimplementasikan pada kasusnya.

Penelitian yang dilakukan oleh Ren dan Sohrab (2013) diperkenalkan pendekatan *term weighting* berdasarkan indeks kelas dan menilai dimensi dalam ruang vektor. *Term wighting* TF.IDF.ICF berbasis indeks kelas diterapkan di mana *inverse class frequency* (ICF) yang tergabung. Dalam percobaannya

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

diselidiki efek dari TF.IDF.ICF terhadap Reuters-21578, 20 Newsgroup, dan dataset RCV1-v2 sebagai patokan, yang menyediakan diskriminasi positif pada *term* langka di ruang vektor. *Inverse class space density frequency* (*ICSδF*) menurut Ren dan Sohrab adalah revisian dari metode sebelumnya yaitu ICF, dan hasilnya metode TF.IDF.*ICSδF* yang menyediakan diskriminasi positif pada *term* yang jarang dan sering muncul. Hasil percobaannya menunjukkan bahwa metode *term weighting* TF.IDF.*ICSδF* berdasarkan indeks kelas diusulkan adalah lebih baik dan mampu mengatasi masalah yang berdimensi tinggi.

Menurut Wang dan Zang (2013) pada penelitiannya, skema pembobotan kata (*term weighting*) seringkali mendominasi kinerja skema klasifikasi teks yang ada. Oleh karena itu Wang dan Zang (2013) memperkenalkan *inverse category frequency* (ICF) dalam pembobotan *term* dan mengusulkan dua pendekatan baru yaitu, TF.IDF dan pembobotan *term supervised* berbasis ICF, yang menggabungkan ICF dengan frekuensi yang relevan. Pengenalan ICF dapat menetapkan kekurangan pembobotan dalam katerogi-kategori yang ada. Hasil penelitiannya berdasarkan tiga *corpus* yaitu, *skewed Reuters*-21578, *balanced* 20 *Newsgroup* and la12, menunjukkan bahwa dua skema *term weighting* yang digunakan mencapai kinerja yang lebih baik dibandingkan dengan tujuh skema *supervised term weighting* dan tiga skema *term weighting* tradisional.

Penelitian yang dilakukan oleh Harrag (2008) membuat aplikasi pengindeksan hadis menurut derajat kemiripannya menggunakan *Vector space model* berbasis *term weighting* TF.IDF. Pada metode tersebut dokumen direpresentasikan sebagai sebuah vektor yang dibentuk dari nilai-nilai *term* yang

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

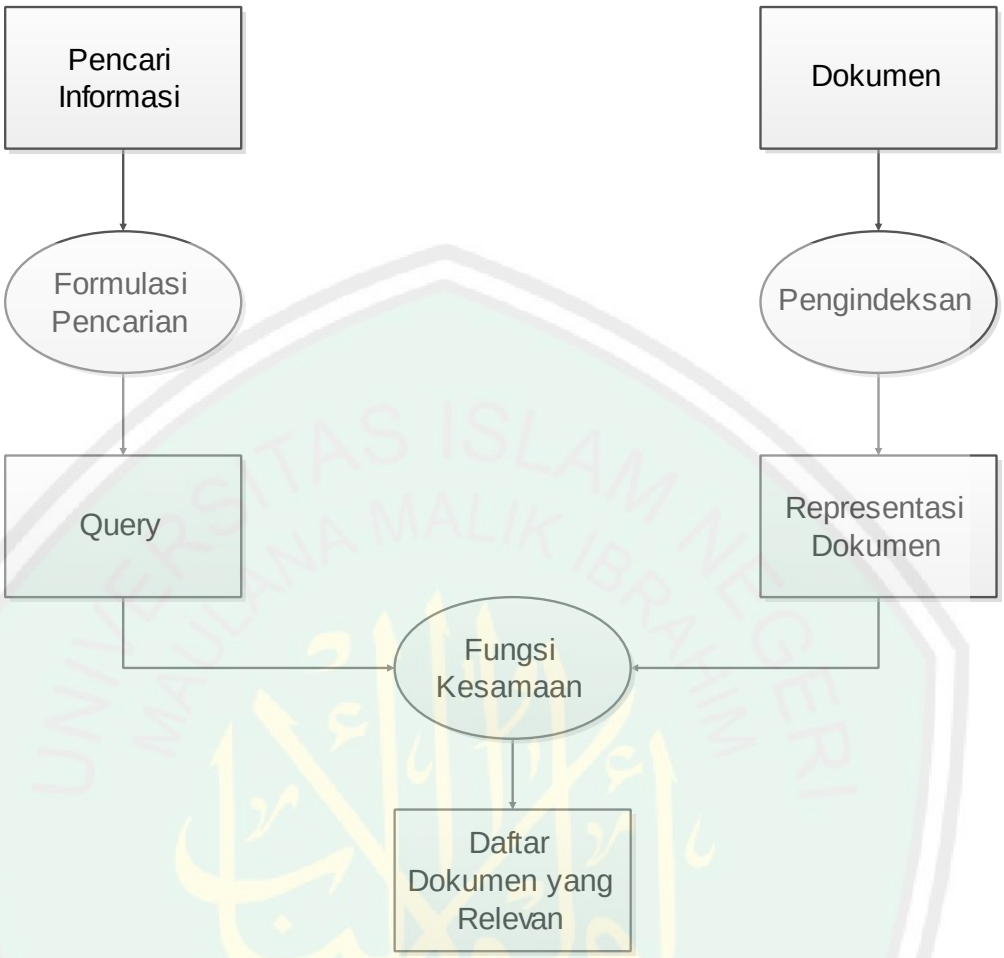
menjadi indeksnya. Nilai-nilai *term* tersebut dihitung dengan menggukanan *term weighting* TF.IDF dan mengukur kemiripannya dengan menggunakan *cosine similarity*. Dan penggunaan teknik tersebut dapat memberikan hasil yang efisien dan menampilkan kerjaan tambahan yang sukses diselesaikan oleh bahasa-bahasa barat. Hasil tersebut menyatakan bahwa ketidaktergantungan antara cara statistik yang digunakan dan bahasa yang sedang dipertimbangkan.

2.2 Sistem Temu Kembali Informasi

Prinsip kerja sistem temu kembali informasi jika ada sebuah kumpulan dokumen dan seorang *user* yang memformulasikan sebuah pertanyaan (*request* atau *query*). Jawaban dari pertanyaan tersebut adalah sekumpulan dokumen yang relevan dan membuang dokumen yang tidak relevan (Rijsbergen, 1976).

Kerangka dari sistem temu kembali informasi sederhana terbagi menjadi dua bagian. Pertama, pencari informasi atau pengguna memasukkan kata kunci yang diproses menjadi sebuah *query* agar dapat dimengerti oleh sistem. Kedua, bagian dari dokumen, dimana dokumen-dokumen direpresentasikan dalam bentuk indeks. Selanjutnya *query* pengguna akan diproses melalui fungsi kesamaan untuk membandingkan *query* dengan indeks dari dokumen untuk mendapatkan dokumen yang relevan. Untuk lebih jelasnya mengenai kerangka sistem temu kembali informasi dapat dilihat pada **gambar 2.1 Kerangka Sistem temu kembali informasi**.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



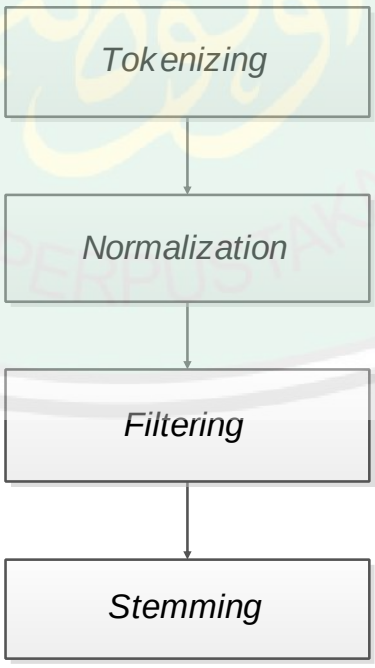
Gambar 2.1 Kerangka Sistem temu kembali informasi (Ingwersen, 1992)

Sistem temu kembali secara garis besar menangani dua pekerjaan yaitu melakukan *preprocessing* terhadap *database* dan kemudian menerapkan metode tertentu untuk menghitung kedekatan antara dokumen didalam *database* yang telah diproses dengan *query user*. *Query* yang dimasukkan oleh *user* akan dikonversi sesuai dengan aturan tertentu untuk mengekstrak *term* penting. *Term* yang sebelumnya telah diekstrak dari dokumen akan dihitung relevansinya antar *query*. Dan sebagai hasilnya, sistem akan mengembalikan suatu daftar dokumen terurut/rangking sesuai dengan nilai kemiripannya dengan *query user*.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

2.3 Preprocessing

Preprocessing atau yang disebut dengan tahap persiapan data, merupakan tahapan awal dalam mengolah data *input* sebelum memasuki proses tahapan pembobotan. Tahap persiapan data bertujuan untuk mempersiapkan data agar formatnya sesuai dengan kebutuhan. Pada tiap dokumen teks terdapat kata-kata, dan kata dalam dokumen disebut dengan *term* yang selanjutnya akan dilakukan *prepocessing*. Pada tahap *preprocessing* ini intinya adalah *indexing*. *Indexing* adalah proses membangun representasi suatu dokumen dengan memberikan suatu pengenal pada item-item kelas. Indeks dokumen adalah himpunan *term* yang menunjukkan isi atau topik yang dikandung oleh dokumen. Proses-proses yang berjalan dalam tahap ini seperti pada **gambar 2.3 Diagram *Preprocessing*** berikut:



Gambar 2.2 Diagram Preprocessing

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

1. *Tokenizing* (pemisahan rangkaian kata)

   Pada tahap ini deret kata didalam kalimat, paragraf maupun halaman akan dipisahkan menjadi *token* atau atau potongan kata tunggal. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti digit, angka, tanda hubung dan tanda baca.

2. *Normalizing and Filtering*

   Pada tahapan ini ditentukan *term* mana mana yang akan digunakan untuk merepresentasikan dokumen sehingga dapat mendeskripsikan isi dokumen dan membedakan dokumen tersebut dengan dokumen lain didalam koleksi. Dan juga penghilangan karakter-karakter yang tidak termasuk dalam huruf *hijaiyah* (ا,ب,ت), kemudian dilakukan penghapusan *harakat*, dan menormalkan teks bahasa Arab ke bentuk dasar. Dan semua teks-teks bukan bahasa Arab dihapus. Setelah itu, baru dilakukan proses *normalization and filtration* data yang dihasilkan berupa potongan kata tunggal tanpa *harakat*, tanpa tanda baca dan dikembalikan ke bentuk normal *hijaiyah*.

3. *Stopword removal*

   Term yang efektif dalam pemisahan dokumen yang relevan dari dokumen yang tidak relevan kemungkinan besar adalah *term* yang muncul pada sedikit dokumen. Yang artinya bahwa *term* dengan frekuensi kemunculan tinggi tidak memberikan nilai informasi tinggi. Dan *term* yang muncul dalam banyak dokumen tidak mencerminkan definisi dari topik atau sub-topik dokumen. Kata fungsi pada bahasa Arab seperti kata ganti dan preposisi akan dihapus. Strategi umum penentuan daftar *stopword* adalah mengurutkan *term*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

berdasarkan frekuensi koleksi (koleksi total kemunculan setiap *term* didalam koleksi dokumen) dan memasukkan *term* yang paling sering muncul sebagai *stopword*. Setelah dilakukan penghapusan pada tahap *stopword* tinggal menyisakan kata-kata yang memiliki pengaruh dalam suatu dokumen.

4. S*temming* (konversi *term* kebentuk dasar).

   Pada tahap *stemming* ini adalah proses konversi *term* ke bentuk dasar. Dalam tahapan ini setiap kata yang dihasilkan dari serangkaian proses sebelumnya kemudian dirubah menjadi kata dasar.

Proses *stemming* mempunyai peranan penting dalam *information retrieval*, karena proses ini tidak hanya mengumpulkan kata yang sama namun juga dapat mengurangi jumlah *term* yang merepresentasikan suatu dokumen tertentu, sehingga penyimpanan data *term* akan lebih sedikit dan proses yang dijalankan akan lebih cepat. Salah satu *stemmer* teks bahasa Arab adalah *light stemming* (Larkey, 2007). *Stemmer* ini yang digunakan pada *library Lucene* apabila menggunakan *class Arabic Analyzer*

2.4 Perangkingan Dokumen

Perangkingan dokumen menggunakan representasi *vector space model* dan kumpulan dataset. Dokumen dalam *vector space model* direpresentasikan dalam matriks yang berisi bobot kata pada dokumen. Bobot tersebut menyatakan kepentingan/kontribusi kata terhadap suatu dokumen dan kumpulan dokumen. Kepentingan suatu kata dalam dokumen dapat dilihat dari frekuensi kemunculannya terhadap dokumen. Kata yang berbeda biasanya memiliki frekuensi yang berbeda pula (Indra, 2015).

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Dari pembobotan akan diperoleh bobot kata pada dokumen yang direpresentasikan sebagai vektor. Dari representasi bobot tersebut dapat dihitung nilai kemiripan suatu dokumen dengan *query*. Nilai kemiripan ini akan dihitung dengan rumus *cosine similarity*, perhitungan tingkat kemiripan ini dibuat dengan berdasarkan besar sudut *cosinus* antara tiap dokumen terhadap *query*, iniah yang digunakan untuk merangkingkan dokumen.

2.4.1 *Term Weighting*

Berikut ini adalah metode pembobotan yang digunakan oleh Kewen dkk, 2016 yaitu, perkembangan pembobotan yang terbaru:

- *Term frequency* (TF)

  Term frequency merupakan metode paling sederhana dalam membobotkan setiap *term*. Setiap *term* diasumsikan memiliki kepentingan yang proposional terhadap jumlah kemunculan *term* pada dokumen. Bobot dari *term* t pada dokumen d, yaitu:

  $$TF\ (t_i\ ,d_j) = f\ (t_i\ ,d_j) \qquad (2.1)$$

  dimana $f(t_i\ ,d_j)$ adalah frekuensi kemunculan *term* t_i pada dokumen d_j.

- *Inverse Gravity Moment* (IGM)

  Metode pembobotan *inverse gravity moment* (IGM) adalah model untuk mengukur kekuatan membedakan kelas *term* untuk mengajukan *term* baru. Untuk mengukur konsentrasi distribusi antar kelas, perlu dipilah semua frekuensi t_i pada masing-masing kelas yang diurutkan. Daftar hasil urutan adalah $f_{i1} \geq f_{i2} \geq \ldots \geq f_{im}$, dimana f_{ir} (r=1, 2, 3, ..., m) adalah frekuensi t_i

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

dari kelas *r* yang telah diurutkan, dan *m* adalah jumlah kelas. Jika f_{ir} dianggap sebagai kelas khusus *gravity*, maka untuk daftar urutannya adalah dari kiri ke kanan dengan angka yang paling besar bagian kiri. Pusat *gravity* antar kelas itu cenderung ke kiri. Untuk kelas khusus "*gravity*" f_{ir}, jika peringkat *r* yang dianggap sebagai jarak ke asal 0, produk dari f_{ir}• *r* disebut "*gravity moment*" (GM) dalam fisika. Untuk total frekuensi *term* dalam korpus, lebih terkonsentrasi distribusi jangka antar kelas, semakin pendek jarak pusat *gravity* adalah asal dan kurang jumlah dari semua *gravity moment* kelas khusus. Jadi, model statistik baru yang disebut "*inverse gravity moment*" (IGM) untuk mengukur memastikan konsentrasi distribusi antar kelas *term*, yang didefinisikan sebagai berikut:

$$igm(t_k) = \frac{f_{i1}}{\sum_{f-1}^{m} f_{ir} \cdot r} \tag{2.2}$$

Pembobotan *term* dalam dokumen harus ditentukan oleh pentingnya dalam dokumen dan kontribusinya terhadap teks klasifikasi, yang sesuai masing-masing dengan faktor bobot lokal dan global dalam pembobotan *term*. kontribusi Sebuah *term* untuk klasifikasi teks tergantung pada daya pembeda kelasnya yang tercermin konsentrasi distribusi antar kelasnya. Semakin tinggi tingkat konsentrasi, bobot yang lebih besar harus diserahkan kepada *term*. Oleh karena itu, bukan faktor pembobotan lokal IDF, namun faktor global yang baru dalam pembobotan *term* didefinisikan berdasarkan IGM metrik *term*, seperti berikut:

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

$$W_g(t_i) = 1 + \lambda \, . \, igm(t_i) \quad (2.3)$$

dimana $W_g(t_i)$ menunjukkan faktor pembobotan global yang berbasis IGM *term* t_k, dan λ adalah koefisien disesuaikan. Tujuan memperkenalkan λ koefisien adalah untuk menjaga keseimbangan relatif antara faktor-faktor global dan lokal dalam bobot *term.* Umumnya, koefisien diatur secara empiris rentang 0.5 sampai 0.9, nilai default λ yang biasanya sesuai untuk dataset yang berbeda dalam tiap percobaan adalah 0.7. Namun, bisa saja nilai optimal λ dapat ditemukan melalui tahap pengujian untuk dataset atau klasifikasi teks tertentu (Kewen et al, 2016).

- TF.IGM

  Pembobotan TF.IGM dalam dokumen *d* adalah hasil dari gabungan dari pembobotan lokal TF yang dikalikan dengan pembobotan globlal *inverse gravity moment* (IGM). Dinyatakan sebagai berikut:

$$W\ (t_k, d) = TF\ (t_i\ , d_j)\ \ (igm(t_k))$$

atau

$$W\ (t_k, d) = TF\ (ti\ , dj)\ \left(\frac{f_{i1}}{\sum_{f-1}^{m} f_{ir}.r}\right) \quad (2.4a)$$

Persamaan 2.4a menunjukkan perhitungan bobot TF.IGM tanpa koefisien penyeimbang, sedangkan **persamaan 2.4b** berikut ini menunjukkan perhitungan bobot TF.IGM dengan koefisien penyeimbang (λ).

$$W\ (t_k, d) = TF\ (t_i\ , d_j)\ \ \left(1\ +\ \lambda\ .\ igm(t_k)\right)$$

atau

$$W\ (t_k, d) = TF\ (ti\ , dj)\ \ \left(1\ +\ \lambda\ .\frac{f_{k1}}{\sum_{f-1}^{m} f_{kr}.r}\right) \quad (2.4b)$$

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

2.4.2 *Vector Space Model* (VSM)

Vector Space Model adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dan *query* dengan mewakili setiap dokumen dalam sebuah koleksi sebagai titik ruang (vektor dalam ruangan vektor) (Turney, 2010). Misalkan terdapat sejumlah *n* kata yang berbeda sebagai kamus kata (*vocabulary*) atau indeks kata (*term index*). Kata-kata ini akan membentuk ruang vektor yang memiliki dimensi sebesar *n*. Setiap kata *i* dalam dokumen atau *query* diberikan bobot sebesar w_i. Baik dokumen maupun *query* direpresentasikan sebagai vektor berdimensi *n* (Mandala, 2006).

Sebagai contoh terdapat 3 buah kata (T_1, T_2 dan T_3), 2 buah dokumen (D_1 dan D_2) serta sebuah *query* Q. Masing-masing bernilai :

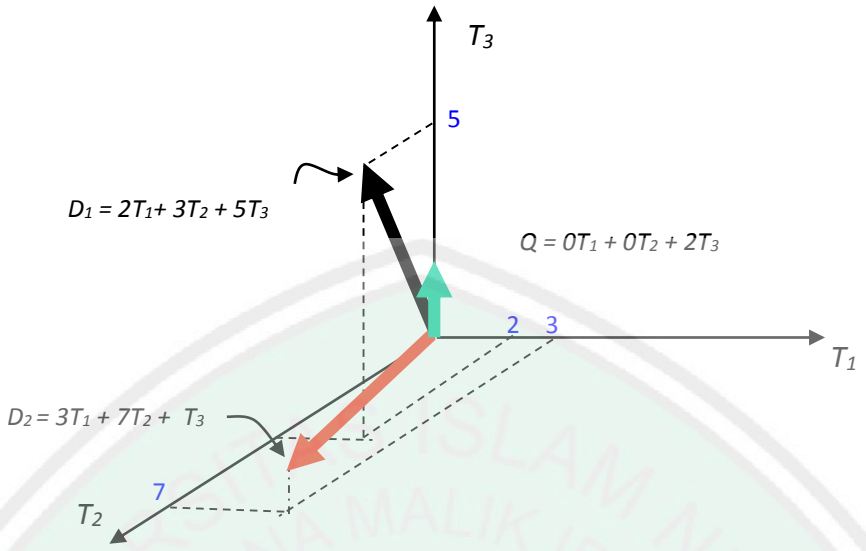
$$D_1 = 2T_1 + 3T_2 + 5T_3$$

$$D_2 = 3T_1 + 7T_2 + 0T_3$$

$$Q = 0T_1 + 0T_2 + 2T_3$$

Maka dapat direpresentasikan grafis dari ketiga vektor ini pada **gambar 2.4 Representasi dokumen dan vektor pada ruang vektor**.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



Gambar 2.3 Representasi dokumen dan vektor pada ruang vektor (Mandala, 2006)

Koleksi dokumen direpresentasi pula dalam ruang vektor sebagai matriks kata-dokumen (*terms-documents matrix*). Nilai dari elemen matriks w_{ij} adalah bobot kata *i* dalam dokumen *j*. Permasalahan pembobotan kata (*term weighting*) dalam dokumen akan dibahas pada bagian selanjutnya. Misalkan terdapat sekumpulan kata *T* sejumlah *n*, yaitu $T = (T_1, T_2, \ldots, T_n)$ dan sekumpulan dokumen *D* sejumlah *m*, yaitu $D = (D_1, D_2, \ldots, D_m)$ serta w_{ij} adalah bobot kata *i* pada dokumen *j*. Maka **gambar 2.5 Representasi matriks kata-dokumen**.

$$\begin{array}{c} \\ D_1 \\ D_2 \\ \vdots \\ \vdots \\ D_m \end{array} \begin{array}{c} \begin{array}{cccc} T_1 & T_2 & \ldots. & T_n \end{array} \\ \begin{pmatrix} w_{11} & w_{21} & \ldots & w_{n1} \\ w_{12} & w_{22} & \ldots & w_{n2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ w_{1m} & w_{2m} & \ldots & w_{nm} \end{pmatrix} \end{array}$$

Gambar 2.4 Representasi matriks kata-dokumen (Mandala, 2006)

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Penentuan relevansi dokumen dengan *query* dipandang sebagai pengukuran kesamaan (*similarity measure*) antara vektor dokumen dengan vektor *query*. Semakin "sama" suatu vektor dokumen dengan vektor *query* maka dokumen dapat dipandang semakin relevan dengan *query*. Salah satu pengukuran kesesuaian yang baik adalah dengan memperhatikan perbedaan arah (*direction difference*) dari kedua vektor tersebut. Perbedaan arah kedua vektor dalam geometri dapat dianggap sebagai sudut yang terbentuk oleh kedua vektor.

2.4.3 *Cosine Similarity*

Cosine Similarity adalah metode untuk menghitung similaritas antara dua dokumen. Penentuan kesesuaian dokumen dengan *query* dipandang sebagai pengukuran (*similarity measure*) antara vektor dokumen (D) dengan vector *query* (Q). Semakin sama suatu vektor dokumen dengan *vector query* maka dokumen dapat dipandang semakin sesuai dengan *query*. Berikut rumus untuk menghitung *cosine similarity:*

$$\cos(q, dj) = \frac{\sum_{ti}[\text{TF.IGM}(t_i,q)].[\text{TF.IGM}(t_i,d_j)]}{\sqrt{\sum|\text{TF.IGM}_q|^2}.\sqrt{\sum|\text{TF.IGM}d_j|^2}} \qquad (2.5)$$

dimana $cos(q, dj)$ adalah nilai *cosinus* antara *query* dan dokumen j. $TF.IGM(t_k, q)$ dan $TF.IGM(t_k, d_j)$ masing-masing menunjukkan pembobotan TF.IGM kata t_k pada *query* dan pembobotan TF.IGM kata t_k pada dokumen j, sedangkan $|TF.IGM_q|$ dan $|TF.IGMd_j|$ masing-masing merupakan panjang dari vektor *query* q dan dan panjang dari dokumen j (Mandala, 2006).

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

2.4.4 Perhitungan Mean Average *Precision* (MAP)

Pada penelitian ini evaluasi yang digunakan adalah *Mean Average Precison* (MAP). MAP memberikan sebuah nilai tunggal terhadap seluruh titik *recall* dari seluruh pengukuran. MAP telah terbukti memiliki diskriminasi dan stabilitas yang sangat baik. Untuk satu kebutuhan informasi, *Average Precision* adalah rata-rata nilai presisi yang diperoleh untuk kumpulan dokumen *k* teratas yang ada setelah setiap dokumen yang relevan diambil, dan nilai ini kemudian dirata-ratakan berdasarkan kebutuhan informasi.

Nilai *precision* untuk *average precision* dihitung dengan memperhatikan urutan dokumen yang diberikan oleh sistem, sehingga nilai *precision* diberikan untuk setiap dokumen yang dihasilkan oleh sistem. **Persamaan 2.6** merupakan persamaan untuk menghitung nilai *mean average precision* (Parwita, 2015).

$$MAP\ (Q) = \frac{1}{|Q|}\sum_{j=1}^{|Q|}\frac{1}{m}\sum_{k=1}^{mj} Precision\ (Rjk) \qquad (2.6)$$

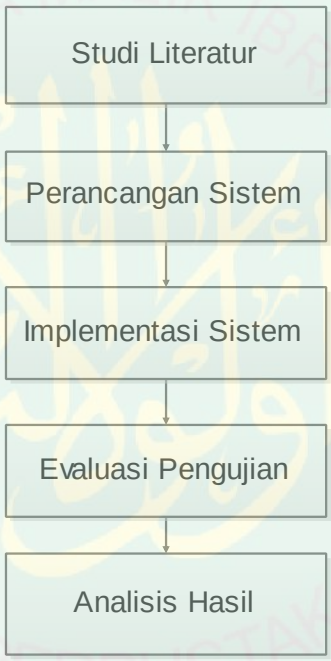
Dimana Q adalah banyaknya *query* atau kumpulan *query* yang di inputkan. R adalah item relevan yang dihasilkan oleh sistem. m adalah jumlah item relevan yang dihasilkan dari *query*. Nilai MAP antara 0 hingga 1. Dan dikatakan baik jika suatu sistem menghasilkan nilai mendekati 1.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB 3

METODOLOGI PENELITIAN

Adapun tahapan-tahapan yag akan dilakukan pada penelitian ini meliputi (1) Studi Literatur, (2) Perancangan Sistem, (3) Implementasi Sistem, (4) Pengujian dan Evaluasi, serta (5) Analisis Hasil. Alur tahapan-tahapan tersebut dapat diliht pada **gambar 3.1 Alur Metodologi Penelitian**.



Gambar 3.1 Alur Metodologi Penelitian

3.1 Studi Literatur

Studi literatur dilakukan untuk mendapatkan informasi yang berkaitan dengan lingkup pembahasan dan penelitian, perkembangan keilmuan terkait, serta metode yang telah ada sebelumnya. Studi literatur yang dilakukan diharapkan dapat memberikan data, informasi, dan fakta mengenai perangkingan dokumen

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

berbahasa Arab. Studi literatur yang dilakukan mencakup pencarian dan mempelajari referensi-referensi yang terkait, seperti:

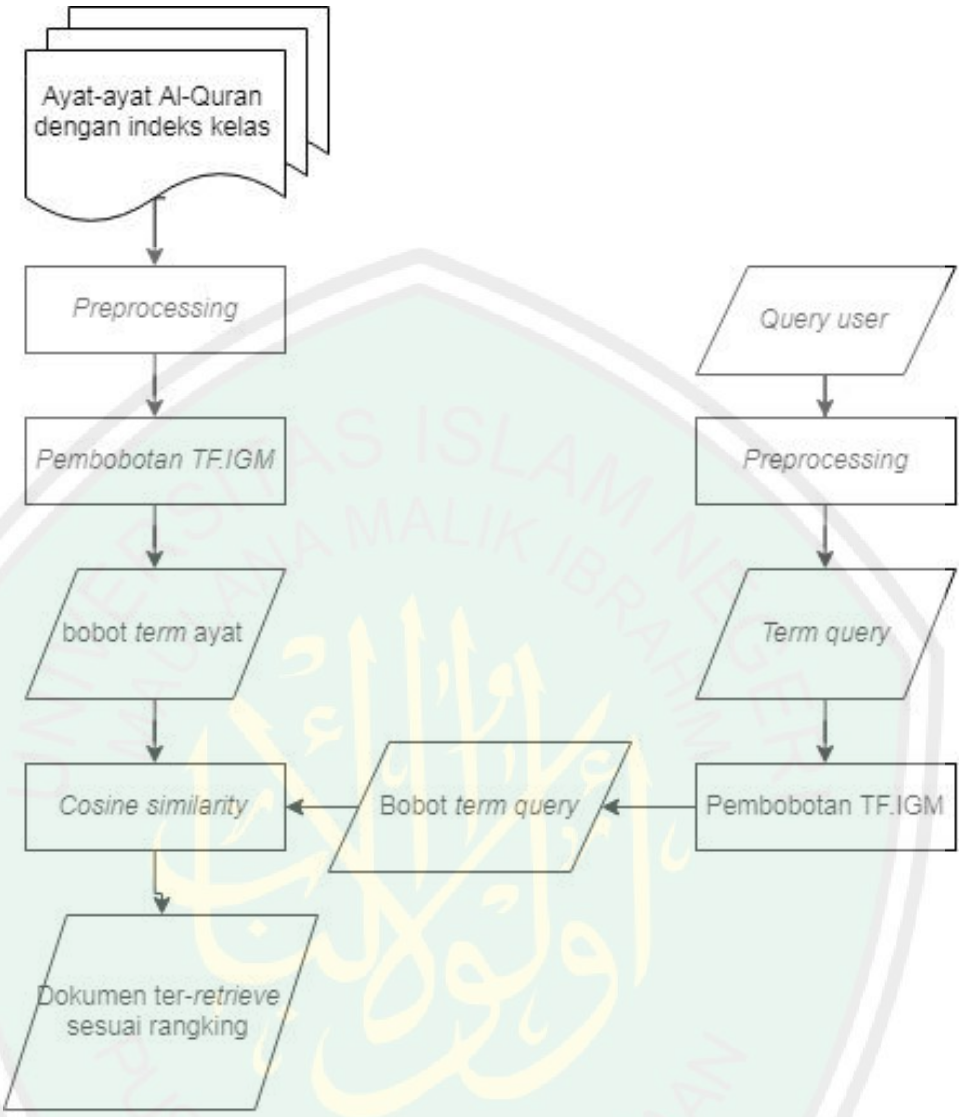
1. *Prepocessing* teks berbahasa Indonesia, yang meliputi:
    a. *Tokenizing*
    b. *Normalizing and filtering*
    c. *Stopword removal*
    d. *Stemming*
2. Metode *term weighting* TF.IGM
3. Metode pengukuran kemiripan dokumen menggunakan *cosine similarity.*

Evalusi hasil perangkingan dokumen dengan menghitung *Mean Average Precision* (MAP).

3.2 Perancangan dan Implementasi Sistem

Proses penelitian yang akan dilakukan terdiri dari beberapa tahapan seperti pada **Gambar 3.2 Blok Diagram Rancangan Sistem**. Ayat-ayat Al-Quran beserta indeks kelas sebagai dataset melalui tahapan awal yaitu *preprocessing*. Pada tahap *preprocessing* dokumen akan dipotong menjadi potongan perkata yang disebut *tokenizing,* dilanjutkan dengan tahap menormalkan dan memfilter dengan menghilangkan harakatnya yang disebut *normalizing* dan *filtering,* kemudian menghilangkan kata yang *term*asuk dalam daftar *stopword* atau yang disebut *stopword removal,* dan akan dirubah menjadi akar kata atau yang disebut *stemming.*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



Gambar 3.2 Blok Diagram Rancangan Sistem

Setelah tahap *stemming*, dilakukan pembobotan kata terhadap *term* dari hasil tahap *preprocessing*, dimulai dari perhitungan TF (*term frequency*) menggunakan **persamaan 2.1** yang dikalikan dengan perhitungan IGM (*inverse gravity moment*) menggunakan **persamaan 2.2**. Perkalian dari bobot TF dan IGM menggunakan **persamaan 2.4**.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Query user juga melalui tahap *preprocessing*, dan pembobotan. Selanjutnya hasil pembobotan dari dokumen dan *query*, masing-masing menghasilkan bobot *term.* Bobot tiap *term* baik dari dataset maupun *query* dimodelkan dalam *vector space model* hingga dilakukan perhitungan jarak cosinus antar sumbu bobot *term* mengguakan ukuran kemiripan *cosine similarity* menggunakan **persamaan 2.5**. Pengukuran kemiripan dilakukan untuk menentukan nilai kesamaan antara dokumen yang dihasilkan pemrosesan dokumen dengan *query*. Dari pengukuran kemiripan akan menghasilkan dokumen yang terangking. Dari hasil dokumen yang terangking akan dilakukan evaluasi yang menggunakan metode evaluasi *Mean Average Precision* (MAP).

3.2.1 Dataset

Data yang digunakan pada penelitian ini adalah dokumen Al-Quran berbahasa Arab yang berisi surat, ayat dan teks isi. Dokumen Al-Quran diambil dari (http://www.qurandatabase.org/) dengan rincian pada **tabel 3.1**. Dokumen yang dimaksudkan adalah ayat-ayat Al-Quran.

Tabel 3.1 Daftar Dokumen Al-Quran

ID	Surat	Ayat	Text
1	1	1	بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ
2	1	2	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ
3	1	3	الرَّحْمَٰنِ الرَّحِيمِ
4	1	4	مَالِكِ يَوْمِ الدِّينِ
5	1	5	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ
6	1	6	اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ
7	1	7	صِرَاطَ الَّذِينَ أَنْعَمْتَ عَلَيْهِمْ غَيْرِ الْمَغْضُوبِ عَلَيْهِمْ وَلَا الضَّالِّينَ
8	2	1	بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ الم
9	2	2	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ ۛ فِيهِ ۛ هُدًى لِّلْمُتَّقِينَ
10	2	3	الَّذِينَ يُؤْمِنُونَ بِالْغَيْبِ وَيُقِيمُونَ الصَّلَاةَ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Adapun data klasifikasi kelas berasal dari indeks Al-Quran pada kitab *Fahturrahman* oleh Syaikh Ilmi Zadeh Fu'ad Abd Al Baqiy yang berisi topik-topik sebagai berikut:

1. Iman
2. Ilmu
3. Bangsa terdahulu
4. Sejarah
5. Al-Quran
6. Akhlaq dan adab
7. Ibadah
8. Makanan dan minuman
9. Pakaian dan perhiasan
10. Hukum privat
11. Muamalat
12. Peradilan dan hakim
13. Hukum pidana,Jinayah
14. Jihad

Data yang digunakan dalam penelitian ini disimpan pada MySQL yang terdiri dari beberapa tabel, yaitu tabel *document* yang berisi ayat-ayat Al-Quran sebanyak 6326 ayat, tabel *class* yang berisi indeks Al-Quran yang telah disebutkan diatas sebanyak 14 indeks, dan tabel surat yang berisi surat-surat dalam Al-Quran sebanyak 114 surat. Proses koneksi ke *database* dilakukan melalui perantara *class* koneksi. *Instance* darikelas Koneksi, digunakan dalam

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

form utama dan dirujuk oleh beberapa kelas yang memerlukan fitur manipulasi seperti fungsi *select, insert, update,* ataupun *delete*. Proses koneksi MySQL dilakukan melalui fungsi-fungsi *library* API JDBC. Pada penelitian ini, *library* JDBC-MySQL yang digunakan yaitu *library* mysql-connector-java-5.1.0-bin.jar.

3.2.2 *Preprocessing*

Semua dokumen akan melalui tahap *preprocessing* yang terdiri dari beberapa tahapan, yang diantaranya adalah *tokenizing, normalizing, filtering, stopword removal* dan *stemming.* Sebagai contoh surat Al-Ikhlas (1-4) berikut akan melalui tahapan-tahapannya:

قُلْ هُوَ ٱللَّهُ أَحَدٌ ﴿١﴾ ٱللَّهُ ٱلصَّمَدُ ﴿٢﴾ لَمْ يَلِدْ وَلَمْ يُولَدْ ﴿٣﴾ وَلَمْ يَكُن لَّهُ كُفُوًا أَحَدٌۢ ﴿٤

Artinya: "*Katakanlah, Dialah Allah Yang Maha Esa (1) Allah tempat bergantung bagi segala sesuatu (2) Dia tidak beranak dan diperanakkan (3) Dan tidak ada suatupun yang menyamai-Nya*"

1. *Tokenizing* dilakukan untuk memisahkan keseluruhan isi dokumen menjadi kata satu tunggal. Proses *tokenizing* dilakukan untuk menghapus karakter tanda baca dan angka, kemudian dilanjutkan dengan pemotongan kata. Dari surat Al-Ikhlas diatas berikut adalah hasil dari *tokenizing* dan **Gambar 3.3** adalah *code* prosesnya.

(*Ahad*) أَحَد - (*Allah*) ٱللَّهُ - (*Huwa*) هُو - (*Qul*) قُلْ

(*Husshomad*) ٱلصَّمَدُ - (*Allah*) ٱللَّهُ

(*Yuulad*) يُولَدْ - (*Walam*) وَلَمْ - (*Yalid*) يَلِد - (*Lam*) لَمْ

(*Ahad*) أَحَد - (*Kufuwan*) كُفُوًا - (*Lahu*) لَّهُ - (*Yakun*) يَكُن - (*Walam*) وَلَمْ

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

```
[1]  public static List<String> parseKeywords(Analyzer analyzer, String field, String keywords,
       List<String> stopword) throws IOException {
[2]    List<String> result = new ArrayList<>();
[3]    TokenStream stream  = analyzer.tokenStream(field, new StringReader(keywords));
[4]    stream.reset();
[5]    try {
[6]        while(stream.incrementToken()) {
[7]           String  temp=stream.getAttribute(CharTermAttribute.class).toString();
[8]            temp=temp.replace("'", "");
[9]            temp=temp.replace("\"", "");
[10]           temp=temp.replace(" ", "");
[11]           temp=temp.replaceAll("[0-9]","");
[12]           temp=temp.trim();
[13]       }
[14]   }
[15] }
```

Gambar 3.3 *Code* Proses *Tokenizing*

2. Pada tahapan *normalizing* dan *filtering* dilakukan untuk menormalkan teks ke dalam bentuk dasar tulisan bahasa Arab (*hijaiyah*) dan menghilangkan karakter-karakter yang bukan termasuk huruf *hijaiyah*. Kemudian dilakukan penghapusan *harakat*. **Gambar 3.4** adalah *code* proses *filtering*.

```
[1]  private boolean removeDiacritics ( String currentWord, StringBuffer modifiedWord )  {
[2]      boolean diacriticFound = false;
[3]      modifiedWord.setLength ( 0 );
[4] Vector diacritics = ( Vector ) staticFiles.elementAt ( 17 );
[5]      for ( int i = 0; i < currentWord.length ( ); i++ )
[6]          if ( ! ( diacritics.contains ( currentWord.substring ( i, i+1 ) ) ) )
[7] modifiedWord.append ( currentWord.substring ( i, i+1 ) );
[8]          else{
[9]              diacriticFound = true;
[10]         }
[11]     return diacriticFound;
[12] }
```

Gambar 3.4 *Code* Proses *Normalizing* dan *Filtering*

Dan berikut hasil dari proses *normalizing* dan *filtering*:

قل – هو – الله – احد

الله – الصمد

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

لم – يلد – لم – يولد

لم – يكن – له – كفوا – احد

3. Pada tahap *stopword removal* dilakukan untuk menghapus *term* yang muncul dalam dokumen yang tidak memiliki kontribusi arti pada dokumen tersebut. Mekanisme penghapusan yaitu dengan cara apakah *term* termasuk *sopword*, jika *term* tidak termasuk *stopword* maka *term* akan disimpan, begitupun sebaliknya jika *term* termasuk dalam *list stopword* maka *term* akan dihapus. Daftar *stoplist* yang dihilangkan seperti pada daftar kata pada (*http://www.ranks.nl/stopwords/arabic*), seperti kata fungsi dalam bahasa Arab, kata ganti dan preposisi yang akan dihapus. Berikut kata yang terdapat pada *stoplist*:

الله – الصمد - قل – **هو** – **الله** – **احد**

لم – **يكن** – **له** – كفوا – **احد** - **لم** – يلد – **لم** – يولد

Kata diatas yang bergaris bawah adalah daftar kata yang harus dihilangkan. Setelah dihilangkan akan tersisa kata yang memiliki pengaruh dalam dokumen. Dan hasilnya sebagai berikut:

قل – الصمد – يلد – يولد – كفوا

Gambar 3.5 adalah *code* penghapusan *term* yang termasuk dalam *stoplist*, dan hasil dari *stopword* hanya akan menyisakan *term* yang memiliki pengaruh dalam suatu dokumen.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

```
[1]  if(!stopword.contains(StringUtils.lowerCase(temp)) && temp.compareTo("")!=0) {
[2]          result.add(temp);
[3]  }
```

Gambar 3.5 *Code* Proses *Stopword Removal*

4. Pada tahap *stemming* dilakukan untuk mengubah *term* menjadi bentuk kata dasar. Kata yang tidak termasuk dalam *stoplist* akan diubah kedalam bentuk kata dasar, seperti yang dijelaskan pada **tabel 3.2**:

Tabel 3.2 *Term* Hasil *Stemming*

Kata		Kata dasar
قل	Menjadi	قال
الصمد	Menjadi	صمد
يلد	Menjadi	ولد
يولد	Menjadi	ولد
كفوا	Menjadi	كفا

Jadi, hasil dari seluruh tahapan *preprocessig* dari dokumen Al-Quran akan tersimpan dalam *database* seperti yang ditunjukkan pada **gambar 3.6 Tabel hasil *preprocessing* pada *database***. Dari 6326 dokumen yang digunakan dalam penelitian ini, diperoleh 3013 *term* yang tersimpan.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

id_doc	id_class	id_surat	ayat	term
1	1	1	1	بسم
1	1	1	1	رحم
1	1	1	1	رحم
2	7	1	2	حمد
2	7	1	2	ربب
2	7	1	2	علم
2	7	1	2	لبل
3	1	1	3	رحم
3	1	1	3	رحم
4	1	1	4	ملك
5	1	1	5	اي
5	1	1	5	اي
5	1	1	5	عبد
5	1	1	5	نستعين
6	7	1	6	صرط
6	7	1	6	قوم
6	7	1	6	هدد
7	5	1	7	صرط
7	5	1	7	ضالين
7	5	1	7	علي
7	5	1	7	علي
7	5	1	7	غضب
7	5	1	7	غور
7	5	1	7	ولي
8	1	2	1	ألم

Gambar 3.6 Tabel hasil *preprocessing* pada *database*

Mekanisme pada proses *stemming* salah satunya adalah penghapusan *prefix*. *Code* penghapusan *prefix* dikerjakan oleh method *checkForPrefixes()*. Dan akan dilakukan juga perintah penghapusan ***suffix*** yang dikerjakan oleh method *checkForSuffixes()* seperti yang ditunjukkan pada **gambar 3.7**. Jika akar kata telah ditemukan dan bukan termasuk dalam daftar *stopword* maka nilai akan dikembalikan ke *modifiedWord*.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

```
[1]  private String checkForSuffixes ( String word ) {
[2]      String suffix = "";
[3]      String modifiedWord = word;
[4]      Vector suffixes = (Vector) staticFiles.elementAt (14);
[5]      fromSuffixes = true;
[6]      for ( int i = 0; i < suffixes.size ( ); i++ ) {
[7]         suffix = ( String ) suffixes.elementAt ( i );
[8]         if( suffix.regionMatches ( 0, modifiedWord, modifiedWord.length ( ) – suffix.length ( ),
                suffix.length ( ) ) ) {
[9]         modifiedWord = modifiedWord.substring ( 0, modifiedWord.length ( ) - suffix.length ( )
                );
[10]        if ( checkStopwords ( modifiedWord ) ) {
[11]              fromSuffixes = false;
[12]              return modifiedWord;
[13]            }
[14]            if ( modifiedWord.length ( ) == 2 ) {
[15]             modifiedWord = isTwoLetters ( modifiedWord );
[16]            }
[17]            else if ( modifiedWord.length ( ) == 3 ) {
[18]             modifiedWord = isThreeLetters (modifiedWord);
[19]            }
[20]            else if ( modifiedWord.length ( ) == 4 ) {
[21]               isFourLetters ( modifiedWord );
[22]           }
[23]          if ( !rootFound && modifiedWord.length( ) > 2 ) {
[24]            modifiedWord = checkPatterns( modifiedWord );
[25]            }
[26]            if ( stopwordFound ) {
[27]               fromSuffixes = false;
[28]               return modifiedWord;
[29]            }
[30]            if ( rootFound )
[31]            {
[32]               fromSuffixes = false;
[33]               return modifiedWord;
[34]            }
[35]         }
[36]     }
[37]     fromSuffixes = false;
[38]     return word;
[39]  }
```

Gambar 3. 7 Code Proses Penghapusan Suffix

3.2.3 *Term Weighting*

Pembobotan kata adalah tahap selanjutnya setelah dilakukan *preprocessing*. Pembobotan kata dilakukan dengan cara menghitung TF (*Term Frequency*) dan IGM (*Inverse Gravity Moment*) dari masing-masing *term* dari keseluruhan dokumen. Setelah diperoleh nilai TF dan IGM, maka akan dikalikan setiap bobot katanya. Perhitungan TF dilakukan dengan menghitung jumlah kemunculan *term* pada masing-masing dokumen. Implementasinya dilakukan

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

dengan kode SQL *count(term)* sebagai TF yang dikelompokkan berdasarkan *id_doc*, seperti yang ditunjukkan pada **Gambar 3.8**.

```
CREATE TABLE weight_tf (select id_doc, term, count (term) as tf FROM term GROUP BY id_doc, term);
```

Gambar 3.8 *Query* SQL Pembobotan TF

Dan hasil dari pembobotan TF ditunjukkan pada **gambar 3.9 Tabel hasil pembobotan TF**, seperti berikut:

id_doc	term	tf
1	بسم	1
1	رحم	2
2	حمد	1
2	ربب	1
2	علم	1
2	ليل	1
3	رحم	2
4	ملك	1
5	اي	2
5	عبد	1
5	نستعين	1
6	صرط	1
6	قوم	1
6	هدد	1
7	صرط	1
7	ضالين	1
7	علي	2
7	غضب	1
7	غور	1
7	ولي	1

Gambar 3.9 Tabel hasil pembobotan TF

Perhitungan bobot IGM dilakukan dengan dua tahapan sebelum dihitung sesuai **persamaan 2.2**. Pertama, pembuatan indeks. Jika pembobotan TF *term* dikelompokkan berdasarkan kemunculannya pada dokumen, maka pembobotan IGM *term* akan dikelompokkan berdasarkan kemunculannya pada kelas, yang dimana sebelumnya telah dijelaskan bahwa dokumen Al-Quran diklasifikasikan 14 kelas. **Gambar 3.10** SQL *code* proses pengindeksan *term*.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

```
SELECT term, id_class, COUNT(*) as CF FROM term GROUP BY term, id_class ORDER by term ASC,
        id_class ASC
```

Gambar 3.10 SQL *Code* pengindeksan *term*

Kedua, pengurutan sesuai jumlah kemunculan *term* pada masing-masing kelas berdasarkan frekuensinya. Setelah pengurutan kelas akan dilanjutkan menghitung nilai IGM sesuai **persamaan 2.2,** yaitu dimana frekuensi atau jumlah kemunculan *term* terbanyak atau dapat disebut dengan fi_1 (frekuensi yang mendapat rangking 1) ditunjukkan pada kode program $max_class dibagi dengan penjumlahan dari tiap frekuensi dikalikan dengan urutan/rangkingnya ditunjukkan pada kode program $sum_class. Untuk perhitungannya ditunjukkan pada kode program $nilai_igm = $max_class / $sum_class;. Lebih jelasnya ditunjukkan pada **Gambar 3.11** ditunjukkan pada baris kode ke-9 sampai baris kode ke-14.

```
[1]     $i = 1;
[2]     $max_class = 0;
[3]     $sum_class = 0;
[4]     foreach ($row as $key => $value) {
[5]         $nilai_igm = 0;
[6]         if ($i == 1) {
[7]             $max_class = $row[$key];
[8]         }
[9]         $sum_class += ($row[$key]*$i); // C(*) x R_C(*)
[10]        if ($i == 14 ) {
[11]            if ($sum_class == 0 ) {
[12]                $nilai_igm = 0;
[13]            } else {
[14]                $nilai_igm = $max_class / $sum_class;
[15]            }
[16]            array_push($row, $nilai_igm);
[17]            $max_class = 0;
[18]            $sum_class = 0;
[19]        }
[20]        $row[$key]= $i;
[21]        $i++;
[22]    }
[23]    print_r($row); // kelas yang sudah terurut dan nilai igm nya
[24]    echo "<hr>";
```

Gambar 3.11 *Code* proses *Sorting Class*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Hasil dari proses *sorting class* ditunjukkan pada **gambar 3.12 Tabel hasil pembobotan IGM**. Dan sesuai **persamaaan 2.4**, setelah perhitungan bobot IGM maka akan dihitung perkalian dari nilai bobot TF dan bobot IGM. **Gambar 3.13** adalah SQL *code* perkalian bobot TF dengan bobot IGM, serta **gambar 3.14** adalah gambar table hasil perkalian TF dan IGM.

```
SELECT `weight_igm`.`igm`, `weight_tf_igm`.`tf` FROM `weight_igm`, `weight_tf_igm` WHERE
        `weight_tf_igm`.`term` = `weight_igm`.`term`'
```

Gambar 3.13 SQL *Code* Perkalian Bobot TF dan Bobot IGM

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

id_igm	term	c1	r_c1	c2	r_c2	c3	r_c3	c4	r_c4	c5	r_c5	c6	r_c6	c7	r_c7	c8	r_c8	c9	r_c9	c10	r_c10	c11	r_c11	c12	r_c12	c13	r_c13	c14	r_c14	igm
1	أل	1	2	0	3	3	1	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	0.6
2	أيق	0	2	0	3	1	1	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	1
3	أبن	4	3	0	9	18	1	0	10	3	4	5	2	3	5	0	11	0	12	2	8	3	6	0	13	0	14	3	7	0.14754098360656
4	أبه	1	2	1	3	6	1	0	7	0	8	0	9	1	4	0	10	0	11	1	5	0	12	1	6	0	13	0	14	0.23076923076923
5	أبى	2	3	0	9	4	1	2	4	4	2	2	5	0	10	0	11	0	12	1	6	1	7	0	13	0	14	1	8	0.070175438596491
6	أتد	2	1	0	3	1	2	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	0.5
7	أتن	16	1	2	5	15	2	1	8	9	3	2	6	8	4	0	10	0	11	2	7	1	9	0	12	0	13	0	14	0.10126582278481
8	أتى	32	1	1	9	12	2	2	7	3	4	8	3	1	10	0	12	0	13	2	8	3	5	1	11	0	14	3	6	0.17297297297297
9	أثث	0	3	0	4	1	1	0	5	1	2	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	0.33333333333333
10	أثر	0	6	0	7	1	2	1	3	0	8	4	1	1	4	0	9	0	10	0	11	0	12	0	13	0	14	1	5	0.22222222222222
11	أثل	1	1	0	2	0	3	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	1
12	أثم	5	1	1	4	0	7	0	8	0	9	1	5	0	10	1	6	0	11	0	12	2	2	0	13	2	3	0	14	0.16666666666667
13	أجب	2	1	0	2	0	3	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	0	14	1
14	أجج	3	1	0	5	1	3	0	6	0	7	0	8	0	9	1	4	0	10	0	11	2	2	0	12	0	13	0	14	0.21428571428571
15	أجر	30	1	0	10	6	3	1	8	3	5	8	2	1	9	0	11	0	12	2	6	4	4	0	13	0	14	2	7	0.21739130434783
16	أجن	0	2	0	3	0	4	0	5	0	6	0	7	0	8	0	9	0	10	0	11	0	12	0	13	1	1	0	14	1
17	أحد	21	1	1	9	6	2	0	12	4	4	5	3	1	10	1	11	0	13	4	5	3	6	2	7	2	8	0	14	0.12962962962963
18	أخا	0	7	0	8	4	1	0	9	1	4	2	2	1	5	0	10	0	11	1	6	0	12	2	3	0	13	0	14	0.13793103448276
19	أخت	1	2	0	6	1	3	0	7	0	8	1	4	0	9	0	10	0	11	3	1	0	12	0	13	1	5	0	14	0.17647058823529
20	أخر	76	1	2	9	24	2	7	5	5	7	19	3	12	4	0	13	0	14	1	10	1	11	3	8	1	12	7	6	0.18269230769231

Gambar 3.12 Tabel hasil pembobotan IGM

LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF

id	term	tf	tf_igm
1	آل	4	2.4
2	أبق	1	1
3	أبن	41	6.049180327869
4	أبه	11	2.5384615384615
5	أبى	17	1.1929824561403
6	أتد	3	1.5
7	أتن	56	5.6708860759494
8	أتى	68	11.762162162162
9	أثث	2	0.66666666666666
10	أثر	8	1.7777777777778
11	أثل	1	1
12	أثه	12	2
13	أجب	2	2
14	أجج	7	1.5
15	أجر	57	12.391304347826
16	أجن	1	1
17	أحد	50	6.4814814814815
18	أخا	11	1.5172413793104
19	أخت	7	1.235294117647
20	أخر	158	28.865384615385
21	أذن	66	7.674418604651
22	أذي	6	1.5
23	أرب	2	0.66666666666666
24	أرج	3	3
25	أرجل	1	1

Gambar 3.14 Tabel hasil perkalian bobot TF dan bobot IGM

Misal dilakukan pencarian terhadap lima dokumen seperti pada **tabel 3.3** contoh dokumen dengan keterangan berikut:

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Query : وَاعْبُدْ رَبَّكَ

Kelas : Iman

Tabel 3.3 Contoh Dokumen

Dokumen	Kelas	Isi Dokumen
D1	Iman	إِنَّ اللَّهَ رَبِّي وَرَبُّكُمْ فَاعْبُدُوهُ هَٰذَا صِرَاطٌ مُسْتَقِيمٌ
D2	Iman	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ
D3	Iman	وَاعْبُدْ رَبَّكَ حَتَّىٰ يَأْتِيَكَ الْيَقِينُ
D4	Ibadah	اهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ
D5	Ibadah	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

Contoh dokumen diatas tersebar dalam dua kelas yaitu kelas iman dan kelas ibadah yang diberi kode C1 untuk kelas iman dan C2 untuk kelas ibadah. Dimana D1 sampai dengan D3 termasuk dalam C1 dan D4 sampai D5 termasuk dalam C2. Kelima dokumen diatas akan melalui tahap *preprocessing*. Hasil dari proses *preprocessing* terdapat 11 kata dasar, yaitu:

رب – نستعين – أيك – يقن – أتي – عبد

عالم – هدد – **حمد – قوم – صرط**

Selanjutnya akan diberi bobot pada setiap *term* dengan pembobotan TF dan IGM. Hasil perhitungan pembobotan TF dan IGM terdapat pada **tabel 3.4**, kemudian pembobotan TF.IGM yaitu perkalian antara bobot TF dan IGM terdapat pada **tabel 3.6**.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 3.4 Pembobotan TF dan IGM

No	*Term*	TF					Fi				IGM fi1/(∑fir.r)
		D1	D2	D3	D4	D5	C1	r(C1)	C2	r(C2)	
1	أتي	0	0	**1**	0	0	1	1	0	2	1
2	أيك	0	**2**	0	0	0	2	1	0	2	1
3	حمد	0	0	0	0	**1**	0	2	1	1	0
4	رب	**2**	0	0	0	**1**	2	1	1	2	0.5
5	صرط	**1**	0	0	**1**	0	1	1	1	2	0.333333
6	عبد	**1**	**1**	**1**	0	0	3	1	0	2	1
7	عالم	0	0	0	0	**1**	0	2	1	1	0
8	قوم	**1**	0	0	**1**	0	1	1	1	2	0.333333
9	نستعين	0	**1**	0	0	0	1	1	0	2	1
10	هدد	0	0	0	**1**	0	0	2	1	1	0
11	يقن	0	0	**1**	0	0	1	1	0	2	1

Pada **tabel 3.4** pemberian urutan pada kelas yang memiliki kemunculan *term* lebih banyak, yang didefinisikan oleh *r*, urutan tergantung banyaknya kelas pada dataset. Seperti contoh diatas terdapat dua kelas, jadi *r* hanya terurut antara 1 dan 2 saja. Kemudian dihitung menggunakan **persamaan 2.2**. dimana fi_1 adalah frekuensi *term* pada urutan pertama atau frekuensi terbanyak dibagi dengan penjumlahan antara perkalian frekuensi kemunculan *term* pada kelas dikalikan dengan *r* (urutannya). Dilanjutkan dengan perkalian bobot setiap dokumen antara nilai TF dan nilai IGM seperti yang ditunjukkan pada **tabel 3.5** berikut:

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 3.5 Pembobotan TF.IGM

No	*Term*	TF					IGM fk1/(∑fkr.r)	TF.IGM				
		D1	D2	D3	D4	D5		D1	D2	D3	D4	D5
1	أتي	0	0	**1**	0	0	1	0	0	**1**	0	0
2	أيك	0	**2**	0	0	0	1	0	**2**	0	0	0
3	حمد	0	0	0	0	**1**	0	0	0	0	0	0
4	رب	**2**	0	0	0	**1**	0.5	**1**	0	0	0	**0.5**
5	صرط	**1**	0	0	**1**	0	0.333	**0.333**	0	0	**0.333**	0
6	عبد	**1**	**1**	**1**	0	0	1	**1**	**1**	**1**	0	0
7	عالم	0	0	0	0	**1**	0	0	0	0	0	0
8	قوم	**1**	0	0	**1**	0	0.333	**0.333**	0	0	**0.333**	0
9	نستعين	0	**1**	0	0	0	1	0	**1**	0	0	0
10	هدد	0	0	0	**1**	0	0	0	0	0	0	0
11	يقن	0	0	**1**	0	0	1	0	0	**1**	0	0

Bukan hanya dataset atau dokumen Al-Quran saja, namun juga *query* yang dimasukkan oleh pengguna akan melalui tahap yang sama yaitu *Preprocessing* dan pembobotan. **Gambar 3.15** menunjukkan proses *preprocessing* pada *query* dan **Gambar 3.16** menunjukkan proses pembobotan *query*.

```
[1] if (tx_input.getText() != null) {
[2] List result = Stemming.parseKeywords(tx_input.getText(), 1, stopword, se);
[3]             for (int j = 0; j < result.size(); j++) {
[4]                 sql = "INSERT INTO Q_term (input_term) VALUES ('" + result.get(j).toString() + "')";
[5]                 theKoneksi.executeUpdate(sql);
[6]             }
```

Gambar 3.15 *Code* prosses *preprocessing* pada *query*

```
[1]sql = "SELECT i.input_term, count(i.input_term) as tf, \n d.igm, (count(i.input_term) * d.igm) as bobot \n
        FROM Q_term i, weight_igm d \n WHERE i.input_term = d.term \n GROUP BY i.input_term";
[2]sql2 = "CREATE TABLE PROSES_BOBOT_INPUT("+sql+")";
```

Gambar 3.16 *Code* prosses pembobotan *query*

Hasil pembobotan *query* seperti pada **tabel 3.6** berikut:

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 3.6 Hasil Pembobotan *query*

No.	*Term*	TF	IGM	TF.IGM
1	عبد	1	1	1
2	رب	1	0.5	0.5

Dan selanjutnya akan dihitung kemiripan antara *query* dengan dokumen. Jumlah masing-masing *term* pada *query* dikalikan dengan bobot IGM dari *term* dokumen sehingga diperoleh TF.IGM. **Gambar 3.17** menunjukkan perkalian bobot *query* dengan bobot *term* pada seluruh dokumen.

```
"INSERT INTO proses_dot_product  "
            + "(SELECT b.id_doc, (sum(i.bobot * b.tf_igm)) "
            + " FROM proses_bobot_input as i, weight as b"
            + " WHERE i.input_term = b.term "
            + " GROUP BY b.id_doc) ";
```

Gambar 3.17 *SQL* code *perkalian bobot query dengan bobot* term *dokumen*

3.2.4 *Vector Space Model* dan *Cosine Similarity*

Setelah dilakukan dilakukan pembobotan pada dataset/dokumen dan *query*, maka tahap selanjutnya adalah perhitungan *cosine similarity*. Pada perhitungan *cosine similarity* yaitu dengan mengalikan antara vektor *query* dan vektor tiap dokumen kemudian menghitung panjang vektor. Nilai perkalian antara vektor *query* dan dokumen akan dibagi dengan panjang vektor sehingga diperoleh nilai kosinus antara vektor *query* dengan dokumen. **Gambar 3.18** menunjukkan perhitungan vektor dokumen dan **Gambar 3.19** menunjukkan perhitungan vektor *query*.

```
"CREATE TABLE vector_length (" SELECT id_doc, sqrt(sum(" + "tf_igm" + "*" + "tf_igm" + ")) as
        vector_length "
    + " FROM weight "
    + " GROUP BY id_doc)";
```

Gambar 3.18 SQL perhitungan vektor dokumen

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

```
sql = "SELECT sqrt(sum(bobot * bobot)) as vectorQ FROM proses_bobot_input ";
```

Gambar 3.19 SQL perhitungan vektor *query*

Sesuai **persamaan 2.5**, perhitungan *cosine* similarity ditunjukkan dengan **Gambar 3.20** kode SQL-nya. Dan contoh perhitungan *cosine* perhitungan *cosine similarity* TF.IGM pada **tabel 3.7**.

```
sql = "INSERT INTO similarity "(SELECT p.id_doc, (p.dot_product /(" + vectorQ + "* v.vector_length))FROM proses_dot_product p, vector_length v WHERE p.id_doc = v.id_doc order by v.id_doc)";
```

Gambar 3.20 SQL *code* perhitungan *similarity*

Tabel 3.7 Perhitungan *cosine similarity*

No.	*Term*	Q	TF.IGM				
			D1	D2	D3	D4	D5
1	أتي	0	0	0	1	0	0
2	أيك	0	0	2	0	0	0
3	حمد	0	0	0	0	0	0
4	رب	0.5	1	0	0	0	0.5
5	صرط	0	0.333	0	0	0.333	0
6	عبد	1	1	1	1	0	0
7	عالم	0	0	0	0	0	0
8	قوم	0	0.333	0	0	0.333	0
9	نستعين	0	0	1	0	0	0
10	هدد	0	0	0	0	0	0
11	يقن	0	0	0	1	0	0
Jumlah kuadrat bobot		1.25	2.222	6	3	0.222	0.25
Panjang vector		1.118	1.491	2.449	1.732	0.471	0.5
Jumlah perkalian bobot *query* dengan bobot dokumen			1.5	1	1	0	0.25
Cosine similarity			0.9	0.365	0.516	0	0.447

Panjang vektor merupakan nilai akar dari jumlah kuadrat bobot masing-masing dokumen dan *query*. Setelah itu dilakukan perhitungan jumlah perkalian

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

bobot *query* dengan setiap dokumen. Kemudian nilai dari *cosine similarity* didapatkan dari membagi jumlah perkalian bobot dengan perkalian panjang vektor dokumen dengan panjang vektor *query*. Berikut perhitungan *cosine similarity*:

- $\cos(q, d_1) = \frac{1.5}{1.491 \ X \ 1.118} = 0.9$
- $\cos(q, d_2) = \frac{1}{2.449 \ X \ 1.118} = 0.365$
- $\cos(q, d_3) = \frac{1}{1.452 \ X \ 1.118} = 0.516$
- $\cos(q, d_4) = \frac{0}{0.471 \ X \ 1.118} = 0$
- $\cos(q, d_5) = \frac{0.25}{0.5 \ X \ 1.118} = 0.447$

Pada **tabel 3.9** adalah hasil dari perhitungan *cosine similarity* yang telah diurutkan dari yang tertinggi sampai yang terendah dengan variasi pembobotan TF.IDF dan TF.IGM. Dari **tabel 3.8** juga diketahui rangking dokumen berdasarkan similaritas dokumen tersebut terhadap *query*. Hasilnya menunjukkan bahwa dengan TF.IGM dokumen D1 berada pada urutan pertama karena memiliki nilai similaritas yang tinggi. Sedangkan nilai similaritas pada D4 adalah 0, karena dalam dokumen tersebut tidak terdapat *term* yang dicari.

Tabel 3.8 Perangkingan Dokumen

Rangking	TF.IGM	
	Dokumen	**Similaritas**
1	**D1**	0.9
2	D3	0.5163978
3	D5	0.4472136
4	D2	0.3651484
5	D4	0

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

3.2.5 Output

Hasil dari penelitian ini adalah perangkingan ayat-ayat Al-Quran berdasarkan pada nilai *Cosine Similarity*, yang diperoleh dari kesesuaian antara dataset dengan *query* dari pengguna. Semakin besar nilai *Cosine Similarity* maka semakin tinggi letak rangkingnya. Kemudian yang ditampilkan kepada pengguna adalah ayat-ayat dari Al-Quran yang terangking tersebut.

3.3 Evaluasi Pengujian dan Anilisa Hasil

Pada penelitian ini, metode evaluasi perangkingan yang digunakan adalah *Mean Average Precision*, MAP memberikan sebuah nilai tunggal terhadap seluruh titik *recall* dari seluruh pengukuran. MAP telah terbukti memiliki diskriminasi dan stabilitas yang sangat baik. Untuk satu kebutuhan informasi, *Average Precision* adalah rata-rata nilai presisi yang diperoleh untuk kumpulan dokumen *k* teratas yang ada setelah setiap dokumen yang relevan diambil, dan nilai ini kemudian dirata-ratakan berdasarkan kebutuhan informasi.

Pengujian dilakukan dengan melakukan pencarian dan mengevaluasi hasil perangkingan dokumen. Pengujian akan dilakukan terhadap sejumlah *input user*, mulai dari *input* ke-1 sampai *input* ke-n. Skenario uji coba yang ditunjukkan pada **tabel 3.9** *query* yang diuji sebanyak klasifikasi kelas dalam dokumen Al-Quran yaitu 14 kelas. Jadi akan ada 14 *query* dalam setiap tema dalam Al-Quran sebagai uji cobanya.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 3.9 Skenario Uji Coba

Query	TF.IGM		
	***MP Top* 10**	***MP Top* 20**	***MAP***
Q-1			
Q-2			
Q-3			
.			
.			
Q-14			

Sebelum menghitung *MAP* akan dihitung nilai *P(k)t, MP(k)* dan *AP(Q)t* yang rumusnya dinyatakan pada **persamaan 2.6**, **persamaan 2.7** dan **persamaan 2.8**. yang kemudian akan dilanjutkan menghitung nilai *MAP* yang dinyatakan pada **persamaan 2.9**, dimana $AP(Q_t)$ adalah presisi rata-rata *query* Q_t dan *B* adalah jumlah *query* total.

3.4 Lingkungan Pengembangan Sistem

Lingkungan pengembangan sistem pada penelitian ini adalah sebagai berikut:

Spesifikasi Perangkat Keras

1. Processor: Intel(R) Core(TM) i3-2328M CPU @ 2.20GHz
2. Memory (RAM): 2.00 GB (1.84 usable)

Spesifikasi Perangkat Lunak

1. Sistem operasi: Windows 7 Pro 64-bit
2. Netbeans IDE 7.3 dengan bahasa pemrograman *Java*
3. *Sublime Text* 3 dengan bahasa pemrograman PHP
4. *Database server* dengan MySql

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Sistem yang dibangun aplikasi berbasis desktop dengan menggunakan bahasa pemrograman *Java* dan *database* MySQL

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB 4

PEMBAHASAN

Pada bagian ini akan menjelaskan mengenai implementasi dari setiap langkah yang telah di jelaskan di bab sebelumnya dan memaparkan hasil uji coba yang telah dilakukan sesuai dengan scenario pengujian yakni dengan menghitung tingkat akurasi dan presisi dengan metode yang di gunakan dalam penelitian ini. Kemudian pada bagian akhir bab akan dipaparkan evaluasi dan pembahasan dari hasil percobaan yang diperoleh. *Input* data dari penelitian ini adalah berupa artikel kesehatan, kemudian dilakukan proses *preprosesing*, pembobotan, dan perhitungan *similarity*. Dan outputnya adalah berupa perangkingan ayat Al-Quran sesuai dengan kata kunci (*query*) yang dilakukan oleh pengguna.

4.1 Implementasi

Sistem yang dibangun adalah aplikasi berbasis desktop menggunakan bahasa pemrograman Java dan MySql sebagai pengolah *database*. Terdapat dua komponen utama pengembangan sistem perangkingan dokumen berbahasa Arab sesuai dengan metode yang diimplementasikan dan pengembangan *interface* sistem sebagai sarana interaksi sistem dengan pengguna. Lingkungan pengembangan penelitian yang digunakan adalah sebagai berikut:

1. Spesifikasi perangkat lunak yang digunakan:
    a. Sistem operasi: Windows 7 Pro 64-bit
    b. Netbeans IDE 7.3 dengan bahasa pemrograman *Java*
    c. *Sublime Text* 3 dengan bahasa pemrograman PHP
    d. *Database server* dengan MySql

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

2. Spesifikasi perangkat keras yang digunakan:
    a. Processor: Intel(R) Core(TM) i3-2328M CPU @ 2.20GHz
    b. Memory (RAM): 2.00 GB (1.84 usable)

Implementasi algoritma dilakukan dengan membuat fungsi-fungsi dari tahapan yang telah dipaparkan pada bab 3. Data yang digunakan dalam uji coba merupakan kumpulan teks dokumen Al-Quran berbahasa Arab, dimana dokumen Al-Quran terdiri dari 6236 ayat. *Database* Al-Quran terdiri dari id_doc, id_class, surat, ayat, dan *text* seperti yang telah ditunjukkan sebelumnya pada **table 3.1**. Dokumen-dokumen tersebut yang diproses melalui beberapa tahap *prepocessing* yaitu *tokenizing*, *normal* dan *filtering*, *stopword removal*, dan pembentukan kata dasar yang selanjutnya diberi nilai bobot setiap *term*, pembobotannya dengan menggunakan metode TF.IGM yang ditunjukkan oleh **persamaan 2.2** kemudian dihitung kemiripan antara *query* dengan dokumen oleh **persamaan 2.5** untuk bisa sampai pada tahap pengurutan atau perangkingan. Evaluasi kemampuan sistem dalam perangkingan dokumen dapat dilakukan dengan melihat posisi dokumen yang relevan. Perhitungan evaluasi dilakukan dengan cara menghitung MAP oleh **persamaan 2.6**

4.2 Analisa Uji coba dan Hasil

Uji coba bertujuan untuk mengukur kemampuan kinerja metode, dimana yang akan dilakukan dengan menggunakan 14 *query* yang ditetapkan dan sesuai tema yang telah dijelaskan pada bab 2. Data *query* yang digunakan dalam uji coba ini merupakan potongan ayat dari tiap tema Al-Quran, dari 6326 ayat atau dokumen menjadi 3013 *term* dari hasil *preprocessing*, dan dibobotkan. *Query*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

diklasifikasikan menjadi dua bagian yaitu *query* pendek dan *query* panjang. Dimana *query* pendek adalah *query* atau kata kunci yang terdiri satu, dua dan tiga kata. Berikut ini adalah daftar *input query* pendek:

Tabel 4.1 Daftar *Input Query* Pendek

NO	Indeks Kelas	*Query* Pendek
1	Iman	وَاعْبُدْ رَبَّكَ
2	Ilmu	عَنِ السَّاعَةِ أَيَّانَ مُرْسَاهَا
3	Bangsa terdahulu	يَا مُوسَىٰ مَسْحُورًا
4	Sejarah	وَجَدْنَا عَلَيْهَا آبَاءَنَا
5	Al-Quran	الشَّمْسَ وَالْقَمَرَ حُسْبَانًا
6	Akhlaq dan adab	إِنَّ اللَّهَ مَعَ الصَّابِرِينَ
7	Ibadah	يُسَبِّحُ بِحَمْدِهِ
8	Makanan dan minuman	حَلَالًا طَيِّبًا
9	Pakaian dan perhiasan	مُتَبَرِّجَاتٍ بِزِينَةٍ
10	Hukum Privat	إِنْ أُمَّهَاتُهُمْ إِلَّا اللَّائِي وَلَدْنَهُمْ
11	Muamalat	أَنْفِقُوا فِي سَبِيلِ اللَّهِ
12	Peradilan dan hakim	وَإِنْ حَكَمْتَ فَاحْكُمْ بَيْنَهُمْ بِالْقِسْطِ
13	Hukum pidana, Jinayah	يَرْمُونَ الْمُحْصَنَاتِ الْغَافِلَاتِ
14	Jihad	هَاجَرُوا فِي سَبِيلِ اللَّهِ

Sedangkan *query* panjang adalah *query* atau kata kunci yang terdiri dari empat, lima, enam kata atau lebih. Dan berikut adalah daftar *input query* panjang:

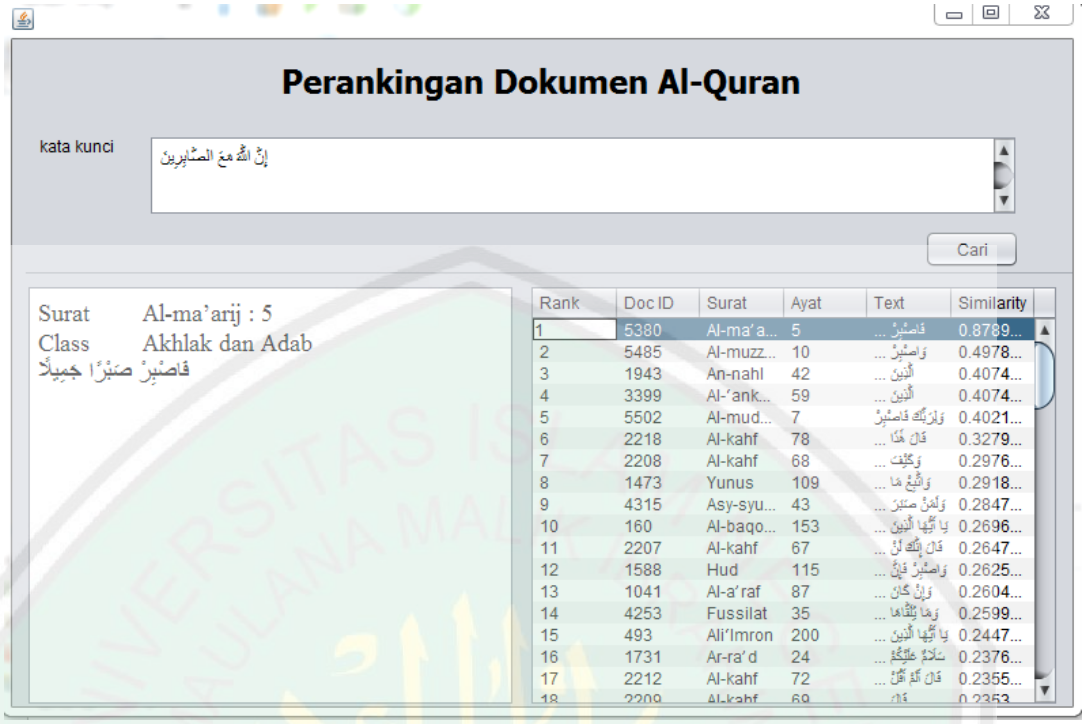
CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 4.2 Daftar *Input Query* Panjang

NO	Indeks Kelas	*Query* Panjang
1	Iman	سُبْحَانَ رَبِّنَا إِنْ كَانَ وَعْدُ رَبِّنَا لَمَفْعُولًا
2	Ilmu	قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا
3	Bangsa terdahulu	كَذَّبَتْ قَوْمُ لُوطٍ الْمُرْسَلِينَ
4	Sejarah	إِنَّا أَرْسَلْنَاكَ شَاهِدًا وَمُبَشِّرًا وَنَذِيرًا
5	Al-Quran	لَوْ أَنْزَلْنَا هَٰذَا الْقُرْآنَ عَلَىٰ جَبَلٍ لَرَأَيْتَهُ
6	Akhlaq dan adab	إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا
7	Ibadah	وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ
8	Makanan dan minuman	حُرِّمَتْ عَلَيْكُمُ الْمَيْتَةُ وَالدَّمُ وَلَحْمُ الْخِنْزِيرِ
9	Pakaian dan perhiasan	آدَمَ قَدْ أَنْزَلْنَا عَلَيْكُمْ لِبَاسًا يُوَارِي سَوْآتِكُمْ وَرِيشًا
10	Hukum Privat	فَعِدَّتُهُنَّ ثَلَاثَةُ أَشْهُرٍ وَاللَّائِي لَمْ يَحِضْنَ
11	Muamalat	مَسَّنَا وَأَهْلَنَا الضُّرُّ وَجِئْنَا بِبِضَاعَةٍ مُزْجَاةٍ
12	Peradilan dan hakim	كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ
13	Hukum pidana, Jinayah	الزَّانِيَةُ وَالزَّانِي فَاجْلِدُوا كُلَّ وَاحِدٍ مِنْهُمَا مِائَةَ جَلْدَةٍ
14	Jihad	خُذُوا حِذْرَكُمْ فَانْفِرُوا ثُبَاتٍ أَوِ انْفِرُوا جَمِيعًا

Setiap *query* juga akan melalui tahap *preprocessing* dan pembobotan, yang kemudian akan dihitung kemiripannya menggunakan *cosine similarity* dan perangkingan pun akan diurutkan dari nilai *similarity* tertinggi. Evaluasi kemampuan sistem dalam perangkingan dokumen dapat dilakukan dengan melihat posisi dokumen yang relevan dimana yang terdapat pada urutan pertama terhadap ke-14 *input query* yang diujikan. Salah satu contoh pencarian dapat ditunjukkan pada **gambar 4.1** berikut.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



Gambar 4.1 Contoh pencarian ayat

Semua *input query* yang diujikan dilakukan seperti yang ditunjukkan oleh **gambar 4.1**. *Input query* akan diproses dan akan tersimpan pada database yang hasilnya berupa kata dasar. Seperti yang ditunjukkan pada **gambar 4.2** yaitu tabel q_term dan akan diberi bobot seperti pada gambar 4.3 yaitu tabel proses_bobot_input. Setelah didapatkan bobot untuk query selanjutnya akan diproses perkalian antara bobot *query* dengan bobot dokumen yang dimana dokumen tersebut memiliki kata yang sesuai dengan *query*. Sebagai kelanjutan contohnya yaitu ditunjukkan pada **gambar 4.4** yaitu tabel dot_product dan terdapat 86 dokumen.

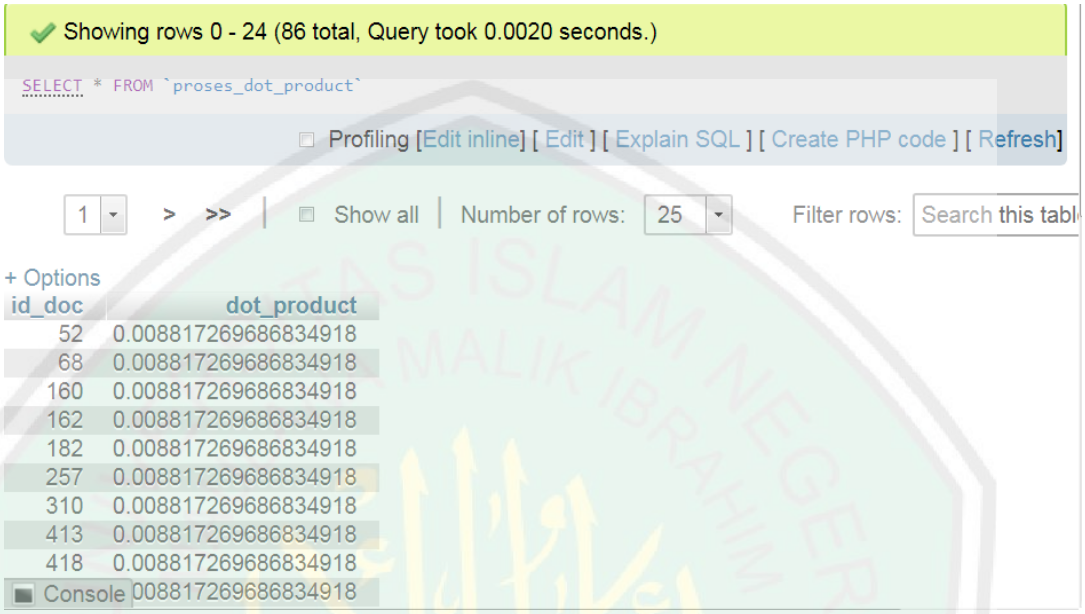
input_term
صبر

Gambar 4.2 Hasil tabel q_term

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

input_term	tf	igm	bobot
صبر	1	0.090909090909091	0.090909090909091

Gambar 4.3 Hasil tabel proses_bobot_input



Gambar 4.4 Hasil tabel dot_product

Selanjutnya yaitu menghitung panjang vector dokumen dan panjang vector *query* untuk mencari kemiripannya pada dokumen Al-Quran. Seperti yang ditunjukkan pada **gambar 4.6** yaitu gambar potongan hasil program jika selesai memproses query sampai didapatkan hasilnya. Setelah program menampilkan pesan bahwa *search* selesai, nilai *similarity* akan tersimpan pada database dan ditunjukkan pada **gambar 4.5** adalah sebagian hasil dari *similarity* yang tersimpan didatabase.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

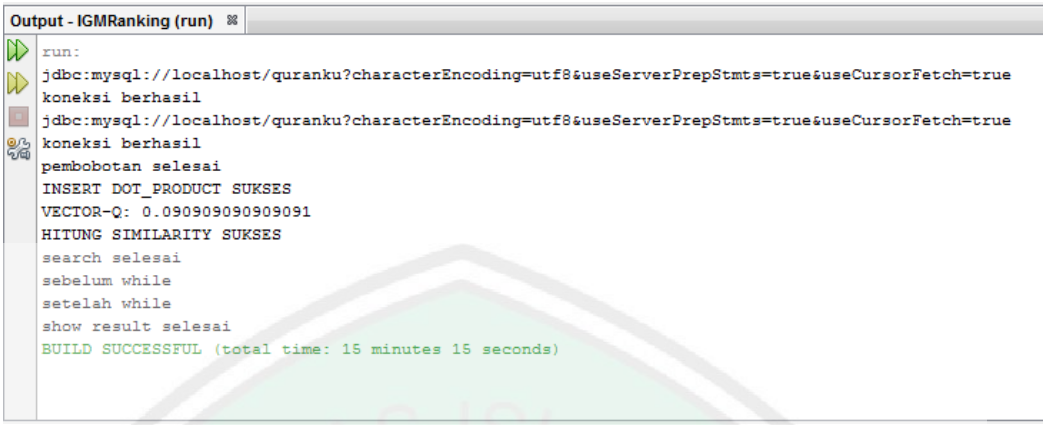
id_doc	similarity
52	0.21496244851899185
68	0.07278533216613689
160	0.26966297820339774
162	0.1490661274879856
182	0.12399170506868497
257	0.08697222460457595
310	0.2144496056697168
413	0.13729858818454815
418	0.0616388021525018
435	0.18848228931014485
439	0.16238391383909723
479	0.08761505535910752
493	0.24474951384850255
823	0.16584778735340217
1041	0.2604229946985852
1080	0.08085227658924395
1082	0.13308936695489385
1091	0.1371112426867367

Gambar 4.5 Hasil tabel *similarity*

Perangkingan diurutkan dari dokumen yang memiliki nilai *similarity* tertinggi sampai ke terendah. Jika pada **gambar 4.6** menampilkan pesan *show result* selesai, maka hasil perangkingan sudah selesai dan siap untuk di evaluasi nilai relevannya. Hasil dokumen yang terurut telah ditunjukkan seperti halnya pada **gambar 4.1**.

Peringkat 10 dan 20 dokumen yang ter-*retrieve* akan dihitung nilai presisinya. Dari nilai presisi tersebut akan dihitung nilai evaluasinya menggunakan MAP. Pengujian ini dilakukan untuk menunjukkan bahwa metode pembobotan TF.IGM dapat digunakan sebagai perangkingan dokumen.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



Gambar 4.6 Hasil proses pada program

Uji coba dilakukan sebanyak 28 kali, sesuai tema atau indeks yang berjumlah 14 kelas dan kategori *query* yaitu *query* pendek dan *query* panjang. Seperti pada perancangan scenario uji coba yang telah dipaparkan sebelumnya nilai presisi diperoleh dengan cara pembagian antara dokumen Al-Quran yang relevan dengan urutan dokumen yang ter-*retrieve.* Penentuan dokumen Al-Quran yang relevan didapat dari pakarnya yang menjadi *user* dan memiliki pengetahuan pada bidang bahasa Arab dan sebagai dosen agama di UIN Maliki Malang yaitu M. Imamuddin, Lc., MA. Perhitungan presisi dilakukan pada setiap dokumen yang ter-*retrieve* pada semua *input query* yang diujikan. Nilai MAP diperoleh dari rata-rata nilai presisi dari seluruh *input query* yang diujikan.

Hasil MAP *query* pendek pada *rank-k* 10 ditunjukkan pada **tabel 4.4** dan *rank-k* 20 ditunjukkan pada **tabel 4.5**. Sedangkan hasil MAP *query* panjang pada *rank-k* 10 ditunjukkan pada **tabel 4.6** *dan rank-k* 20 ditunjukkan pada **tabel 4.7.** Nilai MAP *query* pendek *rank-k* 10 adalah 76.83% dan *rank-k* 20 adalah 75.15%. Sedangkan nilai MAP *query* panjang *rank-k* 10 adalah 84.24% dan dan *rank-k* 20 adalah 74.12%. Dan **tabel 4.8** adalah hasil nilai MAP berdasarkan *rank-k* 10 dari

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

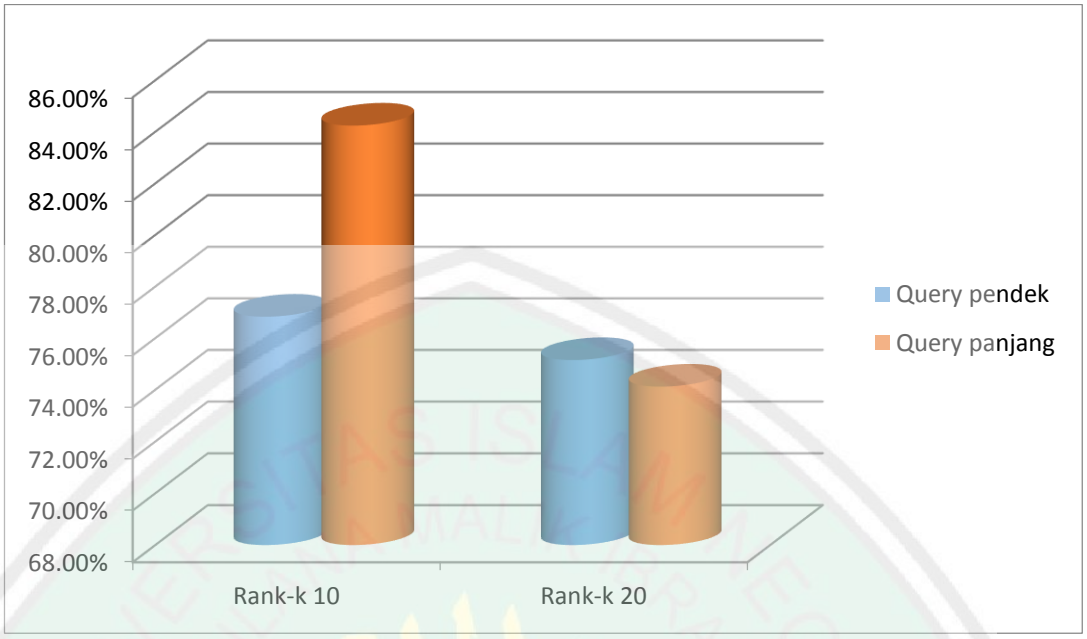
query pendek dan *query* panjang, begitu juga **tabel 4.9** adalah hasil nilai MAP berdasarkan *rank-k* 20 dari *query* pendek dan *query* panjang.

Prosentase nilai MAP ditunjukkan pada **tabel 4.3**, serta grafiknya yang ditunjukkan pada **gambar 4.7**. Jika dilihat dari hasil prosentasenya, maka dapat diambil kesimpulan bahwa, berdasarkan nilai-nilai MAP tersebut semakin panjang *input query* maka semakin tinggi nilai *similarity* dan dokumen yang relevan. Karena pada *query* panjang *term* atau katanya yang berjumlah empat atau lebih, jadi dapat memiliki bobot yang lebih besar dan tingkat relevannya juga lebih tinggi. Dibawah ini adalah tabel prosentasenya, *query* panjang baik pada *rank-k* 10 maupun *rank-k* 20 prosentasenya lebih tinggi dibandingkan dengan *query* pendek. Begitu juga prosentase nilai MAP yaitu nilai rata-rata dari keseluruhan peringkat, nilai prosentase untuk *query* panjang lebih tinggi dibanding nilai prosentase *query query* pendek. Dan untuk nilai prosentase *overall* atau nilai prosentase kesseluruhan *rank-k* 10 dan *rank-k* 20 baik pada *query* pendek maupun *query* panjang, lebih tinggi nilai prosentase untuk *rank-k* 10 karena pada *rank-k* 10 hasil dokumen yang relevan memiliki nilai *similarity* tinggi dan terurut dalam peringkat 10 teratas.

Tabel 4.3 Prosentase Nilai MAP

Kategori *Query*	***Rank-k*=10**	***Rank-k*=20**	**MAP**
Pendek	76.83%	75.15%	75.99%
Panjang	84.24%	74.12%	79.18%
Overall	**81%**	**74.63%**	

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG



Gambar 4.7 Grafik prosentase Nilai MAP

Keunggulan metode TF.IGM ini memberikan bobot pada *term* dimana sesuai indeks yang diberikan. Jadi, metode TF.IGM bisa digunakan untuk perangkinngan dokumen.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 4.4 Perhitungan MAP *rank-k* 10 *Query* Pendek

NO	*Short QUERY*	NILAI *PRECISION*										MP
		1	2	3	4	5	6	7	8	9	10	
1	وَاعْبُدْ رَبَّكَ	1	1	1	1	1		0.857	0.875	0.888	0.9	0.94667
2	عَنِ السَّاعَةِ أَيَّانَ مُرْسَاهَا	1	1	1	1	1	1	1		0.889	0.9	0.97656
3	يَا مُوسَىٰ مَسْحُورًا	1	1	1	1	1	1	1	1	1	1	1
4	وَجَدْنَا عَلَيْهَا آبَاءَنَا	1	1	1	1	1	1	1	1	1	1	1
5	الشَّمْسَ وَالْقَمَرَ حُسْبَانًا	1			0.5	0.6	0.667	0.714		0.667	0.7	0.69257
6	إِنَّ اللَّهَ مَعَ الصَّابِرِينَ	1	1	1	1	1	1	1	1	1	1	1
7	يُسَبِّحُ بِحَمْدِهِ		0.5	0.667	0.75	0.8	0.833	0.857	0.875	0.889	0.9	0.78567
8	حَلَالًا طَيِّبًا			0.333	0.5	0.6	0.667	0.714	0.75		0.7	0.60914
9	مُتَبَرِّجَاتٍ بِزِينَةٍ						0.167				0.2	0.1835
10	إِنْ أُمَّهَاتُهُمْ إِلَّا اللَّائِي وَلَدْنَهُمْ	1								0.222	0.3	0.50733
11	أَنْفِقُوا فِي سَبِيلِ اللَّهِ		0.5		0.75	0.8	0.833		0.75			0.7266
12	وَإِنْ حَكَمْتَ فَاحْكُمْ بَيْنَهُمْ بِالْقِسْطِ	1		0.667		0.6	0.667	0.714	0.75	0.778	0.8	0.747
13	يَرْمُونَ الْمُحْصَنَاتِ الْغَافِلَاتِ	1	1		0.75				0.5			0.8125
14	هَاجَرُوا فِي سَبِيلِ اللَّهِ	1		0.667	0.75	0.8			0.625			0.7684
TOTAL												**10.7559**
MAP												**0.76828**

PUSAT PERPUSTAKAAN
LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 4.5 Perhitungan MAP *rank-k* 20 *Query* Pendek

NO	*Short QUERY*	NILAI *PRECISION*																				MP
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	وَاعْبُدْ رَبَّكَ	1	1	1	1	1		0.857	0.875	0.888	0.9		0.909	0.923	0.928	0.933		0.882	0.889	0.894	0.9	0.92812
2	عن الساعة أيان مرساها	1	1	1	1	1	1	1		0.889	0.9	0.909	0.916	0.923		0.867	0.875	0.882		0.842	0.85	0.93253
3	يا مُوسَىٰ مَسْحُورًا	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				1
4	وَجَدْنَا عَلَيْهَا آبَاءَنَا	1	1	1	1	1	1	1	1	1	1											1
5	الشَّمْسَ وَالْقَمَرَ حُسْبَانًا	1			0.5	0.6	0.667	0.714		0.667	0.7			0.615		0.6	0.625			0.578	0.6	0.6555
6	إِنَّ اللَّهَ مَعَ الصَّابِرِينَ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	يُسَبِّحُ بِحَمْدِهِ		0.5	0.667	0.75	0.8	0.833	0.857	0.875	0.889	0.9	0.909	0.916	0.923	0.928	0.933	0.937	0.941	0.944	0.947	0.95	0.86311
8	حَلَالًا طَيِّبًا			0.333	0.5	0.6	0.667	0.714	0.75		0.7			0.615						0.473	0.5	0.5852
9	مُتَبَرِّجَاتٍ بِزِينَةٍ						0.167				0.2			0.231	0.285	0.333				0.315		0.25517
10	إِنْ أُمَّهَاتُهُمْ إِلَّا اللَّائِي وَلَدْنَهُمْ	1								0.222	0.3							0.235	0.278		0.3	0.38917
11	وَأَنْفِقُوا فِي سَبِيلِ اللَّهِ		0.5		0.75	0.8	0.833		0.75				0.583	0.615	0.642	0.667	0.687				0.6	0.67518
12	وَإِنْ حَكَمْتَ فَاحْكُمْ بَيْنَهُمْ بِالْقِسْطِ	1		0.667		0.6	0.667	0.714	0.75	0.778	0.8	0.818					0.625	0.647	0.667	0.684		0.72438
13	يَرْمُونَ الْمُحْصَنَاتِ الْغَافِلَاتِ	1	1		0.75				0.5													0.8125
14	فَاهْجُرُوا فِي سَبِيلِ اللَّهِ	1		0.667	0.75	0.8			0.625			0.545	0.583	0.615	0.642	0.667	0.687	0.705	0.722	0.736	0.75	0.6996
TOTAL																						**10.5205**
MAP																						**0.75146**

PUSAT PERPUSTAKAAN
LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF M

Tabel 4.6 Perhitungan MAP *rank-k* 10 *Query* Panjang

NO	*Long QUERY*	NILAI *PRECISION*										MP
		1	2	3	4	5	6	7	8	9	10	
1	سُبْحَانَ رَبِّنَا إِنْ كَانَ وَعْدُ رَبِّنَا لَمَفْعُولًا	1	1	1		0.8	0.833	0.857	0.875	0.889	0.9	0.906
2	قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا	1	1	1								1
3	كَذَّبَتْ قَوْمُ لُوطٍ الْمُرْسَلِينَ	1		0.667	0.75		0.667	0.714	0.75	0.778	0.8	0.76575
4	إِنَّا أَرْسَلْنَاكَ شَاهِدًا وَمُبَشِّرًا وَنَذِيرًا	1	1	1	1	1	1	1	1	1	1	1
5	لَوْ أَنْزَلْنَا هَٰذَا الْقُرْآنَ عَلَىٰ جَبَلٍ لَرَأَيْتَهُ	1					0.333	0.428	0.5	0.556		0.5634
6	إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا	1		0.667		0.6	0.667	0.714	0.75			0.733
7	وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ	1	1	1	1	1	1	1	1		0.9	0.98889
8	حُرِّمَتْ عَلَيْكُمُ الْمَيْتَةُ وَالدَّمُ وَلَحْمُ الْخِنْزِيرِ	1	1	1	1							1
9	آدَمَ قَدْ أَنْزَلْنَا عَلَيْكُمْ لِبَاسًا يُوَارِي سَوْآتِكُمْ وَرِيشًا	1										1
10	فَعِدَّتُهُنَّ ثَلَاثَةُ أَشْهُرٍ وَاللَّائِي لَمْ يَحِضْنَ	1										1
11	مَسَّنَا وَأَهْلَنَا الضُّرُّ وَجِئْنَا بِبِضَاعَةٍ مُزْجَاةٍ	1				0.4						0.7
12	كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ	1	1		0.75							0.91667
13	الزَّانِيَةُ وَالزَّانِي فَاجْلِدُوا كُلَّ وَاحِدٍ مِنْهُمَا مِائَةَ جَلْدَةٍ	1						0.285				0.6425
14	خُذُوا حِذْرَكُمْ فَانْفِرُوا ثُبَاتٍ أَوِ انْفِرُوا جَمِيعًا	1					0.333				0.4	0.57767
TOTAL												**11.7939**
MAP												**0.84242**

LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF

Tabel 4.7 Perhitungan MAP *rank-k* 20 *Query* Panjang

NO	*Long QUERY*	NILAI *PRECISION*																				MP
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	سُبْحَانَ رَبِّنَا إِنْ كَانَ وَعْدُ رَبِّنَا لَمَفْعُولًا	1	1	1		0.8	0.833	0.857	0.875	0.889	0.9	0.909	0.916	0.923	0.928	0.933	0.937	0.941	0.944		0.9	0.91583
2	قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا	1	1	1									0.333					0.294				0.7254
3	كَذَّبَتْ قَوْمُ لُوطٍ الْمُرْسَلِينَ	1		0.667	0.75		0.667	0.714	0.75	0.778	0.8	0.818	0.833	0.846	0.857	0.867	0.875	0.882	0.889	0.894	0.9	0.8215
4	إِنَّا أَرْسَلْنَاكَ شَاهِدًا وَمُبَشِّرًا وَنَذِيرًا	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			0.9	0.99444
5	لَوْ أَنْزَلْنَا هَذَا الْقُرْآنَ عَلَى جَبَلٍ لَرَأَيْتَهُ	1					0.333	0.428	0.5	0.556		0.545		0.538					0.444	0.473	0.5	0.5317
6	إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا	1		0.667		0.6	0.667	0.714	0.75													0.733
7	وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ	1	1	1	1	1	1	1	1		0.9	0.909	0.916	0.923	0.928		0.875	0.882	0.889	0.894	0.9	0.94533
8	حُرِّمَتْ عَلَيْكُمُ الْمَيْتَةُ وَالدَّمُ وَلَحْمُ الْخِنْزِيرِ	1	1	1	1																	1
9	ادَمَ قَدْ أَنْزَلْنَا عَلَيْكُمْ لِبَاسًا يُوَارِي سَوْآتِكُمْ وَرِيشًا	1																				1
10	فَعِدَّتُهُنَّ ثَلَاثَةُ أَشْهُرٍ وَاللَّائِي لَمْ يَحِضْنَ	1											0.167						0.167			0.44467
11	مَسَّنَا وَأَهْلَنَا الضُّرُّ وَجِئْنَا بِبِضَاعَةٍ مُزْجَاةٍ	1				0.4																0.7
12	كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ	1	1		0.75											0.267					0.25	0.6534
13	الزَّانِيَةُ وَالزَّانِي فَاجْلِدُوا كُلَّ وَاحِدٍ مِنْهُمَا مِائَةَ جَلْدَةٍ	1						0.285										0.176			0.2	0.41525
14	خُذُوا حِذْرَكُمْ فَانْفِرُوا ثُبَاتٍ أَوِ انْفِرُوا جَمِيعًا	1					0.333				0.4	0.454			0.428			0.411	0.444			0.49571
TOTAL																						**10.3762**
MAP																						**0.74116**

PUSAT PERPUSTAKAAN
LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF

Tabel 4.8 Nilai Perhitungan MAP *query* pendek dan panjang *rank-k* 10

No	*Query*	Kategori	Presisi
1	وَاعْبُدْ رَبَّكَ	Pendek	0.94667
2	عَنِ السَّاعَةِ أَيَّانَ مُرْسَاهَا	Pendek	0.97656
3	يَا مُوسَىٰ مَسْحُورًا	Pendek	1
4	وَجَدْنَا عَلَيْهَا آبَاءَنَا	Pendek	1
5	الشَّمْسَ وَالْقَمَرَ حُسْبَانًا	Pendek	0.69257
6	إِنَّ اللَّهَ مَعَ الصَّابِرِينَ	Pendek	1
7	يُسَبِّحُ بِحَمْدِهِ	Pendek	0.78567
8	حَلَالًا طَيِّبًا	Pendek	0.60914
9	مُتَبَرِّجَاتٍ بِزِينَةٍ	Pendek	0.1835
10	إِنْ أُمَّهَاتُهُمْ إِلَّا اللَّائِي وَلَدْنَهُمْ	Pendek	0.50733
11	أَنْفِقُوا فِي سَبِيلِ اللَّهِ	Pendek	0.7266
12	وَإِنْ حَكَمْتَ فَاحْكُمْ بَيْنَهُمْ بِالْقِسْطِ	Pendek	0.747
13	يَرْمُونَ الْمُحْصَنَاتِ الْغَافِلَاتِ	Pendek	0.8125
14	هَاجَرُوا فِي سَبِيلِ اللَّهِ	Pendek	0.7684
	MAP Jenis *query* pendek		**0.76828**
1	سُبْحَانَ رَبِّنَا إِنْ كَانَ وَعْدُ رَبِّنَا لَمَفْعُولًا	Panjang	0.906
2	قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا	Panjang	1
3	كَذَّبَتْ قَوْمُ لُوطٍ الْمُرْسَلِينَ	Panjang	0.76575
4	إِنَّا أَرْسَلْنَاكَ شَاهِدًا وَمُبَشِّرًا وَنَذِيرًا	Panjang	1
5	لَوْ أَنْزَلْنَا هَٰذَا الْقُرْآنَ عَلَىٰ جَبَلٍ لَرَأَيْتَهُ	Panjang	0.5634
6	إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا	Panjang	0.733
7	وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ	Panjang	0.98889
8	حُرِّمَتْ عَلَيْكُمُ الْمَيْتَةُ وَالدَّمُ وَلَحْمُ الْخِنْزِيرِ	Panjang	1
9	آدَمَ قَدْ أَنْزَلْنَا عَلَيْكُمْ لِبَاسًا يُوَارِي سَوْآتِكُمْ وَرِيشًا	Panjang	1
10	فَعِدَّتُهُنَّ ثَلَاثَةُ أَشْهُرٍ وَاللَّائِي لَمْ يَحِضْنَ	Panjang	1
11	مَسَّنَا وَأَهْلَنَا الضُّرُّ وَجِئْنَا بِبِضَاعَةٍ مُزْجَاةٍ	Panjang	0.7
12	كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ	Panjang	0.91667
13	الزَّانِيَةُ وَالزَّانِي فَاجْلِدُوا كُلَّ وَاحِدٍ مِنْهُمَا مِائَةَ جَلْدَةٍ	Panjang	0.6425
14	خُذُوا حِذْرَكُمْ فَانْفِرُوا ثُبَاتٍ أَوِ انْفِرُوا جَمِيعًا	Panjang	0.57767
	MAP Jenis *query* panjang		**0.84242**

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Tabel 4.9 Perhitungan MAP *query* panjang dan pendek *rank-k* 20

No	*Query*	Kategori	Presisi
1	وَاعْبُدْ رَبَّكَ	Pendek	0.92812
2	عَنِ السَّاعَةِ أَيَّانَ مُرْسَاهَا	Pendek	0.93253
3	يَا مُوسَىٰ مَسْحُورًا	Pendek	1
4	وَجَدْنَا عَلَيْهَا آبَاءَنَا	Pendek	1
5	الشَّمْسَ وَالْقَمَرَ حُسْبَانًا	Pendek	0.6555
6	إِنَّ اللَّهَ مَعَ الصَّابِرِينَ	Pendek	1
7	يُسَبِّحُ بِحَمْدِهِ	Pendek	0.86311
8	حَلَالًا طَيِّبًا	Pendek	0.5852
9	مُتَبَرِّجَاتٍ بِزِينَةٍ	Pendek	0.25517
10	إِنْ أُمَّهَاتُهُمْ إِلَّا اللَّائِي وَلَدْنَهُمْ	Pendek	0.38917
11	أَنْفِقُوا فِي سَبِيلِ اللَّهِ	Pendek	0.67518
12	وَإِنْ حَكَمْتَ فَاحْكُمْ بَيْنَهُمْ بِالْقِسْطِ	Pendek	0.72438
13	يَرْمُونَ الْمُحْصَنَاتِ الْغَافِلَاتِ	Pendek	0.8125
14	هَاجَرُوا فِي سَبِيلِ اللَّهِ	Pendek	0.6996
	MAP Jenis *query* pendek		**0.75146**
1	سُبْحَانَ رَبِّنَا إِنْ كَانَ وَعْدُ رَبِّنَا لَمَفْعُولًا	Panjang	0.91583
2	قَالَ إِنَّكَ لَنْ تَسْتَطِيعَ مَعِيَ صَبْرًا	Panjang	0.7254
3	كَذَّبَتْ قَوْمُ لُوطٍ الْمُرْسَلِينَ	Panjang	0.8215
4	إِنَّا أَرْسَلْنَاكَ شَاهِدًا وَمُبَشِّرًا وَنَذِيرًا	Panjang	0.99444
5	لَوْ أَنْزَلْنَا هَٰذَا الْقُرْآنَ عَلَىٰ جَبَلٍ لَرَأَيْتَهُ	Panjang	0.5317
6	إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا	Panjang	0.733
7	وَأَقَامَ الصَّلَاةَ وَآتَى الزَّكَاةَ	Panjang	0.94533
8	حُرِّمَتْ عَلَيْكُمُ الْمَيْتَةُ وَالدَّمُ وَلَحْمُ الْخِنْزِيرِ	Panjang	1
9	آدَمَ قَدْ أَنْزَلْنَا عَلَيْكُمْ لِبَاسًا يُوَارِي سَوْآتِكُمْ وَرِيشًا	Panjang	1
10	فَعِدَّتُهُنَّ ثَلَاثَةُ أَشْهُرٍ وَاللَّائِي لَمْ يَحِضْنَ	Panjang	0.44467
11	مَسَّنَا وَأَهْلَنَا الضُّرُّ وَجِئْنَا بِبِضَاعَةٍ مُزْجَاةٍ	Panjang	0.7
12	كُونُوا قَوَّامِينَ لِلَّهِ شُهَدَاءَ بِالْقِسْطِ	Panjang	0.6534
13	الزَّانِيَةُ وَالزَّانِي فَاجْلِدُوا كُلَّ وَاحِدٍ مِنْهُمَا مِائَةَ جَلْدَةٍ	Panjang	0.41525
14	خُذُوا حِذْرَكُمْ فَانْفِرُوا ثُبَاتٍ أَوِ انْفِرُوا جَمِيعًا	Panjang	0.49571
	MAP Jenis *query* panjang		**0.74116**

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

4.3 Integrasi dengan Al Quran

Perangkingan dokumen Al-Quran adalah proses pencarian kemiripan ayat Al-Quran berdasarkan kata kunci pengguna. Pencarian ayat yang dilakukan dengan mengukur kemiripan ayat yang dicari dengan dokumen terkait (*document similarity*) yaitu dokumen Al-Quran yang telah terurut peringkatnya. Semakin banyak ciri yang sesuai semakin tinggi nilai kemiripannya. Tujuan dari perangkingan pada objek dokumen Al-Quran adalah karena Al-Quran sebagai sumber hukum islam, selain itu Al-Quran juga merupakan sumber rujukan paling utama dan meyakini Al-Quran adalah bagian dari rukun iman.

Al-Quran sebagai pedoman hidup untuk mengatur semua yang berkaitan dengan perbuatan manusia di dunia. Baik dalam aspek diri sendiri, manusia kepada Tuhannya maupun dengan sesama manusia. Dengan menjadikan A-Quran sebagai pedoman hidup akan menghasilkan kesejahreraan, akhlak mulia dan peradaban bagi manusia. Beriman pada Al-Quran artinya mengikuti ajaran yang terkandung didalamnya, menjadikan panutan dan acuan sebagai referensi. Perintah beriman kepada Al-Quran dan mengikutinya merupakan konsekuensi dari rukun iman yang ketiga. Perintah beriman kepada Al-Quran melalui firman Allah dalam surat *al-An'am* ayat 155, surat *al-A'raf* ayat 3 dan surat *az-Zumar* ayat 55. Berikut dalil naqli untuk menetapkan kewajiban beriman kepada Al-Quran:

1. Firman Allah dalam Surat *al-An'am* ayat 155:

   وَهَذَا كِتَابٌ أَنْزَلْنَاهُ مُبَارَكٌ فَاتَّبِعُوهُ وَاتَّقُوا لَعَلَّكُمْ تُرْحَمُونَ

   Artinya : "*Dan Al Qur'an itu adalah kitab yang Kami turunkan yang diberkati, maka ikutilah dia dan bertakwalah agar kamu diberi rahmat.*" (Q.S. A*l-An'am* : 155)
2. Firman Allah dalam Surat *al-A'raf* ayat 3:

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

اتَّبِعُوا مَا أُنْزِلَ إِلَيْكُمْ مِنْ رَبِّكُمْ وَلَا تَتَّبِعُوا مِنْ دُونِهِ أَوْلِيَاءَ قَلِيلًا مَا مَا تَذَكَّرُونَ

Artinya : "*Ikutilah apa yang diturunkan kepadamu dari Tuhanmu dan janganlah kamu mengikuti pemimpin-pemimpin selain-Nya. Amat sedikitlah kamu mengambil pelajaran (daripadanya).*" (Q.S. *al-A'raf* : 3)

3. Dan firman Allah dalam Surat *az-Zumar* ayat 55:

وَاتَّبِعُوا أَحْسَنَ مَا أُنْزِلَ إِلَيْكُمْ مِنْ رَبِّكُمْ مِنْ قَبْلِ أَنْ يَأْتِيَكُمُ الْعَذَابُ بَغْتَةً وَأَنْتُمْ لَا تَشْعُرُونَ

Artinya : "*Dan ikutilah sebaik-baik apa yang telah diturunkan kepadamu dari Tuhanmu sebelum datang azab kepadamu dengan tiba-tiba, sedang kamu tidak menyadarinya"* (Q.S. *az-Zumar* : 55)

Jika diatas telah dijelaskan tentang alasan Al-Quran sebagai objek perangkingan, maka tujuan dari perangkingan adalah membantu meringankan *user* atau pencari informasi menemukan informasi atau data yang dibutuhkannya.

Dengan perangkingan data atau informasi pencari atau *user* akan lebih mudah mendapatkan informasinya. Seperti halnya Al-Quran pun diturunkan melalui tiga peringkat yaitu:

Peringkat pertama: Penurunan Al-Quran dari Allah SWT ke Luh Mahfudz yaitu satu-satunya tempat yang termula, yang terkumpul didalamnya segala sesuatu tentang ketemtuan Allah untuk semua makhluk-Nya. Ini terbukti dengan firman Allah SWT:

بَلْ هُوَ قُرْآنٌ مَجِيدٌ

Artinya : "*Bahkan yang didustakan mereka itu ialah Al-Quran yang Mulia."* (Q.S. *Al-Buruj* : 21)

فِي لَوْحٍ مَحْفُوظٍ

Artinya : "*Yang tersimpan di Lauh Mahfudz."* (Q.S. *Al-Buruj* : 22)

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Hikmah penurunan Al-Quran peringkat pertama: merupakan kemuliaan bagi Al-Quran itu sendiri karena tempat tersebut merupakan semulia-mulia tempat di sisi Allah. Peringkat ini juga perlu, karena menyempurnakan kewujudan Lauh Mahfudz yaitu tempat bagi segala perancangan yang wujud di alam ini dan Al-Quran adalah termasuk di dalam perancangan Allah ke dunia ini.

Peringkat kedua: Penurunan peringkat ini adalah penurunan Al-Quran dari Lauh Mahfud ke Baitul Izzah dilangit dunia. Bukti adanya peringkat kedua ini ialah dari Firman Allah SWT:

إِنَّا أَنْزَلْنَاهُ فِي لَيْلَةِ الْقَدْرِ

Artinya : "*Sesungguhnya Kami telah menurunkannya (Al-Quran) pada malam kemuliaan.*" (Q.S. *Al-Qadar*: 1)

Begitu juga dengan Firman Allah dalam ayat berikut:

إِنَّا أَنْزَلْنَاهُ فِي لَيْلَةٍ مُبَارَكَةٍ ۚ إِنَّا كُنَّا مُنْذِرِينَ

Artinya : "*Sesungguhnya Kami Menurunkannya pada suatu malam yang diberkahi dan sesungguhnya Kami-lah yang memberi peringatan.*" (Q.S. *Ad-Dukhan*: 3)

Dalam Firman Allah yang lain:

شَهْرُ رَمَضَانَ الَّذِي أُنْزِلَ فِيهِ الْقُرْآنُ هُدًى لِلنَّاسِ وَبَيِّنَاتٍ مِنَ الْهُدَىٰ وَالْفُرْقَانِ ۚ فَمَنْ شَهِدَ مِنْكُمُ الشَّهْرَ فَلْيَصُمْهُ ۖ وَمَنْ كَانَ مَرِيضًا أَوْ عَلَىٰ سَفَرٍ فَعِدَّةٌ مِنْ أَيَّامٍ أُخَرَ ۗ يُرِيدُ اللَّهُ بِكُمُ الْيُسْرَ وَلَا يُرِيدُ بِكُمُ الْعُسْرَ وَلِتُكْمِلُوا الْعِدَّةَ وَلِتُكَبِّرُوا اللَّهَ عَلَىٰ مَا هَدَاكُمْ وَلَعَلَّكُمْ تَشْكُرُونَ

Artinya : "*(Beberapa hari yang ditentukan itu ialah) bulan Ramadhan, bulan yang didalamnya diturunkan (permulaan) Al-Quran sebagai petunjuk bagi manusia dan penjelasan-penjelasan mengenai petunjuk itu dan pembeda (antara yang haq dan bathil). Karena itu, barang siapa diantara kamu hadir (dinegeri tempat tinggalnya) dibulan itu, maka hendaklah ia berpuasa di bulan itu, dan barang siapa sakit atau dalam perjalanan (lalu ia berbuka), maka (wajiblah baginya berpuasa), sebanyak hari yang ditinggalkannya itu, pada hari-hari yang lain. Allah menghendaki kemudahan bagimu, dan tidak menghendaki kesukaran bagimu. Dan hendaklah kamu mencukupkan bilangannya dan hendaklah*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

kamu mengagungkan Allah atas petunjuk-Nya yang diberikan kepadamu, supaya kamu bersyukur." (Q.S. *Al-Baqarah*: 185)

Penurunan peringkat ini adalah masih dalam keadaan sekaligus. Ini dapat dilihat dari hadits, yang artinya : *"Diturunkan Al-Quran ke langit dunia dalam jumlah sekaligus pada malam Lailatul Qadr kemudian diturunkan sesudah itu selama dua puluh tahun."*. Walaupun begitu terdapat juga beberapa pendapat ulama' yang tidak berapa kuat mengatakan bahwa penurunan Al-Quran pada peringkat ini adalah 23, 24, atau 25 malam Lailatul Qadar. Diturunkan Al-Quran untuk kali pertamanya pada malam Lailatur Qadar dan seterusnya secara peringkat. Diturunkan oleh Malaikat Hafadzah kepada Malaikat Jibril dalam jangka 24 malam.

Peringkat ketiga: Peringkatpenurunan dari Baitul Izzah ke dada Rasulullah SAW secara berangsur-angsur selama 22. 23, 24 atau 25 tahhun mengikuti perselisihan Ulama'. Dalil bagi peringkat ini adalah Firman Allah SWT pada ayat berikut:

وَقُرْآنًا فَرَقْنَاهُ لِتَقْرَأَهُ عَلَى النَّاسِ عَلَىٰ مُكْثٍ وَنَزَّلْنَاهُ تَنْزِيلًا

Artinya : "*Al-Quran Kami turunkan secara berangsung-angsur supaya engkau bacakan kepada manusia dengan cara perlahan-lahan dan Kami turunkannya sedikit-sedikit."* (Q.S. *Al-Isra'*: 106)

Firman Allah lagi dalam Al-Quran:

عَلَىٰ قَلْبِكَ لِتَكُونَ مِنَ الْمُنْذِرِينَ

Artinya : "*Diturunkan oleh Roh Suci (Jibril a.s) ke dalam hati engkau (Wahai Muhammad) supaka engkau memberi peringatan ."* (Q.S. *Asy-Syu'ara*: 194)

Begitu juga Firman Allah pada ayat berikut:

وَيَوْمَ نَبْعَثُ فِي كُلِّ أُمَّةٍ شَهِيدًا عَلَيْهِمْ مِنْ أَنْفُسِهِمْ ۖ وَجِئْنَا بِكَ شَهِيدًا عَلَىٰ هَٰؤُلَاءِ ۚ وَنَزَّلْنَا عَلَيْكَ الْكِتَابَ تِبْيَانًا لِكُلِّ شَيْءٍ وَهُدًى وَرَحْمَةً وَبُشْرَىٰ لِلْمُسْلِمِينَ

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Artinya : "*(Dan ingatlah) akan hari (ketika) Kami bangkitkan pada tiap-tiap umat seorang saksi atas mereka dari mereka sendiri dan Kami datangkan kamu (Muhammad) menjadi saksi atas seluruh umat manusia. Dan Kami turunkan kepadamu Al Kitab (Al Quran) untuk menjelaskan segala sesuatu dan petunjuk serta rahmat dan kabar gembira bagi orang-orang yang berserah diri..''* (Q.S. *An-Nahl'*: 89)

Hikmah diturunkan Al-Quran berperingkat peringkat diatas yaitu:

- Peringkat-peringkat tersebut dapat membedakan kitabAl-Quran dengan kitab-kitab Allah yag dahulu. Al-Quran diturunkan secara berperingkat-peringkat sementara yang lain secara sekaligus.
- Peringkatperingkat tersebut dapat engagungkan Kitab Al-Quran yang kedatangannya disebar kepada seluruh penghuni alam dari langit sampai ke bumi.
- Peringkat-peringkat tersebut sesungguhnya bertujuan untuk menambahkan keyakinan kepada seluruh makhluk penghuni semua penjuru alam bahwa Al-Quran sebagai kitab yang terakhir itu adalah dari Allah dan bukan ciptaan Nabi Muhammad dan sebagainya.
- Penurunan secara yang demikian sesungguhnya juga memuliakan kerasulan Nabi Muhammad SAW sebagai penyundah dan penghulu segala Nabi.
- Al-Quran tidak diturunkan sekaligus tetapi secaara berangsur-angsur agar hati Nabi Muhammad menjadi kuat dan tetap teguh.

Begitupun dalil dari tujuan dalam penelitian ini yang ingin mempermudah pengguna atau pencari untuk menemukan informasi yang dibutuhkan. Dari Abu Hurairah ra, Nabi Muhammad SAW, bersabda "*Barang siapa yang melepaskan*

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

satu kesusahan seorang mukmin, pasti Allah melepaskan darinya satu kesusahan pada hari kiamat. Barang siapa yang menjadikan mudah urusan oranglain, pasti Allah akan memudahkannya di dunia dan di akhirat. Barang siapa yang menutupi aib sorang muslim, pasti Allah akan menutupi aibnya di dunia dan di akhirat. Allah senantiasa menolong hamba-Nya selama hamba-Nya suka menolong saudaranya" (HR. Musim, Hadits Arba'in An-Nawawi ke-36).

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

BAB 5

PENUTUP

Pada bab terakhir ini, ditarik beberapa kesimpulan yang didapat dari hasil penelitian, dan juga saran-saran yang dapat digunakan sebagai bahan pertimbangan untuk pengembangan penelitian selanjutnya.

5.1 Kesimpulan

Berdasarkan hasil uji coba yang telah dipaparkan pada bab 4, dimana uji coba *input* 14 *query* yang telah dilakukan pada sistem, maka dapat ditarik kesimpulan bahwa, pada penelitian ini nilai *Mean Average Precision* (MAP) untuk *rank-k* 10 *query* pendek adalah 76.83% dan *rank-k* 20 adalah 75.15%, sedangkan nilai MAP untuk *rank-k* 10 *query* panjang adalah 84.24% dan dan *rank-k* 20 adalah 74.12%. Dan untuk total nilai MAP *query* pendek adalah 75.99% dan total nilai MAP *query* panjang adalah 79.18%.

5.2 Saran

Beberapa saran setelah dilakukan penelitian ini adalah sebagai berikut:

1. Pada penelitian selanjutnya bisa dilakukan perbandingan metode TF.IGM dengan metode pembobotan lainnya.
2. Pada penelitian selanjutnya bisa diimplementasikan pada bahasa Indonesia atau terjemahan Al-Quran.

Kesesuaian hasil pencarian ditentukan dari *input query* pengguna berdasarkan kemiripan *term.* Perlu adanya ekspansi untuk meningkatkan pencarian.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

DAFTAR PUSTAKA

Aly AA. 2008. *Using a Query Technique to Improve Document Retrieval.* Elminia University, Egypt.

Amin, Fatkhul. 2012. *Sistem Temu Kembali Informasi dengan Metode Vector Space Model*. Jurnal Sistem Informasi Bisnis, Universitas Stikubank. Semarang.

Al Haromainy, M. M. 2017. *Pencarian Ayat Berdasarkan Tematik Dalam Terjemahan Al-Quran Menggunakan Metode Vector Space Model dan Ekspansi Query*. Skripsi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim. Malang.

C. Kewen, Z. Zuping, & H. Z. Jun Long. 2016. *Turning from TF-IDF to TF-IGM for term weighting in text clasification.* School of Information Science and Engineering, Central South University. Changsha, Cina.

E.A. Esraa, B.L. Nagwa, dan M.F. Tolba. 2010. *An Efficient Ranking Module for an Arabic Search Engine*. IJCSNS International Journal of Computer Science and Network Security. Shams University.

Fauzi, M. A. 2013. *Term Weighting* berbasis Indeks Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab. *Thesis*, Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember, Indonesia.

Feldman, R. & Sanger, J. (2007). The Text Mining Handbook. New York: Cambridge University Press.

F.H.H. Khadijah. 2014. *Preference Based Term Weighting for Arabic Fiqh Document Ranking.* Thesis Institut Teknologi Sepuluh Nopember. Surabaya.

H.M. Suleiman. 2004. *Character Contiguity in N-gram-based Word Matching: the Case for Arabic Text Searching*, Information and Management. Yarmouk University. Ibrid, Jordan.

Harrag. F., Sherif-Hamdi. A., dan El-Qawasmeh. E. 2008. *Vector Space Model for Arabic Information Retrieval-Application to Hadith Indexing.* Proceedings of the First IEEE Conference on the Applications of Digital Information and Web Technologies.

Holle, K. F. ., Arifin, A. Z., & Purwitasari, D. (2015). Preference Based Term Weighting For Arabic Fiqh Document Rangking, *1*, 45–52.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Ingwersen, Peter. 1992. *INFORMMATION RETRIEVAL INTERACTION*. Royal School of Library and Information Science. Book available download [www.db.dk./pi/iri] Denmark.

Kurniawati. 2016. *Term Weighting Berbasis Indeks Kelas Menggunakan Metode TF.IDF.ICSδF untuk Perangkingan Dokumen Al-Qur;an.* Skripsi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim. Malang.

L. Indra, Z.A. Agus, dan P. Diana. 2015. *Perangkingan Dokumen berbahasa Arab Berdasarkan Susunan Posisi Kata dari Query*. Institut Teknologi Sepuluh Nopember. Surabaya.

Mandala, R. & Setiawan, H. 2006. *Peningkatan Performansi Sistem Temu-kembali Informasi dengan Peruasan Query Secara Otomatis.* Institut Teknologi Bandung. Bandung.

Nadirman, Firnas. 2006. *Sistem Temu-Kembali Informasi dengan Metode Vector Space Model pada Pencarian File Dokumen Berbasis Teks.* Skripsi Universitas Gajah Mada. Jogjakarta.

N. Ahmad, Z.A. Agus & P. Diana. 2011. *Implementasi N-Gram Dalam Pencarian Teks Sebagai Penunjang Aplikasi Perpustakaan Kitab Berbahasa Arab.* Institut Teknologi Sepuluh Nopember. Surabaya.

Parwita, W. G. S. (2015). Hybrid Recommendation System Memanfaatkan Penggalian Frequent Itemset dan Perbandingan Keyword, *9*(2), 19–21.

Rijsbergen, Van. 1976. *Information Retrieval 7 Chapter, Second Edition.* Departement of Computing Science, University of Glasgow. Scotland, United Kingdom.

Ren, F. & Sohrab, M. G. 2013. *Class-indexing-based term weighting for automatic text classification.* Faculty of Engineering, University of Tokushima. Japan.

Salton, G. 1989. *Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer*. Addison– Wesly Publishing Company, Inc. USA.

S Abd Al Baqiy IZF. *Fatur Rahman li Thalibi Ayatil-Quran.* Diponegoro, Bandung.

Turney, P.D. & Pantel, P. 2010. *From Frequency to Meaning: Vector Space Models of Semantic.* Journal of Artifical Intelligence Research. Canada, USA.

Wira, P. B. 2009. *Pengklasifikasian Artikel.* Depok: Ilmu Komunikasi Universitas Indonesia.

W. Adi, H. Andreas, & W. Charistian. 2012. *Implementasi Generalized Vector Space Model Menggunakan WordNet.* Teknik Informatika, Universitas Kristen Petra. Surabaya.

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG

Wang, D. & Zhang, H. 2013. *Inverse-category-frequency based supervised term weighting schemes for text categorization*. State Key Laboratory of Software Development Environment Beihang University. Beijing, Cina

CENTRAL LIBRARY OF MAULANA MALIK IBRAHIM STATE ISLAMIC UNIVERSITY OF MALANG