



TESIS - IF 185401

**PEMBOBOTAN *TERM* BERBASIS KEPADATAN
RUANG KATEGORI DAN KITAB UNTUK
PERANGKINGAN DOKUMEN TERJEMAHAN HADIS**

**Ana Tsalitsatun Ni'mah
NRP. 05111850010004**

**Dosen Pembimbing
Prof. Dr. Agus Zainal Arifin, S.Kom, M.Kom
NIP. 19720809 199512 1 001**

**Departemen Teknik Informatika
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
2020**

[Halaman ini sengaja dikosongkan]

LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik Informatika (M.Kom.)

di

Institut Teknologi Sepuluh Nopember

Oleh:

ANA TSALITSATUN NI'MAH

NRP. 05111850010004

Tanggal Ujian: 14 Juli 2020

Periode Wisuda: September 2020

Disetujui oleh:

Pembimbing:

Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom.
NIP. 19720809 199512 1 001



Penguji:

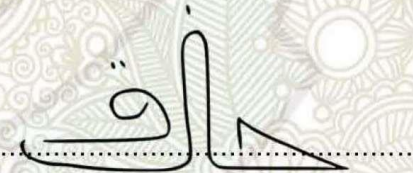
Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D.
NIP. 19490823 197603 2 001



Dr. Ahmad Saikhu, S.Si., M.T.
NIP. 19710718 200604 1 001




Hadziq Fabroyir, S.Kom., Ph.D.
NIP. 19860227 201903 1 006



Kepala Departemen Teknik Informatika
Fakultas Teknologi Elektro dan Informatika Cerdas




Dr. Chastine Fatichah, S.Kom., M.Kom.
NIP. 19751220 200112 2 002

[Halaman ini sengaja dikosongkan]

PEMBOBOTAN *TERM* BERBASIS KEPADATAN RUANG KATEGORI DAN KITAB UNTUK PERANGKINGAN DOKUMEN TERJEMAHAN HADIS

Nama Mahasiswa : Ana Tsalitsatun Ni'mah
NRP : 05111850010004
Pembimbing : Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom.

ABSTRAK

Kitab hadis memiliki karakteristik yang berbeda dibandingkan dokumen konvensional pada umumnya. Setiap kitab memiliki beberapa kategori yang dibentuk oleh pengarang kitab tersebut dengan karakteristik masing-masing. Kumpulan hadis yang memiliki kategori sama dengan pengarang berbeda dapat memiliki kata kunci yang berbeda. Penggunaan metode pembobotan *term* berbasis kategori yang ada saat ini kurang relevan jika diterapkan pada kitab hadis. Hal ini dikarenakan metode tersebut hanya mempertimbangkan kepadatan *term* dalam setiap kategori dan tidak memperhatikan kata kunci pada masing-masing kitab, sehingga dibutuhkanlah pembobotan *term* baru yang mampu mempertimbangkan kepadatan *term* dalam setiap kategori dan memperhatikan *term* kunci pada masing-masing kitab.

Penelitian ini mengusulkan metode pembobotan *term* berbasis kepadatan ruang kategori dan kitab untuk perangkingan dokumen terjemahan hadis. Metode yang diusulkan adalah *Inverse Hadith Space Density Frequency* (IHS_sF). IHS_sF merupakan metode pembobotan *term* yang memperhatikan kata kunci pada masing-masing kitab dengan melakukan perhitungan kepadatan *term* pada masing-masing kitab. IHS_sF digabungkan dengan TF-IDF-ICS_sF untuk mendapatkan pembobotan *term* yang mampu mempertimbangkan kepadatan *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab. Metode ini diterapkan ke dalam pembobotan *term* pada dokumen dari dataset hadis kitab 9 imam yaitu Bukhari, Muslim, Abu Daud, Darimi, Nasa'i, Tirmidzi, Ahmad, Malik, dan Ibnu Majah. Perangkingan dilakukan dengan menggunakan perhitungan *cosine similarity* sebagai perhitungan jarak *query* masukan dengan dataset hadis.

Hasil uji coba menunjukkan bahwa metode yang diusulkan dapat diterapkan ke dalam dataset hadis dan mampu mempertimbangkan densitas (kepadatan) *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab. Rata-rata *f-score* tertinggi didapatkan dari uji coba *short query* dengan *retrieve* 10 teratas adalah sebesar 73%.

Kata Kunci: *Inverse Class Space Density Frequency*, *Inverse Hadith Space Density Frequency*, Pembobotan *Term*, Perangkingan Dokumen, Terjemahan Hadis.

[Halaman ini sengaja dikosongkan]

TERM WEIGHTING BASED ON DENSITY OF THE BOOK AND CLASS OF HADIS TRANSLATION DOCUMENTS

Student Name : Ana Tsalitsatun Ni'mah
Student Identity Number : 05111850010004
Supervisor : Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom.

ABSTRACT

The book of hadith has different characteristics compared to conventional documents in general. Each book has several categories formed by the authors of the book with their respective characteristics. A collection of hadiths that have the same category as different authors can have different keywords. The current use of category-based weighting methods is less relevant if applied to the hadith. This is because the method only considers the word density in each category and does not pay attention to the keywords in each book. So we need a new word weighting that is able to consider the density of words in each category and pay attention to the keywords in each book.

This study proposes a method of weighting words based on the density of categories and books for ranking the hadith translation documents. The proposed method is Inverse Hadith Space Density Frequency (IHS_sF). IHS_sF is a method of weighting words that pay attention to the keywords in each book by calculating the word density in each book. IHS_sF is combined with TF-IDF-ICS_sF to get a weighting of words that is able to consider the density of words in each category and pay attention to the keywords in each book. This method is applied to the weighting of words in documents from the 9 hadith dataset, namely Bukhari, Muslim, Abu Daud, Darimi, Nasa'i, Tirmidzi, Ahmad, Malik, and Ibn Majah. Ranking is done using cosine similarity calculation as the calculation of the input query distance with the hadith dataset.

The trial results show that the proposed method can be applied to the hadith dataset and is able to consider the word density in each category and pay attention to the keywords in each book. The highest average f-score obtained from the short query trial with the top 10 retrieve is 73%.

Keywords : Document Ranking, Hadith Translation, Inverse Class Space Density Frequency, Inverse Hadith Space Density Frequency, Term Weighting

[Halaman ini sengaja dikosongkan]

KATA PENGANTAR

Puji syukur kehadirat Allah Subahanallahu Wa Ta'ala atas segala Rahmat dan Ridho-Nya penulis dapat menyelesaikan tesis dengan judul “Pembobotan *Term* Berbasis Kepadatan Ruang Kategori dan Kitab untuk Perangkingan Dokumen Terjemahan Hadis”. Penulis mengucapkan terima kasih atas doa dan dukungan kepada:

1. Bapak Prof. Dr. Agus Zainal Arifin, S.Kom., M.Kom. sebagai Dosen Pembimbing yang telah membimbing penulis dalam menyusun tesis dengan dukungan semangat dan ilmu melalui progress mingguan di dalam grup Zemi.
2. Ibu Prof. Ir. Handayani Tjandrasa, M.Sc., Ph.D. ; Bapak Dr. Ahmad Saikhu, S.Si., M.T. ; serta Bapak Hadziq Fabroyir, S.Kom., Ph.D. sebagai Dosen Penguji yang telah memberikan masukan yang berharga bagi tesis ini.
3. Suami tercinta Alfian Qomaruddin, anak-anak tersayang Muhammad Harits Al Fatih dan Dwindha Hasna Az Zakiyah, orangtua tercinta Moh. Sholichin dan Ismiati, mertua tersayang M. Rifa'i dan Mariyah, saudara-saudara dan keponakan tersayang Muhammad Ibdau Shulhi, Vina Widyaning Tyas, almh. Itsna Laila Masrurin, Ahmad Abdullah Vicky, Nailul Abyan Fairuz Shulhi, Khayla Az Zahra Shulhi, M. Naufal Fahmi, M. Alfin Yusra, serta Sahabat tersayang Wildan Mubarak dan M. Syaiful Arifin, yang telah memberi dukungan, doa dan semangat bagi penulis untuk menyusun tesis.
4. Seluruh Ustadz dan Ustadzah di Madrasah Aliyah dan Madrasah Tsanawiyah Pondok Pesantren Nurul Islam Pungging Mojokerto, Dosen Sekolah Tinggi Agama Islam Nurul Islam Pungging Mojokerto, seluruh Dosen di Universitas Trunojoyo Madura, serta Grup WA dan Telegram “Konferensi MGMP Nasional Bahasa Arab”, yang telah membantu penulis di dalam pengecekan kebenaran hasil penelitian tesis ini.
5. Teman-teman S2 TC 2018 : Arif, Fahmi, Adam, Dwi, Shaza, *partner* dosen pembimbing sama yang selalu bersama-sama berproses dengan penulis, dan

Mbak Maryamah, Mbak Rizka, Mbak Raras yang sudah membantu dalam memberikan ilmu di dalam grup bimbingan Zemi dan Cahya yang memberikan semangat sejak awal masuk S2 TC serta seluruh jajaran teman-teman S2 TC 2018 yang tidak dapat saya sebut secara satu persatu.

6. Seluruh pihak yang terlibat dalam memberikan semangat kepada penulis untuk menyusun tesis ini dengan berani dan tepat.

Semoga Rahmad dan Ridho Allah menyertai setiap pihak yang telah terlibat dalam memberi motivasi kepada penulis untuk menyelesaikan tesis. Penulis menyadari bahwa penelitian ini masih sangat jauh dari sempurna. Oleh karena itu, kritik dan saran sangat dibutuhkan oleh penulis agar bisa lebih baik lagi. Akhir kata, semoga penelitian yang tertuang dalam buku tesis ini bermanfaat bagi bidang informasi hadis sesuai cita-cita penulis dan dikembangkan lebih lanjut.

Surabaya, 14 Juli 2020

Ana Tsalitsatun Ni'mah

DAFTAR ISI

Abstrak	v
Abstract	vii
Kata Pengantar	ix
Daftar Isi	xi
Daftar Gambar.....	xiii
Daftar Tabel	xv
Daftar Lampiran.....	xvii
Daftar Persamaan	xix
BAB I	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	4
1.4. Manfaat Penelitian.....	4
1.5. Kontribusi Penelitian.....	4
1.6. Batasan Masalah.....	4
BAB II.....	5
2.1. Hadis.....	5
2.2. <i>Preprocessing</i>	9
2.3. <i>Term Weighting</i>	10
2.3.1. <i>Term Frequency (TF)</i>	11
2.3.2. <i>Inverse Document Frequency (IDF)</i>	11
2.3.3. <i>Inverse Class Frequency (ICF)</i>	12
2.3.4. <i>Inverse Class Space Density Frequency (ICS_sF)</i>	13
2.4. <i>Vector space model (VSM)</i>	15
2.5. <i>Cosine similarity</i>	17
2.6. <i>Confusion Matrix</i>	18
BAB III	21
3.1. Rancangan Model Sistem	22
3.1.1. Dataset.....	23
3.1.2. Preprocessing	24
3.1.3. Term Weighting	29
3.1.4. <i>Cosine Similarity</i>	39
3.2. Skenario Uji Coba	39
BAB IV	41
4.1. Persiapan Dataset.....	41
4.2. <i>Preprocessing</i>	43
4.3. <i>Term Weighting</i>	44
4.4. <i>Cosine Similarity</i>	48
4.5. Dokumentasi Uji Coba	49
4.6. Analisa Hasil	63
BAB V.....	65
5.1. Kesimpulan.....	65
5.2. Saran.....	66
DAFTAR PUSTAKA	67

[Halaman ini sengaja dikosongkan]

DAFTAR GAMBAR

Gambar 2.1. Model Ruang Vektor	16
Gambar 2.2. Matriks <i>Term</i> Dokumen	16
Gambar 2.3. Representasi Perumusan <i>Cosine Similarity</i>	17
Gambar 3.1. Tahap Penelitian	21
Gambar 3.2. Rancangan Sistem Secara Keseluruhan	22
Gambar 3.3. <i>Preprocessing</i>	23
Gambar 3.4. <i>Flowchart</i> Proses <i>Case Folding</i>	24
Gambar 3.5. Contoh Proses <i>Case Folding</i>	24
Gambar 3.6. <i>Flowchart</i> Proses <i>Filtering</i>	25
Gambar 3.7. Contoh Proses <i>Filtering</i>	25
Gambar 3.8. <i>Flowchart</i> Proses <i>Stemming</i>	27
Gambar 3.9. Contoh Proses <i>Stemming</i>	28
Gambar 3.10. <i>Flowchart</i> Proses <i>Tokenization</i>	28
Gambar 3.11. Contoh Hasil Proses <i>Tokenization</i>	28
Gambar 3.12. <i>Flowchart</i> Perhitungan <i>Document Frequency</i>	31
Gambar 3.13. <i>Flowchart</i> Perhitungan <i>Class Frequency</i>	32
Gambar 3.14. <i>Flowchart</i> Perhitungan $CS_{\delta F}$	34
Gambar 3.15. <i>Flowchart</i> Perhitungan $HS_{\delta F}$	36
Gambar 4.1. Contoh Dokumen Hadis Yang Diambil	42
Gambar 4.2. Contoh Hasil Input Dataset	43
Gambar 4.3. Contoh Hasil <i>Preprocessing</i>	44
Gambar 4.4. Contoh Hasil Perhitungan <i>Term Frequency</i>	45
Gambar 4.5. Contoh Hasil Perhitungan IDF	45
Gambar 4.6. Contoh Hasil Perhitungan ICF	46
Gambar 4.7. Contoh Hasil Perhitungan $ICS_{\delta F}$	47
Gambar 4.8. Contoh Hasil Perhitungan $IHS_{\delta F}$	47
Gambar 4.9. Contoh Hasil Perkalian Pembobotan	48
Gambar 4.10. Grafik Perbandingan Hasil Rata-rata <i>F-Score Retrieve 10</i> Teratas	52
Gambar 4.11. Grafik Perbandingan Hasil Rata-rata <i>F-Score Retrieve 20</i> Teratas	54
Gambar 4.12. Grafik Perbandingan Hasil Rata-rata <i>F-Score Retrieve 30</i> Teratas	56
Gambar 4.13. Grafik Perbandingan Hasil Rata-rata <i>F-Score Semua Query</i> ...	57
Gambar 4.14. Grafik Perbandingan Hasil Rata-rata <i>F-Score Short Query</i>	59
Gambar 4.14. Grafik Perbandingan Hasil Rata-rata <i>F-Score Long Query</i>	61

[Halaman ini sengaja dikosongkan]

DAFTAR TABEL

Tabel 2.1. <i>Confusion Matrix</i>	18
Tabel 3.1. Jumlah Hadis dan Jumlah Kategori	23
Tabel 3.2. Contoh Stoplist Fadila Tala	25
Tabel 3.3. Contoh Perhitungan <i>Term Frequency</i>	29
Tabel 3.4. Contoh Perhitungan <i>Document Frequency</i> dan <i>Inverse Document Frequency</i>	30
Tabel 3.5. Contoh Hasil TF-IDF	31
Tabel 3.6. Contoh Perhitungan <i>Class Frequency</i> dan <i>Inverse Class Frequency</i>	32
Tabel 3.7. Contoh Perhitungan $CS_{\delta}F$ dan $ICS_{\delta}F$	33
Tabel 3.8. Contoh Hasil TF-IDF- $ICS_{\delta}F$	34
Tabel 3.9. Contoh Perhitungan $HS_{\delta}F$ dan $IHS_{\delta}F$	37
Tabel 3.10. Contoh Hasil TF-IDF- $IHS_{\delta}F$	38
Tabel 3.11. Contoh Hasil TF-IDF-ICF- $IHS_{\delta}F$	38
Tabel 3.12. Contoh Hasil TF-IDF- $ICS_{\delta}F$ - $IHS_{\delta}F$	39
Tabel 3.13. Contoh Perhitungan <i>Cosine Similarity</i>	39
Tabel 3.14. Contoh Perbandingan Hasil <i>Cosine Similarity</i>	40
Tabel 3.15. Rancangan Dokumentasi Hasil Uji Coba	40
Tabel 4.1. Jumlah Hadis dan Jumlah Kategori	43
Tabel 4.2. Pakar Pengecek Kebenaran Hasil	50
Tabel 4.3. Rata-rata Hasil <i>F-Score Retrieve</i> 10 Teratas	51
Tabel 4.4. Rata-rata Hasil <i>F-Score Retrieve</i> 20 Teratas	53
Tabel 4.5. Rata-rata Hasil <i>F-Score Retrieve</i> 30 Teratas	55
Tabel 4.6. Rata-rata Hasil <i>F-Score Semua Query</i>	56
Tabel 4.7. Rata-rata Hasil <i>F-Score Short Query</i>	58
Tabel 4.8. Rata-rata Hasil <i>F-Score Long Query</i>	60
Tabel 4.12. Uji Coba <i>Query Matan</i>	62

[Halaman ini sengaja dikosongkan]

DAFTAR LAMPIRAN

Lampiran 1 Tabel Kategori dalam Kitab Shahih Bukhari	70-71
Lampiran 1 Tabel Kategori dalam Kitab Shahih Muslim	71-72
Lampiran 1 Tabel Kategori dalam Kitab Sunan Abu Dawud	72
Lampiran 1 Tabel Kategori dalam Kitab Ahmad	72-73
Lampiran 1 Tabel Kategori dalam Kitab Tirmidzi	73
Lampiran 1 Tabel Kategori dalam Kitab Darimi	74
Lampiran 1 Tabel Kategori dalam Kitab Nasai	74-75
Lampiran 1 Tabel Kategori dalam Kitab Ibnu Madjah	75
Lampiran 1 Tabel Kategori dalam Kitab Malik	75-76
Lampiran 2 Tabel Contoh Kasus <i>Query</i> "malu sebagian dari iman"	77
Lampiran 3 Tabel 50 <i>Query</i> Pengujian	78-79
Lampiran 4 Gambar Contoh Dokumentasi Hasil Pencarian	80-81
Lampiran 5 Tabel Dokumentasi Hasil <i>Recall</i> , <i>Precision</i> , dan <i>F-Score</i> 50 <i>Query</i> Top 10	82-83
Lampiran 5 Tabel Dokumentasi Hasil <i>Recall</i> , <i>Precision</i> , dan <i>F-Score</i> 50 <i>Query</i> Top 20	84-85
Lampiran 5 Tabel Dokumentasi Hasil <i>Recall</i> , <i>Precision</i> , dan <i>F-Score</i> 50 <i>Query</i> Top 30	86-87
Lampiran 6 Tabel Uji Coba Kedua: <i>Query</i> <i>Matan</i>	88-89

[Halaman ini sengaja dikosongkan]

DAFTAR PERSAMAAN

Persamaan 2.1. <i>Term Frequency</i>	11
Persamaan 2.2. <i>Document Frequency</i>	12
Persamaan 2.3. <i>Inverse Document Frequency</i>	12
Persamaan 2.4. <i>Weight TF-IDF</i>	12
Persamaan 2.5. <i>Class Frequency</i>	13
Persamaan 2.6. <i>Inverse Class Frequency</i>	13
Persamaan 2.7. <i>Weight TF-IDF-ICF</i>	13
Persamaan 2.8. <i>Class Density</i>	14
Persamaan 2.9. <i>Class Space Density</i>	14
Persamaan 2.10. <i>Inverse Class Space Density Frequency</i>	15
Persamaan 2.11. <i>Weight TF-IDF-ICS_δF</i>	15
Persamaan 2.12. <i>Cosine Similarity</i>	18
Persamaan 2.13. <i>Recall</i>	19
Persamaan 2.14. <i>Precision</i>	19
Persamaan 2.15. <i>F-Score</i>	19
Persamaan 3.1. <i>Hadith Density</i>	35
Persamaan 3.2. <i>Hadith Space Density</i>	35
Persamaan 3.3. <i>Inverse Hadith Space Density Frequency</i>	36
Persamaan 3.4. <i>Weight TF-IDF-ICS_δF-ICS_δF</i>	37

[Halaman ini sengaja dikosongkan]

BAB I

PENDAHULUAN

Bab ini merupakan penjelasan mengenai beberapa hal dasar dalam pembuatan laporan penelitian. Hal-hal yang dimaksud meliputi latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, kontribusi penelitian, serta batasan masalah.

1.1. Latar Belakang

Media informasi berkembang dengan sangat pesat diikuti dengan semakin meningkatnya jumlah dokumen yang dapat diambil. Seorang pengguna media informasi membutuhkan informasi yang akan diambil sesuai kebutuhan masing-masing, oleh sebab itu diperlukan pengembangan penelitian dalam bidang *information retrieval* mengenai perangkian dokumen untuk mendapatkan dokumen yang paling relevan sesuai kebutuhan pengguna (Jabri et al., 2018; Kurniawati & Syauqi, 2017). Dokumen yang tersebar di media informasi memiliki berbagai macam jenis, salah satunya adalah dokumen hadis.

Hadis adalah narasi yang berasal dari perkataan, tingkah laku dan ketetapan Nabi Muhammad (Azmi et al., 2019; Mohammad Baharun Edisi Revisi Syamsul Rijal Hamid, 2002). Setiap hadis terdiri dari dua bagian, yaitu sanad dan matan. Matan mewakili isi sebenarnya dari hadis, sementara sanad merupakan rantai periwayat dari hadis tersebut (Abdelaal, Ahmed, et al., 2019; Abdelaal, Elemery, et al., 2019; Abdi et al., 2020; Hassanein, 2018; Rostam & Malim, 2019; Saloot et al., 2016). Ahli-ahli hadis yang mengumpulkan, mendaftar, menyeleksi dan menuliskan hadis-hadis dalam suatu kitab hadis dikenal sebagai *mudawwin* atau *mukharrij*. Bagi umat Islam, hadis adalah sumber hukum Islam terpenting kedua setelah Al Qur'an yang berfungsi sebagai penjelas Al Qur'an (Mohammad Baharun Edisi Revisi Syamsul Rijal Hamid, 2002). Hadis secara struktural maupun fungsional disepakati oleh mayoritas umat Islam dan berbagai madzab Islam sebagai salah satu sumber ajaran Islam. Karena dengan adanya hadis itulah ajaran Islam menjadi jelas, rinci dan spesifik (Arifin, 2013). Oleh sebab itu, diperlukan

sebuah sistem untuk dapat melakukan studi pembelajaran mengenai hadis secara sistematis. Salah satunya adalah mesin pencarian hadis.

Pencarian hadis diperlukan untuk dapat mendapatkan informasi mengenai hadis yang relevan sesuai kebutuhan pengguna. Mesin pencarian hadis dilakukan dengan cara melakukan perankingan hadis sesuai dengan *query* yang dimasukkan oleh pengguna dengan ranking tertinggi adalah hadis yang paling relevan dengan *query*. Perankingan hadis diawali dengan mengubah setiap dokumen menjadi sebuah vector yang dapat dihitung nilai kemiripannya, atau biasa disebut dengan *Vector Space Model* (Lv & El-Gohary, 2016). Algoritma tersebut terlebih dahulu melakukan perhitungan bobot tiap *term* pada masing-masing dokumen (*term weighting*) (Arif et al., 2009).

Term weighting merupakan proses penghitungan bobot tiap *term* yang dicari pada setiap dokumen sehingga dapat diketahui ketersediaan dan kemiripan suatu *term* di dalam dokumen. *Term weighting* adalah tingkat kepentingan *term* t_i dalam dokumen d_j (Sabbah et al., 2017). *Term weighting* yang paling populer adalah *Term Frequency Inverse Document Frequency* (TF-IDF) (Ren & Sohrab, 2013).

TF-IDF melakukan pembobotan *term* pada tiap dokumen dengan memperhatikan kelangkaan *term* pada keseluruhan dokumen. Hasil akhir dari TF-IDF adalah bobot *term* pada tiap-tiap dokumen. TF-IDF memberikan bobot tinggi pada *term* yang jarang muncul pada seluruh dokumen. TF-IDF ini memiliki kekurangan yaitu menghilangkan informasi kategori pada tiap dokumen sehingga dilakukanlah penelitian untuk mengembangkan TF-IDF yaitu *Term Frequency Inverse Document Frequency Inverse Class Frequency* (TF-IDF-ICF) yang lebih memperhatikan informasi kategori di dalam melakukan pembobotan *term*.

TF-IDF-ICF juga melakukan pembobotan *term* pada tiap dokumen seperti halnya TF-IDF dengan hasil bobot *term* pada tiap dokumen pula. Perbedaannya pada perhitungan ICF, jika IDF hanya memperhitungkan kelangkaan *term* pada keseluruhan dokumen, ICF juga menghitung kelangkaan *term* pada keseluruhan kategori. Namun, TF-IDF-ICF tidak memperhatikan kepadatan dokumen tiap kategori terhadap munculnya *term*. Sehingga dikembangkan kembali dengan pendekatan *term weighting* yaitu *Term Frequency Inverse Document Frequency*

Inverse Class Space Density Frequency (TF-IDF- ICS_sF) yang lebih memperhatikan kepadatan dokumen pada ruang kategori terhadap munculnya *term*.

Kitab hadis memiliki karakteristik yang berbeda dibandingkan dokumen konvensional pada umumnya. Setiap kitab memiliki beberapa kategori yang dibentuk oleh pengarang kitab tersebut dengan karakteristik masing-masing. Kumpulan hadis yang memiliki kategori sama dengan pengarang berbeda dapat memiliki kata kunci yang berbeda. Penggunaan metode pembobotan *term* berbasis kategori (TF-IDF-ICS_sF) yang ada saat ini kurang relevan jika diterapkan pada kitab hadis. Hal ini dikarenakan metode tersebut hanya mempertimbangkan kepadatan *term* dalam setiap kategori dan tidak memperhatikan kata kunci pada masing-masing kitab, sehingga dibutuhkanlah pembobotan *term* baru yang mampu mempertimbangkan kepadatan *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab.

Penelitian ini mengusulkan metode pembobotan *term* berbasis kepadatan ruang kategori dan kitab untuk perangkaan dokumen terjemahan hadis. Metode pembobotan *term* yang diusulkan menggabungkan dua aspek di dalam sebuah dokumen hadis yaitu kepadatan ruang kategori dengan kepadatan ruang kitab yaitu *Inverse Class Space Density Frequency* (ICS_sF) dengan *Inverse Hadith Space Density Frequency* (IHS_sF). Metode tersebut diterapkan pada proses pembobotan *term* di dalam sistem pencari hadis berbasis web yang dibangun pada penelitian ini.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana perangkaan hadis dengan menggunakan metode pembobotan *term* berbasis kepadatan ruang kategori pada terjemahan hadis?
2. Bagaimana perangkaan hadis dengan menggunakan metode pembobotan *term* berbasis kepadatan ruang kitab pada terjemahan hadis?
3. Bagaimana perangkaan hadis dengan menggunakan metode pembobotan *term* berbasis kepadatan ruang kitab dan kategori pada terjemahan hadis?

1.3. Tujuan Penelitian

Penelitian ini mengusulkan metode pembobotan *term* berbasis kepadatan ruang kategori dan kitab untuk perangkaan dokumen terjemahan hadis.

1.4. Manfaat Penelitian

Manfaat penelitian ini diantaranya adalah untuk meningkatkan relevansi pencarian dokumen pada kitab seperti Sembilan Kitab Hadis yang memiliki kategori dengan kepadatan istilah yang heterogen.

1.5. Kontribusi Penelitian

Metode pembobotan *term* pada umumnya hanya memperhatikan kepatan *term* pada dokumen dan kategori saja. Kontribusi pada penelitian ini adalah pembobotan *term* baru berbasis kepadatan ruang kategori dan ruang kitab yaitu TF-IDF-ICS_sF-IHS_sF (*Term Frequency Inverse Document Frequency Inverse Class Space Density Frequency Inverse Hadith Space Density Frequency*).

1.6. Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Metode pembobotan *term* yang diusulkan diterapkan pada terjemahan Bahasa Indonesia dari dokumen Hadis.
2. Dokumen yang digunakan adalah 9 Kitab Hadis yaitu: Bukhari, Muslim, Ahmad, Ibnu Majah, Abu Daud, Malik, Nasa'i, Tirmidzi, dan Darimi.

BAB II

KAJIAN PUSTAKA DAN DASAR TEORI

Bab ini merupakan pemaparan mengenai beberapa konsep dasar tentang teori yang dipakai sebagai rujukan dalam melakukan penelitian ini.

2.1. Hadis

Hadis secara harfiah berarti "berbicara", "perkataan" atau "percakapan". Dalam terminologi Islam istilah hadis berarti melaporkan, mencatat sebuah pernyataan dan tingkah laku dari Nabi Muhammad. Menurut istilah ulama ahli hadis, hadis yaitu apa yang diriwayatkan dari Nabi, baik berupa perkataan, perbuatan, ketetapan, sifat jasmani atau sifat akhlak, perjalanan setelah diangkat sebagai Nabi dan terkadang juga sebelumnya.

Setiap hadis mengandung tiga unsur, yaitu:

- Matan (teks atau perkataan yang disampaikan).
- Rawi (disebut juga perawi) adalah orang yang menyampaikan atau yang meriwayatkan hadis yang pernah diterimanya dari seseorang ke dalam suatu kitab.
- Sanad adalah orang-orang yang menjadi sandaran dalam meriwayatkan hadis. Dengan kata lain, sanad adalah orang-orang yang menjadi perantara dari Nabi Muhammad Rosululloh Saw kepada perawi.

Ketiga istilah tersebut lebih jelasnya terdapat dalam suatu hadis berikut ini: “Imam Muslim berkata, telah meriwayatkan kepada kami Sahl bin Utsman Al Askari, Sahl menerima berita dari Yahya bin Zakaria, Yahya mendengar dari Sa’ad bin Thoriq, Sa’ad menerima dari Sa’ad bin Ubaidah, Sa’ad bin Ubaidah menerima dari Ibnu Umar, dan Ibnu Umar mendengar bahwa Nabi Muhammad Rosululloh Saw bersabda, ‘Asas Islam itu ada lima, yakni syahadat, sholat, zakat, puasa dan haji.’”. Yang disebut matan dalam hadis tersebut adalah perkataan “Asas Islam itu...”. Rawi (perawi atau periwayat) yang terakhir bagi kita ialah Imam Muslim. Dan, yang disebut sanad ialah Sahl bin Utsman Al Askari, Yahya bin Zakaria, Sa’ad bin Thoriq, Sa’ad bin Ubaidah, dan Ibnu Umar. Merekalah yang disebut sandaran

atau perantara dari Nabi Muhammad Rosululloh Saw kepada perawi. Terkadang, suatu hadis memang memiliki banyak sanad. Penyebutan sanad secara lengkap itu sangat diperlukan bagi ahli hadis untuk meneliti derajat hadis tersebut. Namun, bagi ulama yang menyusun kitab hadis sebagai petunjuk praktis dalam kehidupan sehari-hari, hanya menyebutkan sahabat yang mendengarkan langsung dari Nabi Muhammad Rosululloh Saw. Untuk hadis di atas misalnya hanya disebutkan, “Ibnu Umar mendengar bahwa Nabi Muhammad Rosululloh Saw bersabda, ‘Asas Islam itu ada lima, yakni syahadat, sholat, zakat, puasa, dan haji.’”

Terkadang suatu hadis juga diriwayatkan oleh beberapa Imam Ahli Hadis. Dalam kata lain, hadis yang sama selain termuat dalam Shohih Bukhori, juga terdapat dalam shohih Muslim, Abu Dawud, Ibnu Majah, Nasa’i, dan Tirmidzi, meskipun ada yang sanadnya berbeda. Ulama-ulama hadis berikutnya, yang menyusunnya berdasarkan karya para imam tersebut tidak perlu menyebutkan seluruh nama imam itu. Untuk menghemat pencantuman nama-nama perawi terakhir, mereka cukup mencantumkan bilangan jumlah periwayatnya. Misalnya, pencantuman Tiga Ahli Hadis, berarti hadis itu diriwayatkan oleh tiga dari imam hadis yang tersebut di atas.

Kitab hadis yang telah beredar di dunia Islam cukup beragam, tetapi tidak semuanya mendapat pengakuan dan dipakai oleh masyarakat secara luas. Bab ini mengetengahkan pengelompokan kitab hadis, enam peringkat kitab hadis, dan biografi singkat ulama penyusunnya. Ahli-ahli hadis yang mengumpulkan, mendaftar, menyeleksi dan menuliskan hadis-hadis dalam suatu kitab hadis dikenal sebagai mudawwin atau mukharrij. Salah satu kumpulan kitab hadis yang terkenal adalah Kitab Sunni atau Kutubut Tis’ah (9 Kitab). Sembilan kitab tersebut adalah:

- Shahih Bukhari, disusun oleh Bukhari (194-256 H).
- Shahih Muslim, disusun oleh Muslim (204-262 H).
- Sunan Abu Dawud, disusun oleh Abu Dawud (202-275 H).
- Sunan at-Turmudzi, disusun oleh At-Turmudzi (209-279 H).
- Sunan an-Nasa’i, disusun oleh an-Nasa’i (215-303 H).
- Sunan Ibnu Majah, disusun oleh Ibnu Majah (209-273).
- Musnad Ahmad, disusun oleh Imam Ahmad bin Hambal (164-241 H).
- Muwatta Malik, disusun oleh Imam Malik (93-179 H).

- Sunan Darimi, disusun oleh Ad-Darimi (181-255 H).

Kitab Shohih Bukhori (Al-Jami' as-Shohih Bukhori), berisi 7.275 hadis yang merupakan seleksi dari 600.000 hadis. Kitab ini juga memuat fatwa sahabat dan tabi'in sebagai penjelasan terhadap hadis yang diketengahkan. Sebagian ulama berpendapat bahwa nilai Shohih Bukhori lebih tinggi dibanding Shohih Muslim. Alasan mereka karena syarat yang ditetapkan oleh Imam Bukhori lebih ketat dibandingkan dengan syarat yang ditetapkan oleh Imam Muslim. Syarat yang dimaksud, di antaranya, Bukhori menetapkan liqoo (bertemu antara rawi yang menyampaikan dengan rawi yang menerima). Sebaliknya, bagi Imam Muslim, syarat tersebut cukuplah mu'aasarah (perawi yang menyampaikan dan perawi yang menerima itu hidup dalam satu masa).

Kitab Shohih Muslim (Al-Jami' as Shohih Muslim). Para ulama berselisih pendapat tentang jumlah hadis yang termuat dalam kitab ini. Muhammad Ajaj al-Khotib (guru besar hadis di Universitas Damsyik) menyatakan, hadis yang termuat dalam kitab ini berjumlah 3.030 hadis tanpa pengulangan, dan jika dengan pengulangan berjumlah 10.000. Menurut al-Khuli (seorang ulama dan ahli hadis di Mesir), Shohih Muslim berisi 4.000 hadis jika tanpa pengulangan, sedangkan jika dengan pengulangan berjumlah 7.275. Ulama wilayah Maghribi menilai peringkat Shohih Muslim lebih tinggi dari kitab Shohih Bukhori. Mengapa? Sebab walau persyaratan yang digunakan oleh Imam Muslim lebih longgar, namun kitab ini dipandang telah memenuhi syarat minimal. Sebaliknya, persyaratan yang diterapkan oleh Imam Bukhori yang lebih ketat, mereka nilai berlebihan.

Kitab Sunan Abu Dawud memuat 4.800 hadis, hasil seleksi dari 50.000 hadis yang dihafal oleh Imam Abu Dawud. Kitab ini dinamakan Kitab Sunan, sebab mengemukakan penjelasan ketidakshohihan hadis-hadis. Itulah sebabnya kitab ini memuat hadis-hadis yang shohih, mendekati shohih, dan menyerupai shohih. Ciri yang menonjol dari karya Abu Dawud ini adalah kaya dengan hadis hadis yang berkaitan dengan masalah hukum.

Kitab Hadis Sunan Tirmidzi (Jaami' at Tirmidzi). Disebut kitab sunan karena menjelaskan rawi dan derajat hadisnya. Sebagaimana kitab-kitab lain yang bernama jaami', cakupan bahasannya masalah-masalah populer, antara lain masalah keimanan (al-Aqidah), pemberi rezeki (ar-Rozzaaq), etika makan-minum

(Adab at-Ta'am wa asy-Syurb), hukum (al-Ahkam), serta tafsir, sejarah, dan biografi (at-Tafsiir wa at-Tariikh wa as-Sayr).

Kitab Hadis Sunan Nasa'i (Sunan as-Sugroo atau Sunan al-Mujtabaa), memuat 5.761 hadis hasil seleksi dari hadishadis yang terdapat dalam kitab As-Sunan al-Kubro karya Imam Nasa'i sebelumnya. Isi As-Sunan al-Kubro diseleksi kembali karena tidak hanya memuat hadis-hadis hasan dan shohih, melainkan juga mengemukakan hadis-hadis dhoif.

Kitab Hadis Sunan Ibnu Majah. Sebenarnya sebagian ulama mengeluarkan kitab ini dari Kutubus Sittah (keenam peringkat kitab hadis) karena selain memuat hadis shohih dan hasan, juga memuat hadis dhoif dan hadis mungkar (hadis yang sangat lemah).

Musnad Ahmad atau ringkasnya dikenali sebagai al-Musnad adalah salah satu kitab hadis Nabi yang terkenal dan terluas, dan kedudukannya menempati posisi yang diutamakan di kalangan Ahlus Sunnah sebagai induk rujukan di kalangan mereka. Selain itu, ia juga dikenal sebagai musnad yang paling terkenal, dan para ahli ilmu hadis meletakkan posisinya no 3 setelah Shahihain dan Sunan yang Empat. Nama Musnad Ahmad didasarkan/dinasabkan dari nama Abu Abdullah Ahmad bin Muhammad bin Hanbal asy-Syaibani adz-Dzuhli (164-241 H/780-855 M).

Perhitungan ahli-ahli hadis menyebutkan ada lebih kurang 40 ribu hadis dengan rincian sebanyak 10 ribunya diulang-ulang, ditulis berurutan sesuai nama para Sahabat Nabi Muhammad yang meriwayatkan hadisnya, yang dalam pengurutannya ia jadikan tiap periwayatan sahabat memiliki satu tempat, dan jumlah sahabat yang diriwayatkan di sini terhitung sebanyak 904 orang. Kitab itu ia bagi dalam 18 bagian, dan bagian permulaannya ialah musnad sepuluh orang yang dijanjikan masuk surga, dan bagian terakhirnya ialah musnad sahabat Nabi yang perempuan (shahabiyah). Dan di sana, banyaklah hadis sahih yang tak didapati dalam Shahihain (yakni riwayat Bukhari dan Muslim).

Ad-Darimi ialah salah satu Imam Ahli Hadis Sunni, Nama lengkapnya Abdullah bin Abdurrahman bin al Fadhl bin Bahram bin Abdush Shamad. Ad Darimi adalah nisbah kepada Darim bin Malik dari kalangan at-Tamimi. Dan

dengan nisbah ini dia terkenal. Ia di lahirkan pada taun 181 H dan wafat pada tahun 255 H, dalam usia 75 tahun.

Al-Muwaththa atau Muwatta Malik merupakan kitab hadis dan fiqih yang disusun oleh Imam Malik bin Anas, merupakan salah satu dari Kutubut Tis'ah (sembilan kitab hadis utama di kalangan Sunni). Imam Malik, Yaitu Abu Abdullah Malik bin Anas bin Malik bin Abu Amir Al-Ashbahi (93 H dan -179 H). Ia banyak tinggal di Madinah. Ia ulama Islam yang terkenal, dan pendiri mazhab Maliki. Ia dikenal mempunyai lebih dari seribu murid di antaranya yang terkenal adalah Imam Syafe'i. Imam Malik senantiasa memperbarui Kitab Muwaththa, sehingga kitab ini mencerminkan pembelajaran dan pengetahuan selama lebih dari empat puluh tahun. Kitab ini mengandung seribuan hadis.

Jumlah hadis pada tiap percetakan dan buku terkadang berbeda, sehingga terkadang penomoran hadis juga berbeda. Lebih baik langsung menggunakan isi ketika hendak mencari tentang hadis, bukan dengan penomoran.

2.2. Preprocessing

Pemrosesan Teks (*Text Preprocessing*) adalah suatu proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan untuk proses yang lebih lanjut (Uysal & Gunal, 2014). *Text Preprocessing* umumnya terdiri dari *case folding*, *tokenization*, *filtering* dan *stemming* (Deposit et al., 2018). Masing-masing proses memiliki manfaat terhadap dokumen yang diolah (Uysal & Gunal, 2014).

Case folding adalah konversi bentuk string menjadi lowercase (Uysal & Gunal, 2014). Pada *case folding* bermanfaat dalam proses selanjutnya pada *stemming* karena pada *stemming* saat penghapusan imbuhan, dia mengenali karakter huruf besar dan huruf kecil.

Filtering adalah pemfilteran *term-term* yang tidak mengandung makna atau biasa disebut *stopword*, pada proses *filtering term* yang termasuk dalam *stopword* dihilangkan (Tala, 2003). Langkah ini melakukan penghapusan *stopword* yang bermanfaat saat proses pembobotan. Karena semakin banyak *term* yang tidak mengandung makna yang ikut dalam proses pembobotan, maka akan semakin terganggu hasil bobot yang mewakili setiap dokumen.

Stemming adalah proses yang menyediakan pemetaan variasi *term* morfologis yang berbeda ke dalam kata dasar / umum mereka (Tala, 2003). *Stemming* untuk bahasa Indonesia telah mengalami banyak perkembangan. Pertama kali pengembangan *stemming* bahasa Indonesia dilakukan oleh Adriani, M dan Nazief, B. pada tahun 1996 dengan diberi nama *Confix Stripping Algorithm* (Adriani, M., Nazief, B., Asian, J., Tahaghoghi, S. M. M., and Williams, 2007). Algoritma tersebut menggunakan tata bahasa Indonesia di dalam proses penghapusan imbuhan katanya, dengan melakukan pengecekan kebenaran kata dasar pada Kamus Besar Bahasa Indonesia. Selanjutnya algoritma *Confix Stripping* dikembangkan kembali oleh Arifin, Z.A. dan Setiono, A.N. pada tahun 2002 (Adriani, M., Nazief, B., Asian, J., Tahaghoghi, S. M. M., and Williams, 2007). Algoritma tersebut melakukan pendekatan dengan menghilangkan *prefiks* terlebih dahulu kemudian *suffiks*. Algoritma selanjutnya pada pengembangan *stemming* bahasa Indonesia adalah *Porter Indonesian* (Tala, 2003). Algoritma ini melakukan proses *stemming* dengan mengadaptasi algoritma *Porter Stemming* pada bahasa Inggris. *Porter stemming* pada bahasa Inggris melakukan penghapusan imbuhan secara sederhana tanpa menggunakan pengecekan terhadap kamus kata dasar. Proses tersebut diadaptasikan terhadap bahasa Indonesia. Algoritma pengembangan *stemming* bahasa Indonesia adalah *Enhanced Confix Stripping* (ECS) (Mahendra, 2007). Kemudian algoritma tersebut mengalami pengembangan pada penelitian (Andita Dwiyoga Tahitoe, 2010).

Tokenization adalah pecahan dokumen teks menjadi satuan *term* (Uysal & Gunal, 2014). Proses tersebut bermanfaat untuk awal mula pembentukan matriks yang nantinya akan dilanjutkan dalam perhitungan matriks pembobotan tiap *term*.

2.3. Term Weighting

Term Weighting merupakan proses penghitungan bobot tiap *term* yang dicari pada setiap dokumen sehingga dapat diketahui ketersediaan dan kemiripan suatu *term* di dalam dokumen (Sabbah et al., 2017). Hal yang dipelajari di sini adalah bagaimana *term weighting* bekerja saat melakukan pembobotan tiap *term* dan bagaimana hasil yang diperoleh.

2.3.1. Term Frequency (TF)

Term frequency (TF) adalah setiap *term* diasumsikan memiliki kepentingan yang proporsional terhadap jumlah kemunculan *term* pada dokumen. *Term frequency* (TF) merupakan metode yang paling sederhana dalam membobotkan *term*. *Term Frequency* merupakan frekuensi dari kemunculan sebuah *term* dalam dokumen yang bersangkutan. Semakin besar jumlah kemunculan suatu *term* (TF tinggi) dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar (Chen et al., 2020).

Term Frequency (TF) memiliki beberapa jenis formula yang dapat digunakan :

- TF biner (*binary TF*), hanya memperhatikan apakah suatu *term* atau *term* ada atau tidak dalam dokumen, jika ada diberi nilai satu (1), jika tidak diberi nilai nol (0).
- TF murni (*raw TF*), nilai TF diberikan berdasarkan jumlah kemunculan suatu *term* di dokumen. Contohnya, jika muncul lima (5) kali maka *term* tersebut akan bernilai lima (5).
- TF logaritmik, hal ini untuk menghindari dominansi dokumen yang mengandung sedikit *term* dalam query, namun mempunyai frekuensi yang tinggi. TF logaritmik divisualisasikan pada persamaan (2.1).

$$TF = \begin{cases} 1 + \log_{10}(f_{t,d}), & f_{t,d} > 0 \\ 0, & f_{t,d} = 0 \end{cases} \quad (2.1)$$

dimana nilai $f_{t,d}$ adalah frekuensi *term* (t) pada dokumen (d). Jadi jika suatu *term* atau *term* terdapat dalam suatu dokumen sebanyak 5 kali maka diperoleh bobot = $1 + \log(5) = 1.699$. Tetapi jika *term* tidak terdapat dalam dokumen tersebut, bobotnya adalah nol (0).

2.3.2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) memperhatikan kemunculan *term* pada kumpulan dokumen (Dogan & Uysal, 2019). Fungsi IDF memberikan skor terendah untuk *term* yang muncul dalam banyak dokumen dalam ruang dokumen

$D = d_1, d_2, \dots, d_n$ (Ren & Sohrab, 2013). Latar belakang pembobotan ini adalah *term* yang jarang muncul pada kumpulan dokumen sangat bernilai. Kepentingan tiap *term* diasumsikan memiliki proporsi yang berkebalikan dengan jumlah dokumen yang mengandung *term*. IDF divisualisasikan dalam Persamaan (2.2) dan (2.3).

$$df_{(t_i)} = \sum d_{(t_i)} \quad (2.2)$$

dimana, $df_{(t_i)}$ merupakan frekuensi dokumen yang mengandung *term* i , $d_{(t_i)}$ merupakan dokumen yang mengandung *term* i .

$$idf = \log \frac{D}{df_{(t_i)}} \quad (2.3)$$

dimana, idf merupakan inverse dari dokumen frekuensi ($df_{(t_i)}$), D merupakan jumlah seluruh dokumen.

Kemudian IDF dikombinasikan dengan TF divisualisasikan pada persamaan (2.4).

$$W_{TF*IDF(t_i,d_j)} = tf_{t_i,d_j} \times \left(\log \frac{D}{df_{(t_i)}} \right) \quad (2.4)$$

dimana, $W_{TF*IDF(t_i,d_j)}$ merupakan bobot istilah i pada dokumen j , tf_{t_i,d_j} merupakan jumlah *term* i di dalam dokumen j , D merupakan jumlah seluruh dokumen, dan $df_{(t_i)}$ merupakan jumlah seluruh dokumen yang mengandung *term* i .

2.3.3. Inverse Class Frequency (ICF)

Inverse Class Frequency (ICF) diadopsi dari metode IDF yaitu dengan menggunakan inverse perbandingan jumlah kategori dengan jumlah kategori yang mengandung istilah (Yang et al., 2019). Dalam hal pengindeksan berorientasi kelas, subset dokumen dari ruang dokumen $D = d_1, d_2, \dots, d_n$ dialokasikan ke kategori tertentu. Sehingga semakin sering istilah muncul dalam dokumen yang ada dalam

kategori tersebut, maka bobot istilah semakin mendekati 0. Fungsi ICF memberikan skor terendah *term* yang muncul di beberapa kategori di ruang kategori $C = C_1, C_2, \dots, C_n$. Persamaan dari ICF dapat dilihat pada Persamaan (2.5) dan (2.6).

$$cf_{(t_i)} = \sum c_{(t_i)} \quad (2.5)$$

dimana, $cf_{(t_i)}$ merupakan frekuensi kategori yang mengandung *term* i , $c_{(t_i)}$ merupakan kategori yang mengandung *term* i .

$$icf = \log \frac{C}{cf_{(t_i)}} \quad (2.6)$$

dimana, icf merupakan inverse dari class frekuensi ($cf_{(t_i)}$), C merupakan jumlah seluruh kategori.

Oleh karena itu, representasi numerik dari suatu *term* adalah produk dari *Term Frequency* (parameter lokal), *IDF* (parameter global), dan *ICF* (parameter global kategori). Persamaan TF x IDF x ICF divisualisasikan pada persamaan (2.7).

$$W_{TF*IDF*ICF(t_i,d_j,c_k)} = tf_{t_i,d_j} \times \left(\log \frac{D}{df_{(t_i)}} \right) \times \left(\log \frac{C}{cf_{(t_i)}} \right) \quad (2.7)$$

dimana, $W_{TF*IDF*ICF(t_i,d_j,c_k)}$ merupakan bobot istilah i pada dokumen j , tf_{t_i,d_j} merupakan jumlah *term* i di dalam dokumen j , D merupakan jumlah seluruh dokumen, $df_{(t_i)}$ merupakan jumlah seluruh dokumen yang mengandung *term* i , C merupakan jumlah seluruh kategori, $cf_{(t_i)}$ merupakan jumlah kategori yang mengandung *term* i .

2.3.4. Inverse Class Space Density Frequency (ICS_sF)

ICS_sF menghitung kepadatan dokumen pada ruang kategori berdasarkan setiap *term*. Karena fungsi ICF memberikan skor terendah untuk istilah-istilah yang muncul di beberapa kategori tanpa memperhatikan tentang ruang kategori, perhitungan ICS_sF kemudian diusulkan. Metode *term weighting* ini untuk

meningkatkan kinerja klasifikasi. Dokumen identik ini yang dikaitkan dengan istilah tertentu mungkin merupakan sub-bagian dari kategori tertentu c_k . Karena itu, penting untuk mengeksplorasi karakteristik kemunculan istilah dalam ruang dokumen $D = \{d_1, d_2, \dots, d_n\}$ dan ruang kategori. Dalam pengindeksan berorientasi kategori, subset dokumen dari ruang dokumen global dialokasikan ke kategori tertentu c_k ($k = 1, 2, \dots, m$) sesuai dengan topik mereka. Karena itu, ruang kategori didefinisikan sebagai $C = \{(d_{11}, d_{12}, \dots, d_{1n}) \text{ anggota } C_1, (d_{21}, d_{22}, \dots, d_{2n}) \text{ anggota } C_2, \dots, (d_{m1}, d_{m2}, \dots, d_{mn}) \text{ anggota } C_m\}$ (Ren & Sohrab, 2013). ICS_δF diawali dengan menghitung kepadatan kelasnya (C_δ) yaitu menghitung dokumen yang mengandung *term* pada kategori tertentu (c_k), divisualisasikan pada persamaan (2.8).

$$C_\delta(t_i) = \frac{n_{c_k}(t_i)}{N_{c_k}} \quad (2.8)$$

dimana, $C_\delta(t_i)$ merupakan kepadatan kategori (*class density*) terhadap *term* i , $n_{c_k}(t_i)$ merupakan jumlah dokumen di dalam kategori c_k yang mengandung *term* i , N_{c_k} merupakan jumlah keseluruhan dokumen di dalam kategori c_k .

Kemudian dilanjutkan dengan menghitung kepadatan ruang kategori, yaitu jumlah dari kepadatan keseluruhan kategori yang ada (CS_δ), divisualisasikan pada persamaan (2.9).

$$CS_\delta(t_i) = \sum_{c_k} C_\delta(t_i) \quad (2.9)$$

dimana, $CS_\delta(t_i)$ merupakan kepadatan ruang kategori (*class space density*) terhadap *term* i , $C_\delta(t_i)$ merupakan kepadatan kategori terhadap *term* i , c_k merupakan kategori ($k = 1, 2, \dots, m$).

Kemudian hasil dari kepadatan ruang kategori ($CS_\delta(t_i)$) dilakukan *inverse* sesuai dengan konsep pada TF-IDF-ICF sebelumnya, divisualisasikan pada persamaan (2.10).

$$ICS_{\delta}F(t_i) = \log\left(\frac{C}{CS_{\delta}(t_i)}\right) \quad (2.10)$$

dimana, $ICS_{\delta}F(t_i)$ merupakan *inverse* kepadatan ruang kategori (*inverse class space density frequency*) terhadap *term* i , C merupakan jumlah keseluruhan kategori, $CS_{\delta}(t_i)$ merupakan kepadatan ruang kategori terhadap *term* i .

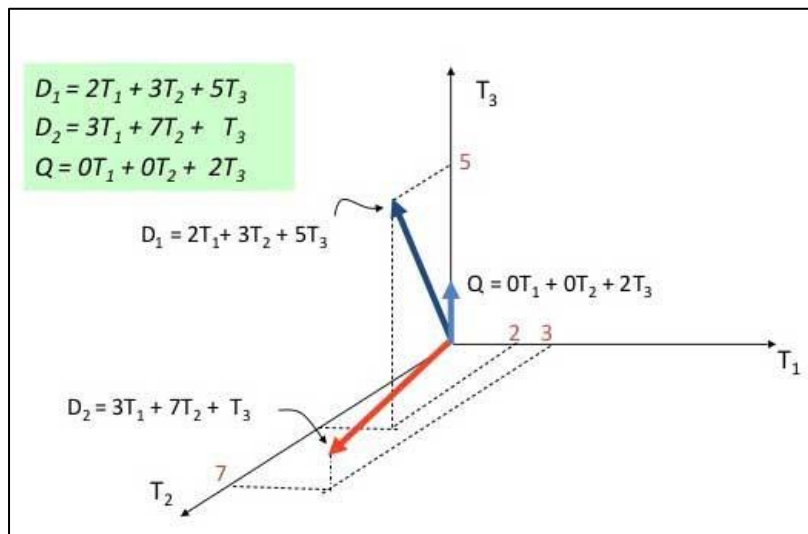
Langkah selanjutnya adalah melakukan perkalian hasil inverse kepadatan ruang kategori terhadap *term* i ($ICS_{\delta}F(t_i)$) dengan TF-IDF, seperti pada persamaan (2.11).

$$W_{TF*IDF*ICS_{\delta}F(t_i,d_j,c_k)} = tf_{t_i,d_j} \times \left(\log\frac{D}{df_{(t_i)}}\right) \times \left(\log\frac{C}{CS_{\delta}(t_i)}\right) \quad (2.11)$$

dimana, $W_{TF*IDF*ICS_{\delta}F(t_i,d_j,c_k)}$ merupakan bobot *term* i pada dokumen j di dalam kategori k , tf_{t_i,d_j} merupakan jumlah *term* i di dalam dokumen j , D merupakan jumlah seluruh dokumen, $df_{(t_i)}$ merupakan jumlah dokumen yang mengandung *term* i , C merupakan jumlah seluruh kategori, $CS_{\delta}(t_i)$ merupakan kepadatan ruang kategori terhadap *term* i .

2.4. Vector space model (VSM)

Pada *Vector space model*, setiap dokumen di dalam database dan query pengguna direpresentasikan oleh suatu vektor multi-dimensi (Lv & El-Gohary, 2016). Dimensi sesuai dengan jumlah *term* dalam dokumen yang terlibat. Contoh dari model ruang vektor tiga dimensi untuk dua dokumen $D1$ dan $D2$, satu *query*



Gambar 2.1. Model Ruang Vektor

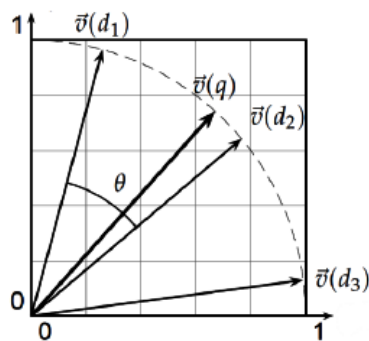
pengguna $Q1$, dan tiga *term* $T1$, $T2$ dan $T3$ diperlihatkan pada Gambar 2.1. Pada model ini:

- Vocabulary merupakan kumpulan semua *term* berbeda yang tersisa dari dokumen setelah preprocessing dan mengandung t *term* index. *Term-term* ini membentuk suatu ruang vektor.
- Setiap *term* i di dalam dokumen atau *query* j , diberikan suatu bobot (*weight*) bernilai real w_{ij} .
- Dokumen dan *query* diekspresikan sebagai vektor t dimensi $d_j = (w_1, w_2, \dots, w_{tj})$ dan terdapat n dokumen di dalam koleksi, yaitu $j = 1, 2, \dots, n$.

Dalam model ruang vektor, koleksi dokumen direpresentasikan oleh matriks *term-document* (atau matriks *term-frequency*). Setiap sel dalam matriks bersesuaian dengan bobot yang diberikan dari suatu *term* dalam dokumen yang ditentukan. Nilai nol berarti bahwa *term* tersebut tidak hadir di dalam dokumen (Lv & El-Gohary, 2016). Pada Gambar 2.2 ditunjukkan contoh matriks *term-document* untuk database dengan n dokumen dan t *term*.

$$\begin{bmatrix}
 & T_1 & T_2 & \dots & T_t \\
 D_1 & w_{11} & w_{21} & \dots & w_{t1} \\
 D_2 & w_{12} & w_{22} & \dots & w_{t2} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 D_n & w_{1n} & w_{2n} & \dots & w_{tn}
 \end{bmatrix}$$

Gambar 2.2. Matriks *Term* Dokumen



Gambar 2.3. Representasi Perumusan *Cosine Similarity*

Keberhasilan dari model VSM ini ditentukan oleh skema pembobotan terhadap suatu *term* baik untuk cakupan lokal maupun global, dan faktor normalisasi (Lv & El-Gohary, 2016). Pembobotan lokal hanya berpedoman pada frekuensi munculnya *term* dalam suatu dokumen dan tidak melihat frekuensi kemunculan *term* tersebut di dalam dokumen lainnya.

2.5. Cosine similarity

Term weighting menghasilkan bobot *term* pada dokumen yang digunakan sebagai representasi vektor. Representasi bobot tersebut dapat dihitung nilai kemiripan suatu dokumen dengan *query*. Nilai kemiripan ini biasa dihitung dengan rumusan *cosine similarity*, perhitungan tingkat kemiripan ini dibuat dengan berdasar pada besar sudut kosinus antara dua vektor, dalam hal ini adalah vektor dokumen. Representasi perumusan ini dalam bidang kartesian seperti diperlihatkan pada Gambar 2.3.

Dimisalkan pada Gambar 2.3. terdapat tiga vektor dokumen d_1 , d_2 dan d_3 dan satu vektor *query* q . cosine similarity menghitung nilai kosinus θ dari *query* dan tiga dokumen lain. Nilai ini menunjukkan derajat kemiripan dokumen dengan *query*. Karena berdasarkan kosinus sudut antara dua vektor, maka nilainya berkisar pada 0 sampai dengan 1, dimana 0 menandakan bahwa kedua dokumen tidak mirip sama sekali, dan 1 menandakan bahwa antara *query* dan dokumen benar-benar identik. *Cosine* dinyatakan pada persamaan (2.12)

$$\cos(q, d_j) = \frac{\sum t_k [W(t_k, q)] \times [W(t_k, d_j)]}{\sqrt{\sum |W_Q|^2} \times \sqrt{\sum |W_J|^2}} \quad (2.12)$$

dimana, $\cos(q, d_j)$ merupakan nilai kosinus antara *query* dan dokumen j , sedangkan $W(t_k, q)$ dan $W(t_k, d_j)$ adalah bobot *term* t_k pada *query* dan dokumen j . $|W_Q|^2$ dan $|W_J|^2$ adalah panjang dari vektor *query* dan dokumen j .

2.6. Confusion Matrix

Information Retrieval bertugas untuk mengembalikan sekumpulan dokumen sebagai jawaban dari *query* pengguna. Terdapat dua kategori dokumen yang dihasilkan oleh *Information Retrieval* terkait pemrosesan *query*, yaitu *relevant documents* (dokumen yang relevan dengan *query*) dan *retrieved documents* (dokumen yang diterima pengguna). Ukuran umum yang digunakan untuk mengukur kualitas dari data *retrieval* adalah kombinasi *precision* dan *recall*. *Precision* merupakan proporsi dari suatu set yang diperoleh yang *relevant*. *Recall* merupakan proporsi dari semua hasil yang relevan dikoleksi termasuk hasil yang diperoleh atau dikembalikan. *F-measure* biasa digunakan pada bidang sistem temu kembali informasi untuk mengukur klasifikasi pencarian dokumen dan performa *query classification*. Evaluasi ini didapatkan dengan mengacu pada *confusion matrix*.

Confusion matrix berisi informasi mengenai klasifikasi yang sebenarnya dan prediksi klasifikasi yang dilakukan oleh sistem. Tabel confusion matrix dapat dilihat pada Tabel 2.1. Nilai *recall* dan *precision* sistem temu kembali informasi tersebut dapat dinyatakan sebagai Persamaan (2.13) dan Persamaan (2.14). Sebagaimana yang ditunjukkan pada Tabel 2.1., TP (*true positive*) merupakan jumlah dokumen relevan yang ter-retrieve secara tepat oleh sistem. FP (*false positive*) merupakan jumlah dokumen tidak relevan yang ter-retrieve. TN (*true negative*) merupakan jumlah dokumen relevan yang tidak ter-retrieve.

Tabel 2.1. *Confusion Matrix*

	Relevant	Non relevant
Retrieved	TP	FP
Not Retrieved	FN	TN

Sedangkan FN (*false negative*) merupakan jumlah dokumen tidak relevan yang tidak ter-retrieve. Berdasarkan evaluasi menggunakan precision, maka efektivitas dari metode yang digunakan dapat diketahui.

$$Recall (R) = \frac{TP}{TP+FN} \quad (2.13)$$

$$Precision (P) = \frac{TP}{TP+FP} \quad (2.14)$$

Pada dasarnya, nilai *recall* dan *precision* berada pada rentang antara 0 sampai dengan 1. Oleh karena itu, suatu sistem temu kembali yang baik adalah yang dapat memberikan nilai *recall* dan *precision* mendekati 1. Nilai *recall* atau *precision* saja belum cukup mewakili kinerja sistem. Oleh karena itu diperlukan metode evaluasi yang mengkombinasikan metode evaluasi recall dan precision metode evaluasi ini adalah *F-score*. Formulasi *F-score* dinyatakan dalam Persamaan (2.15).

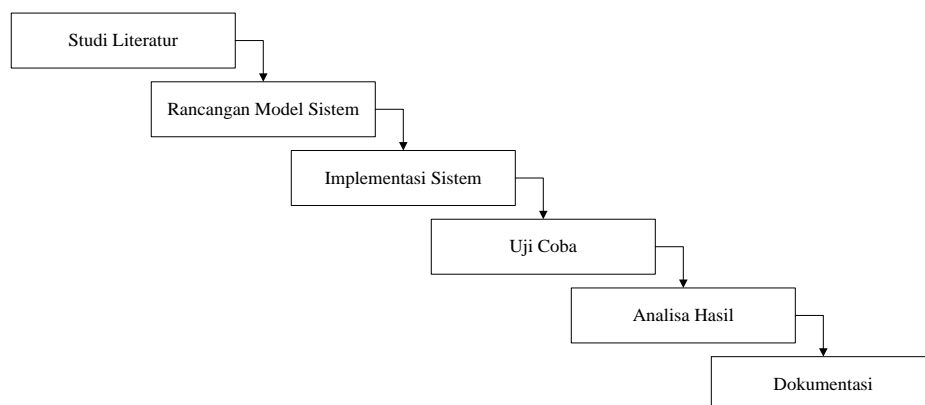
$$F - Score (F) = \frac{2 \times R \times P}{R+P} \quad (2.15)$$

[Halaman ini sengaja dikosongkan]

BAB III

METODE PENELITIAN

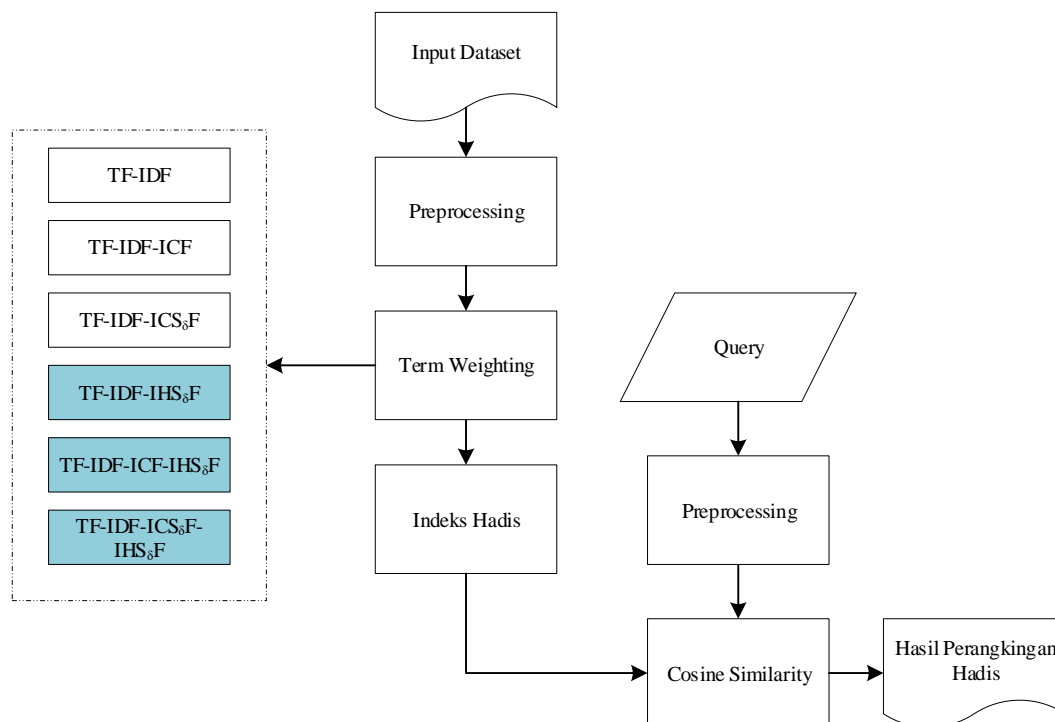
Bab ini merupakan pemaparan rancangan tahap penelitian yang akan dilakukan di dalam penelitian ini. Tahapan penelitian tersebut digambarkan pada Gambar 3.1. Tahapan penelitian yang dipaparkan pada Gambar 3.1. meliputi 5 tahapan yaitu: studi literatur, rancangan model sistem, implementasi sistem, uji coba, analisa hasil, dan dokumentasi. Tahap studi literatur merupakan tahapan pengumpulan kajian pustaka dan dasar teori yang digunakan di dalam penelitian ini. Tahap selanjutnya adalah rancangan model sistem yang merupakan pemaparan rancangan model sistem yang akan dibentuk di dalam penelitian ini. Tahap selanjutnya adalah implementasi sistem yang merupakan implementasi dari seluruh rancangan sistem yang dibentuk pada tahapan sebelumnya ke dalam sistem klasifikasi. Tahapan selanjutnya adalah evaluasi yang merupakan tahapan untuk mengetahui kualitas sistem yang dibangun dan untuk mengetahui apakah sistem telah memenuhi tujuan awal penelitian atau tidak. Tahapan selanjutnya adalah analisa hasil yang merupakan tahapan analisa hasil dari sistem yang dibangun dengan mencari kelebihan dan kekurangannya untuk mendapatkan kesimpulan. Tahapan selanjutnya adalah dokumentasi yang merupakan tahapan penulisan laporan keseluruhan kegiatan penelitian dan dokumentasi hasil.



Gambar 3.1. Tahap Penelitian

3.1. Rancangan Model Sistem

Penelitian ini bertujuan untuk membangun sistem pencarian hadis dengan metode representasi teks yang diusulkan yaitu metode pembobotan *term* yang mengkombinasikan faktor kepadatan ruang kategori dan kitab. Proses sistem yang akan dibangun dimulai dari input dataset hadis yang telah dikumpulkan ke dalam database. Kemudian hadis dilakukan *preprocessing* untuk menyeleksi fitur-fitur yang lebih penting untuk dilakukan pembobotan *term* yaitu: *lowercase* (pengubahan huruf menjadi huruf kecil semua), penghapusan karakter, penghapusan *term-term* yang termasuk ke dalam stopword, dan penghapusan imbuhan kata (*stemming*). Proses selanjutnya adalah proses perhitungan *term weighting* yang terdiri dari beberapa *term weighting* yang akan digunakan sebagai perbandingan hasil. Kemudian setelah bobot tersimpan, dilakukan dengan melakukan uji coba dengan memasukkan beberapa *query* yang telah memiliki *grountruth* sebelumnya. Bobot dari dataset dihitung jaraknya dengan *query* dengan menggunakan *cosine similarity*. Bagian besar proses sistem dipaparkan pada Gambar 3.2.



Gambar 3.2. Rancangan Sistem Secara Keseluruhan

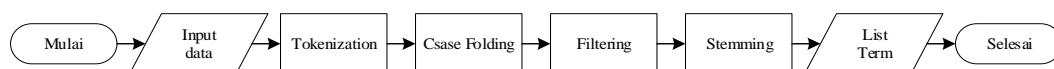
Tabel 3.1. Jumlah Hadis dan Jumlah Kategori

Kitab	Jumlah Hadis	Jumlah Kategori
Shahih Bukhari	7008 hadis	77 kategori
Shahih Muslim	5361 hadis	56 kategori
Sunan Abu Dawud	4590 hadis	35 kategori
Sunan at-Turmudzi	3891 hadis	49 kategori
Sunan an-Nasa'i	5661 hadis	51 kategori
Sunan Ibnu Majah	4331 hadis	32 kategori
Musnad Ahmad	26363 hadis	14 kategori
Muwatta Malik	1594 hadis	32 kategori
Sunan Darimi	3367 hadis	24 kategori

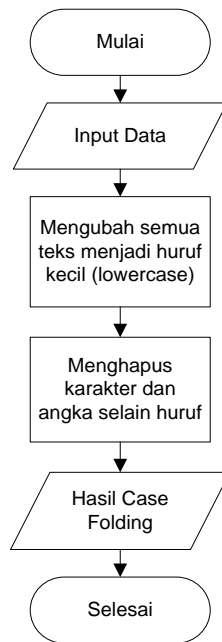
3.1.1. Dataset

Data yang digunakan di dalam penelitian ini merupakan kumpulan Hadis dari 9 Imam yang akan didapatkan dari website <https://tafsirq.com/>, yaitu: Shahih Bukhari, Shahih Muslim, Sunan Abu Dawud, Sunan at-Turmudzi, Sunan an-Nasa'i, Sunan Ibnu Majah, Musnad Ahmad, Muwatta Malik, dan Sunan Darimi. Masing-masing jumlah Hadis yang digunakan pada setiap kitab, yaitu: Shahih Bukhari sebanyak 7008 hadis, Shahih Muslim sebanyak 5361 hadis, Sunan Abu Dawud sebanyak 4590 hadis, Sunan at-Turmudzi sebanyak 3891 hadis, Sunan an-Nasa'i sebanyak 5661 hadis, Sunan Ibnu Majah sebanyak 4331 hadis, Musnad Ahmad sebanyak 26363 hadis, Muwatta Malik sebanyak 1594 hadis, dan Sunan Darimi sebanyak 3367 hadis. Perincian jumlah hadis yang digunakan dapat dilihat pada Tabel 3.1. Kumpulan hadis tersebut dikategorikan sesuai dengan tema yang dituangkan di dalam tiap hadisnya.

Rincian kategori pada tiap kitab dipaparkan pada Lampiran 1. Lampiran 1 menjelaskan detail kategori pada setiap kitab beserta jumlah hadis di dalam setiap kategori. Ada beberapa kitab yang memiliki kategori sama, namun isi hadis di dalamnya berbeda.



Gambar 3.3. *Preprocessing*

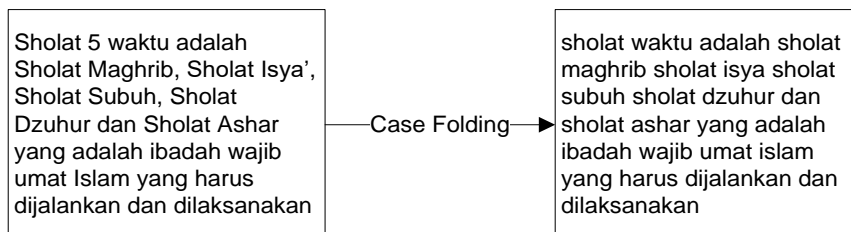


Gambar 3.4. *Flowchart* Proses *Case Folding*

3.1.2. Preprocessing

Penelitian ini melakukan *preprocessing* antara lain: *case folding*, *tokenization*, *filtering* dan *stemming*. Proses *preprocessing* digambarkan dalam *flow chart* pada Gambar 3.3.

Case folding merupakan penyeragaman bentuk string menjadi *lowercase*. Saat *case folding*, dilakukan juga penghapusan karakter dan angka pada setiap dokumen. Proses *case folding* dapat dilihat pada Gambar 3.4. dan simulasi hasil dapat dilihat pada Gambar 3.5.

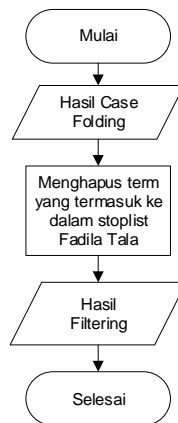


Gambar 3.5. Contoh Proses *Case Folding*

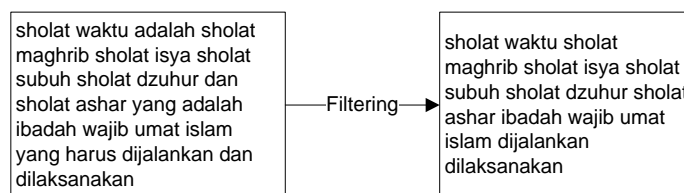
Tabel 3.2. Contoh Stoplist Fadilla Tala

ada	Amatlah	akhiri	agaknya
adalah	Anda	akhirnya	agar
adanya	Andalah	aku	akan
adapun	Antar	akulah	akankah
agak	Antara	amat	akhir

Proses selanjutnya adalah *filtering* yang merupakan proses penghapusan *stopword*. *Stopword* adalah beberapa *term* yang tidak mengandung makna, contohnya: aku, dia, kamu, sebab, karena, oleh, adalah, yang dan lain sebagainya (menggunakan stoplist dari Fadilla Tala). Contoh *term-term* yang termasuk ke dalam stoplist Fadilla Tala ada pada Tabel 3.2. *Filtering* bertujuan untuk mengurangi proses berlebihan dengan menghapus *term* yang tidak berpengaruh terhadap proses sistem selanjutnya. Penggambaran proses *Filtering* dapat dilihat pada Gambar 3.6. dan Contoh penerapannya pada kalimat dapat dilihat pada Gambar 3.7.



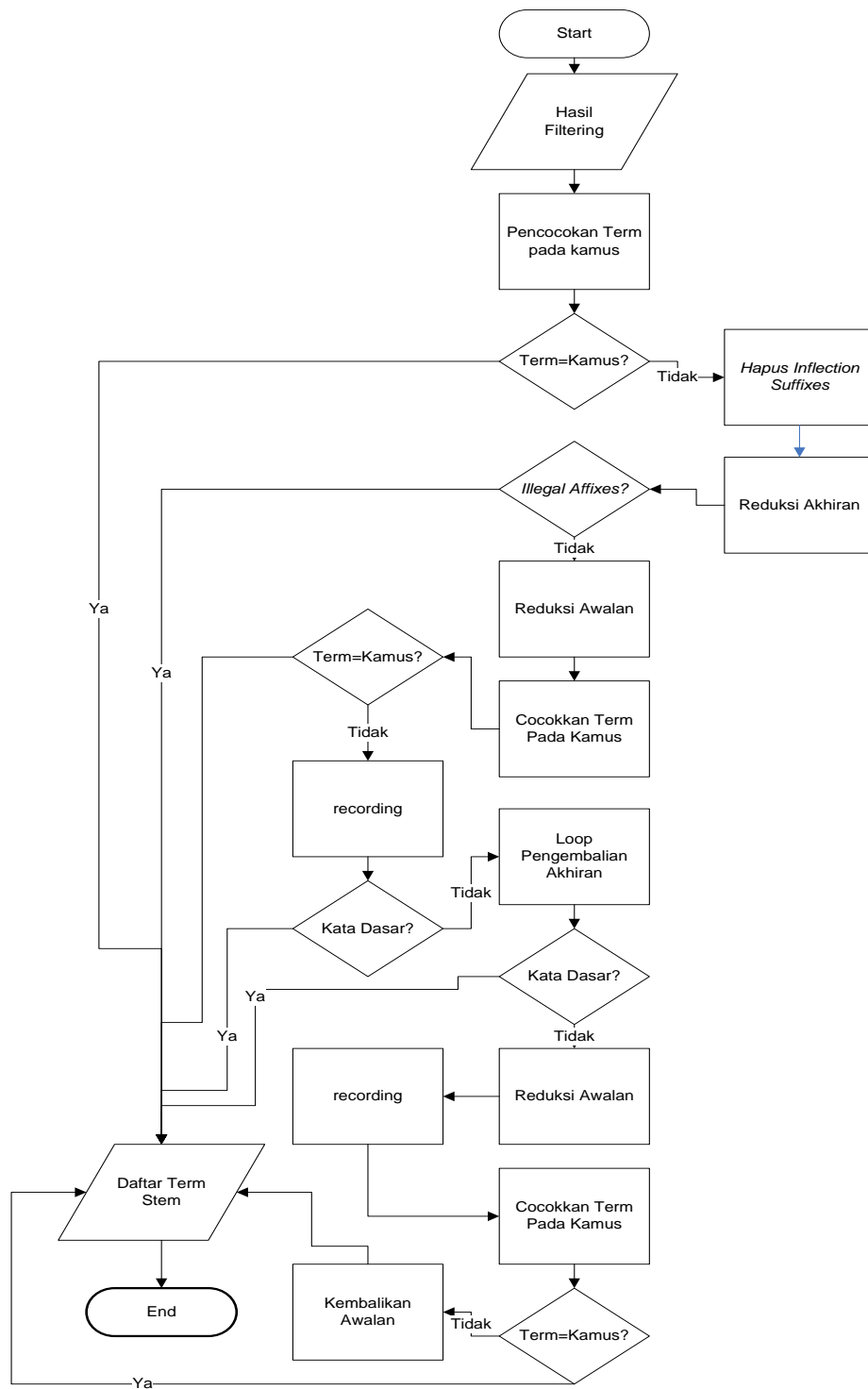
Gambar 3.6. *Flowchart* Proses *Filtering*



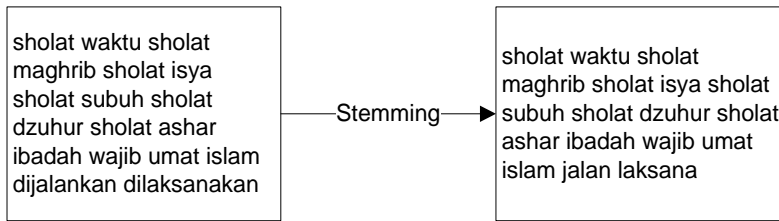
Gambar 3.7. Contoh Proses *Filtering*

Langkah selanjutnya adalah *stemming*. *Stemming* merupakan proses penyederhanaan *term* menjadi kata dasarnya (menggunakan library sastrawi). Penggambaran proses *stemming* pada Gambar 3.8. dan contoh hasil dapat dilihat pada Gambar 3.9. Library sastrawi menggunakan metode *Enhanced Confix Stripping* (ECS) di dalam sistemnya. Proses penghapusan imbuhan pada ECS melakukan aturan morfologi yang ada pada bahasa Indonesia. Urutan proses *stemming* yaitu melakukan pengecekan terhadap kamus terlebih dahulu, jika kata ada di dalam kamus, proses akan berhenti. Namun, jika kata tidak ada di dalam kamus maka proses penghapusan imbuhan akan dilakukan. Penghapusan imbuhan dimulai dengan menghapus partikel (-kah, -lah, -tah, -pun). Kemudian dicocokkan kembali ke dalam kamus, jika kata sudah ada di dalam kamus maka proses penghapusan imbuhan berhenti. Namun, jika tidak ada di dalam kamus maka proses penghapusan imbuhan akan berlanjut. Proses penghapusan imbuhan selanjutnya adalah menghapus imbuhan milik (-ku, -mu, -nya). Proses selanjutnya pengecekan terhadap kamus seperti pada proses sebelumnya. Proses selanjutnya pengecekan terhadap *illegal affix* atau imbuhan yang dilarang oleh tata bahasa baku bahasa Indonesia.

Kemudian proses penghapusan imbuhan akhiran dan awalan dilakukan. Hingga kata dasar yang telah sesuai ada di dalam kamus kata dasar bahasa Indonesia ditemukan di dalam sistem. Proses *stemming* menggunakan algoritma ini masih sangat bergantung terhadap kamus kata dasar Bahasa Indonesia sehingga memerlukan waktu yang agak lama dalam prosesnya. Karena proses ini menerapkan algoritma kata perkata. Namun, dalam penelitian terkait dengan *stemming* bahasa Indonesia yang ada teknik ini merupakan teknik yang menghasilkan kata dasar paling akurat karena hasil akhir dari *stemming* ini akan selalu dicek terhadap kamus terlebih dahulu.

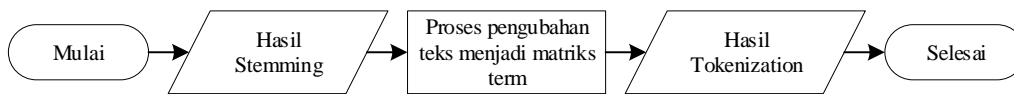


Gambar 3.8. Flowchart Proses Stemming

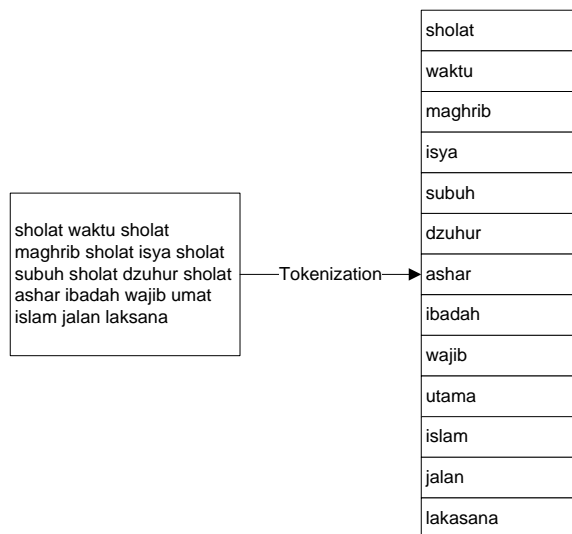


Gambar 3.9. Contoh Proses *Stemming*

Proses selanjutnya adalah *tokenization* yang merupakan proses pemecahan teks menjadi satuan *term*, tujuan dilakukannya *tokenization* adalah untuk mempermudah dalam perhitungan matrix pada proses selanjutnya. Saat proses *tokenization* dilakukan pula penghapusan *term* yang sama muncul dalam seluruh dokumen. Proses *tokenization* dapat dilihat pada Gambar 3.10. dan contoh hasil proses ada pada Gambar 3.11. Proses ini bertujuan untuk memudahkan dalam pembentukan matriks pembobotan tiap *term* pada proses selanjutnya. Selain itu, proses ini juga melakukan penggabungan *term* yang muncul bersamaan.



Gambar 3.10. *Flowchart* Proses *Tokenization*



Gambar 3.11. Contoh Hasil Proses *Tokenization*

Tabel 3.3. Contoh Perhitungan *Term Frequency*

term	Term Frequency												
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1		c2	c3			c4	c5	c6	c7	c8	c9
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12
malu	1	1	2	3	1	1	2	1	2	1	2	1	1
iman	1	1	1	1	2	2	1	2	1	2	1	2	2
milik	0	1	0	0	0	0	0	0	0	0	0	0	0
enam	0	1	0	0	0	2	0	0	0	0	0	0	0
puluh	0	1	0	0	2	4	0	1	0	0	0	1	1
cabang	0	1	0	0	1	1	0	1	0	0	0	2	2
rasulullah	0	0	1	1	0	0	0	0	1	0	2	0	0
...
...
lewat	0	0	1	1	0	0	1	0	1	0	1	0	0

3.1.3. Term Weighting

Penelitian ini melakukan perbandingan metode *term weighting*. Pada subbab-subbab di bawah ini memaparkan penjelasan dan contoh proses yang akan dilaksanakan pada penelitian. Subbab *term weighting* pada penelitian ini disertai dengan contoh perhitungan dengan beberapa dokumen hadis. Hadis yang akan digunakan dalam contoh perhitungan berada pada Lampiran 2. *Query* yang digunakan adalah “malu adalah sebagian dari iman”.

3.1.3.1. Term Frequency (TF)

Pertama adalah melakukan perhitungan TF (*Term Frequency*) yaitu jumlah kemunculan *term* pada tiap dokumen. Penelitian ini menggunakan metode TF murni (*raw TF*) untuk perhitungan *Term Frequency*-nya, nilai TF diberikan berdasarkan jumlah kemunculan suatu *term* di dokumen. Contohnya, jika muncul lima (5) kali maka *term* tersebut akan bernilai lima (5). Penggambaran potongan perhitungan *term frequency* dapat dilihat pada Tabel 3.3. Tabel 3.3. terdiri dari beberapa kolom, diantaranya adalah kolom *q* yaitu kolom *query* yang berisi jumlah kemunculan *term* pada *query*. Kolom c1 merupakan kolom kategori 1 yang berisi beberapa dokumen yang ada pada kategori 1, begitu pula pada kolom c2 hingga c9. Selanjutnya adalah kolom d1 merupakan kolom dokumen 1 yang berisi jumlah kemunculan *term* pada dokumen 1, begitu pula pada kolom d2 hingga d12. Jumlah

Tabel 3.4. Contoh Perhitungan DF dan IDF

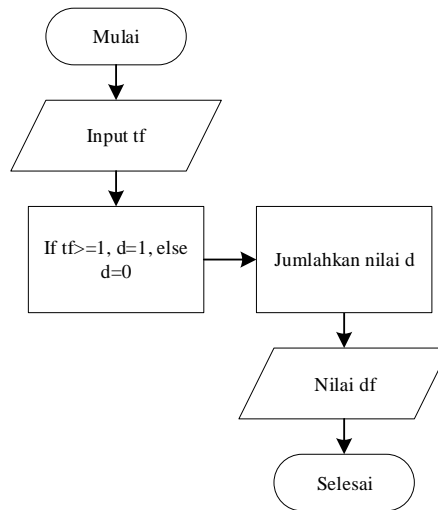
term	tf												df	idf = log(N/df)		
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai			Ahmad	
		c1		c2	c3			c4	c5	c6		c7			c8	c9
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11			d12	
malu	1	1	2	3	1	1	2	1	2	1	2	1	1	12	3.714365	
iman	1	1	1	1	2	2	1	2	1	2	1	2	2	12	3.714365	
milik	0	1	0	0	0	0	0	0	0	0	0	0	0	1	4.793546	
enam	0	1	0	0	0	2	0	0	0	0	0	0	0	2	4.492516	
puluh	0	1	0	0	2	4	0	1	0	0	0	1	1	6	4.015395	
cabang	0	1	0	0	1	1	0	1	0	0	0	2	2	6	4.015395	
rasulullah	0	0	1	1	0	0	0	0	1	0	2	0	0	4	4.191486	
....	
...	
lewat	0	0	1	1	0	0	1	0	1	0	1	0	0	5	4.094576	

kemunculan *term* dihitung berdasarkan kemunculannya pada tiap dokumen dan *query*. Tabel 3.3. menunjukkan bahwa jumlah kemunculan *term* rata-rata adalah 1, 0, dan 2.

3.1.3.2. Inverse Document Frequency (IDF)

Langkah kedua adalah melakukan penghitungan IDF (*Inverse Document Frequency*). Sebelum proses perhitungan IDF dilakukan terlebih dahulu perhitungan DF (*Document Frequency*), karena IDF merupakan hasil *Invers* dari DF. DF dihitung dengan menjumlahkan keseluruhan dokumen yang mengandung *term*. *Flowchart* perhitungan DF dipaparkan pada Gambar 3.12. Kemudian IDF dihitung dengan melakukan logaritma dari perbandingan jumlah keseluruhan kelas dengan jumlah dokumen yang mengandung *term*. Perhitungan IDF dilakukan dengan rumus yang telah dipaparkan pada Bab II persamaan (2.2) dan (2.3). Contoh perhitungannya dapat dilihat pada Tabel 3.4. Kolom baru yang ada pada Tabel 3.4. adalah kolom df dan idf. Kolom df atau *document frequency* berisi jumlah dokumen yang mengandung *term*, kemudian kolom idf atau *inverse document frequency* berisi hasil inverse dari hasil df dengan rumus yang telah tertera pada tabel.

Langkah setelah perhitungan IDF adalah mengkalikan hasil IDF dengan TF sesuai dengan rumus $TF \times IDF$ pada Bab II Persamaan (2.4). Contoh hasil perhitungan perkalian nilai TF dengan IDF dipaparkan pada Tabel 3.5.



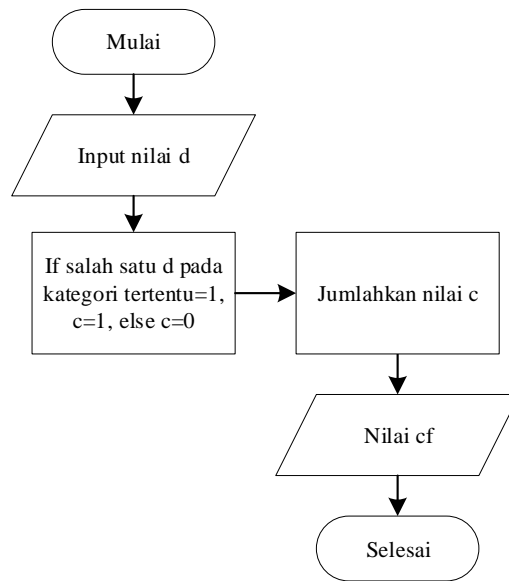
Gambar 3.12. Flowchart Perhitungan DF

3.1.3.3. Inverse Class Frequency (ICF)

ICF (*Inverse Class Frequency*) memiliki cara perhitungan yang hampir sama dengan IDF, hanya saja ICF menghitung jumlah kategori. Pada ICF proses pertama yang dilakukan adalah menghitung CF (*Class Frequency*) terlebih dahulu, CF menghitung jumlah kategori yang mengandung *term*. Flowchart perhitungan CF dipaparkan pada Gambar 3.13. Kemudian ICF dihitung dengan melakukan logaritma perbandingan antara jumlah keseluruhan kelas dengan jumlah kategori yang mengandung *term*.

Tabel 3.5. Contoh Hasil TF-IDF

term	TF-IDF												
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1		c2	c3			c4	c5	c6	c7	c8	c9
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12
malu	3,71	3,71	7,43	11,14	3,71	3,71	7,43	3,71	7,43	3,71	7,43	3,71	3,71
iman	3,71	3,71	3,71	3,71	7,43	7,43	3,71	7,43	3,71	7,43	3,71	7,43	7,43
milik	0,00	4,79	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
enam	0,00	4,49	0,00	0,00	0,00	8,99	0,00	0,00	0,00	0,00	0,00	0,00	0,00
puluh	0,00	4,02	0,00	0,00	8,03	16,06	0,00	4,02	0,00	0,00	0,00	4,02	4,02
cabang	0,00	4,02	0,00	0,00	4,02	4,02	0,00	4,02	0,00	0,00	0,00	8,03	8,03
rasulullah	0,00	0,00	4,19	4,19	0,00	0,00	0,00	0,00	4,19	0,00	8,38	0,00	0,00
....
...
lewat	0,00	0,00	4,09	4,09	0,00	0,00	4,09	0,00	4,09	0,00	4,09	0,00	0,00



Gambar 3.13. Flowchart Perhitungan CF

Rumus yang digunakan di dalam perhitungan CF dan ICF dipaparkan pada Bab II persamaan (2.5) dan (2.6). Contoh perhitungannya dapat dilihat pada Tabel 3.6. Kolom baru yang ada pada Tabel 3.6. adalah kolom cf dan icf. Kolom cf atau *class frequency* berisi jumlah kategori yang mengandung *term*, kemudian kolom icf atau *inverse class frequency* berisi hasil inverse dari hasil cf dengan rumus yang telah tertera pada tabel.

Tabel 3.6. Contoh Perhitungan CF dan ICF

term	tf												cf	icf = log(C/cf)	
	Bukhari		Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad				
	c1	c2	c3		c4	c5	c6	c7	c8	c9					
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12			
Malu	1	1	2	3	1	1	2	1	2	1	2	1	1	9	1.613959
iman	1	1	1	1	2	2	1	2	1	2	1	2	2	9	1.613959
milik	0	1	0	0	0	0	0	0	0	0	0	0	0	1	2.568202
enam	0	1	0	0	0	2	0	0	0	0	0	0	0	2	2.267172
puluh	0	1	0	0	2	4	0	1	0	0	0	1	1	5	1.869232
cabang	0	1	0	0	1	1	0	1	0	0	0	2	2	5	1.869232
rasulullah	0	0	1	1	0	0	0	0	1	0	2	0	0	4	1.966142
....
...
lewat	0	0	1	1	0	0	1	0	1	0	1	0	0	5	1.869232

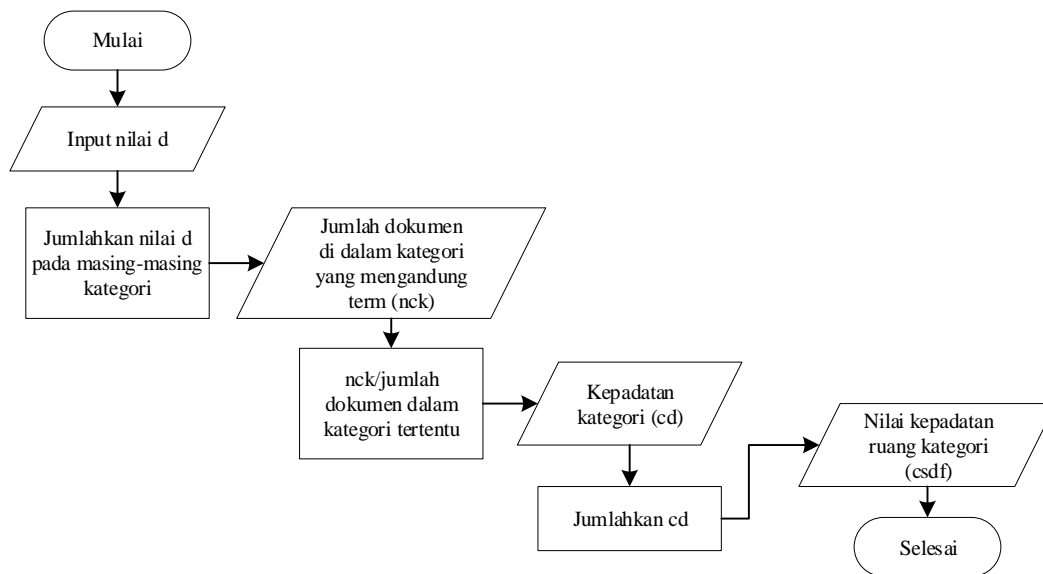
Tabel 3.7. Contoh Perhitungan CS_sF dan ICS_sF

term	tf												cs _s f	ics _s f = log(C/cs _s f)	
	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad			
	c1		c2	c3		c4	c5	c6	c7	c8	c9				
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12			
malu	1	1	2	3	1	1	2	1	2	1	2	1	1	0.148442	3.396643
iman	1	1	1	1	2	2	1	2	1	2	1	2	2	0.148442	3.396643
milik	0	1	0	0	0	0	0	0	0	0	0	0	0	0.02	4.267172
enam	0	1	0	0	0	2	0	0	0	0	0	0	0	0.023205	4.202618
puluh	0	1	0	0	2	4	0	1	0	0	0	1	1	0.085388	3.636804
cabang	0	1	0	0	1	1	0	1	0	0	0	2	2	0.085388	3.636804
rasulullah	0	0	1	1	0	0	0	0	1	0	2	0	0	0.052706	3.846338
...
...
lewat	0	0	1	1	0	0	1	0	1	0	1	0	0	0.055912	3.8207

Langkah setelah perhitungan ICF adalah mengkalikan hasil ICF dengan TF-IDF sesuai dengan rumus TF x IDF x ICF pada Bab II Persamaan (2.7). Contoh hasil perhitungan perkalian nilai TF, IDF, dan ICF dipaparkan pada Tabel 3.7.

3.1.3.4. Inverse Class Space Density Frequency (ICS_sF)

ICS_sF (*Inverse Class Space Density Frequency*) memiliki cara perhitungan yang hampir sama dengan ICF, hanya saja ICS_sF menghitung kepadatan ruang kategori. Di CS_sF atau *class space density* melakukan pemetaan kategori berdasarkan setiap istilah, *inverse class space density* (ICS_sF) dikalikan dengan TF-IDF untuk menghasilkan TF-IDF-ICS_sF. Karena fungsi ICF memberikan skor terendah untuk istilah-istilah yang muncul di beberapa kategori tanpa pengetahuan sebelumnya tentang ruang kategori, pemetaan kategori CS_sF ini memperbaiki kekurangan CF. Bobot masing-masing istilah tidak hanya bergantung pada karakteristik kemunculannya di dokumen dan kategori tetapi juga pada kepadatan kategorisnya (C_s) di mana satu set dokumen yang relevan dihubungkan dengan istilah tertentu. *Flowchart* perhitungan CS_sF dipaparkan pada Gambar 3.14. Rumus yang digunakan di dalam perhitungan ICS_sF ini dipaparkan pada Bab II persamaan (2.8), (2.9), dan (2.10). Contoh perhitungannya dipaparkan pada Tabel 3.7. Kolom baru yang ada pada Tabel 3.7. adalah kolom cs_sf dan ics_sf. Kolom cs_sf atau *class space density frequency* berisi jumlah kepadatan ruang kategori yang



Gambar 3.14. *Flowchart* Perhitungan CS_δF

dihitung sesuai dengan proses flowchart yang dipaparkan pada Gambar 3.14., kemudian kolom icssf atau *inverse class space density frequency* berisi hasil inverse dari hasil csdf dengan rumus yang telah tertera pada tabel.

Langkah setelah perhitungan ICS_δF adalah mengkalikan hasil ICS_δF dengan TF-IDF sesuai dengan rumus $TF \times IDF \times ICS_{\delta}F$ pada Bab II Persamaan (2.10). Contoh hasil perhitungan perkalian nilai TF, IDF, dan ICS_δF dipaparkan pada Tabel 3.8.

Tabel 3.8. Contoh Hasil TF-IDF- ICS_δF

term	q	TF-IDF- ICS _δ F											
		Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1	c2	c3	c4	c5	c6	c7	c8	c9			
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	
malu	12,62	12,62	25,23	37,85	12,62	12,62	25,23	12,62	25,23	12,62	25,23	12,62	12,62
iman	12,62	12,62	12,62	12,62	25,23	25,23	12,62	25,23	12,62	25,23	12,62	25,23	25,23
milik	0,00	20,45	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
enam	0,00	18,88	0,00	0,00	0,00	37,76	0,00	0,00	0,00	0,00	0,00	0,00	0,00
puluh	0,00	14,60	0,00	0,00	29,21	58,41	0,00	14,60	0,00	0,00	0,00	14,60	14,60
cabang	0,00	14,60	0,00	0,00	14,60	14,60	0,00	14,60	0,00	0,00	0,00	29,21	29,21
rasulullah	0,00	0,00	16,12	16,12	0,00	0,00	0,00	0,00	16,12	0,00	32,24	0,00	0,00
...
...
lewat	12,62	12,62	25,23	37,85	12,62	12,62	25,23	12,62	25,23	12,62	25,23	12,62	12,62

3.1.3.5. Inverse Hadith Space Density Frequency (IHS_δF)

Jika ICS_δF memperhatikan kepadatan ruang kategori terhadap suatu *term*, maka IHS_δF lebih memperhatikan kepadatan ruang Kitab Hadis terhadap suatu *term*. Kepadatan ruang Kitab Hadis dihitung untuk mengetahui seberapa besar bobot *term* jika dihitung pula kelangkaan *term* tersebut dari keseluruhan Kitab Hadis. Semakin *term* tersebut jarang muncul pada banyak Kitab Hadis, maka *term* tersebut memiliki nilai *invers* yang tinggi. Kemudian Proses pertama dalam perhitungan IHS_δF adalah menghitung kepadatan hadis terlebih dahulu seperti pada persamaan (3.1).

$$H_{\delta}(t_i) = \frac{n_{h_l}(t_i)}{N_{h_l}} \quad (3.1)$$

dimana, $H_{\delta}(t_i)$ merupakan kepadatan kitab terhadap *term* i , $n_{h_l}(t_i)$ merupakan jumlah hadis di dalam kitab l yang mengandung *term* i , N_{h_l} merupakan jumlah keseluruhan hadis pada kitab l , dan h_l adalah kitab ($l = 1, 2, \dots, 9$).

Kepadatan kategori (C_{δ}) dengan kepadatan Kitab Hadis (H_{δ}) sangatlah berbeda, C_{δ} menghitung seberapa besar kepadatan/kemunculan *term* pada kategori tertentu dengan melakukan perbandingan jumlah dokumen yang memiliki *term* pada kategori tertentu dibanding jumlah dokumen keseluruhan pada kategori tertentu sedangkan H_{δ} menghitung seberapa besar kepadatan/kemunculan *term* pada Kitab Hadis tertentu dengan melakukan perbandingan jumlah dokumen yang memiliki *term* pada kitab hadis tertentu dibanding jumlah dokumen keseluruhan pada kitab hadis tertentu. Kepadatan hadis terhadap *term* i dijumlahkan untuk mendapatkan kepadatan ruang kitab hadis terhadap *term* i seperti pada persamaan (3.2)

$$HS_{\delta}(t_i) = \sum_{h_l} H_{\delta}(t_i) \quad (3.2)$$

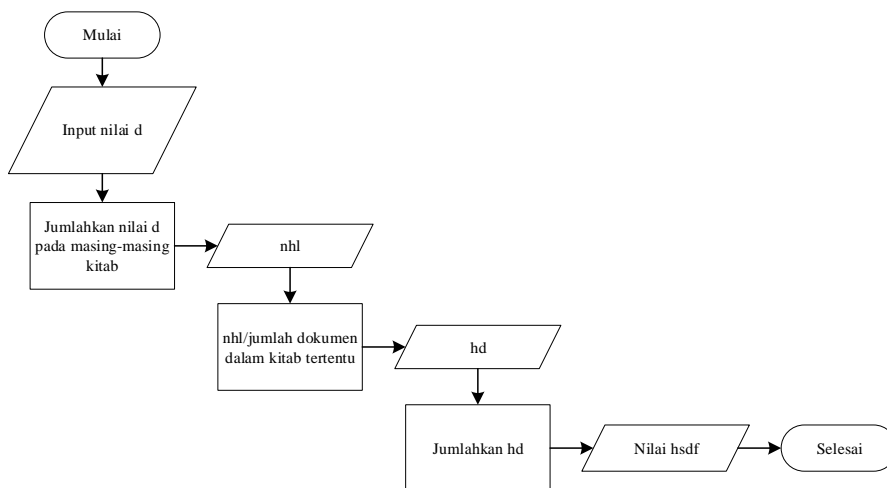
dimana, $HS_{\delta}(t_i)$ merupakan kepadatan ruang kitab hadis terhadap *term i*, $H_{\delta}(t_i)$ merupakan kepadatan kitab hadis terhadap *term i*, dan h_l adalah kitab hadis yang berjumlah 9 ($l = 1, 2, \dots, 9$).

Kepadatan ruang kelas (CS_{δ}) berbeda dengan kepadatan ruang Hadis (HS_{δ}), CS_{δ} menghitung seberapa besar kepadatan kemunculan *term* pada keseluruhan kelas sedangkan HS_{δ} menghitung seberapa besar kepadatan kemunculan *term* pada keseluruhan kitab Hadis. *Flowchart* perhitungan $HS_{\delta}F$ dipaparkan pada Gambar 3.15.

Selanjutnya hasil kepadatan ruang hadis terhadap *term i* dilakukan *inverse* untuk mengetahui tingkat kelangkaan *term* terhadap ruang hadis seperti pada persamaan (3.3).

$$IHS_{\delta}F(t_i) = \log \left(\frac{H}{HS_{\delta}(t_i)} \right) \quad (3.3)$$

dimana, $IHS_{\delta}F(t_i)$ merupakan inverse kepadatan ruang kitab terhadap *term i*, H merupakan jumlah keseluruhan kitab (9), dan $HS_{\delta}(t_i)$ merupakan kepadatan ruang kitab terhadap *term i*.



Gambar 3.15. *Flowchart* Perhitungan $HS_{\delta}F$

Selanjutnya hasil inverse dilakukan perkalian dengan $TF * IDF * ICS_{\delta}F$ untuk mengetahui bobot *term* yang memperhatikan kepadatan ruang kelas dan juga kepadatan ruang hadis seperti pada persamaan (3.4)

$$W_{TF*IDF*ICS_{\delta}F*IHS_{\delta}F}(t_i,d_j,c_k,h_l) = tf_{t_i,d_j} \times \left(\log \frac{D}{df_{(t_i)}} \right) \times \left(\log \frac{C}{CS_{\delta}(t_i)} \right) \times \left(\log \left(\frac{H}{HS_{\delta}(t_i)} \right) \right) \quad (3.4)$$

dimana, $W_{TF*IDF*ICS_{\delta}F*IHS_{\delta}F}(t_i,d_j,c_k,h_l)$ adalah Bobot *term* *i* pada hadis *j* di dalam kategori *k* pada kitab *l*, tf_{t_i,d_j} merupakan jumlah *term* *i* di dalam hadis *j*, *D* merupakan jumlah seluruh hadis, $df_{(t_i)}$ merupakan jumlah hadis yang mengandung *term* *i*, *C* merupakan jumlah seluruh kategori, $CS_{\delta}(t_i)$ merupakan kepadatan ruang kategori terhadap *term* *i*, *H* merupakan jumlah keseluruhan kitab (9), dan $HS_{\delta}(t_i)$ merupakan kepadatan ruang kategori terhadap *term* *i*. Hasil pembobotan *term* dengan memperhatikan kepadatan ruang kategori dan kepadatan ruang hadis ini memberikan nilai yang lebih tinggi terhadap *term* yang langka pada kategori dan hadis, dan memberikan nilai yang rendah pada *term* yang sering muncul.

Tabel 3.9. Contoh Perhitungan HS_δF dan IHS_δF

term	tf												hs _δ f	ihs _δ f = log(B/hsdf)	
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai			Ahmad
		c1	c2	c3		c4	c5	c6	c7	c8	c9				
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12			
malu	1	1	2	3	1	1	2	1	2	1	2	1	1	0.002152	3.621401
iman	1	1	1	1	2	2	1	2	1	2	1	2	2	0.002152	3.621401
milik	0	1	0	0	0	0	0	0	0	0	0	0	0	0.000143	4.799837
enam	0	1	0	0	0	2	0	0	0	0	0	0	0	0.000329	4.436748
puluh	0	1	0	0	2	4	0	1	0	0	0	1	1	0.000948	3.977341
cabang	0	1	0	0	1	1	0	1	0	0	0	2	2	0.000948	3.977341
rasulullah	0	0	1	1	0	0	0	0	1	0	2	0	0	0.00076	4.073282
....
...
lewat	0	0	1	1	0	0	1	0	1	0	1	0	0	0.000947	3.977989

Inverse kepadatan ruang kategori dan ruang hadis ini tetap dikalikan dengan banyaknya *term* tersebut muncul pada tiap dokumen, sehingga jika *term* tersebut muncul sangat banyak di dalam suatu dokumen maka bobot *term* tersebut tetap tinggi, karena umumnya inti dari sebuah dokumen lebih sering disebutkan di dalam dokumen tersebut. Nilai *inverse* kepadatan ruang kategori dan ruang hadis ini digunakan untuk menyeleksi *term* yang terlalu banyak muncul di keseluruhan kategori dan keseluruhan hadis. Contoh perhitungannya dipaparkan pada Tabel 3.9. Kolom baru yang ada pada Tabel 3.9. adalah kolom $hs_{\delta f}$ dan $ihs_{\delta f}$. Kolom $hs_{\delta f}$ atau *hadith space density frequency* berisi jumlah kepadatan ruang hadis yang dihitung sesuai dengan proses *flowchart* yang dipaparkan pada Gambar 3.14., kemudian kolom $ihs_{\delta f}$ atau *inverse hadith space density frequency* berisi hasil inverse dari hasil $hs_{\delta f}$ dengan rumus yang telah tertera pada tabel.

Langkah setelah perhitungan $IHS_{\delta F}$ adalah melakukan kombinasi perkalian hasil $IHS_{\delta F}$ dengan TF, IDF, ICF, dan $ICS_{\delta F}$ sesuai dengan rumus $TF \times IDF \times ICS_{\delta F} \times IHS_{\delta F}$ pada Persamaan (3.4). Contoh hasil perhitungan perkalian dengan nilai $IHS_{\delta F}$ dipaparkan pada Tabel 3.10., Tabel 3.11., dan Tabel 3.12.

Tabel 3.11. Contoh Hasil TF-IDF-IHS_δF

term	TF-IDF- IHS _δ F												
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
malu	13,45	13,45	26,90	40,35	13,45	13,45	26,90	13,45	26,90	13,45	26,90	13,45	13,45
iman	13,45	13,45	13,45	13,45	26,90	26,90	13,45	26,90	13,45	26,90	13,45	26,90	26,90
milik	0,00	23,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
enam	0,00	19,93	0,00	0,00	0,00	39,86	0,00	0,00	0,00	0,00	0,00	0,00	0,00
puluh	0,00	15,97	0,00	0,00	31,94	63,88	0,00	15,97	0,00	0,00	0,00	15,97	15,97
cabang	0,00	15,97	0,00	0,00	15,97	15,97	0,00	15,97	0,00	0,00	0,00	31,94	31,94
rasulullah	0,00	0,00	17,07	17,07	0,00	0,00	0,00	0,00	17,07	0,00	34,15	0,00	0,00
....
...
lewat	0,00	0,00	16,29	16,29	0,00	0,00	16,29	0,00	16,29	0,00	16,29	0,00	0,00

Tabel 3.12. Contoh Hasil TF-IDF-ICF-IHS_δF

term	TF-IDF-ICF- IHS _δ F												
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
malu	21,71	21,71	43,42	65,13	21,71	21,71	43,42	21,71	43,42	21,71	43,42	21,71	21,71
iman	21,71	21,71	21,71	21,71	43,42	43,42	21,71	43,42	21,71	43,42	21,71	43,42	43,42
milik	0,00	59,09	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
enam	0,00	45,19	0,00	0,00	0,00	90,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00
puluh	0,00	29,85	0,00	0,00	59,71	119,41	0,00	29,85	0,00	0,00	0,00	29,85	29,85
....
...
lewat	0,00	0,00	30,45	30,45	0,00	0,00	30,45	0,00	30,45	0,00	30,45	0,00	0,00

Tabel 3.12. Contoh Hasil TF-IDF-ICS₈F-IHS₈F

term	TF-IDF-ICS ₈ F-IHS ₈ F												
	q	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
		c1	c2	c3	d4	d5	d6	c4	c5	c6	c7	c8	c9
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	
malu	45,69	45,69	91,38	137,07	45,69	45,69	91,38	45,69	91,38	45,69	91,38	45,69	45,69
iman	45,69	45,69	45,69	45,69	91,38	91,38	45,69	91,38	45,69	91,38	45,69	91,38	91,38
milik	0,00	98,18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
enam	0,00	83,77	0,00	0,00	0,00	167,53	0,00	0,00	0,00	0,00	0,00	0,00	0,00
puluh	0,00	58,08	0,00	0,00	116,16	232,33	0,00	58,08	0,00	0,00	0,00	58,08	58,08
cabang	0,00	58,08	0,00	0,00	58,08	58,08	0,00	58,08	0,00	0,00	0,00	116,16	116,16
rasulullah	0,00	0,00	65,67	65,67	0,00	0,00	0,00	0,00	65,67	0,00	131,34	0,00	0,00
...
...
lewat	0,00	0,00	62,23	62,23	0,00	0,00	62,23	0,00	62,23	0,00	62,23	0,00	0,00

3.1.4. Cosine Similarity

Setelah dilakukan perhitungan bobot, dilakukan perhitungan jarak terdekat hadis dengan *query* masukan. Penelitian ini menggunakan *cosine similarity* sebagai penghitung jarak dokumennya. Persamaan yang digunakan telah dipaparkan di dalam Bab II persamaan (2.12). Contoh perhitungan *cosine similarity* dipaparkan pada Tabel 3.13. dan perbandingan hasil *cosine similarity* dari beberapa *term weighting* yang telah dipaparkan pada subbab sebelumnya dirangkum di dalam Tabel 3.14.

3.2. Skenario Uji Coba

Pengujian dilakukan dengan mengambil *retrieve* 10 teratas, *retrieve* 20 teratas, dan *retrieve* 30 teratas. Pengujian juga membandingkan hasil dari *short query* dengan *long query*. Kemudian pengujian membandingkan hasil rata-rata dari

Tabel 3.13. Contoh Perhitungan *Cosine Similarity*

q	Panjang Vector											
	Bukhari			Muslim			Abu Dawud		Tirmidzi		Nasai	Ahmad
	c1	c2	c3	c4	c5	c6	c7	c8	c9			
	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12
86,99	86,99	8349,92	18787,32	86,99	86,99	8349,92	86,99	8349,92	86,99	8349,92	86,99	86,99
86,99	86,99	86,99	8349,92	8349,92	8349,92	86,99	8349,92	86,99	8349,92	86,99	8349,92	8349,92
..
..
..
..
jumlah	27578,27	85192,41	11575,48	69158,12	24553,83	11438,15	79909,74	93880,2	16314,95	15921,48	10745,13	10603,44
sqrt	16606,71	29187,74	34022,76	26297,93	49551,82	33820,33	28268,31	30639,88	40391,77	39901,73	32779,77	32562,92
	0,194542	0,221374	0,284871	0,12285	0,065198	0,191051	0,114287	0,210882	0,079984	0,161933	0,098558	0,099214

beberapa kombinasi *term weighting* yaitu TF-IDF, TF-IDF-ICF, TF-IDF-ICS_δF, TF-IDF-ICF-IHS_δF, TF-IDF-IHS_δF, dan TF-IDF-ICS_δF-IHS_δF.

Metode evaluasi pengujian menggunakan *precision*, *recall*, dan *F-measure*. *F-measure* digunakan pada temu kembali informasi dengan mengkombinasikan konsep *recall* dan *precision*. Metode sistem temu kembali informasi diaplikasikan untuk mendapatkan informasi yang relevan dengan keinginan pengguna. Dari kumpulan hadis yang ada penelitian ini melakukan perangkingan berdasarkan *query*. Evaluasi dilakukan untuk menganalisa performa dari metode perangkingan yang digunakan. Analisa performa ini menggambarkan efektifitas metode yang direlasikan dengan kemampuan metode untuk mengembalikan dokumen yang relevan dengan *query*. Rancangan dokumentasi hasil uji coba dipaparkan pada Tabel 3.15.

Tabel 3.14. Contoh Perbandingan Hasil *Cosine Similarity*

TF-IDF		TF-IDF-ICF		TF-IDF-ICS _δ F		TF-IDF-IHS _δ F		TF-IDF-ICF-IHS _δ F		TF-IDF-ICS _δ F-IHS _δ F	
0,28	d3	0,32	d3	0,33	d3	0,35	d3	0,27	d3	0,4	d3
0,22	d2	0,23	d2	0,26	d2	0,26	d2	0,20	d8	0,3	d2
0,21	d8	0,23	d8	0,24	d8	0,25	d8	0,19	d2	0,29	d8
0,19	d1	0,21	d6	0,23	d1	0,22	d6	0,18	d6	0,26	d1
0,19	d6	0,20	d1	0,21	d6	0,22	d1	0,17	d1	0,25	d6
0,16	d10	0,16	d10	0,19	d10	0,19	d10	0,13	d10	0,23	d10
0,12	d4	0,13	d4	0,14	d4	0,14	d4	0,11	d4	0,16	d4
0,11	d7	0,12	d7	0,13	d7	0,13	d7	0,10	d7	0,15	d7
0,10	d12	0,11	d12	0,12	d12	0,12	d11	0,09	d11	0,14	d12
0,10	d11	0,10	d11	0,12	d11	0,11	d12	0,08	d12	0,13	d11
0,08	d9	0,09	d9	0,10	d9	0,11	d9	0,07	d9	0,13	d9
0,07	d5	0,07	d5	0,08	d5	0,08	d5	0,06	d5	0,09	d5

Tabel 3.15. Rancangan Dokumentasi Hasil Uji Coba

ID_Q	Recall, Precision dan F-Measure																	
	TF-IDF			TF-IDF-ICF			TF-IDF-ICS _δ F			TF-IDF-IHS _δ F			TF-IDF-ICF-IHS _δ F			TF-IDF-ICS _δ F-IHS _δ F		
	r	p	f	r	p	f	r	p	f	r	p	f	r	p	f	r	p	f
Q1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Q2	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
...

BAB IV

IMPLEMENTASI DAN PEMBAHASAN

Bagian ini memaparkan hasil implementasi dari setiap langkah tertera pada Bab 3. Selanjutnya dipaparkan hasil dari uji coba sesuai dengan skenario pengujian yakni perhitungan *precision*, *recall*, dan *F-Measure* pada metode yang diusulkan dibandingkan dengan beberapa metode sebelumnya. Setelah itu dipaparkan evaluasi dan pembahasan hasil yang diperoleh pada bagian akhir bab ini.

Metode yang diusulkan diimplementasikan dengan menggunakan bahasa pemrograman Python 3. Database MySQL dengan desain antarmuka web dengan PHP versi 7.3.9, dan library yang digunakan antara lain: Sastrawi (*stemming*), scikit-multilearn, pandas, mysql-connector, dan beberapa library lain untuk tampilan mesin pencariannya. Aplikasi ini dibangun di atas platform Microsoft Windows 10 Pro 64-bit, dengan spesifikasi processor Core i7 dan Memory DDR III 2 GB. Implementasi algoritma dilakukan dengan membuat fungsi-fungsi dari setiap proses atau tahap yang telah dipaparkan pada Bab 3. Pada bagian ini ditampilkan hasil implementasi di setiap langkahnya beserta potongan-potongan script yang penting dalam setiap bagiannya.

4.1. Persiapan Dataset

Persiapan dataset dilakukan dengan cara melakukan *crawl* menggunakan aplikasi *crawler* yang terdapat pada website <https://www.octoparse.com/>. Proses pengambilan dilakukan dengan menganalisa beberapa alamat website yang akan dijaring kumpulan Hadisnya dari website <https://tafsirq.com/>. Hadis yang digunakan merupakan hadis dari 9 Imam, yaitu:

- Shahih Bukhari : <https://tafsirq.com/hadits/bukhari>
- Shahih Muslim : <https://tafsirq.com/hadits/muslim>
- Sunan Abu Dawud : <https://tafsirq.com/hadits/abu-daud>
- Sunan at-Turmudzi : <https://tafsirq.com/hadits/tirmidzi>
- Sunan an-Nasa'i : <https://tafsirq.com/hadits/nasai>
- Sunan Ibnu Majah : <https://tafsirq.com/hadits/ibnu-majah>

Sahih Bukhari
Hadits No. : 1
Kitab 1 : Permulaan Wahyu
Bab : Permulaan Wahyu

حَدَّثَنَا الْحُمَيْدِيُّ عَبْدُ اللَّهِ بْنُ الزُّبَيْرِ قَالَ حَدَّثَنَا سُفْيَانُ قَالَ حَدَّثَنَا
يَحْيَى بْنُ سَعِيدٍ الْأَنْصَارِيُّ قَالَ أَخْبَرَنِي مُحَمَّدُ بْنُ إِبْرَاهِيمَ التَّمِيمِيُّ
أَنَّهُ سَمِعَ عَلْقَمَةَ بْنَ وَقَّاصٍ اللَّيْثِيَّ يَقُولُ سَمِعْتُ عُمَرَ بْنَ الْخَطَّابِ رَضِيَ
اللَّهُ عَنْهُ عَلَى الْمِنْبَرِ قَالَ سَمِعْتُ رَسُولَ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ يَقُولُ
إِنَّمَا الْأَعْمَالُ بِالنِّيَّاتِ وَإِنَّمَا لِكُلِّ امْرِئٍ مَا نَوَى فَمَنْ كَانَتْ هِجْرَتُهُ
إِلَى دُنْيَا يُصِيبُهَا أَوْ إِلَى امْرَأَةٍ يَنْكِحُهَا فَهِجْرَتُهُ إِلَى مَا هَاجَرَ
إِلَيْهِ

Telah menceritakan kepada kami {Al Humaidi Abdullah bin Az Zubair} dia berterm, Telah menceritakan kepada kami {Sufyan} yang berterm, bahwa Telah menceritakan kepada kami {Yahya bin Sa'id Al Anshari} berterm, telah mengabarkan kepada kami {Muhammad bin Ibrahim At Taimi}, bahwa dia pernah mendengar {Alqamah bin Waqash Al Laitsi} berterm; saya pernah mendengar {Umar bin Al Khaththab} diatas mimbar berterm; saya mendengar Rasulullah saw. bersabda: **"Semua perbuatan tergantung niatnya, dan (balasan) bagi tiap-tiap orang (tergantung) apa yang diniatkan; Barangsiapa niat hijrahnya karena dunia yang ingin digapainya atau karena seorang perempuan yang ingin dinikahnya, maka hijrahnya adalah kepada apa dia diniatkan"**

Link: <https://tafsirq.com/hadits/bukhari/1>

Gambar 4.1. Contoh Dokumen Hadis Yang Diambil

- Musnad Ahmad : <https://tafsirq.com/hadits/ahmad>
- Muwatta Malik : <https://tafsirq.com/hadits/malik>
- Sunan Darimi : <https://tafsirq.com/hadits/darimi>

Perincian jumlah hadis yang digunakan dapat dilihat pada Tabel 4.1. Masing-masing jumlah Hadis yang digunakan pada setiap kitab, yaitu: Shahih Bukhari sebanyak 7008 hadis, Shahih Muslim sebanyak 5361 hadis, Sunan Abu Dawud sebanyak 4590 hadis, Sunan at-Turmudzi sebanyak 3891 hadis, Sunan an-Nasa'i sebanyak 5661 hadis, Sunan Ibnu Majah sebanyak 4331 hadis, Musnad Ahmad sebanyak 26363 hadis, Muwatta Malik sebanyak 1594 hadis, dan Sunan Darimi sebanyak 3367 hadis. Salah satu isi dari dokumen hadis dicantumkan dalam Gambar 4.1.

Tabel 4.1. Jumlah Hadis dan Jumlah Kategori

Hadis	Jumlah Hadis	Jumlah Kategori
Shahih Bukhari	7008 hadis	77 kategori
Shahih Muslim	5361 hadis	56 kategori
Sunan Abu Dawud	4590 hadis	35 kategori
Sunan at-Turmudzi	3891 hadis	49 kategori
Sunan an-Nasa'i	5661 hadis	51 kategori
Sunan Ibnu Majah	4331 hadis	32 kategori
Musnad Ahmad	26363 hadis	14 kategori
Muwatta Malik	1594 hadis	32 kategori
Sunan Darimi	3367 hadis	24 kategori

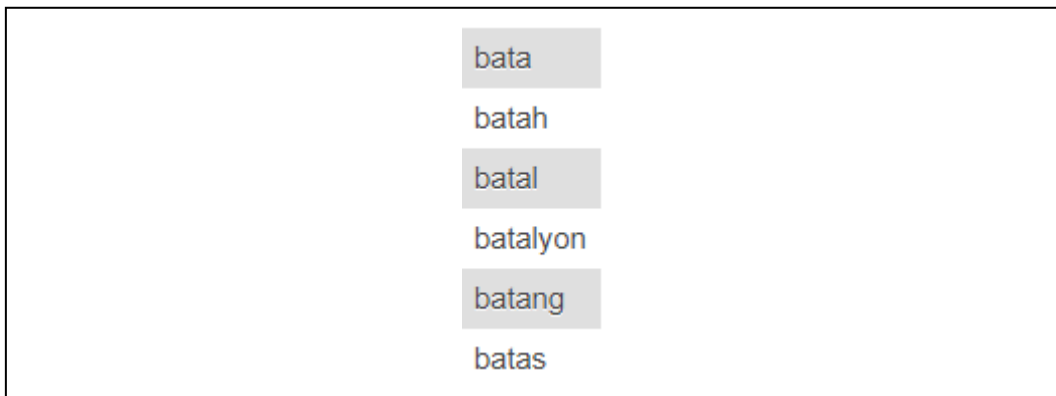
Setelah data terkumpul kemudian dataset diinputkan ke dalam database yang akan digunakan oleh sistem pencarian. Hasil inputan yang tersimpan di dalam database dapat dilihat pada Gambar 4.2.

4.2. Preprocessing

Preprocessing terhadap data yang dikumpulkan dilakukan dalam 4 tahap yaitu: *case folding*, *filtering*, *stemming*, dan *tokenization*. Proses ini sangat menentukan banyak dan tidaknya fitur yang akan muncul saat proses selanjutnya. Fitur-fitur tersebut yang nantinya akan digunakan di dalam perangkingan. Sehingga, *preprocessing* di dalam sistem perangkingan dokumen sangat diperlukan. Akan lebih baik lagi jika ada optimasi pada bagian *preprocessing* ini, namun pada penelitian ini tidak melakukan optimasi tersebut. Penelelitian ini hanya

hadis_id	hadis_index	hadis_arab	hadis_terjemah	hadis_keyword
1	1	حَدَّثَنَا الْخَمِيدِيُّ عِنْدَ... اللَّهُ بْنُ الرَّ...	Telah menceritakan kepada kami (Al Humaidi Abdulla...	gantung niat balas orang gantung niat barangsiapa ...
2	2	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ... يُوسُفَ قَالَ أَحَبُّ...	Telah menceritakan kepada kami (Abdullah bin Yusuf...	terkadang suara gerincing lonceng berat henti erti...
3	3	حَدَّثَنَا يَحْيَى بْنُ بُكَيْرٍ... فَالْحَدِيثُ...	Telah menceritakan kepada kami (Yahya bin Bukair) ...	baca nabi saw malaikat pegang peluk kuat lepas kat...
4	4	حَدَّثَنَا مُوسَى بْنُ... إِسْمَاعِيلَ قَالَ... حَدَّثَنَا...	Telah menceritakan kepada kami (Musa bin Isma'il) ...	cerita firman allah ta ala gerak lidah baca al qur...
5	5	حَدَّثَنَا عِدْنَانُ قَالَ... الْحَيْرَانُ عِنْدَ اللَّهِ...	Telah menceritakan kepada kami (Abdan] dia berkata...	abar abar riwayat rasulullah saw manusia lembut ut...
6	6	حَدَّثَنَا أَبُو الْيَمَانِ... الْحَكَمُ بْنُ نَافِعٍ...	Telah menceritakan kepada kami (Abu Al Yaman Al Ha...	abar abar abar abar heraclius terima rombongan dagan...
7	7	حَدَّثَنَا غَيْثُ اللَّهِ بْنُ... مُوسَى قَالَ أَع...	Telah menceritakan kepada kami (Abdullah bin Musa)...	islam bangun landas saksi ilah allah sungguh muham...
8	8	حَدَّثَنَا عَبْدُ اللَّهِ بْنُ... مُحَمَّدٍ الْخُضَيْبِ...	Telah menceritakan kepada kami (Abdullah bin Muham...	iman milik enam puluh cabang malu iman
			Telah menceritakan	cerita nabi saw sabda

Gambar 4.2. Contoh Hasil Input Dataset



Gambar 4.3. Contoh Hasil *Preprocessing*

menggunakan library sastrawi untuk *stemming* dan menggunakan dataset stopwords dari Fadila Tala untuk *filtering* stoplist. Khususnya untuk dataset hadis ini, banyak sekali *term* yang merupakan terjemahan bahasa arab. Saat *preproceesing* ini berlangsung akan banyak sekali ditemukan *term-term* yang tidak dapat *distemming* juga tidak dapat dihapus dengan *filtering*. Selain itu, dokumen hadis selalu terdiri dari *matan* dan *sanad*. Matan adalah isi hadis, dan sanad adalah periwayatnya. Pada contoh hadis yang diampil dalam Gambar 4.1. matan adalah huruf yang diberikan *bold*, sedangkan sanad adalah huruf yang tidak *dibold*. Sanad tentunya akan merusak sistem perangkingan jika diikutkan di dalam proses selanjutnya setelah *preprocessing*. Oleh sebab itu, pada penelitian ini mencoba melakukan penghapusan sanad dengan bantuan *regular expression*. Contoh hasil *preprocessing* yang tersimpan di dalam database dapat dilihat pada Gambar 4.3.

4.3. Term Weighting

Subbab-subbab di dalam *term weighting* ini akan menampilkan beberapa contoh hasilnya yang tersimpan dari proses *term weighting* yang selanjutnya akan digunakan di dalam proses perangkingan. Penelitian ini melakukan perbandingan kombinasi *term weighting* di antaranya adalah: TF-IDF, TF-IDF-ICF, TF-IDF-ICS₈F, TF-IDF-IHS₈F, TF-IDF-ICF-IHS₈F, dan TF-IDF-ICS₈F-IHS₈F.

Langkah pertama di dalam *term weighting* adalah perhitungan *Term Frequency* (TF). Perhitungan TF dilakukan dengan cara menghitung jumlah kemunculan *term* pada setiap dokumen. Contoh hasil perhitungan frekuensi *term*

word	hadis_id	tf	df
bagus	61870	1	1
bagus	62042	2	1
bagus	62125	1	1
bagus	62146	2	1

Gambar 4.4. Contoh Hasil Perhitungan *Term Frequency*

yang tersimpan di dalam database terlihat pada Gambar 4.4. Pada Gambar 4.4. kolom word adalah kolom *term* hasil preprocessing dari tiap dokumen hadis. Kolom dokumen adalah id dokumen, dan kolom tf adalah kolom yang berisi *term frequency term* pada setiap dokumen.

Tahap selanjutnya pada *Term Weighting* di dalam penelitian ini adalah perhitungan IDF (*Inverse Document Frequency*). Tahap ini dilakukan dengan cara menghitung DF (*Document Frequency*) terlebih dahulu yaitu jumlah dokumen yang mengandung tiap *term*. Kemudian DF tersebut di-*inverse*-kan dengan rumus IDF yang tertera pada Persamaan (2.2) dan (2.3) pada Bab II. Contoh hasil yang tersimpan di dalam database dicantumkan dalam Gambar 4.5.

Tahap selanjutnya di dalam penelitian ini adalah perhitungan *Class Frequency*. Perhitungan CF dilakukan dengan cara menghitung jumlah kemunculan *term* pada setiap kategori terlebih dahulu kemudian dihitung jumlah dari dokumen yang mengadung *term* di dalam kategori tertentu. Kemudian hasil CF diinverse kan dengan menambahkan *log* seperti yang telah dipaparkan pada persamaan (2.5) dan (2.6) di Bab II sehingga didapatkanlah ICF. Contoh hasil CF yang tersimpan di dalam databse dapat dilihat pada Gambar 4.6.

word	idf
sepertiitu	10.3393
sepertimemerdekakan	11.0324
sepertinganya	11.0324
sepet	8.5475
sepi	8.95296
sepoi	11.0324
sepperti	11.0324
seprei	11.0324
seprti	10.3393
sepsang	11.0324

Gambar 4.5. Contoh Hasil Perhitungan IDF

word	kitab_id	cf	cs	ctf	_csdf
kafir	238	1	2	2	0.0137931
kafir	241	1	2	2	0.0322581
kafir	246	1	1	1	0.030303
kafir	248	1	1	2	0.0263158
kafir	250	1	3	3	0.032967
kafir	251	1	2	3	0.0208333
kafir	253	1	5	5	0.0367647
kafir	254	1	6	11	0.01875
kafir	255	1	1	1	0.00793651
kafir	258	1	1	1	0.00310559
kafir	262	1	1	2	0.0217391

word	idf	icf
taqnathuu	11.0324	5.9135
taqorrub	11.0324	5.9135
taqrabuuhunna	11.0324	5.9135
taqtulnaa	10.3393	5.22036
taqullaaha	11.0324	5.9135
taqullaahal	11.0324	5.9135
taquu	11.0324	5.9135
taqwa	7.89691	3.34855

Gambar 4.6. Contoh Hasil Perhitungan ICF

Tahap selanjutnya adalah perhitungan *Class Space Density Frequency* (ICS_δF). Perhitungan ICS_δF diawali dengan menghitung kepadatan kategorinya (C_{δ}) yaitu menghitung dokumen yang mengandung *term* pada kategori tertentu (c_k), kemudian dilanjutkan dengan menghitung kepadatan ruang kategori, yaitu jumlah dari kepadatan keseluruhan kategori yang ada (CS_{δ}), kemudian hasil dari kepadatan ruang kategori ($CS_{\delta}(t_i)$) dilakukan *inverse*. Persamaan yang digunakan di dalam proses ini terdapat pada Bab II persamaan (2.8), (2.9), dan (2.10). Contoh hasil pada Gambar 4.7.

word	idf	icf	icsdf
mukminim	11.0324	5.9135	11.3117
mukminin	5.11082	1.16857	4.94247
mukminun	8.95296	3.96759	8.91516
mukminuuna	11.0324	5.9135	9.049
mukram	9.93379	4.81489	9.70959
mukran	11.0324	5.9135	14.3855
muktib	10.3393	5.9135	10.7257
mukul	11.0324	5.9135	10.1032
mul	11.0324	5.9135	11.6272
mulaa	10.3393	5.9135	11.3645

Gambar 4.7. Contoh Hasil Perhitungan ICS_δF

Tahap selanjutnya adalah perhitungan *Inverse Hadis Space Density Frequency* (IHS_δF) yang merupakan metode usulan di dalam penelitian ini. Proses pertama dalam perhitungan IHS_δF adalah menghitung kepadatan Hadis terhadap *term i* $H_{\delta}(t_i)$. Kepadatan hadis terhadap *term i* dijumlahkan untuk mendapatkan kepadatan ruang hadis terhadap *term i* $HS_{\delta}(t_i)$. Selanjutnya hasil kepadatan ruang hadis terhadap *term i* dilakukan *inverse* untuk mengetahui tingkat kelangkaan *term* terhadap ruang hadis. Persamaan yang digunakan di dalam proses ini terdapat pada Bab III persamaan (3.1), (3.2), dan (3.3). Contoh hasil pada Gambar 4.8.

word	idf	icf	icsdf	ihsdf
mukminim	11.0324	5.9135	11.3117	11.0611
mukminin	5.11082	1.16857	4.94247	4.89192
mukminun	8.95296	3.96759	8.91516	8.83404
mukminuuna	11.0324	5.9135	9.049	9.22454
mukram	9.93379	4.81489	9.70959	9.88513
mukran	11.0324	5.9135	14.3855	13.4418
muktib	10.3393	5.9135	10.7257	10.9012
mukul	11.0324	5.9135	10.1032	10.3759
mul	11.0324	5.9135	11.6272	11.8999
mulaa	10.3393	5.9135	11.3645	12.005

Gambar 4.8. Contoh Hasil Perhitungan IHS_δF

tf_idf	tf_idf_icf	tf_idf_icsdf	tf_idf_ihsdf
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
6.965587139129639	3.78137331122403	23.462138728663717	24.119457158811315
10.448380708694458	5.672059966836045	35.193208092995576	36.17918573821697
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
3.4827935695648193	1.890686655612015	11.731069364331859	12.059728579405657
6.965587139129639	3.78137331122403	23.462138728663717	24.119457158811315
6.965587139129639	3.78137331122403	23.462138728663717	24.119457158811315

Gambar 4.9. Contoh Hasil Perkalian Pembobotan

Proses selanjutnya yang dilakukan oleh sistem adalah melakukan perkalian bobot TF-IDF, TF-IDF-ICF, TF-IDF-ICS_sF, TF-IDF-IHS_sF, TF-IDF-ICF-IHS_sF, dan TF-IDF-ICS_sF-IHS_sF. contoh hasilnya di dalam database ditampilkan pada Gambar 4.9.

4.4. Cosine Similarity

Setelah dilakukan perhitungan bobot, dilakukan pembentukan matriks *term weighting*. Penelitian ini menggunakan *vector space model* dalam merepresentasikan dokumen atau *query* sebagai sebuah vector dalam sebuah ruang *term*. Ruang ini memiliki dimensi sebanyak jumlah *term* atau dengan *term* lain untuk dokumen yang memiliki N *term* maka dibutuhkan N dimensi. Setiap vektor direpresentasikan sesuai bobot dari *term-term* yang ada. *Term-term* yang ada menjadi sumbu-sumbu koordinat sebanyak jumlah *term*, sedangkan vektornya adalah sebuah titik yang posisinya berdasarkan nilai dari sumbu-sumbu tersebut. Misal untuk dokumen dengan dua dimensi, maka apabila direpresentasikan dalam bidang kartesian akan menjadi sebuah titik yang terdiri dari x dan y . Nilai x dan y ini tergantung pada bobot *term* x dan y . Dalam implementasinya, *Vector Space Model* ini direpresentasikan sebagai matriks dua dimensi dengan kolom sebagai *term* dan dokumen sebagai baris, serta bobot *term* sebagai isi matriks.

Pembuatan *Vector Space Model* untuk dokumen adalah dengan cara membuat sebuah matriks dua dimensi dengan kolom sebagai *term* dan dokumen sebagai baris. Isi dari matriks tersebut merupakan nilai bobot *term* (*term weight*) dari masing-masing *term* terhadap masing-masing dokumen. *Term* dalam *Vector Space Model* ini diambil dari seluruh *term* unik pada keseluruhan dokumen. Dari bobot *term* yang termuat dalam *vector space model* dilakukan perhitungan kemiripan menggunakan *cosine similarity*, dengan *cosine similarity* tersebut dilakukan perhitungan jarak terdekat hadis dengan *query* masukan. Persamaan yang digunakan telah dipaparkan di dalam Bab II persamaan (2.12). Perhitungan *cosine similarity* memberikan hasil akhir nilai 0 hingga 1. Nilai 0 menunjukkan bahwa dokumen tersebut tidak sesuai dengan *query* dan semakin mendekati 1 menunjukkan bahwa dokumen tersebut sesuai dengan *query*. Dari nilai kemiripan tersebut dokumen dapat dirangkingkan mulai dari nilai tertinggi, yang paling sesuai hingga terendah, yang paling tidak sesuai.

4.5. Dokumentasi Uji Coba

Uji coba penelitian ini menggunakan 50 *query* yang telah dikumpulkan dari beberapa situs hadis, diantaranya <https://www.facebook.com/TAHDIS>, <https://www.facebook.com/tafsirqcom/>, aplikasi ensiklopedia 9 imam dari lidwa.com, <https://tafsirq.com/>, <http://carihadis.com/>, dan <http://www.pissktb.com/>. 50 *query* yang digunakan dikumpulkan di dalam Tabel pada Lampiran 3. 50 *query* tersebut terdiri dari 2 jenis *query* yaitu *short query* dan *long query*. *Short query* memiliki kurang dan sama dengan 4 *term*, sedangkan *long query* memiliki lebih dari 4 *term*. Uji coba yang dilakukan pada 50 *query* ini dibandingkan dengan menggunakan 6 metode kombinasi *term weighting* yaitu TF-IDF, TF-IDF-ICF, TF-IDF-ICS_F, TF-IDF-IHS_F, TF-IDF-ICF-IHS_F, dan TF-IDF-ICS_F-IHS_F. Uji coba dilakukan dengan mengambil *retrieve document* 10 teratas, *retrieve document* 20 teratas, dan *retrieve document* 30 teratas. Dua contoh hasil *retrieve document* 10 teratas dilampirkan pada Lampiran 4, yaitu hasil dari *query* yang berbunyi “umrah bulan ramadhan setara ibadah haji”, dan “malu bagian dari iman”.

Tabel 4.2. Pakar Pengecek Kebenaran Hasil

No.	Pakar	Query yang Dicek
1	Guru Bahasa Arab yang tergabung di dalam grup WA dan Telegram “Konferensi Nasional FMGMP”	20 <i>query</i> top 10 (1-20)
2	Guru Bahasa Arab Madrasah Tsanawiyah Pondok Pesantrem Nurul Islam Pungging Mojokerto	20 <i>query</i> top 10 (21-40)
3	Guru Al Qur’an Hadis Madrasah Tsanawiyah Pondok Pesantrem Nurul Islam Pungging Mojokerto	10 <i>query</i> top 10 (41-50) + 10 <i>query</i> top 20 (1-10)
4	Guru Bahasa Arab Madrasah Aliyah Pondok Pesantrem Nurul Islam Pungging Mojokerto	20 <i>query</i> top 20 (11-30)
5	Guru Al Qur’an Hadis Madrasah Aliyah Pondok Pesantrem Nurul Islam Pungging Mojokerto	20 <i>query</i> top 20 (31-50)
6	Dosen Pendidikan Agama Islam Sekolah Tinggi Ilmu Agama Pondok Pesantrem Nurul Islam Pungging Mojokerto	20 <i>query</i> top 30 (1-20)
7	Dosen Fakultas Keislaman Universitas Trunojoyo Madura	30 <i>query</i> top 30 (21-50)

Hasil *retrieve* teratas dari perankingan 50 *query* dikumpulkan, selanjutnya dilakukan pengecekan oleh pakar Hadis untuk keakuratan hasil kebenarannya yang kemudian akan dievaluasi menggunakan *recall*, *precision*, dan *f-score*. Pakar-pakar tersebut merupakan beberapa guru dan dosen yang memiliki bidang dalam keagamaan Islam khususnya mengenai Hadis. Detail pakar dan jumlah hasil *retrieve* teratas dari perankingan 50 *query* yang dicek oleh masing-masing pakar dapat dilihat pada Tabel 4.2.

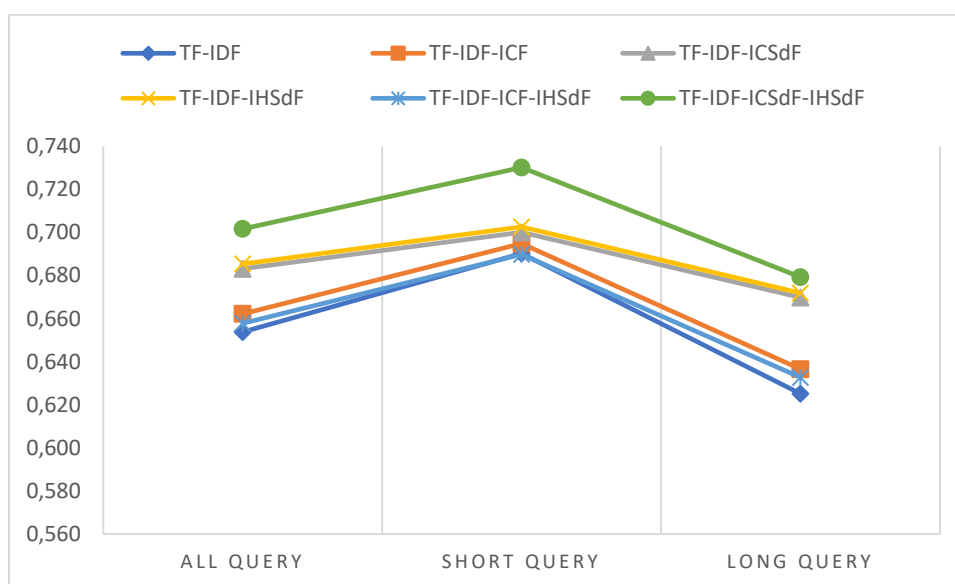
50 *query* yang digunakan merupakan beberapa kalimat yang sedang hangat diperbincangkan oleh masyarakat. Salah satu contohnya adalah *query* nomor 50 yang berbunyi “setiap penyakit ada obatnya”. *Query* tersebut mencerminkan permasalahan masyarakat yang sedang membuat resah saat ini yaitu persebaran virus Covid-19, virus yang saat ini telah menjadi pandemi dunia. Hadis-hadis yang berhasil *terretrieve* dan relevan dengan *query* nomor 50, ada beberapa yang menjelaskan bahwa Allah selalu memberikan penyakit beserta obatnya, Allah yang memiliki hak untuk memberi atau mencabut penyakit tersebut dari setiap manusia. Salah satu contoh hadis yang dapat *diretrieve* dari *query* tersebut adalah Hadis dari Shahih Bukhari Nomor 5246, yaitu: “Allah tidak akan menurunkan penyakit melainkan menurunkan obatnya juga.” Juga hadis dari Shahih Muslim Nomor 4084 yaitu: “Setiap penyakit ada obatnya. Apabila ditemukan obat yang tepat untuk suatu

penyakit, maka akan sembuhlah penyakit itu dengan izin Allah ‘azza wajalla.”. Dua hadis tersebut merupakan hadis dari 2 imam yang terkenal dengan keshahihan hadis-hadisnya. Dua hadis tersebut merupakan salah satu yang relevan dari beberapa hadis yang berhasil *diretrieve* oleh sistem dari inputan *query* “setiap penyakit ada obatnya”. Hadis Shahih Bukhari Nomor 5246 mendapat peringkat ke-4 pada kombinasi metode yang diusulkan (TF-IDF-ICS₈F-IHS₈F) dan Hadis Shahih Muslim Nomor 4084 mendapat peringkat ke-2 pada kombinasi metode yang diusulkan. Uji coba pada *query* ke-50 yang merupakan salah satu dari *short query* ini mendapatkan hasil yang seimbang pada *retrieve* 10 teratas dari seluruh kombinasi *term weighting* dengan *recall* 0,615, *precision* 0,8, dan *f-score* sebesar 0,696. Contoh permasalahan yang terjadi di masyarakat lainnya adalah pada *query* ke-44 yang berbunyi “masa nifas selama empat puluh hari”. *Query* tersebut merupakan permasalahan yang mewakili wanita setelah melahirkan, bahwa masa nifas (masa keluarnya darah setelah melahirkan) rata-rata adalah empat puluh hari. Uji coba pada *query* ke-44 yang merupakan salah satu dari *long query* ini mendapatkan hasil terbaik pada *retrieve* 10 teratas dari kombinasi *term weighting* yang diusulkan yaitu *recall* 0,467, *precision* 0,7, dan *f-score* sebesar 0,560. Hal ini menunjukkan bahwa pada beberapa *short query*, sistem bekerja dengan baik karena akan banyak hadis relevan yang akan *diretrieve* pada semua metode kombinasi *term weighting*. Hal tersebut memberikan akibat bahwa hasil rata-rata *recall*, *precision* dan *f-score short query* lebih tinggi dari rata-rata *long query*.

Perbandingan hasil rata-rata dari semua *query*, *short query* dan *long query* pada hasil *retrieve* 10 teratas dapat dilihat pada Tabel 4.3. dan Gambar 4.10. Hasil *retrieve* 10 teratas tersebut melakukan rekap hasil dari pengembalian sistem 10

Tabel 4.3. Rata-rata Hasil *F-Score Retrieve* 10 Teratas

Top 10	All query	Short query	Long query
TF-IDF	0,654	0,690	0,625
TF-IDF-ICF	0,662	0,695	0,637
TF-IDF-ICS₈F	0,683	0,700	0,670
TF-IDF-IHS₈F	0,685	0,703	0,672
TF-IDF-ICF-IHS₈F	0,658	0,690	0,633
TF-IDF-ICS₈F-IHS₈F	0,702	0,730	0,679



Gambar 4.10. Grafik Perbandingan Hasil Rata-rata F -Score Retrieve 10 Teratas

teratas. Pada beberapa tipe *query* yang telah dimasukkan ke dalam sistem terlihat bahwa pada *query* bertipe *short* memiliki banyak hasil hadis yang *terretrieve* dan relevan. TF-IDF mendapatkan f -score sebesar 0,645 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,690, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,625. TF-IDF-ICF mendapatkan f -score sebesar 0,662 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,695, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,637. TF-IDF-ICS δ F mendapatkan f -score sebesar 0,683 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,7, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,67. TF-IDF-IHS δ F mendapatkan f -score sebesar 0,685 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,703, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,672. Hasil TF-IDF-ICS δ F dengan TF-IDF-IHS δ F memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS δ F mendapatkan f -score sebesar 0,658 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,690, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,633. TF-IDF-ICS δ F-IHS δ F

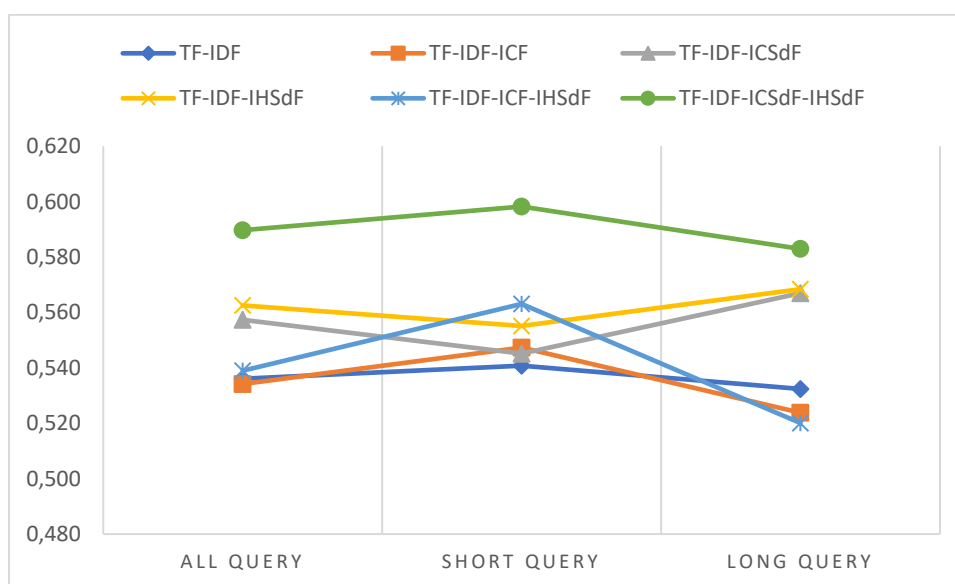
mendapatkan *f-score* sebesar 0,702 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,730, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,679.

Kesimpulan dari tabel perbandingan rata-rata hasil *f-score retrieve* 10 teratas adalah metode TF-IDF-ICS₈F-IHS₈F mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *query* bertipe *short* yaitu sebesar 0,730. Grafik pada Gambar 4.10. juga menunjukkan bahwa metode TF-IDF-ICS₈F-IHS₈F memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan pada semua metode *query* bertipe *short* memiliki hasil rata-rata *f-score* paling tinggi.

Perbandingan hasil rata-rata dari semua *query*, *short query* dan *long query* pada hasil *retrieve* 20 teratas dapat dilihat pada Tabel 4.4. dan Gambar 4.11. Hasil *retrieve* 20 teratas tersebut melakukan rekap hasil dari pengembalian sistem 20 teratas. Pada beberapa tipe *query* yang telah dimasukkan ke dalam sistem terlihat bahwa pada *query* bertipe *short* memiliki banyak hasil hadis yang ter*retrieve* dan relevan. TF-IDF mendapatkan *f-score* sebesar 0,536 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,541, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,532. TF-IDF-ICF mendapatkan *f-score* sebesar 0,534 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,547, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,524. TF-IDF-ICS₈F mendapatkan *f-score* sebesar 0,557 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,545, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,567. TF-IDF-IHS₈F mendapatkan

Tabel 4.4. Rata-rata Hasil *F-Score Retrieve* 20 Teratas

	All query	Short query	Long query
TF-IDF	0,536	0,541	0,532
TF-IDF-ICF	0,534	0,547	0,524
TF-IDF-ICS₈F	0,557	0,545	0,567
TF-IDF-IHS₈F	0,563	0,555	0,568
TF-IDF-ICF-IHS₈F	0,539	0,563	0,520
TF-IDF-ICS₈F-IHS₈F	0,590	0,598	0,583



Gambar 4.11. Grafik Perbandingan Hasil Rata-rata F -Score Retrieve 20 Teratas

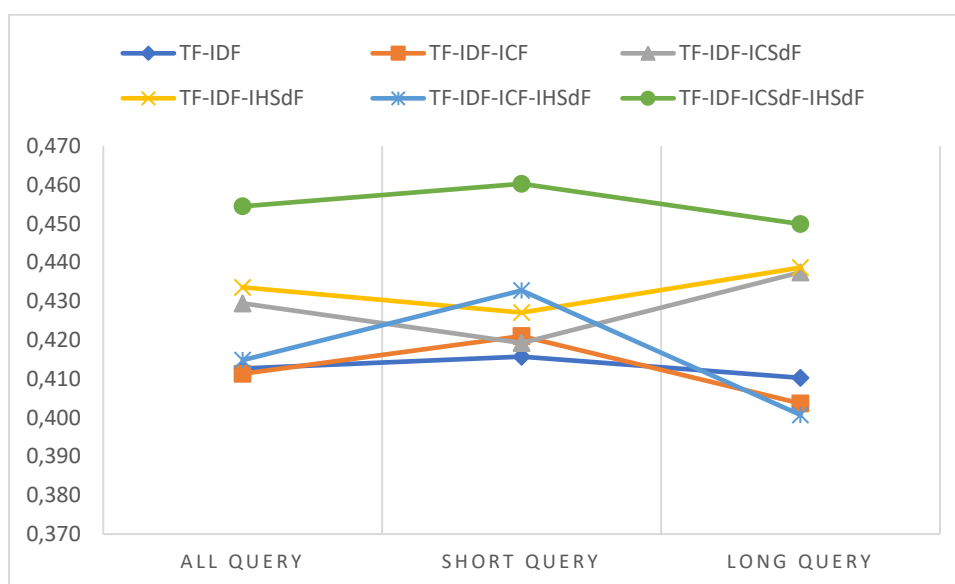
f -score sebesar 0,563 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,555, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,568. Hasil TF-IDF-ICS₈F dengan TF-IDF-IHS₈F memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS₈F mendapatkan f -score sebesar 0,539 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,563, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,520. TF-IDF-ICS₈F-IHS₈F mendapatkan f -score sebesar 0,590 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan f -score sebesar 0,598, dan pada hasil rata-rata *long query* didapatkan f -score sebesar 0,583.

Kesimpulan dari tabel perbandingan rata-rata hasil f -score *retrieve* 20 teratas adalah metode TF-IDF-ICS₈F-IHS₈F mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *query* bertipe *short* yaitu sebesar 0,598. Grafik pada Gambar 4.11. juga menunjukkan bahwa metode TF-IDF-ICS₈F-IHS₈F memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan pada semua metode *query* bertipe *short* memiliki hasil rata-rata f -score paling tinggi.

Tabel 4.5. Rata-rata Hasil *F-Score Retrieve* 30 Teratas

	All query	Short query	Long query
TF-IDF	0,413	0,416	0,410
TF-IDF-ICF	0,411	0,421	0,404
TF-IDF-ICS_sF	0,429	0,419	0,437
TF-IDF-IHS_sF	0,434	0,427	0,439
TF-IDF-ICF-IHS_sF	0,415	0,433	0,401
TF-IDF-ICS_sF-IHS_sF	0,454	0,460	0,450

Perbandingan hasil rata-rata dari semua *query*, *short query* dan *long query* pada hasil *retrieve* 30 teratas dapat dilihat pada Tabel 4.5. dan Gambar 4.12. Hasil *retrieve* 30 teratas tersebut melakukan rekap hasil dari pengembalian sistem 30 teratas. Pada beberapa tipe *query* yang telah dimasukkan ke dalam sistem terlihat bahwa pada *query* bertipe *short* memiliki banyak hasil hadis yang ter*retrieve* dan relevan. TF-IDF mendapatkan *f-score* sebesar 0,413 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,416, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,410. TF-IDF-ICF mendapatkan *f-score* sebesar 0,411 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,421, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,404. TF-IDF-ICS_sF mendapatkan *f-score* sebesar 0,429 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,419, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,437. TF-IDF-IHS_sF mendapatkan *f-score* sebesar 0,434 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,427, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,439. Hasil TF-IDF-ICS_sF dengan TF-IDF-IHS_sF memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS_sF mendapatkan *f-score* sebesar 0,415 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,433, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,401. TF-IDF-ICS_sF-IHS_sF mendapatkan *f-score* sebesar 0,454 pada pengukuran hasil rata-rata seluruh *query*, pada hasil rata-rata *short query* didapatkan *f-score* sebesar 0,460, dan pada hasil rata-rata *long query* didapatkan *f-score* sebesar 0,450.



Gambar 4.12. Grafik Perbandingan Hasil Rata-rata F -Score Retrieve 30 Teratas

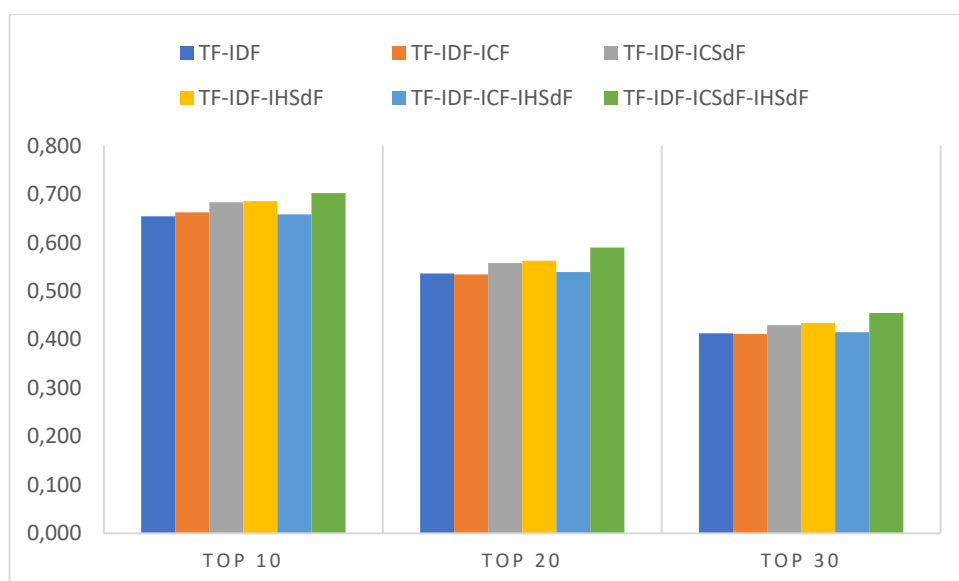
Kesimpulan dari tabel perbandingan rata-rata hasil f -score retrieve 30 teratas adalah metode TF-IDF-ICS δ F-IHS δ F mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *query* bertipe *short* yaitu sebesar 0,335. Grafik pada Gambar 4.12. juga menunjukkan bahwa metode TF-IDF-ICS δ F-IHS δ F memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan pada semua metode *query* bertipe *short* memiliki hasil rata-rata f -score paling tinggi.

Perbandingan hasil rata-rata dari *retrieve* 10, 20, dan 30 teratas pada semua *query* dapat dilihat pada Tabel 4.6. dan Gambar 4.13. Hasil *retrieve* dari 10, 20, dan 30 teratas pada semua *query* tersebut melakukan rekap hasil dari pengembalian sistem 10, 20, dan 30 teratas. Pada beberapa rata-rata *retrieve* antara 10, 20, dan 30 teratas dapat dilihat bahwa *retrieve* 10 teratas memiliki hasil rata-rata f -score ter-

Tabel 4.6. Rata-rata Hasil F -Score Semua *Query*

	Top 10	Top 20	Top 30
TF-IDF	0,654	0,536	0,413
TF-IDF-ICF	0,662	0,534	0,411
TF-IDF-ICSδF	0,683	0,557	0,429
TF-IDF-IHSδF	0,685	0,563	0,434
TF-IDF-ICF-IHSδF	0,658	0,539	0,415
TF-IDF-ICSδF-IHSδF	0,702	0,590	0,454

tinggi. TF-IDF mendapatkan *f-score* sebesar 0,654 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,536, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,413. TF-IDF-ICF mendapatkan *f-score* sebesar 0,662 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,534, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,411. TF-IDF-ICS_sF mendapatkan *f-score* sebesar 0,683 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,557, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,429. TF-IDF-IHS_sF mendapatkan *f-score* sebesar 0,658 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,563, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,434. Hasil TF-IDF-ICS_sF dengan TF-IDF-IHS_sF memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS_sF mendapatkan *f-score* sebesar 0,658 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,539, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,415. TF-IDF-ICS_sF-IHS_sF mendapatkan *f-score* se-



Gambar 4.13. Grafik Perbandingan Hasil Rata-rata *F-Score* Semua *Query*

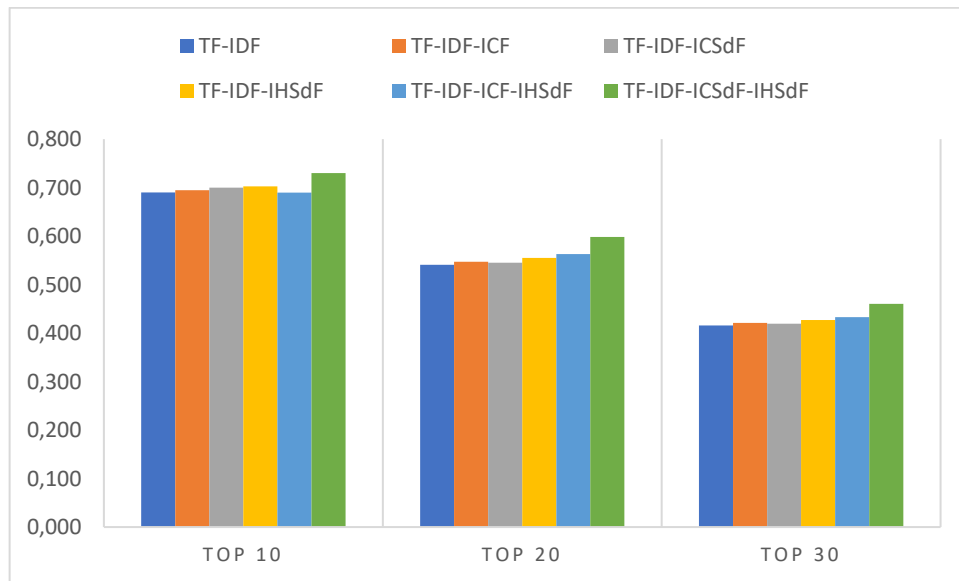
Tabel 4.7. Rata-rata Hasil *F-Score Short Query*

	Top 10	Top 20	Top 30
TF-IDF	0,690	0,541	0,416
TF-IDF-ICF	0,695	0,547	0,421
TF-IDF-ICS_sF	0,700	0,545	0,419
TF-IDF-IHS_sF	0,703	0,555	0,427
TF-IDF-ICF-IHS_sF	0,690	0,563	0,433
TF-IDF-ICS_sF-IHS_sF	0,730	0,598	0,460

besar 0,702 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,590, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,454.

Kesimpulan dari tabel perbandingan rata-rata hasil *f-score retrieve* 10, 20, dan 30 teratas pada semua *query* adalah metode TF-IDF-ICS_sF-IHS_sF mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *retrieve* 10 teratas yaitu sebesar 0,702. Grafik pada Gambar 4.13. juga menunjukkan bahwa metode TF-IDF-ICS_sF-IHS_sF memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan dari semua batas teratas *retrieve* 10 teratas adalah yang paling tinggi.

Perbandingan hasil rata-rata dari *retrieve* 10, 20, dan 30 teratas pada *short query* dapat dilihat pada Tabel 4.7. dan Gambar 4.14. Hasil *retrieve* dari 10, 20, dan 30 teratas pada *short query* tersebut melakukan rekap hasil dari pengembalian sistem 10, 20, dan 30 teratas. Pada beberapa rata-rata *retrieve* antara 10, 20, dan 30 teratas dapat dilihat bahwa *retrieve* 10 teratas memiliki hasil rata-rata *f-score* tertinggi. TF-IDF mendapatkan *f-score* sebesar 0,690 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,541, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,302. TF-IDF-ICF mendapatkan *f-score* sebesar 0,695 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,547, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,306. TF-IDF-ICS_sF mendapatkan *f-score* sebesar 0,700 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,545, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,305. TF-IDF-IHS_sF mendapatkan *f-score* sebesar 0,703 pada pengukuran



Gambar 4.14. Grafik Perbandingan Hasil Rata-rata *F-Score Short Query*

hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,555, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,311. Hasil TF-IDF-ICS_δF dengan TF-IDF-IHS_δF memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS_δF mendapatkan *f-score* sebesar 0,690 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,563, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,314. TF-IDF-ICS_δF-IHS_δF mendapatkan *f-score* sebesar 0,730 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,598, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,335.

Kesimpulan dari tabel perbandingan rata-rata hasil *f-score retrieve* 10, 20, dan 30 teratas pada *short query* adalah metode TF-IDF-ICS_δF-IHS_δF mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *retrieve* 10 teratas yaitu sebesar 0,730. Grafik pada Gambar 4.13. juga menunjukkan bahwa metode TF-IDF-ICS_δF-IHS_δF memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan dari semua batas teratas *retrieve* 10 teratas adalah yang paling tinggi.

Perbandingan hasil rata-rata dari *retrieve* 10, 20, dan 30 teratas pada *long query* dapat dilihat pada Tabel 4.8. dan Gambar 4.15. Hasil *retrieve* dari 10, 20, dan

Tabel 4.8. Rata-rata Hasil *F-Score Long Query*

	Top 10	Top 20	Top 30
TF-IDF	0,625	0,532	0,410
TF-IDF-ICF	0,637	0,524	0,404
TF-IDF-ICS₈F	0,670	0,567	0,437
TF-IDF-IHS₈F	0,672	0,568	0,439
TF-IDF-ICF-IHS₈F	0,633	0,520	0,401
TF-IDF-ICS₈F-IHS₈F	0,679	0,583	0,450

30 teratas pada *long query* tersebut melakukan rekap hasil dari pengembalian sistem 10, 20, dan 30 teratas. Pada beberapa rata-rata *retrieve* antara 10, 20, dan 30 teratas dapat dilihat bahwa *retrieve* 10 teratas memiliki hasil rata-rata *f-score* tertinggi. TF-IDF mendapatkan *f-score* sebesar 0,625 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,532, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,410. TF-IDF-ICF mendapatkan *f-score* sebesar 0,637 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,524, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,404. TF-IDF-ICS₈F mendapatkan *f-score* sebesar 0,670 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,567, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,437. TF-IDF-IHS₈F mendapatkan *f-score* sebesar 0,672 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,568, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,439. Hasil TF-IDF-ICS₈F dengan TF-IDF-IHS₈F memiliki jarak yang sangat rendah, oleh sebab itu banyak dari hasil *retrieve* 2 metode tersebut memiliki jumlah hadis yang relevan hampir sama. TF-IDF-ICF-IHS₈F mendapatkan *f-score* sebesar 0,633 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,520, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,401. TF-IDF-ICS₈F-IHS₈F mendapatkan *f-score* sebesar 0,679 pada pengukuran hasil rata-rata *retrieve* 10 teratas, pada hasil rata-rata *retrieve* 20 teratas didapatkan *f-score* sebesar 0,583, dan pada hasil rata-rata *retrieve* 30 teratas didapatkan *f-score* sebesar 0,450.



Gambar 4.14. Grafik Perbandingan Hasil Rata-rata *F-Score Long Query*

Kesimpulan dari tabel perbandingan rata-rata hasil *f-score retrieve* 10, 20, dan 30 teratas pada *long query* adalah metode TF-IDF-ICS_sF-IHS_sF mampu mengungguli seluruh perbandingan *term weighting* dengan perolehan terbesar pada *retrieve* 10 teratas yaitu sebesar 0,679. Grafik pada Gambar 4.14. juga menunjukkan bahwa metode TF-IDF-ICS_sF-IHS_sF memiliki posisi tertinggi dibandingkan dengan metode lainnya, dan dari semua batas teratas *retrieve* 10 teratas adalah yang paling tinggi. Dokumentasi hasil rata-rata *precision*, *recall*, dan *f-score* untuk *retrieve* 10, 20, dan 30 teratas pada setiap *query* dirangkum di dalam tabel pada Lampiran 5.

Uji coba selanjutnya adalah uji coba pada *query* yang merupakan *matan* dari sebuah hadis yang telah diambil dari masing-masing kitab hadis yaitu 9 *query*. *Query* yang digunakan dapat dilihat pada Lampiran 6. Pada tabel tersebut ditampilkan 9 *query* yang merupakan *matan* dari perwakilan hadis masing-masing kitab dengan kategori yang sama yaitu Shalat, namun ada 2 kitab yang memiliki kategori berbeda karena tidak ada kategori Shalat di dalam kitabnya. Peneliti mengambil kategori yang hampir mirip, yaitu pada Kitab Malik diambil hadis pada kategori Adzan, dan pada kitab Ahmad diambil hadis pada kategori paling awal. Hasil perangkaian dari 9 *query* tersebut dirangkum di dalam tabel 4.12. Tabel tersebut mendokumentasikan hasil perangkaian pada keseluruhan metode *term*

Tabel 4.12. Uji Coba *Query Matan*

Id Q		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Rata-rata
TF-IDF	Ranking	2	5	1	25	1	1	1	1	1	0,67
	%	0%	0%	100%	0%	100%	100%	100%	100%	100%	67%
TF-IDF-ICF	Ranking	2	14	1	20	11	1	1	1	1	0,56
	%	0%	0%	100%	0%	100%	100%	100%	100%	100%	56%
TF-IDF-ICS _s F	Ranking	1	2	1	2	1	1	1	1	1	0,78
	%	100%	0%	100%	0%	100%	100%	100%	100%	100%	78%
TF-IDF-IHS _s F	Ranking	1	2	1	2	1	1	1	1	1	0,78
	%	100%	0%	100%	0%	100%	100%	100%	100%	100%	78%
TF-IDF-ICF-IHS _s F	Ranking	2	18	1	23	1	1	1	1	1	0,67
	%	0%	0%	100%	0%	100%	100%	100%	100%	100%	67%
TF-IDF-ICS _s F-IHS _s F	Ranking	1	2	1	1	1	1	1	1	1	0,89
	%	100%	0%	100%	100%	100%	100%	100%	100%	100%	89%

weighting yang dibandingkan. Hasil ranking akan diberi nilai presentase 100% jika mendapat ranking 1, dan diberi nilai 0% untuk yang tidak mendapatkan ranking 1. Nilai presentase tersebut selanjutnya akan dirata-rata untuk mendapatkan persentase uji coba kedua ini. Salah satu contoh hasilnya dapat dilihat pada *Query* ke 4 yang *matannya* berbunyi “Tiga orang yang shalatnya tidak akan melampaui telinga mereka; seorang budak yang kabur hingga ia kembali, seorang istri yang bermalam sementara suaminya dalam keadaan marah dan seorang imam bagi suatu kaum sedangkan mereka tidak suka” dengan *ground truth* kitab Hadis At Turmudzi kategori Shalat nomor hadis 328 mampu mendapatkan ranking 1 pada metode *term weighting* TF-IDF-ICS_sF-IHS_sF, sedangkan pada *term weighting* lainnya tidak dapat mendapatkan ranking 1. Sedangkan contoh hadis yang tidak dapat mendapatkan ranking 1 pada metode *term weighting* TF-IDF-ICS_sF-IHS_sF adalah *Query* ke 2 yang *matannya* berbunyi “Keadaan seorang hamba yang paling dekat dari Rabbnya adalah ketika dia sujud, maka perbanyaklah doa” dengan *ground truth* kitab Hadis Muslim kategori Shalat nomor 744.

Tujuan dilakukannya uji coba yang mengambil *matan* hadis sebagai *query* ini adalah untuk membuktikan bahwa metode yang diusulkan mampu mencapai tujuan utama dari penelitian ini. Mengulas kembali masalah utama yang ingin dihadapi di dalam penelitian ini yaitu Kitab hadis memiliki karakteristik yang berbeda dibandingkan dokumen konvensional pada umumnya. Setiap kitab memiliki beberapa kategori yang dibentuk oleh pengarang kitab tersebut dengan

karakteristik masing-masing. Kumpulan hadis yang memiliki kategori sama dengan pengarang berbeda dapat memiliki *term* kunci yang berbeda. Penggunaan metode pembobotan berbasis kategori (TF-IDF-ICS_F) yang ada saat ini kurang relevan jika diterapkan pada kitab hadis. Hal ini dikarenakan metode tersebut hanya mempertimbangkan densitas *term* dalam setiap kategori dan tidak memperhatikan kata kunci pada masing-masing kitab. Sehingga dibutuhkanlah pembobotan *term* baru yang mampu mempertimbangkan densitas *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab. Hasil uji coba kedua ini mampu untuk memberikan jawaban bahwa metode IHS_F yang telah diusulkan jika disandingkan dengan metode TF-IDF-ICS_F akan memberikan hasil yang lebih optimal jika dibandingkan dengan metode *term weighting* lainnya.

4.6. Analisa Hasil

Hasil penelitian yang telah dilakukan menunjukkan bahwa metode yang diusulkan mampu untuk mencapai tujuan utama penelitian yaitu membentuk sebuah pembobotan *term* baru yang mampu mempertimbangkan densitas/kepadatan *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab. Uji coba pada 50 *query* memberikan nilai terbaik pada hasil *retrieve* 10 teratas dengan memperoleh persentase 70% pada metode TF-IDF-ICS_F-IHS_F. Nilai tersebut kemudian turun menjadi 59% pada hasil *retrieve* 20 teratas, dan 45% pada hasil *retrieve* 30 teratas. Hal ini menunjukkan bahwa semakin banyak hasil *retrieve* yang diambil sebagai parameter uji coba, maka persentase *f-score* akan semakin turun. Begitu halnya dengan hasil *f-score* dari metode pembobotan yang dibandingkan lainnya, nilai *f-score* juga mengalami penurunan ketika hasil *retrieve* yang diambil semakin besar. Kemudian jarak *f-score* antara metode TF-IDF-ICS_F-IHS_F dengan metode *term weighting* pembanding lainnya memiliki jarak rata-rata sebesar 3%.

Uji coba pada 50 *query* ini kemudian dilakukan perhitungan hasil *f-score* kembali dengan membedakan jenis *query*. Dari 50 *query* tersebut dibedakan menjadi *short query* atau *query* dengan jumlah *term* kurang dan sama dengan 4, dan *long query* atau *query* dengan jumlah *term* lebih dari 4. *Short query* mendapatkan nilai terbaik pada hasil *retrieve* 10 teratas dengan memperoleh *f-score* sebesar 73%

pada metode TF-IDF-ICS_sF-IHS_sF. Nilai tersebut tetap turun ketika hasil *retrieve* yang diambil semakin besar. Kemudian pada *long query* mendapatkan nilai terbaik pada hasil *retrieve* 10 teratas juga dengan memperoleh *f-score* sebesar 68%. Hasil dari *long query* menunjukkan penurunan jika dibandingkan dengan hasil *f-score* 50 *query* dan *short query*. Penurunan kinerja ini dapat disebabkan oleh fakta bahwa ketika *query* menjadi lebih panjang, jumlah dokumen yang relevan dalam koleksi menjadi lebih kecil dan dengan demikian dokumen yang relevan menjadi lebih sulit didapatkan kembali. Uji coba kedua menggunakan potongan *matan* sebuah hadis juga menunjukkan hal tersebut. Hasil hadis yang *diretrieve* oleh sistem, hanya didapatkan beberapa saja yang mirip dengan *query* yang berupa *matan* tersebut. Kekurangan di dalam sistem pencarian hadis ini adalah hanya dapat menemukan hadis dengan *query* yang sama dengan isi hadis, hal tersebut dikarenakan sistem ini hanya melakukan kombinasi pengkategorian secara manual dari pengarang kitab dan perhitungan statistik bobot *term* dan *cosine similarity* tiap hadis dengan *query*. Kesimpulan dari uji coba pada penelitian ini adalah performa sistem (nilai *f-score*) pencarian hadis pada penelitian ini sangat dipengaruhi oleh dua hal, yaitu jumlah parameter *retrieve* yang digunakan di dalam uji coba serta jumlah dan jenis *term* pada *query* yang digunakan di dalam uji coba.

BAB V

KESIMPULAN DAN SARAN

Bab terakhir ini merupakan beberapa kesimpulan yang didapat dari hasil penelitian ini, juga saran-saran yang dapat digunakan sebagai bahan pertimbangan untuk pengembangan atau riset selanjutnya.

5.1. Kesimpulan

Penelitian yang mengusulkan metode pembobotan *Term Frequency Inverse Document Frequency Inverse Class Space Density Frequency Inverse Hadith Space Density Frequency* (TF-IDF-ICS_δF-IHS_δF) untuk perangkaian terjemahan hadis ini mendapatkan hasil sesuai tujuan. Penerapan metode yang diusulkan terhadap dataset hadis 9 kitab yang berjumlah 62165 hadis ini mampu menandingi beberapa kombinasi metode *term weighting* sebelumnya. Beberapa kesimpulan dari hasil uji coba adalah sebagai berikut:

1. Rata-rata hasil *f-score* terhadap hasil *retrieve* 10, 20, dan 30 teratas dari 50 *query* uji coba yang memperoleh nilai tertinggi adalah hasil *retrieve* 10 teratas pada metode *term weighting* TF-IDF-ICS_δF-IHS_δF yaitu 70%.
2. Rata-rata hasil *f-score* terhadap hasil *retrieve* 10, 20, dan 30 teratas dari *short query* pada 50 *query* uji coba yang memperoleh nilai tertinggi adalah hasil *retrieve* 10 teratas pada metode *term weighting* TF-IDF-ICS_δF-IHS_δF yaitu 73%.
3. Rata-rata hasil *f-score* terhadap hasil *retrieve* 10, 20, dan 30 teratas dari *long query* pada 50 *query* uji coba yang memperoleh nilai tertinggi adalah hasil *retrieve* 10 teratas pada metode *term weighting* TF-IDF-ICS_δF-IHS_δF yaitu 68%.
4. Rata-rata hasil persentase terhadap hasil *retrieve* 9 *query* dari *matan* masing-masing kitab yang memperoleh nilai tertinggi adalah metode TF-IDF-ICS_δF-IHS_δF yaitu 89%.

Kesimpulan yang dapat diambil dari beberapa hasil uji coba tersebut adalah metode *Term Frequency Inverse Document Frequency Inverse Class Space Density Frequency Inverse Hadith Space Density Frequency* (TF-IDF-ICS_δF-IHS_δF) mampu untuk mencapai tujuan utama dari penelitian ini, yaitu mengusulkan

metode pembobotan *term* baru yang mampu mempertimbangkan densitas *term* dalam setiap kategori dan memperhatikan kata kunci pada masing-masing kitab. Performa sistem (*f-score*) pencarian hadis pada penelitian ini sangat dipengaruhi oleh dua hal, yaitu jumlah parameter *retrieve* yang digunakan di dalam uji coba dan jumlah dan jenis *term* pada *query* yang digunakan di dalam uji coba.

5.2. Saran

Saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Penerapan metode pembobotan *term* yang diusulkan kepada dataset lainnya.
2. Penerapan metode pembobotan *term* yang diusulkan pada studi kasus selain perangkaan, misalnya untuk klasifikasi, cek plagiarisme, dan lain-lain.

DAFTAR PUSTAKA

- Abdelaal, H. M., Ahmed, A. M., Ghribi, W., & Youness Alansary, H. A. (2019). Knowledge Discovery in the Hadith According to the Reliability and Memory of the Reporters Using Machine Learning Techniques. *IEEE Access*, 7, 157741–157755. <https://doi.org/10.1109/ACCESS.2019.2944118>
- Abdelaal, H. M., Elemary, B. R., & Youness, H. A. (2019). Classification of Hadith According to Its Content Based on Supervised Learning Algorithms. *IEEE Access*, 7, 152379–152387. <https://doi.org/10.1109/ACCESS.2019.2948159>
- Abdi, A., Hasan, S., Arshi, M., Shamsuddin, S. M., & Idris, N. (2020). A question answering system in hadith using linguistic knowledge. *Computer Speech and Language*, 60, 101023. <https://doi.org/10.1016/j.csl.2019.101023>
- Adriani, M., Nazief, B., Asian, J., Tahaghoghi, S. M. M., and Williams, H. E. (2007). Stemming Indonesian. *ACM J. Educ. Resour. Comput.* 6, 4, 38(December), 307–314. <https://doi.org/10.1145/1316457.1316459>
- Andita Dwiwoga Tahitoe, D. P. (2010). Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia Dengan Metode Corpus Based Stemming. *Jurnal Ilmiah*, 1–15.
- Arif, A. S. M., Rahman, M. M., & Mukta, S. Y. (2009). Information retrieval by modified term weighting method using random walk model with query term position ranking. *2009 International Conference on Signal Processing Systems, ICSPS 2009*, 526–530. <https://doi.org/10.1109/ICSPS.2009.122>
- Arifin, Z. (2013). Studi Kitab Hadis. In *Srudi*. <https://doi.org/10.1364/OL.36.000127>
- Azmi, A. M., Al-Qabbany, A. O., & Hussain, A. (2019). Computational and natural language processing based studies of hadith literature: a survey. *Artificial Intelligence Review*, 52(2), 1369–1414. <https://doi.org/10.1007/s10462-019-09692-w>
- Chen, J., Kudjo, P. K., Mensah, S., Brown, S. A., & Akorfu, G. (2020). An Automatic Software Vulnerability Classification Framework Using Term Frequency-Inverse Gravity Moment and Feature Selection. *Journal of Systems and Software*, 167, 110616. <https://doi.org/10.1016/j.jss.2020.110616>
- Deposit, C., Shi, L. I., & Jianping, C. (2018). Prospecting Information Extraction by Text Mining Based on Convolutional Neural Networks — A Case Study of the Lala. *IEEE Access*, 6, 52286–52297. <https://doi.org/10.1109/ACCESS.2018.2870203>
- Dogan, T., & Uysal, A. K. (2019). Improved inverse gravity moment term

- weighting for text classification. *Expert Systems with Applications*, 130, 45–59. <https://doi.org/10.1016/j.eswa.2019.04.015>
- Hassanein, H. (2018). Discourse functions of opposition in Classical Arabic: The case in Ḥadīth Genre. In *Lingua* (Vol. 201, pp. 18–44). <https://doi.org/10.1016/j.lingua.2017.08.009>
- Jabri, S., Dahbi, A., Gadi, T., & Bassir, A. (2018). Ranking of text documents using TF-IDF weighting and association rules mining. *Proceedings of the 2018 International Conference on Optimization and Applications, ICOA 2018*, 1–6. <https://doi.org/10.1109/ICOA.2018.8370597>
- Kurniawati, & Syauqi, A. (2017). Term weighting based class indexes using space density for Al-Qur'an relevant meaning ranking. *2016 International Conference on Advanced Computer Science and Information Systems, ICACISIS 2016*, 460–463. <https://doi.org/10.1109/ICACISIS.2016.7872753>
- Lv, X., & El-Gohary, N. M. (2016). Enhanced context-based document relevance assessment and ranking for improved information retrieval to support environmental decision making. *Advanced Engineering Informatics*, 30(4), 737–750. <https://doi.org/10.1016/j.aei.2016.08.004>
- Mahendra. (2007). Enhanced Confix Stripping Stemmer And Ants Algorithm For Classifying News Document in Representation of Textual. *Technology*, April, 149–158.
- Mohammad Baharun Edisi Revisi Syamsul Rijal Hamid, O. H. (2002). *Kamus Hadis Nabi*.
- Ren, F., & Sohrab, M. G. (2013). *Class-indexing-based term weighting for automatic text classification*. 236, 109–125.
- Rostam, N. A. P., & Malim, N. H. A. H. (2019). Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting. *Journal of King Saud University - Computer and Information Sciences*, xxxx. <https://doi.org/10.1016/j.jksuci.2019.03.007>
- Sabbah, T., Selamat, A., Selamat, M. H., Al-Anzi, F. S., Viedma, E. H., Krejcar, O., & Fujita, H. (2017). Modified frequency-based term weighting schemes for text classification. *Applied Soft Computing Journal*, 58, 193–206. <https://doi.org/10.1016/j.asoc.2017.04.069>
- Saloot, M. A., Idris, N., Mahmud, R., Ja'afar, S., Thorleuchter, D., & Gani, A. (2016). Hadith data mining and classification: a comparative analysis. *Artificial Intelligence Review*, 46(1), 113–128. <https://doi.org/10.1007/s10462-016-9458-x>
- Tala, F. Z. (2003). A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *M.Sc. Thesis, Appendix D*, pp, 39–46.

- Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing and Management*, 50(1), 104–112. <https://doi.org/10.1016/j.ipm.2013.08.006>
- Yang, K., Cai, Y., Leung, H. fung, Lau, R. Y. K., & Li, Q. (2019). ITWF: A framework to apply term weighting schemes in topic model. *Neurocomputing*, 350, 248–260. <https://doi.org/10.1016/j.neucom.2019.02.048>

Lampiran 1

Tabel Kategori dalam Kitab

Shahih Bukhari (H1)

No	ID_Kategori	Kategori	Jumlah Hadis
1	C1	CBu_Wahyu	6
2	C2	CBu_Iman	50
3	C3	CBu_Ilmu	75
4	C4	CBu_Wudlu	108
5	C5	CBu_Mandi	45
6	C6	CBu_Haidl	37
7	C7	CBu_Tayamum	14
8	C8	CBu_Shalat	155
9	C9	CBu_Waktu-Shalat	77
10	C10	CBu_Adzan	259
11	C11	CBu_Jum'at	334
12	C12	CBu_Jenazah	147
13	C13	CBu_Zakat	109
14	C14	CBu_Haji	341
15	C15	CBu_Puasa	111
16	C16	CBu_Tarawih	16
17	C17	CBu_I'tikaf	21
18	C18	CBu_Jual-Beli	179
19	C19	CBu_JB-As-Salam	12
20	C20	CBu_Asy-Syuf'ah	3
21	C21	CBu_Sewa	25
22	C22	CBu_Hutang	10
23	C23	CBu_Perwakilan	17
24	C24	CBu_Pertanian	28
25	C25	CBu_Tanaman	30
26	C26	CBu_Hutang	22
27	C27	CBu_Persengketaan	16
28	C28	CBu_Temuan	12
29	C29	CBu_Zhalim	43
30	C30	CBu_Usaha	22
31	C31	CBu_Gadai	8
32	C32	CBu_Budak	45
33	C33	CBu_Hibah	65
34	C34	CBu_Kesaksian	50
35	C35	CBu_Perdamaian	18
36	C36	CBu_Syarat	22
37	C37	CBu_Wasiyat	41
38	C38	CBu_Jihad	287
39	C39	CBu_Seperlima	62
40	C40	CBu_Jizyah	29
41	C41	CBu_Penciptaan-Makhluk	81
42	C42	CBu_Riwayat-Nabi	124
43	C43	CBu_Perilaku	328
44	C44	CBu_Peperangan	324
45	C45	CBu_Tafsir-Al-Qur'an	482
46	C46	CBu_Keutamaan-Al Qur'an	79
47	C47	CBu_Nikah	175
48	C48	CBu_Talaq	82
49	C49	CBu_Nafkah	22
50	C50	CBu_Makanan	91
51	C51	CBu_Aqiqah	8
52	C52	CBu_Penyembelihan	66
53	C53	CBu_Qurban	28
54	C54	CBu_Minuman	62
55	C55	CBu_Sakit	37
56	C56	CBu_Pengobatan	91
57	C57	CBu_Pakaian	176
58	C58	CBu_Adab	246

No	ID_Kategori	Kategori	Jumlah Hadis
59	C59	CBu_Izin	70
60	C60	CBu_Do'a	104
61	C61	CBu_Hati	152
62	C62	CBu_Qadar	26
63	C63	CBu_Sumpah-Nadzar	75
64	C64	CBu_Kafarat-Sumpah	14
65	C65	CBu_Fara'idl	46
66	C66	CBu_Hudud	79
67	C67	CBu_Diyat	53
68	C68	CBu_Taubat	20
69	C69	CBu_Keterpaksanaan	12
70	C70	CBu_Siasat	28
71	C71	CBu-Ta'bir	59
72	C72	CBu_Fitnah	78
73	C73	CBu_Hukum	81
74	C74	CBu_Mengharap	20
75	C75	CBu_Ahad	17
76	C76	CBu_Kitab-Sunnah	82
77	C77	CBu_Tauhid	169

Shahih Muslim (H2)

No	ID_Kategori	Kategori	Jumlah Hadis
1	C78	Mukadimah	8
2	C79	Iman	312
3	C80	Thaharah	112
4	C81	Haid	123
5	C82	Shalat	231
6	C83	Masjid dan Tempat Shalat	278
7	C84	Shalat Musafir dan Qashar	247
8	C85	Jum'at	70
9	C86	Shalat Dua Hari Raya	22
10	C87	Shalat Istisqa	13
11	C88	Shalat Kusuf	24
12	C89	Jenazah	98
13	C90	Zakat	168
14	C91	Puasa	180
15	C92	I'tikaf	10
16	C93	Haji	426
17	C94	Nikah	114
18	C95	Menyusui	60
19	C96	Talak	53
20	C97	Li'an	10
21	C98	Memerdekakan Budak	22
22	C99	Jual Beli	95
23	C100	Pengairan	121
24	C101	Waris	17
25	C102	Hibah	28
26	C103	Wasiat	18
27	C104	Nadzar	12
28	C105	Sumpah	51
29	C106	Qusama, Pemberontak, Qishah, dan Diyat	32
30	C107	Hudud	39
31	C108	Peradilan	19
32	C109	Barang Temuan	13
33	C110	Jihad dan Ekspedisi	129
34	C111	Kepemimpinan	164
35	C112	Buruan, Sembelihan, dan Hewan yang Dimakan	55
36	C113	Hewan Qurban	36
37	C114	Minuman	153
38	C115	Pakaian dan Perhiasan	105
39	C116	Adab	45
40	C117	Salam	133
41	C118	Lafadz Adab	20
42	C119	Syair	10

No	ID_Kategori	Kategori	Jumlah Hadis
43	C120	Mimpi	25
44	C121	Keutamaan	156
45	C122	Keutamaan Shalat	223
46	C123	Berbuat Baik, Menyambut Silaturahmi, dan Adab	158
47	C124	Takdir	33
48	C125	Ilmu	15
49	C126	Dzikir, Doa, Taubat, dan Istighfar	91
50	C127	Taubat	45
51	C128	Sifat Munafik dan Hukumnya	10
52	C129	Sifat Hari Kiamat, Surga, dan Neraka	58
53	C130	Surga, Sifat, dan Penghuninya	67
54	C131	Fitnah dan Tanda Kiamat	104
55	C132	Zuhud dan Kelembutan Hati	48
56	C133	Tafsir	21

Sunan Abu Dawud (H3)

No	ID_Kategori	Kategori	Jumlah Hadis
1	C134	Thaharah	318
2	C135	Shalat	964
3	C136	Zakat	115
4	C137	Barang Temuan	10
5	C138	Haji	277
6	C139	Nikah	110
7	C140	Talak	109
8	C141	Puasa	149
9	C142	Jihad	288
10	C143	Sembelihan	55
11	C144	Buruan	16
12	C145	Washiyat	21
13	C146	Waris	40
14	C147	Pakak, Kepemimpinan dan Fai	136
15	C148	Jenazah	136
16	C149	Sumpah dan Nadzar	67
17	C150	Jual Beli	210
18	C151	Peradilan	57
19	C152	Ilmu	27
20	C153	Minuman	62
21	C154	Makanan	110
22	C155	Pengobatan	62
23	C156	Pembebasan Budak	30
24	C157	Huruf dan Bacaan	10
25	C158	Pemandian Umum	10
26	C159	Pakaian	124
27	C160	Merapikan Rambut	53
28	C161	Cincin	21
29	C162	Fitnah dan Peperangan Besar	27
30	C163	Al Mahdi	9
31	C164	Peperangan Besar	47
32	C165	Hudud	103
33	C166	Diyat	82
34	C167	Sunnah	134
35	C168	Adab	430

Ahmad (H4)

No	ID_Kategori	Kategori	Jumlah Hadis
1	C169	Musnad Sepuluh Sahabat yang Dijamin Surga	100
2	C170	Musnad Sahabat setelah Sepuluh Sahabat yang Dijamin Surga	14
3	C171	Musnad Ahli Bait	30
4	C172	Dari Musnad Bani Hasyim	23
5	C173	Musnad Sahabat yang banyak meriwayatkan Hadis	50
6	C174	Sisa Musnad Sahabat yang banyak meriwayatkan Hadis	30
7	C175	Musnad Penduduk Makkah	625

No	ID Kategori	Kategori	Jumlah Hadis
8	C176	Musnad Penduduk Madinah	433
9	C177	Musnad Penduduk Syam	671
10	C178	Musnad Penduduk Kufah	526
11	C179	Musnad Penduduk Bashrah	390
12	C180	Musnad Sahabat Anshar	287
13	C181	Sisa Musnad Sahabat Anshar	809
14	C182	Musnad beberapa Kabilah	317

Tirmidzi (H5)

No	ID Kategori	Kategori	Jumlah
1	C183	Bersuci	137
2	C184	Shalat	312
3	C185	Jum'at	110
4	C186	Zakat	58
5	C187	Puasa	119
6	C188	Haji	151
7	C189	Jenazah	112
8	C190	Nikah	65
9	C191	Penyusun	30
10	C192	Cerai	31
11	C193	Jual Beli	99
12	C194	Hukum	82
13	C195	Diyat	36
14	C196	Hudud	41
15	C197	Berburu	8
16	C198	Penyembelihan	1
17	C199	Makanan	9
18	C200	Faedah	11
19	C201	Hewan Qurban	31
20	C202	Nadzar dan Sumpah	24
21	C203	Ekspedisi	76
22	C204	Keutamaan Jihad	49
23	C205	Jihad	49
24	C206	Pakaian	68
25	C207	Makanan	74
26	C208	Minuman	35
27	C209	Berbakti dan Silaturahmi	140
28	C210	Kedokteran	57
29	C211	Fara'idl	26
30	C212	Wasiat	9
31	C213	Pemberian	8
32	C214	Qadar	25
33	C215	Fitnah	104
34	C216	Mimpi	23
35	C217	Persaksian	7
36	C218	Zuhud	113
37	C219	Qiamat	48
38	C220	Sifat Surga	50
39	C221	Sifat Jahannam	35
40	C222	Iman	38
41	C223	Ilmu	43
42	C224	Izin	48
43	C225	Adab	126
44	C226	Permisalan	14
45	C227	Keutamaan Al-Qur'an	49
46	C228	Qira'at	28
47	C229	Tafsir Al-Qur'an	311
48	C230	Do'a	203
49	C231	Budi Pekerti	302

Darimi (H6)

ID	ID_Kategori	Kategori	Jumlah Hadis
1	C232	Mukaddimah	414
2	C233	Taharah	391
3	C234	Shalat	404
4	C235	Zakat	57
5	C236	Puasa	98
6	C237	Manasik	145
7	C238	Udhiyah	55
8	C239	Buruan	16
9	C240	Makanan	62
10	C241	Minuman	47
11	C242	Mimpi	25
12	C243	Nikah	92
13	C244	Talak	32
14	C245	Hukuman	33
15	C246	Nadzar dan Iman	18
16	C247	Diyat	38
17	C248	Jihad	45
18	C249	Sejarah	91
19	C250	Jual Beli	96
20	C251	Meminta Izin	75
21	C252	Kitab Budak	136
22	C253	Faraid	290
23	C254	Wasiat	126
24	C255	Keutamaan Alquran	163

Nasai (H7)

No	ID_Kategori	Kategori	Jumlah
1	C256	Thaharah	321
2	C257	Air	23
3	C258	Haidl dan Istihadoh	47
4	C259	Mandi dan Tayamum	51
5	C260	Shalat	46
6	C261	Waktu Shalat	132
7	C262	Adzan	59
8	C263	Masjid	53
9	C264	Kiblat	35
10	C265	Keimanan	97
11	C266	Iftitah (Pembukaan)	152
12	C267	Pelaksanaan	144
13	C268	Sahwi (Lupa)	180
14	C269	Jum'at	66
15	C270	Meng-qashar Shalat saat Safar	26
16	C271	Shalat Gerhana	38
17	C272	Shalat Istisqa'	25
18	C273	Shalat Khauf (Shalat dalam Perang)	10
19	C274	Shalat Idul Adha Idul Fitri	42
20	C275	Qiyamul Lail dan Shalat Sunnah Siang	202
21	C276	Jenazah	268
22	C277	Puasa	289
23	C278	Zakat	180
24	C279	Manasik Haji	457
25	C280	Jihad	109
26	C281	Pernikahan	191
27	C282	Talaq	161
28	C283	Kuda Perang	33
29	C284	Waqaf	15
30	C285	Wasiat	59
31	C286	Pemberian	10
32	C287	Hibah	18
33	C288	Ruqba	13
34	C289	Umra	38
35	C290	Iman dan Nadzar	112

No	ID_Kategori	Kategori	Jumlah
36	C291	Menggauli Wanita	25
37	C292	Kesucian Darah	130
38	C293	Pembagian Fai'	10
39	C294	Bai'ah	61
40	C295	Aqiqah	9
41	C296	Far' dan 'Atirah	36
42	C297	Buruan dan Sembelihan	95
43	C298	Hewan Sembelihan	88
44	C299	Jual Beli	254
45	C300	Qussamah	154
46	C301	Potong Tangan	88
47	C302	Iman dan Syariatnya	55
48	C303	Perhiasan	315
49	C304	Adab Hakim	49
50	C305	Meminta Perlindungan	108
51	C306	Minuman	185

Ibnu Majah (H8)

No	ID_Kategori	Kategori	Jumlah
1	C307	Mukadimah	230
2	C308	Thaharah	396
3	C309	Shalat	39
4	C310	Adzan	29
5	C311	Masjid	68
6	C312	Mendirikan Shalat	627
7	C313	Jenazah	203
8	C314	Puasa	145
9	C315	Zakat	62
10	C316	Nikah	171
11	C317	Talak	75
12	C318	Kafarah	47
13	C319	Perdagangan	171
14	C320	Hukum	225
15	C321	Hudud	81
16	C322	Diyat	80
17	C323	Wasiat	25
18	C324	Waris	33
19	C325	Jihad	130
20	C326	Manasik	238
21	C327	Qurban	42
22	C328	Sembelihan	38
23	C329	Buruan	51
24	C330	Makanan	120
25	C331	Minuman	65
26	C332	Pengobatan	113
27	C333	Pakaian	107
28	C334	Adab	169
29	C335	Doa	66
30	C336	Takbir Mimpi	34
31	C337	Fitnah	170
32	C338	Zuhud	235

Malik (H9)

No	ID_Kategori	Kategori	Jumlah
1	C339	Waktu Shalat	27
2	C340	Thaharah	105
3	C341	Adzan	327
4	C342	Jenazah	48
5	C343	Zakat	44
6	C344	Puasa	48
7	C345	I'tikaf	12
8	C346	Haji	232

No	ID Kategori	Kategori	Jumlah
9	C347	Jihad	46
10	C348	Nadzar dan Iman	17
11	C349	Sembelihan Qurban	13
12	C350	Sembelihan	9
13	C351	Buruan	11
14	C352	Aqiqah	6
15	C353	Faraidl	13
16	C354	Nikah	45
17	C355	Talaq	92
18	C356	Penyusunan	16
19	C357	Jual-beli	76
20	C358	Pinjaman	2
21	C359	Penyiraman	2
22	C360	Sewa Tanah	4
23	C361	Syufah	2
24	C362	Peradilan	59
25	C363	Memerdekakan Budak	19
26	C364	Budak	3
27	C365	Mudabbar	2
28	C366	Hukum Pelanggaran	36
29	C367	Minuman	14
30	C368	Tebusan	34
31	C369	Qussamah	2
32	C370	Lain-lain	221

Lampiran 2

Tabel Contoh Kasus Query “malu sebagian dari iman”

id hadis	kategori	Dokumen
Sahih Bukhari Hadis No. : 8	Iman Bab : Tentang Perkara-Perkara Iman	Iman memiliki lebih dari enam puluh cabang, dan malu adalah bagian dari iman.
Sahih Bukhari Hadis No. : 23	Iman Bab : Malu Bagian Dari Iman	Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: “Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman”.
Sahih Bukhari Hadis No. : 5653	Adab Bab : Malu	“Nabi saw. pernah melewati seorang laki-laki yang tengah mencela saudaranya karena malu, kata laki-laki itu; “Sesungguhnya kamu selalu malu hingga hal itu akan membahayakan bagimu.” Maka Rasulullah saw. bersabda: “Biarkanlah ia, karena sesungguhnya sifat malu itu termasuk dari iman.”
Sahih Muslim Hadis No. : 50	Iman Bab : Penjelasan Tentang Jumlah Cabang Keimanan, yang Paling Utama dan yang Paling Rendah, dan Keutamaan Malu	Iman itu ada tujuh puluh tiga sampai tujuh puluh sembilan cabang, dan malu adalah termasuk iman.
Sahih Muslim Hadis No. : 51	Iman Bab : Penjelasan Tentang Jumlah Cabang Keimanan, yang Paling Utama dan yang Paling Rendah, dan Keutamaan Malu	Iman itu ada tujuh puluh tiga sampai tujuh puluh sembilan, atau enam puluh tiga sampai enam puluh sembilan cabang. Yang paling utama adalah perkataan, LAA ILAHA ILLALLAHU (Tidak ada tuhan yang berhak disembah selain Allah). Dan yang paling rendah adalah menyingkirkan gangguan dari jalan. Dan malu itu adalah sebagian dari iman.
Sahih Muslim Hadis No. : 52	Iman Bab : Penjelasan Tentang Jumlah Cabang Keimanan, yang Paling Utama dan yang Paling Rendah, dan Keutamaan Malu	Nabi mendengar seorang laki-laki menasihati saudaranya karena malu, maka beliau pun bersabda: “Malu itu adalah sebagian dari iman.” “Beliau melewati seorang laki-laki dari kalangan Anshar yang sedang menasihati saudaranya.”
Sunan Abu Dawud Hadis No. : 4056	Sunnah Bab : Bantahan Terhadap Kelompok Murji` Ah	Iman itu ada tujuh puluh cabang, yang paling tinggi adalah ucapan LAA ILAHA ILLAALLAH dan yang paling rendah adalah menyingkirkan tulang dari jalan, dan malu adalah bagian dari keimanan.
Sunan Abu Dawud Hadis No. : 4162	Adab Bab : Penjelasan Tentang Malu	Nabi saw. pernah melewati seorang laki-laki Anshar yang sedang menasihati saudaranya karena sikap malu. Rasulullah saw. kemudian bersabda: “Biarkanlah ia, sesungguhnya malu itu bagian dari iman.”
Sunan At Turmudzi Hadis No. : 1932	Berbakti dan Menyambung Silaturahmi Bab : Malu	Sifat malu adalah bagian dari iman, sedangkan iman itu tempatnya di dalam surga. Perkataan yang keji itu berasal dari watak dan perangai yang keras, sedangkan kekerasan itu tempatnya di dalam neraka.
Sunan At Turmudzi Hadis No. : 2540	Iman Bab : Malu Bagian Iman	Rasulullah shallallahu ‘alaihi wasallam melewati seorang laki-laki yang sedang menasihati saudaranya dalam masalah malu, maka Rasulullah shallallahu ‘alaihi wasallam bersabda: “Malu adalah sebagian dari iman.”
Sunan An Nasa’i Hadis No. : 4919	Iman dan Syareatnya Bab : Cabang-Cabang Keimanan	“Iman itu ada tujuh puluh sekian cabang yang paling utama adalah persaksian bahwa tidak ada Tuhan yang berhak disembah kecuali Allah dan yang paling rendah adalah menyingkirkan gangguan dari jalan. Dan rasa malu adalah salah satu cabang dari keimanan.”
Musnad Ahmad Hadis No. : 8993	Sisa Musnad Sahabat yang Banyak Meriwayatkan Hadis Bab : Musnad Abu Hurairah ra.	“Iman itu mempunyai tujuh puluh sekian cabang pintu, yang paling utama adalah ucapan; LAA ILAHA ILLA ALLAH, dan yang paling ringan adalah membuang duri dari jalan, dan malu adalah cabang dari iman.”

Lampiran 3
Tabel 50 Query Pengujian

Nomor Query	Query	Klasifikasi Type Query
1	umrah bulan ramadhan setara ibadah haji	Long Query
2	malu bagian dari iman	Short Query
3	tanggung jawab seorang pemimpin	Short Query
4	penyesalan adalah taubat	Short Query
5	tetangga lebih berhak terhadap dindingnya	Long Query
6	tidak diterima shalat seseorang tanpa bersuci	Long Query
7	tangan diatas lebih baik dari tangan di bawah	Long Query
8	laut itu suci airnya dan halal bangkai (hewan)nya	Long Query
9	pahala memberi makan buka puasa	Long Query
10	tidak masuk surga orang yang suka mengadu domba	Long Query
11	siwak pembersih mulut	Short Query
12	muslim sejati adalah orang yang selamat dari bahaya lisan dan tangannya	Long Query
13	menuntut ilmu wajib bagi setiap muslim	Long Query
14	menahan diri memakan harta anak yatim	Long Query
15	datangi shalat dengan berjalan tenang tidak terburu-buru	Long Query
16	larangan istinja dengan tangan kanan atau menghadap kiblat	Long Query
17	keutamaan belajar dan mengajarkan al qur'an	Long Query
18	wanita dilarang menyambung rambut	Long Query
19	Tidak Sah Shalat Tanpa Fatihah	Long Query
20	satu bulan dua puluh sembilan hari	Long Query
21	tidurlah miring ke kanan	Short Query
22	pahala puasa sunnah hari arafah	Long Query
23	larangan tampakkan paha	Short Query
24	Shalat Jika Melihat Gerhana	Short Query
25	perintah berbakti kepada orangtua	Short Query
26	barisan pertama dalam shalat didoakan para malaikat	Long Query
27	boleh makan daging kuda	Short Query
28	doa berlindung dari siksa kubur	Long Query
29	larangan duduk di atas kuburan	Short Query
30	laki-laki dilarang memakai emas	Short Query
31	mengqadha puasa ramadhan	Short Query
32	perbanyak langkah ke masjid	Short Query
33	Allah menyukai kelemah lembutan	Short Query
34	tutuplah mulut saat menguap	Short Query
35	larangan mendiamkan saudaranya lebih dari tiga hari	Long Query
36	makanlah dengan tangan kanan	Short Query
37	Durhaka kepada Orangtua	Short Query
38	hak bertamu tiga hari	Short Query
39	larangan meminum susu binatang pemakan kotoran	Long Query
40	Larangan Makan Minum sambil Berdiri	Long Query
41	tengkurap hal yang dimurkai Allah	Long Query
42	ternak sapi untuk qurban tujuh orang	Long Query

Nomor Query	Query	Klasifikasi Type Query
43	mimpi melihat rasulullah	Short Query
44	masa nifas selama empat puluh hari	Long Query
45	menikah salah satu sunnah rasul	Long Query
46	pengobatan bagian dari takdir Allah	Long Query
47	menunaikan haji	Short Query
48	Makan dengan Bersandar	Short Query
49	membaca doa qunut dalam shalat subuh	Long Query
50	setiap penyakit ada obatnya	Short Query

Lampiran 4

Gambar Contoh Dokumentasi Hasil Pencarian

<p>TFIDF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 24826) Rasulullah saw. pernah berwudhu setara dengan satu mud air dan mandi dengan setara satu sha' air</p> <p>(Ahmad No. 24642) Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Malik No. 862) Orang-orang (kaum muslimin) jika membagi harta ghanimah dari peperangan, mereka mensetarakan satu ekor unta dengan sepuluh kambing</p> <p>(Ahmad No. 5300) Saya menjual unta di Baq', terkadang saya menjualnya dengan mata uang dinar dan saya menerimanya dengan perak (yang setara dengan dinar yang saya tetapkan). Namun terkadang saya menjualnya dengan mata uang perak dan saya menerimanya dengan dinar (yang setara dengan perak yang saya tetapkan). Lalu kutemui nabi saw. Kutemui beliau sedang keluar dari rumah Hafshah. Masalahku kusampaikan kepadanya, dan beliau menjawab "Tidak masalah asalkan setara</p> <p>(Ahmad No. 3301) Barang siapa menjual makanan maka janganlah menjualnya hingga menerimanya." la berter; Ibnu Abbas berter; Dan aku mengira setiap sesuatu setara dengan ukuran makanan</p> <p>(Ibnu Madjah No. 4112) Orang-orang fakir dari kaum mukminin akan masuk surga sebelum orang-orang kaya dengan jarak setengah hari yang setara dengan lima ratus tahun</p> <p>(Ibnu Madjah No. 4113) Sesungguhnya orang-orang fakir miskin dari kaum Mukminin akan masuk surga sebelum orang-orang kaya dengan jarak setengah hari yang setara dengan lima ratus tahun</p> <p>(Ahmad No. 16938) Umrah pada bulan ramadhan menyamai ibadah haji</p> <p>(Ahmad No. 24783) Nabi saw. mandi dengan satu sha' air dan berwudhu dengan sekitar satu mud air. Yazid meriwayatkan; "Setara dengan satu mud."</p>
<p>TFIDFICF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 24826) Rasulullah saw. pernah berwudhu setara dengan satu mud air dan mandi dengan setara satu sha' air</p> <p>(Ahmad No. 24642) Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Ahmad No. 16938) Umrah pada bulan ramadhan menyamai ibadah haji</p> <p>(Ahmad No. 26028) "Umrah di bulan ramadhan seperti melaksanakan ibadah haji."</p> <p>(Ahmad No. 16940) "Umrah pada bulan ramadhan menyamai pahala ibadah haji."</p> <p>(Ibnu Madjah No. 2985) "Umrah di bulan Ramadhan menyamai (pahala) ibadah haji."</p> <p>(Ahmad No. 16939) "Aku duduk di sisi Rasulullah saw., lalu datanglah seorang wanita kepada beliau seraya berter, "Wahai Rasulullah, pada bulan apakah seharusnya aku melakukan umrah?" beliau bersabda: "Lakukanlah umrah pada bulan ramadhan. Karena umrah yang dilakukan pada bulan ramadhan menyamai ibadah haji."</p> <p>(Muslim No. 2201) Rasulullah saw. bertanya kepada seorang wanita dari kalangan Anshar -Ibnu Abbas menyebutkan namanya, tetapi aku lupa: "Apa yang menghalangimu untuk melaksanakan haji bersama kami?" wanita itu menjawab, "Kami tidak mempunyai apa-apa kecuali dua ekor Unta, yang satu ekor dipakai suamiku pergi haji bersama anaknya sedangkan yang satu lagi ia tinggalkan agar dipakai menyiram kebun." Beliau bersabda: "Kalau bulan ramadhan tiba, maka tunaikanlah umrah, sebab umrah di bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 17001) "Ibadah Umrah pada bulan ramadhan menyamai ibadah haji."</p>
<p>TFIDFICSdF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 16938) Umrah pada bulan ramadhan menyamai ibadah haji</p> <p>(Ahmad No. 26028) "Umrah di bulan ramadhan seperti melaksanakan ibadah haji."</p> <p>(Ahmad No. 17001) "Ibadah Umrah pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 16940) "Umrah pada bulan ramadhan menyamai pahala ibadah haji."</p> <p>(Ibnu Madjah No. 2985) "Umrah di bulan Ramadhan menyamai (pahala) ibadah haji."</p> <p>(Ahmad No. 16939) "Aku duduk di sisi Rasulullah saw., lalu datanglah seorang wanita kepada beliau seraya berter, "Wahai Rasulullah, pada bulan apakah seharusnya aku melakukan umrah?" beliau bersabda: "Lakukanlah umrah pada bulan ramadhan. Karena umrah yang dilakukan pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 24826) Rasulullah saw. pernah berwudhu setara dengan satu mud air dan mandi dengan setara satu sha' airHadis Musnad Ahmad No. 24642 : Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Muslim No. 2201) Rasulullah saw. bertanya kepada seorang wanita dari kalangan Anshar -Ibnu Abbas menyebutkan namanya, tetapi aku lupa: "Apa yang menghalangimu untuk melaksanakan haji bersama kami?" wanita itu menjawab, "Kami tidak mempunyai apa-apa kecuali dua ekor Unta, yang satu ekor dipakai suamiku pergi haji bersama anaknya sedangkan yang satu lagi ia tinggalkan agar dipakai menyiram kebun." Beliau bersabda: "Kalau bulan ramadhan tiba, maka tunaikanlah umrah, sebab umrah di bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 17169) "Wahai Rasulullah, sesungguhnya Ummu Ma'qil telah tertinggal untuk menunaikan haji bersamamu, lalu ia keluar untuk haji untuk menggantikan haji bersamamu." Beliau bersabda: "Hendaklah ia melakukan umrah pada bulan ramadhan, karena pahala umrah pada bulan ramadhan seperti pahalanya ibadah haji."</p>
<p>TFIDFIHSdF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 16938) Umrah pada bulan ramadhan menyamai ibadah haji</p> <p>(Ahmad No. 26028) "Umrah di bulan ramadhan seperti melaksanakan ibadah haji."</p> <p>(Ahmad No. 17001) "Ibadah Umrah pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 16940) "Umrah pada bulan ramadhan menyamai pahala ibadah haji."</p> <p>(Ibnu Madjah No. 2985) "Umrah di bulan Ramadhan menyamai (pahala) ibadah haji."</p> <p>(Ahmad No. 16939) "Aku duduk di sisi Rasulullah saw., lalu datanglah seorang wanita kepada beliau seraya berter, "Wahai Rasulullah, pada bulan apakah seharusnya aku melakukan umrah?" beliau bersabda: "Lakukanlah umrah pada bulan ramadhan. Karena umrah yang dilakukan pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 24826) Rasulullah saw. pernah berwudhu setara dengan satu mud air dan mandi dengan setara satu sha' airHadis Musnad Ahmad No. 24642 : Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Muslim No. 2201) Rasulullah saw. bertanya kepada seorang wanita dari kalangan Anshar -Ibnu Abbas menyebutkan namanya, tetapi aku lupa: "Apa yang menghalangimu untuk melaksanakan haji bersama kami?" wanita itu menjawab, "Kami tidak mempunyai apa-apa kecuali dua ekor Unta, yang satu ekor dipakai suamiku pergi haji bersama anaknya sedangkan yang satu lagi ia tinggalkan agar dipakai menyiram kebun." Beliau bersabda: "Kalau bulan ramadhan tiba, maka tunaikanlah umrah, sebab umrah di bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 17169) "Wahai Rasulullah, sesungguhnya Ummu Ma'qil telah tertinggal untuk menunaikan haji bersamamu, lalu ia keluar untuk haji untuk menggantikan haji bersamamu." Beliau bersabda: "Hendaklah ia melakukan umrah pada bulan ramadhan, karena pahala umrah pada bulan ramadhan seperti pahalanya ibadah haji."</p>
<p>TFIDFICFIHSdF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 24826) Rasulullah saw. pernah berwudhu setara dengan satu mud air dan mandi dengan setara satu sha' airHadis Musnad Ahmad No. 24642 : Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Ahmad No. 24642) Rasulullah saw. pernah membaca dalam shalat sambil duduk, apabila beliau ingin ruku' beliau berdiri dan membaca yang lamanya setara dengan orang yang membaca empat puluh ayat</p> <p>(Ahmad No. 3301) Barang siapa menjual makanan maka janganlah menjualnya hingga menerimanya." la berter; Ibnu Abbas berter; Dan aku mengira setiap sesuatu setara dengan ukuran makanan</p> <p>(Ahmad No. 5300) Saya menjual unta di Baq', terkadang saya menjualnya dengan mata uang dinar dan saya menerimanya dengan perak (yang setara dengan dinar yang saya tetapkan). Namun terkadang saya menjualnya dengan mata uang perak dan saya menerimanya dengan dinar (yang setara dengan perak yang saya tetapkan). Lalu kutemui nabi saw. Kutemui beliau sedang keluar dari rumah Hafshah. Masalahku kusampaikan kepadanya, dan beliau menjawab "Tidak masalah asalkan setara</p> <p>(Malik No. 862) Orang-orang (kaum muslimin) jika membagi harta ghanimah dari peperangan, mereka mensetarakan satu ekor unta dengan sepuluh kambingHadis Musnad Ahmad No. 5300 : Saya menjual unta di Baq', terkadang saya menjualnya dengan mata uang dinar dan saya menerimanya dengan perak (yang setara dengan dinar yang saya tetapkan). Namun terkadang saya menjualnya dengan mata uang perak dan saya menerimanya dengan dinar (yang setara dengan perak yang saya tetapkan). Lalu kutemui nabi saw. Kutemui beliau sedang keluar dari rumah Hafshah. Masalahku kusampaikan kepadanya, dan beliau menjawab "Tidak masalah asalkan setara</p> <p>(Ibnu Madjah No. 4112) Orang-orang fakir dari kaum mukminin akan masuk surga sebelum orang-orang kaya dengan jarak setengah hari yang setara dengan lima ratus tahun</p> <p>(Ibnu Madjah No. 4113) Sesungguhnya orang-orang fakir miskin dari kaum Mukminin akan masuk surga sebelum orang-orang kaya dengan jarak setengah hari yang setara dengan lima ratus tahun</p> <p>(Ahmad No. 7873) "Allah 'azza wajalla berfirman: 'Hamba-Ku telah mendustakan diri-ku padahal dia tidak punya bukti akan hal itu, dan hamba-Ku juga telah mencela-Ku padahal dia tidak punya bukti akan hal itu; pendustaannya kepada-Ku adalah ketika dia berter; 'Sekali-kali Allah tidak akan bisa mengembalikan kami sebagaimana la mencipta kami.' Dan celanya kepada-Ku adalah ketika dia berter; 'Allah mempunyai anak.' Padahal aku adalah Tuhan yang segala sesuatu bergantung kepada diri-Ku. Aku tiada beranak dan tidak pula diperanakkan, dan tidak ada seorangpun yang setara dengan Aku."</p> <p>(Ahmad No. 21727) "Tidaklah seseorang yang terluka jasadnya lalu menyedekhkannya melainkan Allah menghapus dosanya setara yang ia sedekahkan."</p>
<p>TFIDFICSdFIHSdF</p> <p>(Darimi No. 1785) Berumrahlah pada bulan Ramadhan, karena Umrah pada bulan Ramadhan setara dengan ibadah haji</p> <p>(Ahmad No. 16938) Umrah pada bulan ramadhan menyamai ibadah haji</p> <p>(Ahmad No. 17001) "Ibadah Umrah pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 16940) "Umrah pada bulan ramadhan menyamai pahala ibadah haji."</p> <p>(Ibnu Madjah No. 2985) "Umrah di bulan Ramadhan menyamai (pahala) ibadah haji."</p> <p>(Ahmad No. 26028) "Umrah di bulan ramadhan seperti melaksanakan ibadah haji."</p> <p>(Muslim No. 2201) Rasulullah saw. bertanya kepada seorang wanita dari kalangan Anshar -Ibnu Abbas menyebutkan namanya, tetapi aku lupa: "Apa yang menghalangimu untuk melaksanakan haji bersama kami?" wanita itu menjawab, "Kami tidak mempunyai apa-apa kecuali dua ekor Unta, yang satu ekor dipakai suamiku pergi haji bersama anaknya sedangkan yang satu lagi ia tinggalkan agar dipakai menyiram kebun." Beliau bersabda: "Kalau bulan ramadhan tiba, maka tunaikanlah umrah, sebab umrah di bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 16939) "Aku duduk di sisi Rasulullah saw., lalu datanglah seorang wanita kepada beliau seraya berter, "Wahai Rasulullah, pada bulan apakah seharusnya aku melakukan umrah?" beliau bersabda: "Lakukanlah umrah pada bulan ramadhan. Karena umrah yang dilakukan pada bulan ramadhan menyamai ibadah haji."</p> <p>(Ahmad No. 14268) "umrah pada bulan ramadhan menyamai haji".</p> <p>(Ahmad No. 14353) "Umrah pada (bulan) ramadhan menyamai haji"</p>

TFIDF

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 24392) "Saya tidak pernah melihat kemaluan Rasulullah saw. sama sekali."
(Ibnu Madjah No. 1912) "Aku tidak pernah melihat kemaluan Rasulullah saw. sama sekali."

TFIDFIC

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 24392) "Saya tidak pernah melihat kemaluan Rasulullah saw. sama sekali."
(Ibnu Madjah No. 1912) "Aku tidak pernah melihat kemaluan Rasulullah saw. sama sekali."

TFIDFICsDf

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 24392) "Saya tidak pernah melihat kemaluan Rasulullah saw. sama sekali."
(Ibnu Madjah No. 1912) "Aku tidak pernah melihat kemaluan Rasulullah saw. sama sekali."

TFIDFIHSdF

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 24392) "Saya tidak pernah melihat kemaluan Rasulullah saw. sama sekali."
(Ibnu Madjah No. 1912) "Aku tidak pernah melihat kemaluan Rasulullah saw. sama sekali."

TFIDFICFIHSdF

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 24392) "Saya tidak pernah melihat kemaluan Rasulullah saw. sama sekali."
(Ibnu Madjah No. 1912) "Aku tidak pernah melihat kemaluan Rasulullah saw. sama sekali."

TFIDFICsDFIHSdF

(Ahmad No. 4326) Nabi saw. mendengar seseorang memberi nasehat kepada saudaranya tentang rasa malu, lalu beliau bersabda: "Rasa malu itu adalah bagian dari iman."
(Ibnu Madjah No. 57) "Nabi saw. mendengar seorang lelaki memberi nasehat kepada saudaranya tentang malu. Maka Rasulullah saw. bersabda: " Sesungguhnya malu adalah bagian iman."
(Bukhari No. 23) Rasulullah saw. berjalan melewati seorang sahabat Anshar yang saat itu sedang memberi pengarahannya tentang malu. Maka Rasulullah saw. bersabda: "Tinggalkanlah dia, karena sesungguhnya malu adalah bagian dari iman."
(An Nasa'i No. 4947) Rasulullah saw. melewati seorang laki-laki yang memberikan nasehat kepada saudaranya mengenai rasa malu. Maka beliau bersabda: "Tinggalkanlah dia, sesungguhnya rasa malu itu sebagian dari keimanan."
(Ahmad No. 19058) "Malu itu semuanya baik."
(Ahmad No. 19157) "Malu itu baik semuanya."
(Darimi No. 71) Rasulullah saw. seorang pemalu, dan tidaklah beliau diminta sesuatu melainkan akan memberinya.
(Ahmad No. 19126) Aku mendengar Rasulullah saw. bersabda: "Malu itu semuanya baik." Kemudian ia menyebutkan hadits.
(Ahmad No. 25261) "Pada suatu hari Rasulullah shallallahu 'alaihi wa sallam meletakkan kainnya diantara kedua paha beliau, lalu datanglah Abu Bakar dan ia meminta izin, beliau pun mengizinkan dan masih seperti keadaan semula. Lalu (datanglah) Umar seperti yang dilakukan terhadap Abu Bakar, kemudian Ali kemudian beberapa orang dari sahabatnya dan Nabi shallallahu 'alaihi wa sallam masih bersikap seperti semula. Lalu datanglah Utsman meminta izin, beliau pun mengizinkan serta mengambil kainnya lalu menutupi paha beliau. Kemudian mereka berbincang-bincang hingga mereka selesai. Aku bertanya: "Wahai Rasulullah! Abu Bakar, Umar, dan Ali, dan juga beberapa sahabatmu datang dan keadaan engkau masih seperti semula. Namun, ketika datang Utsman, engkau menutup paha engkau dengan kainmu." Beliau menjawab: "Tidaklah aku malu dengan orang yang para malaikat malu kepadanya."
(Ahmad No. 25262) "Pada suatu hari Rasulullah shallallahu 'alaihi wa sallam pernah menemuiku, beliau pun meletakkan kainnya di antara kedua paha beliau, lalu datanglah Abu Bakar dan ia meminta izin. Beliau mengizinkan dan masih seperti keadaan semula. Lalu (datanglah) Umar dan ia meminta izin. Beliau mengizinkan dan masih seperti keadaan semula. Kemudian beberapa orang dari sahabatnya dan beliau pun mengizinkan. Kemudian datanglah Ali meminta izin. Beliau mengizinkan dan masih seperti keadaan semula. Lalu datanglah Utsman meminta izin dan beliau mengizinkan serta mengambil kainnya lalu menutupi paha beliau. Kemudian mereka berbincang-bincang beberapa waktu. Kemudian setelah mereka keluar, aku bertanya: "Wahai Rasulullah! Abu Bakar, Umar, dan Ali, dan juga beberapa sahabatmu datang dan keadaan engkau masih seperti semula dan tidak bergerak. Namun, ketika datang Utsman, engkau menutup paha engkau dengan kainmu. Beliau bersabda: "Tidaklah aku malu dengan orang yang para malaikat malu kepadanya."

Lampiran 5
Tabel Dokumentasi Hasil Recall, Precision, dan F-Score 50 Query

Top 10

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q1	0,105	0,200	0,138	0,421	0,800	0,552	0,474	0,900	0,621	0,474	0,900	0,621	0,053	0,100	0,069	0,526	1,000	0,690
Q2	0,381	0,800	0,516	0,381	0,800	0,516	0,381	0,800	0,516	0,381	0,800	0,516	0,381	0,800	0,516	0,381	0,800	0,516
Q3	0,375	0,600	0,462	0,313	0,500	0,385	0,563	0,900	0,692	0,563	0,900	0,692	0,250	0,400	0,308	0,625	1,000	0,769
Q4	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q5	0,176	0,300	0,222	0,176	0,300	0,222	0,353	0,600	0,444	0,353	0,600	0,444	0,118	0,200	0,148	0,353	0,600	0,444
Q6	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909
Q7	0,556	1,000	0,714	0,500	0,900	0,643	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714
Q8	0,563	0,900	0,692	0,500	0,800	0,615	0,625	1,000	0,769	0,625	1,000	0,769	0,500	0,800	0,615	0,625	1,000	0,769
Q9	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q10	0,500	1,000	0,667	0,500	1,000	0,667	0,500	1,000	0,667	0,500	1,000	0,667	0,500	1,000	0,667	0,500	1,000	0,667
Q11	0,364	0,800	0,500	0,364	0,800	0,500	0,455	1,000	0,625	0,455	1,000	0,625	0,273	0,600	0,375	0,455	1,000	0,625
Q12	0,240	0,600	0,343	0,200	0,500	0,286	0,320	0,800	0,457	0,360	0,900	0,514	0,240	0,600	0,343	0,400	1,000	0,571
Q13	0,083	0,100	0,091	0,167	0,200	0,182	0,167	0,200	0,182	0,167	0,200	0,182	0,167	0,200	0,182	0,250	0,300	0,273
Q14	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Q15	0,313	0,500	0,385	0,313	0,500	0,385	0,375	0,600	0,462	0,375	0,600	0,462	0,375	0,600	0,462	0,375	0,600	0,462
Q16	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870
Q17	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q18	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714	0,556	1,000	0,714
Q19	0,313	0,500	0,385	0,313	0,500	0,385	0,313	0,500	0,385	0,313	0,500	0,385	0,313	0,500	0,385	0,375	0,600	0,462
Q20	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769
Q21	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q22	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q23	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909
Q24	0,571	0,800	0,667	0,643	0,900	0,750	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833
Q25	0,909	1,000	0,952	0,818	0,900	0,857	0,909	1,000	0,952	0,909	1,000	0,952	0,818	0,900	0,857	0,909	1,000	0,952
Q26	0,429	0,600	0,500	0,429	0,600	0,500	0,500	0,700	0,583	0,500	0,700	0,583	0,500	0,700	0,583	0,429	0,600	0,500
Q27	0,909	1,000	0,952	0,818	0,900	0,857	0,909	1,000	0,952	0,818	0,900	0,857	0,909	1,000	0,952	0,818	0,900	0,857
Q28	0,200	0,200	0,200	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300	0,300
Q29	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Q30	0,222	0,400	0,286	0,111	0,200	0,143	0,278	0,500	0,357	0,278	0,500	0,357	0,111	0,200	0,143	0,444	0,800	0,571

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q31	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909
Q32	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833	0,714	1,000	0,833
Q33	0,500	0,800	0,615	0,500	0,800	0,615	0,438	0,700	0,538	0,500	0,800	0,615	0,500	0,800	0,615	0,500	0,800	0,615
Q34	0,533	0,800	0,640	0,533	0,800	0,640	0,533	0,800	0,640	0,533	0,800	0,640	0,533	0,800	0,640	0,667	1,000	0,800
Q35	0,500	0,700	0,583	0,500	0,700	0,583	0,571	0,800	0,667	0,571	0,800	0,667	0,500	0,700	0,583	0,500	0,700	0,583
Q36	0,588	1,000	0,741	0,588	1,000	0,741	0,588	1,000	0,741	0,588	1,000	0,741	0,588	1,000	0,741	0,588	1,000	0,741
Q37	0,235	0,400	0,296	0,235	0,400	0,296	0,176	0,300	0,222	0,235	0,400	0,296	0,235	0,400	0,296	0,235	0,400	0,296
Q38	0,571	0,800	0,667	0,571	0,800	0,667	0,500	0,700	0,583	0,500	0,700	0,583	0,571	0,800	0,667	0,571	0,800	0,667
Q39	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870	0,769	1,000	0,870
Q40	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q41	0,467	0,700	0,560	0,400	0,600	0,480	0,333	0,500	0,400	0,333	0,500	0,400	0,467	0,700	0,560	0,333	0,500	0,400
Q42	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769	0,625	1,000	0,769
Q43	0,333	0,500	0,400	0,333	0,500	0,400	0,333	0,500	0,400	0,333	0,500	0,400	0,333	0,500	0,400	0,333	0,500	0,400
Q44	0,333	0,500	0,400	0,333	0,500	0,400	0,400	0,600	0,480	0,400	0,600	0,480	0,400	0,600	0,480	0,467	0,700	0,560
Q45	0,091	0,100	0,095	0,091	0,100	0,095	0,091	0,100	0,095	0,091	0,100	0,095	0,091	0,100	0,095	0,091	0,100	0,095
Q46	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909
Q47	0,222	0,400	0,286	0,556	1,000	0,714	0,111	0,200	0,143	0,111	0,200	0,143	0,556	1,000	0,714	0,167	0,300	0,214
Q48	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952	0,909	1,000	0,952
Q49	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909	0,833	1,000	0,909
Q50	0,615	0,800	0,696	0,615	0,800	0,696	0,615	0,800	0,696	0,615	0,800	0,696	0,538	0,700	0,609	0,615	0,800	0,696

Top 20

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q1	0,105	0,100	0,103	0,421	0,400	0,410	0,474	0,450	0,462	0,474	0,450	0,462	0,053	0,050	0,051	0,526	0,500	0,513
Q2	0,381	0,400	0,390	0,381	0,400	0,390	0,381	0,400	0,390	0,381	0,400	0,390	0,381	0,400	0,390	0,381	0,400	0,390
Q3	0,375	0,300	0,333	0,313	0,250	0,278	0,563	0,450	0,500	0,563	0,450	0,500	0,250	0,200	0,222	0,625	0,500	0,556
Q4	1,273	0,700	0,903	1,182	0,650	0,839	1,273	0,700	0,903	1,182	0,650	0,839	1,182	0,650	0,839	1,182	0,650	0,839
Q5	0,176	0,150	0,162	0,176	0,150	0,162	0,353	0,300	0,324	0,353	0,300	0,324	0,118	0,100	0,108	0,353	0,300	0,324
Q6	1,250	0,750	0,938	1,167	0,700	0,875	1,250	0,750	0,938	1,250	0,750	0,938	1,083	0,650	0,813	1,250	0,750	0,938
Q7	0,556	0,500	0,526	0,500	0,450	0,474	0,556	0,500	0,526	0,556	0,500	0,526	0,556	0,500	0,526	0,556	0,500	0,526
Q8	0,563	0,450	0,500	0,500	0,400	0,444	0,625	0,500	0,556	0,625	0,500	0,556	0,500	0,400	0,444	0,625	0,500	0,556
Q9	1,273	0,700	0,903	1,182	0,650	0,839	1,273	0,700	0,903	1,273	0,700	0,903	1,182	0,650	0,839	1,273	0,700	0,903
Q10	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550	0,550
Q11	0,364	0,400	0,381	0,364	0,400	0,381	0,455	0,500	0,476	0,455	0,500	0,476	0,273	0,300	0,286	0,455	0,500	0,476
Q12	0,240	0,300	0,267	0,200	0,250	0,222	0,320	0,400	0,356	0,360	0,450	0,400	0,240	0,300	0,267	0,400	0,500	0,444
Q13	0,083	0,050	0,063	0,167	0,100	0,125	0,167	0,100	0,125	0,167	0,100	0,125	0,167	0,100	0,125	0,250	0,150	0,188
Q14	1,300	0,650	0,867	1,300	0,650	0,867	1,300	0,650	0,867	1,300	0,650	0,867	1,300	0,650	0,867	1,400	0,700	0,933
Q15	0,313	0,250	0,278	0,313	0,250	0,278	0,375	0,300	0,333	0,375	0,300	0,333	0,375	0,300	0,333	0,375	0,300	0,333
Q16	1,154	0,750	0,909	1,154	0,750	0,909	1,000	0,650	0,788	1,000	0,650	0,788	1,000	0,650	0,788	1,231	0,800	0,970
Q17	1,273	0,700	0,903	1,273	0,700	0,903	1,273	0,700	0,903	1,273	0,700	0,903	1,273	0,700	0,903	1,273	0,700	0,903
Q18	0,722	0,650	0,684	0,722	0,650	0,684	0,722	0,650	0,684	0,722	0,650	0,684	0,722	0,650	0,684	0,722	0,650	0,684
Q19	0,313	0,250	0,278	0,313	0,250	0,278	0,313	0,250	0,278	0,313	0,250	0,278	0,313	0,250	0,278	0,375	0,300	0,333
Q20	0,813	0,650	0,722	0,813	0,650	0,722	0,813	0,650	0,722	0,813	0,650	0,722	0,813	0,650	0,722	0,813	0,650	0,722
Q21	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968
Q22	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968	1,364	0,750	0,968
Q23	1,083	0,650	0,813	1,083	0,650	0,813	1,083	0,650	0,813	1,083	0,650	0,813	1,250	0,750	0,938	1,333	0,800	1,000
Q24	0,571	0,400	0,471	0,643	0,450	0,529	0,714	0,500	0,588	0,714	0,500	0,588	0,714	0,500	0,588	0,714	0,500	0,588
Q25	0,909	0,500	0,645	0,818	0,450	0,581	0,909	0,500	0,645	0,909	0,500	0,645	0,818	0,450	0,581	0,909	0,500	0,645
Q26	0,429	0,300	0,353	0,429	0,300	0,353	0,500	0,350	0,412	0,500	0,350	0,412	0,500	0,350	0,412	0,429	0,300	0,353
Q27	0,909	0,500	0,645	0,818	0,450	0,581	0,909	0,500	0,645	0,818	0,450	0,581	0,909	0,500	0,645	0,818	0,450	0,581
Q28	0,200	0,100	0,133	0,300	0,150	0,200	0,300	0,150	0,200	0,300	0,150	0,200	0,300	0,150	0,200	0,300	0,150	0,200
Q29	1,300	0,650	0,867	1,300	0,650	0,867	1,300	0,650	0,867	1,300	0,650	0,867	1,500	0,750	1,000	1,600	0,800	1,067
Q30	0,222	0,200	0,211	0,111	0,100	0,105	0,278	0,250	0,263	0,278	0,250	0,263	0,111	0,100	0,105	0,444	0,400	0,421
Q31	1,000	0,600	0,750	1,000	0,600	0,750	0,917	0,550	0,688	1,000	0,600	0,750	1,000	0,600	0,750	1,083	0,650	0,813
Q32	0,786	0,550	0,647	0,857	0,600	0,706	0,786	0,550	0,647	1,000	0,700	0,824	1,071	0,750	0,882	1,071	0,750	0,882
Q33	0,500	0,400	0,444	0,500	0,400	0,444	0,438	0,350	0,389	0,500	0,400	0,444	0,500	0,400	0,444	0,500	0,400	0,444
Q34	0,533	0,400	0,457	0,533	0,400	0,457	0,533	0,400	0,457	0,533	0,400	0,457	0,533	0,400	0,457	0,667	0,500	0,571
Q35	0,500	0,350	0,412	0,500	0,350	0,412	0,571	0,400	0,471	0,571	0,400	0,471	0,500	0,350	0,412	0,500	0,350	0,412
Q36	0,647	0,550	0,595	0,647	0,550	0,595	0,647	0,550	0,595	0,647	0,550	0,595	0,647	0,550	0,595	0,647	0,550	0,595

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q37	0,235	0,200	0,216	0,235	0,200	0,216	0,176	0,150	0,162	0,235	0,200	0,216	0,235	0,200	0,216	0,235	0,200	0,216
Q38	0,571	0,400	0,471	0,571	0,400	0,471	0,500	0,350	0,412	0,500	0,350	0,412	0,571	0,400	0,471	0,571	0,400	0,471
Q39	0,846	0,550	0,667	0,846	0,550	0,667	0,923	0,600	0,727	1,000	0,650	0,788	0,846	0,550	0,667	1,077	0,700	0,848
Q40	0,909	0,500	0,645	0,909	0,500	0,645	0,909	0,500	0,645	0,818	0,450	0,581	0,909	0,500	0,645	0,818	0,450	0,581
Q41	0,467	0,350	0,400	0,400	0,300	0,343	0,333	0,250	0,286	0,333	0,250	0,286	0,467	0,350	0,400	0,333	0,250	0,286
Q42	1,000	0,800	0,889	0,688	0,550	0,611	1,063	0,850	0,944	1,063	0,850	0,944	0,875	0,700	0,778	1,000	0,800	0,889
Q43	0,333	0,250	0,286	0,333	0,250	0,286	0,333	0,250	0,286	0,333	0,250	0,286	0,333	0,250	0,286	0,333	0,250	0,286
Q44	0,333	0,250	0,286	0,333	0,250	0,286	0,400	0,300	0,343	0,400	0,300	0,343	0,400	0,300	0,343	0,467	0,350	0,400
Q45	0,091	0,050	0,065	0,091	0,050	0,065	0,091	0,050	0,065	0,091	0,050	0,065	0,091	0,050	0,065	0,091	0,050	0,065
Q46	0,917	0,550	0,688	0,917	0,550	0,688	0,917	0,550	0,688	0,917	0,550	0,688	0,917	0,550	0,688	0,917	0,550	0,688
Q47	0,222	0,200	0,211	0,556	0,500	0,526	0,111	0,100	0,105	0,111	0,100	0,105	0,556	0,500	0,526	0,167	0,150	0,158
Q48	1,000	0,550	0,710	1,091	0,600	0,774	1,000	0,550	0,710	1,000	0,550	0,710	1,091	0,600	0,774	1,000	0,550	0,710
Q49	1,000	0,600	0,750	0,917	0,550	0,688	1,083	0,650	0,813	1,083	0,650	0,813	0,917	0,550	0,688	1,083	0,650	0,813
Q50	0,615	0,400	0,485	0,615	0,400	0,485	0,615	0,400	0,485	0,615	0,400	0,485	0,538	0,350	0,424	0,615	0,400	0,485

Top 30

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q1	0,105	0,067	0,082	0,421	0,267	0,327	0,474	0,300	0,367	0,474	0,300	0,367	0,053	0,033	0,041	0,526	0,333	0,408
Q2	0,381	0,267	0,314	0,381	0,267	0,314	0,381	0,267	0,314	0,381	0,267	0,314	0,381	0,267	0,314	0,381	0,267	0,314
Q3	0,375	0,200	0,261	0,313	0,167	0,217	0,563	0,300	0,391	0,563	0,300	0,391	0,250	0,133	0,174	0,625	0,333	0,435
Q4	1,273	0,467	0,683	1,182	0,433	0,634	1,273	0,467	0,683	1,182	0,433	0,634	1,182	0,433	0,634	1,182	0,433	0,634
Q5	0,176	0,100	0,128	0,176	0,100	0,128	0,353	0,200	0,255	0,353	0,200	0,255	0,118	0,067	0,085	0,353	0,200	0,255
Q6	1,250	0,500	0,714	1,167	0,467	0,667	1,250	0,500	0,714	1,250	0,500	0,714	1,083	0,433	0,619	1,250	0,500	0,714
Q7	0,556	0,333	0,417	0,500	0,300	0,375	0,556	0,333	0,417	0,556	0,333	0,417	0,556	0,333	0,417	0,556	0,333	0,417
Q8	0,563	0,300	0,391	0,500	0,267	0,348	0,625	0,333	0,435	0,625	0,333	0,435	0,500	0,267	0,348	0,625	0,333	0,435
Q9	1,273	0,467	0,683	1,182	0,433	0,634	1,273	0,467	0,683	1,273	0,467	0,683	1,182	0,433	0,634	1,273	0,467	0,683
Q10	0,550	0,367	0,440	0,550	0,367	0,440	0,550	0,367	0,440	0,550	0,367	0,440	0,550	0,367	0,440	0,550	0,367	0,440
Q11	0,364	0,267	0,308	0,364	0,267	0,308	0,455	0,333	0,385	0,455	0,333	0,385	0,273	0,200	0,231	0,455	0,333	0,385
Q12	0,240	0,200	0,218	0,200	0,167	0,182	0,320	0,267	0,291	0,360	0,300	0,327	0,240	0,200	0,218	0,400	0,333	0,364
Q13	0,083	0,033	0,048	0,167	0,067	0,095	0,167	0,067	0,095	0,167	0,067	0,095	0,167	0,067	0,095	0,250	0,100	0,143
Q14	1,300	0,433	0,650	1,300	0,433	0,650	1,300	0,433	0,650	1,300	0,433	0,650	1,300	0,433	0,650	1,400	0,467	0,700
Q15	0,313	0,167	0,217	0,313	0,167	0,217	0,375	0,200	0,261	0,375	0,200	0,261	0,375	0,200	0,261	0,375	0,200	0,261
Q16	1,154	0,500	0,698	1,154	0,500	0,698	1,000	0,433	0,605	1,000	0,433	0,605	1,000	0,433	0,605	1,231	0,533	0,744
Q17	1,273	0,467	0,683	1,273	0,467	0,683	1,273	0,467	0,683	1,273	0,467	0,683	1,273	0,467	0,683	1,273	0,467	0,683
Q18	0,722	0,433	0,542	0,722	0,433	0,542	0,722	0,433	0,542	0,722	0,433	0,542	0,722	0,433	0,542	0,722	0,433	0,542
Q19	0,313	0,167	0,217	0,313	0,167	0,217	0,313	0,167	0,217	0,313	0,167	0,217	0,313	0,167	0,217	0,375	0,200	0,261
Q20	0,813	0,433	0,565	0,813	0,433	0,565	0,813	0,433	0,565	0,813	0,433	0,565	0,813	0,433	0,565	0,813	0,433	0,565
Q21	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732
Q22	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732	1,364	0,500	0,732
Q23	1,083	0,433	0,619	1,083	0,433	0,619	1,083	0,433	0,619	1,083	0,433	0,619	1,250	0,500	0,714	1,333	0,533	0,762
Q24	0,571	0,267	0,364	0,643	0,300	0,409	0,714	0,333	0,455	0,714	0,333	0,455	0,714	0,333	0,455	0,714	0,333	0,455
Q25	0,909	0,333	0,488	0,818	0,300	0,439	0,909	0,333	0,488	0,909	0,333	0,488	0,818	0,300	0,439	0,909	0,333	0,488
Q26	0,429	0,200	0,273	0,429	0,200	0,273	0,500	0,233	0,318	0,500	0,233	0,318	0,500	0,233	0,318	0,429	0,200	0,273
Q27	0,909	0,333	0,488	0,818	0,300	0,439	0,909	0,333	0,488	0,818	0,300	0,439	0,909	0,333	0,488	0,818	0,300	0,439
Q28	0,200	0,067	0,100	0,300	0,100	0,150	0,300	0,100	0,150	0,300	0,100	0,150	0,300	0,100	0,150	0,300	0,100	0,150
Q29	1,300	0,433	0,650	1,300	0,433	0,650	1,300	0,433	0,650	1,300	0,433	0,650	1,500	0,500	0,750	1,600	0,533	0,800
Q30	0,222	0,133	0,167	0,111	0,067	0,083	0,278	0,167	0,208	0,278	0,167	0,208	0,111	0,067	0,083	0,444	0,267	0,333
Q31	1,000	0,400	0,571	1,000	0,400	0,571	0,917	0,367	0,524	1,000	0,400	0,571	1,000	0,400	0,571	1,083	0,433	0,619
Q32	0,786	0,367	0,500	0,857	0,400	0,545	0,786	0,367	0,500	1,000	0,467	0,636	1,071	0,500	0,682	1,071	0,500	0,682
Q33	0,500	0,267	0,348	0,500	0,267	0,348	0,438	0,233	0,304	0,500	0,267	0,348	0,500	0,267	0,348	0,500	0,267	0,348
Q34	0,533	0,267	0,356	0,533	0,267	0,356	0,533	0,267	0,356	0,533	0,267	0,356	0,533	0,267	0,356	0,667	0,333	0,444
Q35	0,500	0,233	0,318	0,500	0,233	0,318	0,571	0,267	0,364	0,571	0,267	0,364	0,500	0,233	0,318	0,500	0,233	0,318
Q36	0,647	0,367	0,468	0,647	0,367	0,468	0,647	0,367	0,468	0,647	0,367	0,468	0,647	0,367	0,468	0,647	0,367	0,468

Id query	TF-IDF			TF-IDF-ICF			TF-IDF-ICSδF			TF-IDF-IHSδF			TF-IDF-ICF-IHSδF			TF-IDF-ICSδF-IHSδF		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
Q37	0,235	0,133	0,170	0,235	0,133	0,170	0,176	0,100	0,128	0,235	0,133	0,170	0,235	0,133	0,170	0,235	0,133	0,170
Q38	0,571	0,267	0,364	0,571	0,267	0,364	0,500	0,233	0,318	0,500	0,233	0,318	0,571	0,267	0,364	0,571	0,267	0,364
Q39	0,846	0,367	0,512	0,846	0,367	0,512	0,923	0,400	0,558	1,000	0,433	0,605	0,846	0,367	0,512	1,077	0,467	0,651
Q40	0,909	0,333	0,488	0,909	0,333	0,488	0,909	0,333	0,488	0,818	0,300	0,439	0,909	0,333	0,488	0,818	0,300	0,439
Q41	0,467	0,233	0,311	0,400	0,200	0,267	0,333	0,167	0,222	0,333	0,167	0,222	0,467	0,233	0,311	0,333	0,167	0,222
Q42	1,000	0,533	0,696	0,688	0,367	0,478	1,063	0,567	0,739	1,063	0,567	0,739	0,875	0,467	0,609	1,000	0,533	0,696
Q43	0,333	0,167	0,222	0,333	0,167	0,222	0,333	0,167	0,222	0,333	0,167	0,222	0,333	0,167	0,222	0,333	0,167	0,222
Q44	0,333	0,167	0,222	0,333	0,167	0,222	0,400	0,200	0,267	0,400	0,200	0,267	0,400	0,200	0,267	0,467	0,233	0,311
Q45	0,091	0,033	0,049	0,091	0,033	0,049	0,091	0,033	0,049	0,091	0,033	0,049	0,091	0,033	0,049	0,091	0,033	0,049
Q46	0,917	0,367	0,524	0,917	0,367	0,524	0,917	0,367	0,524	0,917	0,367	0,524	0,917	0,367	0,524	0,917	0,367	0,524
Q47	0,222	0,133	0,167	0,556	0,333	0,417	0,111	0,067	0,083	0,111	0,067	0,083	0,556	0,333	0,417	0,167	0,100	0,125
Q48	1,000	0,367	0,537	1,091	0,400	0,585	1,000	0,367	0,537	1,000	0,367	0,537	1,091	0,400	0,585	1,000	0,367	0,537
Q49	1,000	0,400	0,571	0,917	0,367	0,524	1,083	0,433	0,619	1,083	0,433	0,619	0,917	0,367	0,524	1,083	0,433	0,619
Q50	0,615	0,267	0,372	0,615	0,267	0,372	0,615	0,267	0,372	0,615	0,267	0,372	0,538	0,233	0,326	0,615	0,267	0,372

Lampiran 6
Tabel Uji Coba Kedua: *Query Matan*

Id Query	Query	Ground Truth		
		No. Hadis	Kategori	Kitab
Q1	Barangsiapa lewat dengan membawa panah di masjid atau pasar kita, maka hendaklah dipegang ujung panahnya dengan tangannya agar tidak melukai seorang muslim	433	Shalat	Bukhari
Q2	Keadaan seorang hamba yang paling dekat dari Rabbnya adalah ketika dia sujud, maka perbanyaklah doa	744	Shalat	Muslim
Q3	Rasulullah shallallahu 'alaihi wasallam shalat dzuhur pada musim dingin disaat sekitar panjang bayangan tiga kaki hingga lima kaki, dan apabila musim panas sekitar lima kaki hingga tujuh kaki	339	Shalat	Abu Daud
Q4	Tiga orang yang shalatnya tidak akan melampai telinga mereka; seorang budak yang kabur hingga ia kembali, seorang istri yang bermalam sementara suaminya dalam keadaan marah dan seorang imam bagi suatu kaum sedangkan mereka tidak suka	328	Shalat	Tirmidzi
Q5	Perintahkan suami-suami kalian bersuci dengan air, karena aku malu untuk mengatakan kepada mereka bahwa Rasulullah shallallahu 'alaihi wasallam melakukan hal tersebut	46	Shalat	Nasa'i

Id Query	Query	Ground Truth		
		No. Hadis	Kategori	Kitab
Q6	Barangsiapa diberi fatwa dengan fatwa yang tidak benar, maka dosanya atas orang yang memberi fatwa	52	Shalat	Ibnu Majah
Q7	Akupun mengikuti mulutnya bergerak ke arah sini dan ke arah sana saat mengumandangkan adzan	1172	Shalat	Darimi
Q8	Ada dua waktu yang di dalamnya dibukakan pintu-pintu langit, dan sedikit orang yang tertolak doanya; saat adzan dan saat berperang di jalan Allah	140	Adzan	Malik
Q9	barangsiapa berbuat suatu kejahatan, maka Allah akan membalasnya didunia	23	Musnad Sepuluh Sahabat yang Dijamin Masuk Surga	Ahmad

BIODATA PENULIS



Ana Tsalitsatun Ni'mah merupakan anak ketiga dari empat bersaudara dari pasangan Moh. Sholichin dengan Ismiati. Penulis merupakan seorang istri dari Alfian Qomaruddin dan seorang ibu dari Muhammad Harits Al Fatih dan Dwindha Hasna Az Zakiyah. Penulis menempuh pendidikan di TK Dharma Wanita Kusuma Mulia Bakalan Grogol Kediri (1994-1996), Madrasah Ibtidaiyah Miftahul Huda Bakalan Grogol Kediri (1996-2002), SMP Negeri 1 Grogol Kediri (2002-2005), SMA Negeri 1 Grogol Kediri (2005-2008), Program Studi Pendidikan Matematika, Fakultas Teknik, Universitas Nusantara PGRI Kediri (2008-2009), dan Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas

Trunojoyo Madura (2009-2013). Saat menempuh skripsi (Tugas Akhir S1), penulis mengambil minat *information retrieval* dan mengangkat topik mengenai pengembangan *stemming*, yang mengantarkan penulis pada kegemaran terhadap bidang ilmu komputer dan melanjutkan studi di Pascasarjana Teknik Informatika ITS. Penulis memiliki ketertarikan dalam penelitian pada bidang *Information Retrieval*. Penulis dapat dihubungi via email di anatsalits@gmail.com