



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

SISTEM PENCARIAN AYAT AL-QURAN BERBASIS KEMIRIPAN FONETIS

MUHAMMAD ABRAR ISTIADI



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2012**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

SISTEM PENCARIAN AYAT AL-QURAN BERBASIS KEMIRIPAN FONETIS

MUHAMMAD ABRAR ISTIADI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2012**



ABSTRACT

MUHAMMAD ABRAR ISTIADI. Holy Quran Search System Based on Phonetic Similarity. Supervised by AHMAD RIDHA.

Searching Arabic text in the Holy Quran is not easy for users that do not have sufficient knowledge about Arabic language and script. Therefore, phonetic search can be used to facilitate users to search in the Holy Quran with their pronunciation represented in Latin script using regular alphabets. This research aims to build such search system, specifically for speakers of Bahasa Indonesia. A phonetic coding method regarding Quran recitation rules (tajweed) is proposed to match between Quran texts in Arabic script and user's queries in Latin script. Indexed trigram is used for approximate string matching. The system uses two search schemes: vowelized and non-vowelized search; two ranking methods: trigram count and trigram position ranking; and two search purposes: pronunciation and topic search. Experiment using user-generated queries and given relevance judgments shows that vowelized search performs better than non-vowelized search with 0.651 average precision (AVP) at the cost of longer search time. Trigram count ranking performs better than trigram position ranking with 0.668 AVP, although trigram position ranking can obtain higher precision at lower recall level. Pronunciation search queries perform better than topic search queries with 0.751 AVP.

Keywords: Arabic, Holy Quran, phonetic search, trigram

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Judul Skripsi : Sistem Pencarian Ayat Al-Quran Berbasis Kemiripan Fonetis
Nama : Muhammad Abrar Istiadi
NRP : G64080008

Menyetujui:
Pembimbing

Ahmad Ridha, S.Kom, MS
NIP.19800507 200501 1 001

Mengetahui:
Ketua Departemen Ilmu Komputer

Dr. Ir. Agus Buono, M.Si, M.Kom
NIP. 19660702 199302 1 001

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

Tanggal Lulus :

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



KATA PENGANTAR

Segala puji bagi Allah *subhanahu wata'ala* atas segala limpahan rahmat serta karunia-Nya sehingga penulis mampu menyelesaikan penelitian ini dengan baik. *Shalawat* dan salam penulis sampaikan kepada Nabi Muhammad *shallallahu 'alaihi wasallam* serta kepada keluarganya, sahabatnya, serta para pengikutnya yang selalu berpegang kepada Al-Quran dan As-Sunnah. Penulis juga menyampaikan terima kasih kepada seluruh pihak yang telah berperan dalam penelitian ini, yaitu:

- 1 Ayahanda Oni Setiyadi, Ibunda Istinah Laksiastuti, serta Adik Khairunnisa Anbar Istiadi atas doa, kasih sayang, dukungan, serta motivasi kepada penulis untuk penyelesaian penelitian ini.
 - 2 Bapak Ahmad Ridha, S.Kom, MS selaku dosen pembimbing yang telah memberi banyak ide, saran, bantuan, serta dukungan sampai selesainya penelitian ini.
 - 3 Bapak Firman Ardiansyah, S.Kom, M.Si dan Ir. Julio Adisantoso, M.Kom yang telah bersedia menjadi penguji.
 - 4 Rekan-rekan seperjuangan di Ilmu Komputer IPB angkatan 45 atas segala kebersamaan, bantuan, dukungan, serta kenangan bagi penulis selama menjalani masa studi. Semoga kita bisa berjumpa kembali kelak sebagai orang-orang sukses.
 - 5 Abdul Qifli Sangadji, Isnan Mulia, Alif Kurniawan, Arief Hidayatulloh, dan sahabat lainnya yang telah menjadi mitra dan menemani penulis dalam menjalani kehidupan sebagai mahasiswa.
 - 6 Rekan-rekan satu bimbingan, Fajar Surya Dharma dan Roni Rahmon, semoga lancar dalam melanjutkan penelitiannya.
 - 7 Rekan-rekan satu kontrakan di Wisma At-Tauhid, Mas Auriza, Angga, Dendy, Hasan, Rusmanto, Frendy, dan Mujianto yang telah memberikan dukungan untuk segera menyelesaikan studi.
 - 8 Masyarakat Internet penghuni milis serta jejaring sosial yang telah melakukan uji coba, memberikan masukan, serta bersedia menjadi responden sebagai salah satu bagian dari penelitian.
- Penulis berharap penelitian ini dapat memberikan manfaat, khususnya bagi umat Islam di Indonesia dan masyarakat pada umumnya.

Bogor, April 2012

Muhammad Abrar Istiadi

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



RIWAYAT HIDUP

Penulis dilahirkan di Klaten, Jawa Tengah pada tanggal 29 Agustus 1990. Penulis merupakan anak pertama dari pasangan Oni Setiyadi dan Istinah Laksiastuti. Pada tahun 2008, penulis menamatkan pendidikan di SMA Negeri 1 Metro, Lampung. Penulis lulus seleksi masuk Institut Pertanian Bogor (IPB) pada tahun yang sama melalui jalur Undangan Seleksi Masuk IPB dan diterima sebagai mahasiswa di Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam.

Selama aktif menjadi mahasiswa, penulis menjadi salah satu pengurus Himpunan Mahasiswa Ilmu Komputer (Himalkom) pada tahun 2010. Penulis juga menjadi asisten praktikum pada Mata Kuliah Algoritme dan Pemrograman (2010), Bahasa Pemrograman (2011), Komputer Grafik (2011), Komunikasi Data dan Jaringan Komputer (2012), serta Pengembangan Sistem Berorientasi Objek (2012). Selain itu, penulis melaksanakan kegiatan Praktik Kerja Lapangan di Kantor Komunikasi dan Informatika Kota Bogor pada tahun 2011.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



DAFTAR ISI

	Halaman
DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN	vi
PENDAHULUAN	
Latar Belakang.....	1
Tujuan	2
Ruang Lingkup	2
METODE PENELITIAN	
Gambaran Umum Sistem	2
Data Teks Al-Quran	2
Pengodean Fonetis	2
Tokenisasi Trigram	3
Pengeindeksan Trigram	3
Pencocokan Trigram	3
Pemeringkatan Dokumen	3
<i>Relevance Judgment</i>	3
Pengukuran Kinerja	4
Implementasi	4
HASIL DAN PEMBAHASAN	
Praproses Data	4
Pengodean Fonetis Teks Al-Quran.....	4
Pengodean Fonetis Teks Latin.....	6
Tokenisasi Trigram	8
Pengeindeksan Trigram	8
Pemrosesan <i>Query</i>	10
Pemeringkatan Dokumen	10
Pembuatan Antarmuka Pengguna.....	11
Pembentukan Koleksi Pengujian	12
Pengujian Sistem	13
KESIMPULAN DAN SARAN	
Kesimpulan.....	16
Saran	16
DAFTAR PUSTAKA	16
LAMPIRAN	18

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



DAFTAR TABEL

Halaman

1	Aturan pemadanan aksara Arab ke kode fonetis	5
2	Contoh pengodean fonetis dari teks Al-Quran	6
3	Aturan pemadanan aksara Latin ke kode fonetis	7
4	Contoh pengodean fonetis dari teks Latin	8
5	Perbandingan hasil pengindeksan dengan vokal dan tanpa vokal	9
6	Nilai AVP per <i>query</i>	13
7	Waktu komputasi setiap unit percobaan (dalam detik)	16

DAFTAR GAMBAR

Halaman

1	Diagram alir proses pada sistem	2
2	Contoh skema indeks trigram (Kim <i>et al.</i> 2005)	3
3	Hustrasi tokenisasi trigram	8
4	Hustrasi pembangunan indeks	9
5	Hustrasi pemrosesan <i>query</i>	10
6	Antarmuka awal sistem	12
7	Antarmuka hasil pencarian	12
8	Grafik perbandingan pencarian dengan vokal dan tanpa vokal	14
9	Grafik perbandingan pemeringkatan jumlah dan posisi	14
10	Grafik perbandingan pengujian kelompok <i>query</i> A dan B	15
11	Grafik perbandingan keempat unit percobaan	15

DAFTAR LAMPIRAN

Halaman

1	Daftar <i>fawatihush shuwar</i> dan bacaannya	19
2	Contoh isi dari <i>file</i> penyimpanan indeks	20
3	Algoritme untuk mencari LIS (Liben-Nowell & Vee 2006)	21
4	Daftar <i>query</i> (aksara Arab) untuk pengujian pencarian lafal	22
5	Daftar <i>query</i> (aksara Arab) untuk pengujian pencarian topik	23
6	Daftar variasi penulisan <i>query</i> untuk pengujian	24
7	Grafik <i>precision-recall</i> untuk masing-masing <i>query</i>	26

PENDAHULUAN

Latar Belakang

Al-Quran adalah kitab suci yang menjadi rujukan utama bagi umat Islam di seluruh dunia. Al-Quran diturunkan seluruhnya dalam bahasa Arab dan dituliskan dengan aksara Arab (*hijaiyyah*). Secara statistik, Al-Quran terdiri atas 114 surat, 6236 ayat, dan 77 845 kata (Hammo *et al.* 2007). Dengan jumlah surat, ayat, dan kata yang cukup banyak, pencarian kata pada teks Al-Quran secara manual sulit dilakukan. Oleh karena itu, komputer dapat digunakan untuk membantu melakukan pencarian di dalam Al-Quran.

Penelitian tentang pengembangan sistem untuk membantu pencarian ayat Al-Quran telah dilakukan sejak lama, salah satunya oleh Ahmad (1988) yang menggunakan metode pencarian secara sekuensial. Saat ini juga telah banyak dikembangkan aplikasi perangkat lunak untuk mempelajari Al-Quran. Salah satu contoh yang berbasis *desktop* ialah Zekr, sedangkan yang berbasis web ialah Tanzil Quran Navigator (ZEKR 2010; TANZ 2011).

Pada aplikasi-aplikasi perangkat lunak Al-Quran yang telah disebutkan, tersedia fasilitas pencarian ayat yang mengharuskan pengguna untuk memasukkan kata kunci pencarian dalam bahasa dan aksara Arab. Hal ini cukup menyulitkan pengguna yang tidak bisa berbahasa Arab maupun menulis dengan aksara Arab. Selain itu, untuk mengetik aksara Arab di komputer dibutuhkan *keyboard* khusus atau perangkat lunak tambahan.

Oleh karena itu, untuk mengatasi kesulitan pengguna dalam melakukan pencarian dengan aksara Arab, perlu dikembangkan sistem pencarian berbasis kemiripan fonetis untuk teks Al-Quran. Pencarian berbasis kemiripan fonetis memanfaatkan pencocokan fonetis untuk membuat padanan antara aksara Arab dan pelafalan dalam aksara Latin dan sebaliknya.

Pencocokan fonetis biasanya digunakan untuk pencarian nama orang yang ejaannya bisa berbeda meskipun pelafalannya serupa (Zobel & Dart 1996). Beberapa metode atau algoritme pengodean fonetis telah dikembangkan, antara lain yang paling populer ialah Soundex untuk pencocokan nama dalam bahasa Inggris. Aqeel *et al.* (2006) mengembangkan algoritme ASoundex yang mengadaptasi dari Soundex untuk mencari transliterasi (alihaksara) nama asing dalam bahasa Arab. Nwesri (2008) mengembangkan dua algoritme serupa, yaitu algoritme NORM yang melakukan normalisasi

huruf-huruf dalam bahasa Arab yang ekuivalen dan algoritme Soutex yang mirip dengan ASoundex dengan beberapa modifikasi.

Dengan pencocokan fonetis, pencarian pada teks Al-Quran dapat dilakukan dengan kata kunci pencarian berupa pelafalan kata dalam aksara Latin. Selain itu, pencarian bersifat toleran terhadap perbedaan cara pelafalan atau penulisan lafal yang mungkin terjadi.

Untuk metode pencocokan *string* (dalam hal ini ayat Al-Quran) secara tidak tepat sama, dapat digunakan metode *n*-gram yang digabungkan dengan pengodean fonetis. Nilai *n* atau jumlah gram yang sering digunakan ialah 2 dan 3 (bigram dan trigram) (Aqeel *et al.* 2006). Beberapa pendekatan statistik untuk temu kembali informasi bahasa Arab telah diuji, dan trigram adalah jumlah gram terbaik untuk mengindeks teks berbahasa Arab (Nwesri 2008). Keuntungan menggunakan trigram atau *n*-gram secara umum ialah dapat dilakukan pengindeksan dengan tokenisasi *n*-gram pada korpus untuk membangun struktur data *inverted index* sehingga pencarian *term* tertentu dapat dilakukan dengan cepat (Kim *et al.* 2005). Nurhanifah (2001) menggunakan metode trigram terindeks untuk pencocokan data secara *inexact* pada istilah-istilah berbahasa Inggris dari dokumen pertanian dan memberikan hasil yang baik.

IslamiCity (IC 2001) telah mengembangkan sistem pencarian Al-Quran secara fonetis yang berbasis web. Namun, sistem tersebut menggunakan pemadanan aksara Arab-Latin internasional yang sedikit berbeda dengan yang digunakan di Indonesia. Perbedaan tersebut antara lain pemadanan aksara Arab *th* dan *dh* dengan aksara Latin *th*, sedangkan yang umum digunakan di Indonesia adalah pemadanan *th* dengan *ts* dan *dh* dengan *dz*. Perbedaan ini dapat menjadi kendala bagi pengguna yang tidak terbiasa dengan pemadanan aksara Arab-Latin internasional.

Oleh karena itu, pada penelitian ini dibangun sistem pencarian ayat Al-Quran berbasis kemiripan fonetis yang lebih sesuai dengan representasi pelafalan orang Indonesia. Untuk tujuan itu, dikembangkan metode pengodean fonetis yang didasarkan pada pemadanan aksara Arab-Latin yang digunakan di Indonesia serta kemiripan cara pelafalan huruf-huruf dalam Al-Quran. Metode pencarian yang digunakan adalah pencarian dengan trigram yang diterapkan pada kode fonetis dengan ukuran kesamaan yang ditentukan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Pengujian terhadap sistem juga dilakukan untuk mengetahui seberapa baik sistem dapat melakukan pencarian. Ukuran yang digunakan untuk menguji kinerja sistem ialah waktu pencarian dan *average precision*.

Tujuan

Tujuan penelitian ini, yaitu:

1. Membangun sistem pencarian ayat Al-Quran berbasis kemiripan fonetis.
2. Mengembangkan metode pengodean fonetis untuk teks Al-Quran yang sesuai untuk pembicara bahasa Indonesia.
3. Mengukur kinerja sistem pencarian ayat Al-Quran berbasis kemiripan fonetis.

Ruang Lingkup

Ruang lingkup penelitian ini, yaitu:

1. Menggunakan korpus teks Al-Quran dengan *harakat* (tanda vokal) lengkap.
2. Menggunakan aturan alihaksara Arab-Latin yang umum digunakan di Indonesia.
3. Menggunakan aturan cara membaca Al-Quran (*qira'ah* dan *tajwid*) yang umum digunakan di Indonesia, yaitu *qira'ah* Imam Ashim riwayat Hafsh.

METODE PENELITIAN

Gambaran Umum Sistem

Gambaran umum proses pada sistem diberikan pada Gambar 1. Secara umum, proses pada sistem terbagi menjadi 2, yaitu pemrosesan secara *offline* dan pemrosesan secara *online*.

Pemrosesan *offline* hanya dilakukan satu kali untuk membentuk indeks. Pemrosesan *online* dilakukan setiap kali *query* dimasukkan ke sistem dan menggunakan indeks yang telah dibentuk untuk melakukan pencarian. Kedua pemrosesan ini secara umum memiliki langkah-langkah yang sama, di antaranya pengodean fonetis dan tokenisasi trigram. Proses pencarian yang sesungguhnya ada pada pemrosesan *online*, yaitu pencocokan trigram dan pemeringkatan dokumen.

Data Teks Al-Quran

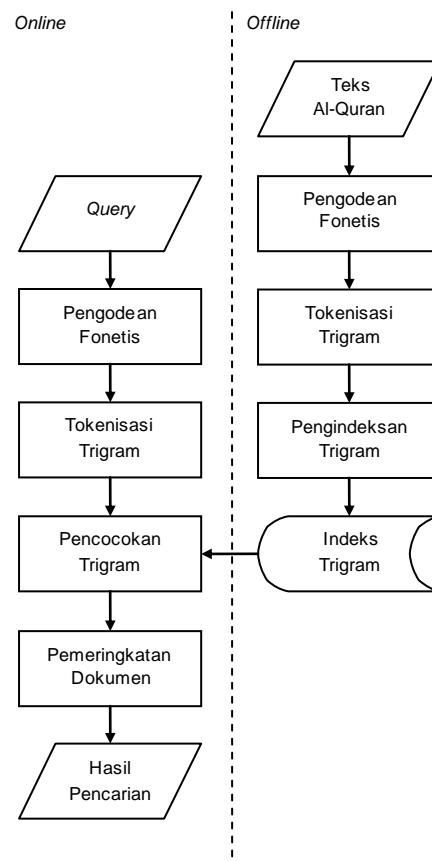
Data teks Al-Quran sebagai korpus diperoleh dari Tanzil Project*. Data ini berupa *file* teks yang setiap barisnya berupa 1 ayat

dalam Al-Quran. Teks ini berisi ayat-ayat Al-Quran dalam bahasa dan aksara Arab dengan tanda vokal lengkap dan tanpa tanda berhenti dalam ayat.

Pengodean Fonetis

Algoritme pengodean fonetis yang telah disebutkan (Aqeel *et al.* 2006; Nwesri 2008) tidak dapat digunakan secara langsung karena ditujukan untuk nama orang dalam aksara Arab, bukan teks berbahasa Arab secara umum. Oleh karena itu, dibuat suatu algoritme serupa yang dapat memetakan aksara Arab dan penulisan lafal dalam aksara Latin ke suatu kode yang sama bila pelafalannya serupa.

Algoritme pengodean fonetis yang dibuat juga mempertimbangkan cara membaca Al-Quran (*tajwid*) yang sedikit berbeda dengan cara membaca teks berbahasa Arab biasa. Selain itu, untuk mengantisipasi kesalahan cara pelafalan Al-Quran yang sering terjadi,



Gambar 1 Diagram alir proses pada sistem.

* Alamat web: <http://tanzil.net/download>

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

dipertimbangkan juga kesalahan yang umum terjadi tersebut berdasarkan penelitian Junaedi (2009). Contoh kesalahan tersebut ialah pelafalan huruf ش (*syin*) dengan lafal س (*sin*).

Pengodean fonetis diterapkan baik pada teks Al-Quran maupun pada *query* yang dimasukkan oleh pengguna. Hasil dari proses pengodean adalah *string* kode fonetis.

Tokenisasi Trigram

Pada *string* kode fonetis yang dihasilkan, baik dari *query* maupun teks Al-Quran, dilakukan tokenisasi untuk mengambil trigram. Trigram yang diambil tidak memerlukan penanda awal atau akhir *string* karena *query* dapat berupa *substring* dari teks Al-Quran.

Proses tokenisasi menggunakan *overlapping window* sepanjang 3 karakter (Nwesri 2008). Sebagai contoh, trigram dari *string* "ARABIC" adalah "ARA", "RAB", "ABI", dan "BIC".

Pengindeksan Trigram

Pada kode fonetis teks Al-Quran yang telah ditokenisasi, dilakukan pembentukan *inverted index*. *Inverted index* menggunakan trigram sebagai *term* dan ayat Al-Quran sebagai dokumen. Satu dokumen pada indeks adalah satu ayat pada Al-Quran. Informasi yang disimpan pada indeks adalah *identifier* dokumen, jumlah trigram tertentu pada dokumen, serta posisi kemunculan pertama trigram pada dokumen.

Contoh indeks dengan trigram dapat dilihat pada Gambar 2. Misalkan terdapat sekumpulan dokumen dengan *identifier* 7, 12, 44, dan 97 yang mengandung *string* "string" dan "data". Dari "string" dapat diambil 4 trigram, yaitu "str", "tri", "rin", dan "ing", sedangkan dari "data" dapat diambil 2 trigram, yaitu "dat" dan "ata".

trigram	identifier dokumen [posisi]		
ata	7 [59]	12 [4, 28]	44 [84]
dat	7 [58]	12 [5, 27]	44 [83]
ing	7 [9, 54]	44 [15]	97 [7, 90]
rin	7 [8, 53]	44 [14]	97 [6, 89]
tri	7 [7, 52]	44 [13]	97 [5, 88]
str	7 [6, 51]	44 [12]	97 [4, 87]

Gambar 2 Contoh skema indeks trigram (Kim et al. 2005).

Pada proses pengindeksan diterapkan 2 metode, yaitu pengindeksan dengan menggunakan huruf vokal dan pengindeksan tanpa huruf vokal. Hasil dari kedua metode pengindeksan ini dibandingkan untuk dicari yang paling baik.

Pencocokan Trigram

Trigram dari *query* dibandingkan dengan trigram yang ada pada indeks. Pada tahap ini, dihitung jumlah trigram dari dokumen yang sama dengan trigram dari *query*. Perhitungan dilakukan dengan memanfaatkan informasi yang tersimpan dalam indeks.

Pemeringkatan Dokumen

Pemeringkatan dokumen dilakukan dengan 2 metode, yaitu:

1. Pemeringkatan yang hanya menghitung jumlah trigram yang sama antara dokumen dan *query*. Semakin banyak jumlah trigram yang sama, semakin tinggi peringkatnya (pemeringkatan jumlah).
2. Pemeringkatan yang menghitung jumlah trigram yang sama antara dokumen dengan *query* serta memperhitungkan posisi kemunculan *term* dari *query* pada dokumen. Semakin terurut dan rapat posisi kemunculan *term*, semakin tinggi peringkatnya (pemeringkatan posisi).

Penggunaan kedua metode ini untuk mengetahui metode yang paling baik dengan memperhitungkan waktu komputasi serta kualitas hasil pencarian yang diketahui pada saat pengujian.

Relevance Judgment

Karena tidak ada *relevance judgment* yang baku untuk sistem pencarian ayat Al-Quran berbasis kemiripan fonetis, maka perlu dibuat *relevance judgment* sendiri. Daftar *query* dan daftar dokumen yang relevan dengan *query* dibuat sesuai tujuan pengujian.

Daftar *query* pengujian (dalam aksara Arab) dibuat berdasarkan beberapa aspek yang diuji dari sistem, yaitu mengandung setiap huruf dalam Al-Quran dan hukum-hukum *tajwid* tertentu. Daftar *query* ini kemudian diberikan kepada beberapa responden yang mampu membaca Al-Quran untuk dituliskan cara penulisan lafalnya menurut masing-masing responden. Cara penulisan lafal menurut responden inilah yang menjadi *query* yang sesungguhnya untuk pengujian.

Daftar *query* pengujian dibagi menjadi 2, yaitu:

- 1 Kelompok *query* A (pencarian lafal) yang digunakan untuk menguji sistem dalam mencari lafal tanpa memerhatikan makna dari *query* maupun hasil pencarian.
- 2 Kelompok *query* B (pencarian topik) yang digunakan untuk menguji sistem dalam mencari topik atau informasi yang memiliki makna tertentu sesuai makna dari *query*.

Pengukuran Kinerja

Terdapat 2 faktor yang diuji, yaitu metode pengindeksan (dengan vokal dan tanpa vokal) dan metode pemeringkatan (pemeringkatan jumlah dan pemeringkatan posisi). Dengan demikian, terdapat 4 unit percobaan (skema), yaitu:

- pengindeksan dengan vokal dan pemeringkatan jumlah (kode: VJ),
- pengindeksan dengan vokal dan pemeringkatan posisi (kode: VP),
- pengindeksan tanpa vokal dan pemeringkatan jumlah (kode: NJ), dan
- pengindeksan tanpa vokal dan pemeringkatan posisi (kode: NP).

Untuk setiap unit percobaan, dihitung kinerja sistem untuk setiap *query* pengujian. Kinerja yang diukur ialah waktu pencarian dan kualitas hasil pencarian. Untuk pengukuran kualitas hasil pencarian, sesuai Manning *et al.* (2008), digunakan 11 titik *recall* standar dan interpolasi maksimum *precision* untuk menghitung nilai *average precision*. Hasil pengujian dapat digambarkan dengan kurva *recall-precision* untuk membandingkan kinerja setiap unit percobaan.

Implementasi

Implementasi sistem, yaitu pengindeks dan pencari, dilakukan dengan bahasa pemrograman PHP. Antarmuka pencarian berupa aplikasi berbasis web. Spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam pengembangan sebagai berikut:

- prosesor Intel Core Duo 1.66 GHz,
- memori 1 GB,
- *hard disk* 120 GB,
- sistem operasi Ubuntu versi 11.10,
- *web server* Apache versi 2.2.20,
- bahasa pemrograman PHP versi 5.3.6, dan
- lingkungan pengembangan (IDE) NetBeans versi 7.1.

HASIL DAN PEMBAHASAN

Praproses Data

Pada permulaan beberapa surat dalam Al-Quran, terdapat rangkaian huruf yang disebut *fawatihush shuwar* (Ismail & Nawawi 1995). Rangkaian huruf ini memiliki cara pembacaan khusus, yaitu dengan membaca nama huruf alih-alih membunyikan huruf.

Rangkaian huruf ini harus diubah secara manual agar dapat diproses dengan benar pada langkah selanjutnya. Sebagai contoh, rangkaian huruf *الم* pada permulaan Surat Al-Baqarah diubah menjadi *أَلِفٌ لَامٌ مِيمٌ* (*alif laam miim*). Daftar lengkap aturan pengubahan terlampir pada Lampiran 1.

Pengodean Fonetis Teks Al-Quran

Sebelum dilakukan prosedur pengodean fonetis, data yang dikodekan harus diidentifikasi ciri-cirinya terlebih dahulu. Data teks Al-Quran yang diperoleh dari Tanzil Project memiliki beberapa ciri, yaitu:

- tidak ada tanda berhenti di tengah ayat,
- tidak ada tanda vokal yang menunjukkan bacaan panjang (*mad badal*),
- huruf konsonan yang digunakan untuk memperpanjang bacaan (ا, ي, dan و) tidak memiliki tanda vokal,
- penggunaan huruf *alif madd* (آ) untuk menunjukkan bacaan *a* panjang, dan
- huruf *lam* (ل) pada susunan *alif lam* bila seharusnya dibaca maka bertanda vokal *sukun*, sedangkan bila tidak dibaca maka tidak memiliki tanda vokal apapun.

Dengan memerhatikan ciri-ciri pada data tersebut, prosedur pengodean fonetis untuk teks Al-Quran dijelaskan pada langkah-langkah berikut. Prosedur ini dilakukan pada setiap ayat dan harus dilakukan secara berurutan.

1 Penghilangan spasi

Seluruh karakter spasi pada teks Al-Quran dihilangkan karena tidak berpengaruh terhadap cara pelafalan.

2 Penghilangan *syaddah*

Tanda vokal *syaddah* yang menunjukkan bunyi huruf konsonan ganda dihilangkan karena *syaddah* yang tidak dibaca termasuk kesalahan pelafalan yang umum terjadi (Junaedi 2009).

3 Penggabungan konsonan mati

Dua huruf konsonan yang sama yang berdampingan, jika huruf konsonan pertama

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

mati (bertanda vokal *sukun*), maka huruf konsonan yang mati tersebut dihilangkan. Langkah ini disebut juga aturan *tajwid idgham mutamatsilain* dan setara dengan penghilangan *syaddah*.

4 Penanganan akhir ayat

Huruf yang berada pada akhir ayat memiliki cara pembacaan khusus dengan beberapa aturan yang diterapkan berdasarkan Ismail dan Nawawi (1995), yaitu:

- mengganti tanda vokal *fathatain* dengan *fathah* jika diikuti huruf *alif* (ا),
- mengganti huruf *ta marbutah* (ة) dengan huruf *ha* (ه), dan
- mengganti tanda vokal huruf terakhir dengan *sukun* kecuali jika huruf terakhir *alif* (ا) atau *alif maqshurah* (ي).

5 Substitusi *tanwin*

Seluruh tanda vokal yang termasuk *tanwin* diganti dengan huruf *nun* mati (نْ) untuk penyeragaman.

6 Penghilangan bacaan panjang

Bacaan atau bunyi vokal panjang dihilangkan (dipendekkan) karena termasuk kesalahan pelafalan yang umum terjadi (Junaedi 2009). Pemendekan meliputi bacaan panjang dengan sebab adanya huruf konsonan yang memanjangkan maupun huruf *alif madd* (اْ).

7 Penghilangan huruf yang tidak dibaca

Huruf-huruf yang tidak memiliki tanda vokal tidak dilafalkan sehingga harus dihilangkan.

8 Substitusi bacaan *iqlab*

Aturan *tajwid iqlab* berlaku saat huruf *nun* mati (نْ) bertemu dengan huruf *ba* (ب), yaitu mengganti huruf *nun* mati dengan *mim* mati (مْ).

9 Substitusi bacaan *idgham*

Aturan *tajwid idgham* berlaku saat huruf *nun* mati (نْ) bertemu dengan salah satu dari huruf-huruf *idgham*, yaitu *ya*, *nun*, *mim*, *wau*, *lam*, dan *ra* (ي , ن , م , و , ل , dan ر). Aturannya ialah menghilangkan bunyi *nun* tersebut. Pengecualian ada pada beberapa kata tertentu.

10 Pemadanan ke kode fonetis

Setelah cara pembacaan diolah, setiap karakter (huruf serta tanda vokal) pada teks Al-Quran dipadankan dengan kode fonetis

dalam aksara Latin. Huruf-huruf konsonan yang pelafalannya sering keliru (berdasarkan Junaedi (2009)) dikelompokkan menjadi 1. Aturan pemadanan tercantum pada Tabel 1.

Prosedur pengodean tersebut belum mencakup seluruh aturan *tajwid*, terutama aturan tingkat lanjut. Beberapa aturan *tajwid* yang belum dapat ditangani antara lain *idgham mutajanisain* dan *idgham mutaqaribain* yang membutuhkan analisis fonetik yang lebih mendalam. Selain itu, ada juga aturan *nun washal* yang membutuhkan analisis linguistik serta aturan bacaan *imalah*, *isymam*, dan *saktah* yang hanya terdapat pada beberapa tempat saja di dalam Al-Quran (Ismail & Nawawi 1995).

Tabel 1 Aturan pemadanan aksara Arab ke kode fonetis

Aksara Arab	Padanan
ج ز ظ ذ	Z
ح خ ه	H
اء ائ ؤ و ع	X
ص س ش ث	S
د ض	D
ت ة ط	T
ق ك	K
غ	G
ف	F
م	M
ن	N
ل	L
ب	B
ي	Y
و	W
ر	R
ا	A
ي	I
و	U
ا	(dihapus)

Contoh penerapan prosedur pengodean fonetis dari teks Al-Quran ini tercantum pada Tabel 2 dengan menggunakan teks dari Surat Al-Baqarah (2) ayat 2. Dari contoh, dapat

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

dilihat bahwa prosedur pengodean ini adalah suatu transformasi dari teks Al-Quran menjadi teks Latin yang merepresentasikan cara pembacaannya, atau dengan kata lain prosedur untuk membuat transkripsi dari teks Al-Quran.

Prosedur pengodean ini ditujukan khusus untuk teks Al-Quran dengan aturan cara pembacaan tertentu. Akan tetapi, prosedur yang sama dapat digeneralisasi untuk teks Arab bertanda vokal secara umum dengan menghilangkan beberapa aturan khusus *tajwid*, antara lain Langkah 8 dan Langkah 9. Selain itu, untuk teks Arab secara umum, Langkah 4 perlu diubah agar dapat berlaku pada akhir kata (tidak hanya akhir kalimat atau ayat).

Tabel 2 Contoh pengodean fonetis dari teks Al-Quran

Langkah	Hasil
Teks asli	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
1	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
2	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
3	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
4	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
5	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
6	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
7	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
8	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
9	ذَٰلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ
10	ZALIKALKITABULARAY BAFIHIHUDALILMUTAKIN

Prosedur ini diimplementasikan dalam bahasa pemrograman PHP dengan menggunakan fungsi-fungsi ekspresi reguler (*regular expression*) dan pemrosesan *string*. Prosedur pengodean ini hanya perlu dilakukan 1 kali untuk menghasilkan korpus kode fonetis yang akan diproses pada tahap pengindeksan. Karena hanya perlu dilakukan 1 kali, implementasi tidak perlu mempertimbangkan kecepatan pemrosesan. Implementasi prosedur ini untuk memproses seluruh dokumen atau ayat (sebanyak 6236 ayat) berlangsung dalam waktu sekitar 18 detik.

Pengodean Fonetis Teks Latin

Query dari pengguna yang berupa teks Latin harus dikodekan dengan prosedur serupa sehingga dapat dicocokkan dengan hasil pengodean teks Al-Quran. Sebelum diproses,

teks dari *query* harus dinormalkan terlebih dahulu, antara lain dengan mengubah seluruh huruf menjadi huruf kapital, menghapus spasi ganda, dan menghapus seluruh karakter kecuali huruf dan beberapa tanda baca (tanda kutip tunggal, apostrof, dan tanda hubung).

Prosedur pengodean fonetis untuk teks Latin dijelaskan pada langkah-langkah berikut. Prosedur ini dilakukan pada *query* dari pengguna dan harus dilakukan secara berurutan.

1 Substitusi vokal

Pada aksara Arab, hanya terdapat 3 jenis vokal, yaitu *A*, *I*, dan *U*, sedangkan pada aksara Latin ada tambahan *E* dan *O*. Oleh karena itu, vokal harus disubstitusi. Huruf vokal *O* (misalnya pada “*furqon*”) diganti dengan huruf vokal *A*, sedangkan huruf vokal *E* (misalnya pada “*muttaqien*”) diganti dengan huruf vokal *I*.

2 Penggabungan konsonan

Huruf-huruf konsonan yang sama yang berdampingan digabungkan atau dijadikan satu. Langkah ini setara dengan penghilangan *syaddah*.

3 Penggabungan vokal

Huruf-huruf vokal yang sama yang berdampingan digabungkan atau dijadikan satu. Langkah ini setara dengan penghilangan bacaan panjang.

4 Substitusi diftong

Huruf vokal berdampingan bila membentuk diftong yang serupa dengan diftong pada bahasa Arab (*ay* dan *aw*), maka disesuaikan dengan penulisan diftong pada bahasa Arab. Aturannya ialah mengubah *AI* menjadi *AY* dan *AU* menjadi *AW*.

5 Penandaan huruf hamzah

Huruf Arab *hamzah* (ء) yang tidak dituliskan secara eksplisit dalam tulisan Latin (misalnya dengan tanda kutip tunggal) diduga dari huruf vokal berdampingan yang tidak memenuhi syarat untuk menjadi diftong. Huruf vokal ditandai sebagai *hamzah* bila memenuhi salah satu aturan berikut:

- huruf *A*, *I*, atau *U* yang berada pada awal kata atau setelah spasi,
- susunan vokal *IA* atau *IU* (huruf setelah *I* ditandai *hamzah*), atau
- susunan vokal *UA* atau *UI* (huruf setelah *U* ditandai *hamzah*).

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

6 Substitusi bacaan *ikhfa*

Bacaan *ikhfa* (bunyi *N* samar-samar bila diikuti huruf konsonan tertentu) terkadang dituliskan dengan bunyi *NG* (misalnya pada “angfusakum”). Pada kasus demikian, huruf *G* harus dihilangkan agar setara dengan pengodean dari teks Al-Quran.

7 Substitusi bacaan *iqlab*

Setara dengan Langkah 8 pada prosedur pengodean dari teks Al-Quran, bacaan *iqlab* diproses dengan mengubah susunan huruf konsonan *NB* menjadi *MB*.

8 Substitusi bacaan *idgham*

Setara dengan Langkah 9 pada prosedur pengodean dari teks Al-Quran, bacaan *idgham* diproses dengan menghilangkan huruf konsonan *N* jika bertemu dengan huruf-huruf *idgham* (*Y, N, M, W, L, R*).

9 Pemadanan ke kode fonetis

Pemadanan ke kode fonetis dari teks Latin perlu mempertimbangkan huruf Arab yang direpresentasikan dalam 2 huruf konsonan Latin, misalnya huruf *ẓ* yang dituliskan *DZ*. Oleh karena itu, pengodean dilakukan pada susunan 2 huruf konsonan terlebih dahulu. Aturan pemadanan tercantum pada Tabel 3. Aturan pemadanan ini disesuaikan dengan aturan pemadanan pada prosedur pengodean teks Al-Quran dengan pedoman alihaksara berdasarkan Ismail dan Nawawi (1995). Huruf yang sudah sesuai tidak perlu dipadankan lagi karena kode fonetis sudah berupa huruf Latin.

10 Penghilangan spasi

Seluruh spasi dihilangkan agar setara dengan hasil pada prosedur pengodean dari teks Al-Quran.

Prosedur pengodean tersebut masih memiliki kekurangan pada penandaan huruf *hamzah* (Langkah 5). Masih terdapat ambiguitas untuk menentukan 2 huruf vokal berdampingan termasuk diftong, bacaan panjang, atau merupakan huruf *hamzah*. Sebagai contoh bila pengguna menuliskan “ULAIKA” (أُولَيْكَ) maka huruf vokal *AI* akan terdeteksi sebagai diftong (menjadi “ULAYKA”), padahal seharusnya huruf vokal *I* sebagai huruf *hamzah* (menjadi “ULA’IKA”). Contoh lain misalnya bila pengguna menuliskan “SAALA” (سَالَا), maka huruf vokal *AA* akan terdeteksi sebagai bacaan panjang dan akan dipendekkan, padahal seharusnya huruf vokal *A* yang kedua sebagai huruf *hamzah* (menjadi “SA’ALA”).

Kekurangan lain ialah huruf *hamzah* mati yang terkadang dituliskan dengan huruf konsonan *K*, misalnya bila pengguna menuliskan “MUKMININ” (مُؤْمِنِينَ) yang seharusnya dituliskan “MU’MININ”. Pada kondisi-kondisi seperti ini huruf *hamzah* tidak dapat teridentifikasi.

Kondisi sebaliknya juga dapat terjadi, yaitu yang seharusnya bukan huruf *hamzah* namun teridentifikasi sebagai huruf *hamzah*. Contohnya bila pengguna menuliskan “QUL HUALLAH” (قُلْ هُوَ اللَّهُ) maka deretan huruf *HUA* akan dianggap mengandung huruf *hamzah* menjadi “HU’A”, padahal seharusnya adalah “HUWA”.

Selain masalah terkait huruf *hamzah*, terdapat masalah terkait ambiguitas saat pemadanan ke kode fonetis untuk 2 konsonan. Sebagai contoh, bila pengguna menuliskan “ASHABU” (أَصْحَابُ) maka pasangan konsonan *SH* akan dianggap satu kesatuan dan dikodekan ke huruf *S* (huruf *H* hilang). Hal ini dapat diperbaiki dengan menambahkan spasi untuk memisahkan kedua huruf menjadi “AS HABU”.

Dengan demikian, dapat dilihat bahwa terdapat banyak ambiguitas dan kemungkinan kesalahan dalam pengodean dari teks Latin. Namun demikian, metode pencarian trigram diharapkan dapat bersifat toleran terhadap kesalahan-kesalahan pengodean *query* yang telah disebutkan.

Tabel 3 Aturan pemadanan aksara Latin ke kode fonetis

Aksara Latin	Padanan
2 konsonan	
SH, TS, SY	S
KH, CH	H
ZH, DZ	Z
DH	D
TH	T
GH	G
NG (‘ain)	X
1 konsonan	
F, V, P	F
Q, K	K
J, Z	Z
‘, ` (apostrof)	X

Contoh penerapan prosedur pengodean fonetis dari teks Latin ini tercantum pada Tabel 4. Sebagai contoh, pengguna menuliskan *query*

“hudan lil muttaqien” (potongan lafal dari Surat Al-Baqarah (2) ayat 2). Dapat dilihat bahwa prosedur pengodean fonetis dari teks Latin menghasilkan kode fonetis yang cocok dengan prosedur pengodean fonetis dari teks Al-Quran.

Tabel 4 Contoh pengodean fonetis dari teks Latin

Langkah	Hasil
Teks asli	hudan lil muttaqien
1	HUDAN LIL MUTTAQIIN
2	HUDAN LIL MUTAQIIN
3	HUDAN LIL MUTAQIN
4	HUDAN LIL MUTAQIN
5	HUDAN LIL MUTAQIN
6	HUDAN LIL MUTAQIN
7	HUDAN LIL MUTAQIN
8	HUDALIL MUTAQIN
9	HUDALIL MUTAKIN
10	HUDALILMUTAKIN

Prosedur ini diimplementasikan dalam bahasa pemrograman PHP untuk memproses *query* dari pengguna dan dijalankan setiap kali pengguna memasukkan *query*. Serupa dengan prosedur pengodean dari teks Al-Quran, implementasi prosedur pengodean ini juga menggunakan fungsi-fungsi ekspresi reguler dan pemrosesan *string*.

Tokenisasi Trigram

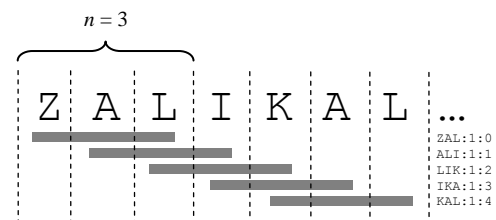
Proses tokenisasi dilakukan pada teks dari dokumen sebelum diindeks. Proses ini adalah proses pemotongan dokumen menjadi *token* yang kemudian dimasukkan sebagai *term* pada indeks (Manning *et al.* 2008). Pada proses tokenisasi trigram (atau *n*-gram secara umum), tidak diperlukan segmentasi kalimat pada teks menjadi sejumlah kata, tetapi hanya perlu memotong teks dengan *overlapping window* sepanjang *n*.

Secara formal, proses tokenisasi trigram didefinisikan sebagai berikut (Kim *et al.* 2005). Misalkan *d* sekuens dengan panjang *N* dari karakter c_1, c_2, \dots, c_N , maka trigram ke-*i* dari *d* adalah sekuens karakter c_i, c_{i+1}, c_{i+2} . Proses tokenisasi berlangsung sampai $i = N - 2$. Dengan demikian, untuk *string* sepanjang *N* dapat diambil sejumlah $N - 2$ trigram nonunik.

Sebagai contoh, untuk hasil pengodean fonetis dari contoh pada Tabel 2, trigram-trigram yang dapat diambil yaitu {ZAL, ALI, LIK, IKA, KAL, ALK, LKI, KIT, ITA, TAB,

ABU, BUL, ULA, LAR, ARA, RAY, AYB, YBA, BAF, AFI, FIH, IHI, HIH, IHU, HUD, UDA, DAL, ALI, LIL, ILM, LMU, MUT, UTA, TAK, AKI, KIN} sebanyak 36 trigram. Ilustrasinya tercantum pada Gambar 3 dengan menyertakan perhitungan jumlah serta posisi kemunculan trigram.

Prosedur tokenisasi trigram ini digunakan untuk memotong teks dokumen (dalam hal ini kode fonetis dari teks Al-Quran) menjadi *term* yang akan diindeks. Pada proses pengindeksan, setelah dilakukan pemotongan teks menjadi trigram-trigramnya perlu dihitung jumlah kemunculan trigram unik serta posisi kemunculan pertama trigram tersebut pada dokumen.



Gambar 3 Ilustrasi tokenisasi trigram.

Pengindeksan Trigram

Proses pengindeksan dilakukan dengan 2 metode, yaitu dengan vokal dan tanpa vokal. Dengan demikian, terdapat 2 indeks sebagai hasil proses pengindeksan.

Pada metode pengindeksan dengan vokal, korpus kode fonetis dapat langsung ditokenisasi. Sementara itu, pada pengindeksan tanpa vokal, seluruh huruf vokal (A, I, U) yang terdapat pada korpus dihilangkan terlebih dahulu sebelum ditokenisasi. Seperti halnya algoritme-algoritme pengodean fonetis yang telah disebutkan (Soundex, ASoundex, NORM, dan Soutex), pada metode pengindeksan ini seluruh vokal diabaikan dengan tujuan agar lebih toleran terhadap kesalahan pemberian vokal.

Inverted index yang dibangun terdiri atas 2 bagian, yaitu *vocabulary* yang menyimpan seluruh *term* yang terindeks serta daftar *posting* yang menyimpan daftar dokumen serta informasi lain yang mengandung *term* yang bersesuaian. Proses konstruksi indeks dilakukan di dalam memori dengan mengimplementasikan *single-pass in-memory indexing* (SPIMI) sesuai Manning *et al.* (2008). Pengindeksan dengan metode SPIMI dilakukan di dalam memori utama dan hanya perlu membaca korpus dan menulis indeks ke media penyimpanan (*hard*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

disk) sebanyak 1 kali, sehingga akan mengurangi *overhead* baca dan tulis dari dan ke media penyimpanan. Algoritme pengindeksan dilakukan dalam 2 tahap menggunakan struktur data *hash table* atau pasangan *key-value*. Proses pembangunan indeks diilustrasikan pada Gambar 4.

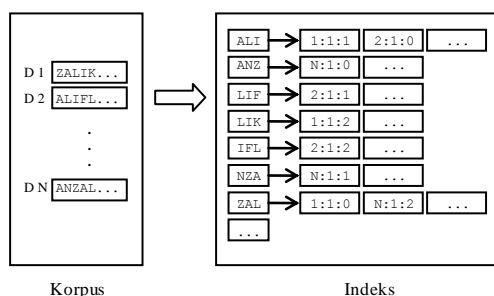
1 Tahap pembangunan indeks

- 1 Inisialisasi *hash table H*
- 2 Untuk setiap teks dokumen *d* pada korpus:
 - 1 Lakukan tokenisasi trigram
 - 2 Untuk setiap trigram *t* dalam *d*:
 - 1 Tambahkan *t* sebagai *key* dari *H*
 - 2 Tambahkan identifier *d* beserta jumlah dan posisi kemunculan *t* dalam *d* sebagai *value* dari *H* untuk *key t*
- 3 Urutkan *key* dari *H*

2 Tahap penulisan indeks

- Untuk setiap entri pada *H*:
 - 1 Tulis *key* sebagai *vocabulary*
 - 2 Tulis seluruh *value* sebagai *posting list*
- 2 Hapus *H* dari memori.

Hasil dari proses pengindeksan ini, yaitu *vocabulary* atau daftar *term* serta *posting list* disimpan dalam media penyimpanan berupa *file* teks biasa tanpa kompresi. Baeza-Yates dan Ribeiro-Neto (1999) menyebutkan bahwa ukuran *vocabulary* biasanya tidak terlalu besar, sedangkan ukuran *posting list* jauh lebih besar. Oleh karena itu, untuk menghindari pembacaan *file* indeks yang terlalu besar, perlu dipisahkan antara *file* untuk menyimpan *vocabulary* dan *file* untuk menyimpan *posting list*. Pada proses pembacaan indeks nantinya, *file* untuk *vocabulary* dibaca seluruhnya ke memori, sedangkan *file* untuk *posting list* hanya dibaca pada bagian yang diperlukan saja (Baeza-Yates & Ribeiro-Neto 1999; Manning *et al.* 2008).



Gambar 4 Ilustrasi pembangunan indeks.

Implementasi algoritme pengindeksan dilakukan dengan bahasa pemrograman PHP untuk 2 metode pengindeksan (dengan vokal dan tanpa vokal). Perbandingan hasil kedua metode pengindeksan tersebut tercantum pada Tabel 5 dengan jumlah dokumen pada korpus sebanyak 6236 dokumen.

Tabel 5 Perbandingan hasil pengindeksan dengan vokal dan tanpa vokal

Parameter	Dengan Vokal	Tanpa Vokal
Ukuran korpus (dalam KiB*)	498	291
Jumlah <i>term</i> pada <i>vocabulary</i>	2086	3747
Ukuran <i>vocabulary</i> (dalam KiB)	24	42
Ukuran <i>posting list</i> (dalam KiB)	3645	2183
Rata-rata panjang <i>posting list</i>	184	63
Waktu pengindeksan (dalam detik)	8.1	4.8
Penggunaan memori (dalam KiB)	197 672	121 831

Contoh isi *file* indeks (dengan vokal) terlampir pada Lampiran 2. Pada *file vocabulary*, selain disimpan *term* disimpan juga *offset* dalam byte yang menyatakan posisi awal pembacaan *file posting list* untuk *term* yang bersesuaian.

Dari hasil pengindeksan dapat dilihat bahwa secara umum ukuran indeks (ukuran *posting list* dan *vocabulary*) pada pengindeksan dengan vokal sekitar 1.6 kali lebih besar daripada pengindeksan tanpa vokal. Hal yang sama terjadi pada waktu yang dibutuhkan untuk mengindeks dan jumlah memori yang digunakan. Perbedaan ini terjadi akibat ukuran korpus yang harus diindeks lebih besar untuk pengindeksan dengan vokal. Dengan demikian, parameter waktu pengindeksan, memori yang digunakan, serta ukuran indeks berbanding lurus dengan ukuran korpus yang diindeks.

Namun, berbeda halnya dengan jumlah *term* pada *vocabulary*. Pada pengindeksan dengan vokal, jumlah *term* yang ada pada *vocabulary* lebih sedikit (sebanyak 2086 *term*) dibandingkan dengan pengindeksan tanpa vokal

* KiB = kibibita; 1 kibibita = 1024 bita (*byte*)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

(sebanyak 3747 *term*). Hal ini dapat dipahami bahwa untuk ukuran korpus yang besar, jumlah trigram yang unik semakin sedikit, atau dengan kata lain banyak trigram yang berulang. Pada pengindeksan korpus yang menggunakan vokal, peluang suatu trigram berulang semakin besar. Konsekuensi lain dari hal ini adalah rata-rata panjang *posting list* pada pengindeksan dengan vokal sekitar 3 kali lebih banyak dibandingkan dengan pengindeksan tanpa vokal.

Pemrosesan Query

Pada pemrosesan *query* atau proses pencarian, dihitung jumlah trigram yang sama antara *query* dan seluruh dokumen. Pembobotan untuk tiap *term* (trigram) pada indeks menggunakan pembobotan yang paling sederhana, yaitu pembobotan TF (*term frequency*).

Perhitungan jumlah trigram yang sama dilakukan secara sekuensial untuk tiap dokumen jika dokumen tidak terindeks. Namun, karena dokumen yang dicari telah terindeks, informasi yang tersimpan dalam indeks dapat dimanfaatkan untuk menghitung jumlah trigram yang sama dengan lebih cepat.

Algoritme untuk pencarian dalam *inverted index* secara umum terbagi menjadi 3 langkah, yaitu pencarian dalam *vocabulary*, pengambilan *posting list* yang sesuai, serta manipulasi *posting list* tersebut (Baeza-Yates & Ribeiro-Neto 1999). Proses manipulasi *posting list* yang dilakukan adalah penggabungan (*merge*) dari seluruh *posting list* yang cocok dengan trigram-trigram dari *query* untuk dihitung jumlah trigram yang sama. Algoritme pencarian dilakukan dalam langkah-langkah berikut yang menggunakan struktur data *hash table*.

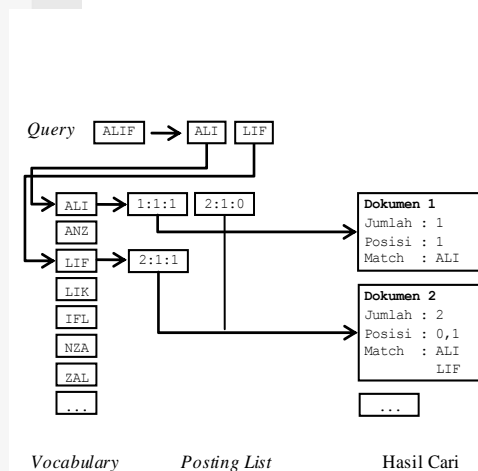
- 1 Inisialisasi *hash table V* (untuk *vocabulary*)
- 2 Inisialisasi *hash table D* (untuk dokumen yang ditemukan)
- 3 Baca seluruh *file vocabulary*, dan masukkan *term* sebagai *key* dari *V* dan *offset* sebagai *value* dari *V*
- 4 Lakukan tokenisasi trigram pada *query*
- 5 Untuk setiap trigram *t* dari *query*:
 - 1 Ambil *offset f* sebagai *value* dari *V* untuk *key t*
 - 2 Baca *file posting list* mulai dari *byte ke-f* sampai akhir baris
 - 3 Untuk setiap *posting p* dari *posting list*:
 - 1 Masukkan *id* dari *p* (*p.id*) sebagai *key* dari *D*
 - 2 Tambahkan posisi kemunculan dari *p* sebagai *value* posisi dari *D* untuk *key p.id*
 - 3 Tambahkan jumlah kemunculan dari *p* sebagai *value* frekuensi dari *D* untuk *key p.id*
- 6 Berikan *D* sebagai hasil pencarian (dokumen yang ditemukan).

Ilustrasi tercantum pada Gambar 5 dengan *query* yang dikodekan “ALIF”. Pada pemrosesan *query* ini, didapatkan himpunan dokumen yang mengandung paling sedikit 1 trigram dari *query*. Pada tahap ini belum dilakukan pemeringkatan maupun pengurutan hasil pencarian. Masing-masing dokumen hasil pencarian memiliki beberapa informasi, yaitu:

- *identifier* dari dokumen (suatu angka),
- jumlah trigram yang sama dengan *query*, dan
- posisi kemunculan trigram yang cocok.

Pemeringkatan Dokumen

Untuk memeringkat dokumen, perlu dilakukan pemberian skor terhadap seluruh dokumen yang didapat pada proses pencarian. Pada pengukuran kesamaan dengan trigram (atau *n-gram* secara umum), cara yang paling sederhana ialah dengan memberi skor sebanyak jumlah trigram yang sama antara *query* dan masing-masing dokumen (Zobel & Dart 1996). Cara inilah yang digunakan pada jenis pemeringkatan pertama, yaitu pemeringkatan jumlah. Skor masing-masing dokumen adalah jumlah trigram yang sama antara *query* dengan dokumen. Dengan demikian, skor maksimum ialah sebanyak jumlah trigram pada *query*, sedangkan skor minimum ialah 1 (karena proses pencarian hanya mengambil dokumen yang memiliki paling sedikit 1 trigram yang sama dengan *query*).



Gambar 5 Ilustrasi pemrosesan *query*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Namun, pemeringkatan seperti ini memiliki kekurangan. Untuk teks dokumen yang panjang, trigram-trigram yang cocok bisa saja ditemukan dalam posisi yang tidak berurutan atau terpisah cukup jauh. Sebagai contoh, pada suatu pencarian dengan *query* tertentu, posisi kemunculan trigram pada sebuah dokumen dengan panjang 409 karakter ialah {31, 32, 212, 16, 214, 34, 223, 2, 169, 8, 307}. Dokumen yang ditemukan ini memiliki jumlah trigram yang sama dengan *query*, yaitu 11 trigram. Dengan kata lain, dokumen ini cocok secara sempurna dengan *query*, namun posisi kemunculan trigramnya tidak berurutan dan terpisah cukup jauh. Dokumen seperti ini dapat dikatakan tidak relevan dengan *query* yang diberikan.

Untuk mengatasi masalah demikian, perlu dibuat mekanisme pemberian skor sedemikian sehingga dokumen yang posisi kemunculan trigramnya teratur dan rapat memiliki skor yang lebih tinggi daripada yang posisi kemunculan trigramnya teracak dan terpisah jauh.

Untuk memberi skor keterurutan, digunakan *longest increasing subsequence* (LIS) yang diterapkan pada posisi kemunculan trigram. LIS dari sebuah sekuens S adalah subsekuens monoton naik dari S dengan panjang maksimum (Liben-Nowell & Vee 2006). Sekuens posisi kemunculan trigram yang teratur sempurna akan bernilai maksimum, yaitu panjang LIS sama dengan panjang sekuens. Algoritme yang efisien untuk mencari LIS terlampir pada Lampiran 3.

Untuk memberi skor kerapatan, dihitung rata-rata dari invers dari selisih antareleman berdampingan pada LIS. Misalkan suatu LIS sepanjang n adalah $\{s_1, s_2, \dots, s_n\}$, maka skor kerapatannya (c) adalah:

$$c = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{1}{s_{i+1} - s_i}$$

Penggunaan fungsi invers atau *reciprocal* bertujuan agar selisih yang besar tidak berdampak besar terhadap skor keseluruhan (memperhalus selisih). Dengan demikian, untuk suatu LIS yang rapat (selisih antareleman berdampingan sebesar 1) skor kerapatannya sebesar 1, yaitu skor maksimum.

Skor akhir dihitung dengan mengalikan skor panjang LIS dengan skor kerapatan dari LIS. Skor inilah yang diberikan kepada setiap dokumen hasil pencarian untuk metode pemeringkatan posisi. Sebagai contoh, dokumen yang memiliki posisi kemunculan trigram cocok

{31, 32, 212, 16, 214, 34, 223, 2, 169, 8, 307}, LIS-nya ialah {31, 32, 212, 214, 223, 307} sepanjang 6, sedangkan skor kerapatan dari LIS ialah 0.33. Dengan demikian dokumen tersebut diberi skor akhir $6 \times 0.33 = 1.96$.

Metode pemberian skor berdasarkan posisi ini masih memiliki kelemahan terutama untuk dokumen-dokumen panjang dan trigram-trigram nonunik. Karena posisi kemunculan trigram yang ada pada indeks ialah posisi kemunculan pertama dari suatu trigram unik, posisi kemunculan dapat dihitung dengan tidak tepat. Sebagai contoh, trigram "ALI" muncul pada suatu dokumen panjang pada posisi 3 dan 350, dan suatu trigram dari *query* cocok dengan dokumen ini pada posisi 350. Karena yang dihitung adalah posisi kemunculan pertama, maka yang disimpan dalam indeks adalah posisi 3 alih-alih 350. Hal ini dapat membuat skor keterurutan dan skor kerapatan menjadi rendah walaupun sebenarnya dokumen tersebut relevan dengan *query*.

Selanjutnya, untuk kedua metode pemeringkatan (pemeringkatan jumlah dan posisi), skor tambahan diberikan pada dokumen yang posisi kemunculan trigram terakhirnya jatuh pada karakter spasi. Motivasinya bahwa pencocokan trigram tidak mengenal batas kata, sedangkan pengguna biasanya memberi *query* yang bagian akhirnya merupakan akhir dari suatu kata. Oleh karena itu, dokumen yang posisi trigram terakhirnya jatuh pada karakter spasi, yaitu yang tepat pada akhir suatu kata dapat dianggap lebih relevan dan diberi skor tambahan dengan suatu bilangan kecil sebesar 0.001.

Setelah seluruh dokumen diberi skor, baik skor jumlah maupun skor posisi, seluruh dokumen hasil pencarian diurutkan berdasarkan skor terbesar. Jika beberapa dokumen memiliki skor yang sama, dokumen-dokumen tersebut diurutkan berdasarkan nomor surat dan ayat.

Pembuatan Antarmuka Pengguna

Antarmuka pengguna termasuk hal yang penting dalam sebuah sistem pencarian. Sistem harus dapat membantu mengarahkan pengguna dalam menggunakan sistem dan mendapatkan hasil pencarian yang dibutuhkan pengguna (Baeza-Yates & Ribeiro-Neto 1999).

Antarmuka halaman awal sistem dibuat sederhana dengan 1 kotak *input* teks untuk memasukkan *query* dan 1 tombol untuk memulai pencarian seperti terdapat pada Gambar 6. Terdapat tombol untuk melihat bantuan bagi pengguna dalam memasukkan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

query. Terdapat juga tombol untuk pengaturan pencarian, yaitu memperhitungkan vokal atau tanpa vokal, serta menggunakan pemeringkatan jumlah atau posisi.

Pada halaman hasil pencarian, ditampilkan ayat Al-Quran yang dianggap cocok dengan *query* (terurut skor). Hasil pencarian ditampilkan dalam beberapa halaman dengan tampilan hasil sebanyak 10 hasil tiap halaman. Selain teks ayat, ditampilkan juga nama surat serta nomor ayat. Data yang ditampilkan diambil dari *file* teks yang sama dengan yang digunakan untuk membuat korpus kode fonetis dengan mengacu pada *identifier* dokumen yang diberikan oleh algoritme pemrosesan *query*. Pengecualian untuk ayat-ayat yang mengandung *fawatihush shuwar*, kata-kata yang pada praproses data telah diubah dikembalikan seperti semula sesuai Lampiran 1.

Pada halaman hasil pencarian, diperlukan suatu *highlighting* (sorotan) untuk menandai bagian dari teks yang cocok dengan *query*. Sorotan ini berguna agar pengguna dapat fokus ke bagian dari teks yang cocok (Baeza-Yates &

Ribeiro-Neto 1999). Sorotan diberikan pada rentang yang sesuai dengan memanfaatkan informasi posisi kemunculan trigram dari *query* pada dokumen.

Contoh tampilan hasil pencarian tercantum pada Gambar 7 dengan *query* “alhamdulillah” dan menggunakan pilihan pencarian dengan menggunakan vokal dan pemeringkatan posisi. Bagian dari ayat yang cocok dengan *query* diberi sorotan dengan warna tertentu. Selain itu, ditampilkan juga angka persentase yang menyatakan persentase skor tiap ayat terhadap skor maksimum. Skor maksimum untuk setiap metode pemeringkatan ialah sebanyak jumlah trigram pada *query*.

Pembentukan Koleksi Pengujian

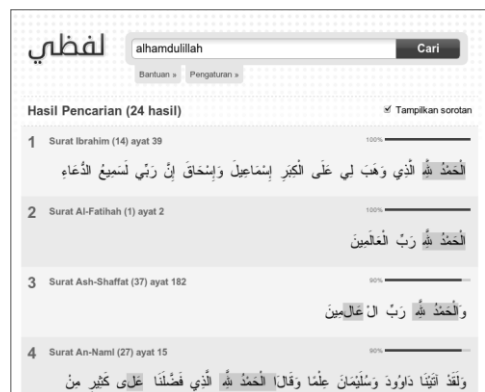
Untuk kelompok *query* A, dibuat sebanyak 16 *query*. Jumlah ini mencakup seluruh huruf Arab yang ada serta fitur-fitur pada proses pengodean fonetis termasuk hukum bacaan *tajwid* (*ikhfa*, *idgham*, dan *iqlab*). Setelah ditentukan 16 *query* tersebut, dicari daftar dokumen atau ayat yang relevan. Pencarian daftar ayat yang relevan menggunakan fasilitas pencarian tidak tepat sama pada aplikasi Zekr dengan *query* dalam aksara Arab. Daftar *query* beserta kodenya untuk kelompok *query* A terlampir pada Lampiran 4.

Untuk kelompok *query* B, dibuat sebanyak 5 *query*. Jumlah ini lebih sedikit daripada jumlah *query* untuk pengujian pencarian lafal karena lebih sulit menentukan ayat-ayat yang relevan dengan *query*. Makna dari *query* serta makna dari ayat yang ditemukan perlu diketahui. Pencarian daftar ayat yang relevan juga menggunakan aplikasi Zekr, namun *query* yang dimasukkan ke Zekr adalah akar kata dari suatu kata dalam bahasa Arab. Sebagai contoh, untuk *query* مُتْلِحُونَ (arti: orang-orang yang beruntung) akar katanya ialah تَلَعَ . Daftar *query* beserta kodenya untuk kelompok *query* B terlampir pada Lampiran 5.

Seluruh *query* dalam aksara Arab ini diberikan kepada responden dalam bentuk kuesioner untuk dituliskan cara penulisan lafalnya menurut masing-masing responden. Tujuan pembuatan kuesioner ini ialah mencari sebanyak mungkin variasi penulisan lafal Al-Quran oleh pembicara bahasa Indonesia. Kuesioner yang dibuat adalah kuesioner *online* berbasis web yang disebar di *mailing list* serta situs jejaring sosial dan dibuka selama 1 pekan (7 hari). Dari hasil kuesioner ini didapatkan daftar *query* pengujian yang menjadi *input* dalam pengujian sistem.



Gambar 6 Antarmuka awal sistem.



Gambar 7 Antarmuka hasil pencarian.

Pengujian Sistem

Dari hasil kuesioner, didapatkan sebanyak 499 variasi penulisan *query* untuk kelompok *query* A dan 52 variasi penulisan *query* untuk kelompok *query* B yang terlampir pada Lampiran 6. Sebagai contoh, untuk *query* dengan kode A15, beberapa variasi penulisan lafal oleh responden antara lain:

- tanziiil,
- tanzil,
- tangziil,
- tanziyl,
- tanjiil,
- tanzhil, dan
- tandziil.

Hasil perhitungan *average precision* (AVP) masing-masing *query* untuk setiap unit percobaan tercantum pada Tabel 6.

Tabel 6 Nilai AVP per *query*

Kode <i>Query</i>	NJ	VJ	NP	VP
A1	0.629	0.780	0.521	0.680
A2	0.995	0.895	0.900	0.868
A3	0.892	0.917	0.920	0.978
A4	0.865	0.510	0.841	0.496
A5	0.949	1.000	0.916	0.832
A6	0.723	0.876	0.444	0.714
A7	1.000	1.000	1.000	1.000
A8	0.976	0.991	0.895	0.897
A9	0.831	0.953	0.834	0.946
A10	0.917	0.996	0.829	0.717
A11	0.572	0.621	0.693	0.784
A12	0.787	0.627	0.710	0.754
A13	0.197	0.188	0.197	0.374
A14	0.924	0.984	0.759	0.754
A15	0.485	0.910	0.425	0.965
A16	0.452	0.419	0.285	0.292
B1	0.490	0.503	0.476	0.437
B2	0.599	0.375	0.555	0.504
B3	0.525	0.678	0.512	0.591
B4	0.378	0.408	0.374	0.384
B5	0.824	0.815	0.786	0.600

Dari hasil perhitungan AVP, terlihat bahwa masing-masing *query* memberikan kinerja yang berbeda untuk tiap unit percobaan. Hal ini disebabkan oleh karakteristik setiap *query* yang juga berbeda. Berdasarkan panjang, ada *query* yang pendek (misalnya A13), ada pula yang

panjang (misalnya A7). Berdasarkan letak kemunculan, ada *query* yang biasanya terletak pada awal dokumen atau ayat (misalnya A3), ada pula yang biasanya terletak pada akhir ayat (misalnya A10). Berdasarkan keunikan, ada *query* yang mengandung kata yang sering muncul beberapa kali dalam satu ayat (misalnya A4 yang mengandung kata “ALLAH”), ada pula yang mengandung kata yang biasanya hanya muncul satu kali dalam ayat (misalnya A15). Berdasarkan kemungkinan ambiguitas, ada *query* yang kemungkinan ambiguitas penulisannya kecil (misalnya A2), ada pula yang kemungkinan ambiguitas penulisannya besar (misalnya A16 yang memungkinkan ambiguitas penandaan *hamzah*).

Grafik *precision* terhadap *recall* untuk analisis per *query* terlampir pada Lampiran 7. Secara umum, sistem memberikan hasil yang baik, ditandai dengan nilai AVP yang tinggi pada masing-masing unit percobaan.

Pengaruh Panjang *Query*

Pada *query* tertentu, sistem menghasilkan kinerja sempurna yang ditandai dengan nilai AVP sebesar 1. *Query* yang menghasilkan kinerja sempurna untuk seluruh skema ialah A7. Hal ini karena A7 merupakan *query* yang panjang (paling panjang dari seluruh *query* untuk pengujian), sehingga kesalahan penulisan dapat ditangani dengan baik oleh metode pencarian trigram. Selain itu, A7 merupakan satu ayat penuh, tidak seperti *query* lainnya yang merupakan potongan ayat. Kinerja yang baik juga dihasilkan oleh *query* panjang lain seperti A5, A8, dan A10.

Berbeda halnya dengan *query* yang paling pendek, yaitu A13 yang menghasilkan nilai-nilai AVP rendah. Hal ini disebabkan kesalahan penulisan akan menurunkan skor kecocokan secara signifikan. Sebagai contoh, A13 akan dikodekan fonetis tanpa vokal menjadi “DLN” yang hanya memiliki 1 trigram. Kesalahan 1 huruf saja pada kode fonetis ini (misalnya menjadi “ZLN”) akan menghasilkan hasil pencarian yang tidak tepat.

Pengaruh Letak Kemunculan *Query*

Query yang biasanya terletak pada awal ayat, yaitu A3 dan A11 menghasilkan kinerja yang lebih baik untuk metode pemeringkatan dengan posisi (NP dan VP). Sementara itu, untuk *query* yang biasanya terletak pada akhir ayat, yaitu A8 dan A10, metode pemeringkatan dengan jumlah (NJ dan VJ) memberikan hasil yang lebih baik.

Pengaruh Keunikan Trigram pada Query

Query yang mengandung trigram yang sering muncul, misalnya A2, A4, A10, dan B1 yang mengandung kata “ALLAH”, memberikan hasil yang baik untuk pemeringkatan dengan jumlah (NJ dan VJ), namun buruk untuk pemeringkatan dengan posisi (NP dan VP). Alasannya serupa dengan penyebab rendahnya kinerja pemeringkatan posisi untuk query yang terletak pada akhir ayat, yaitu adanya trigram nonunik yang tercatat posisi kemunculannya secara tidak tepat.

Contoh lain ialah A14. Pada A14, meskipun query panjang, trigram dari query ada yang tidak unik pada ayat yang ditemukan, sehingga pada pemeringkatan posisi (NP dan VP) nilai *precision*-nya turun secara signifikan di sekitar titik *recall* 0.7.

Pengaruh Kemungkinan Ambiguitas Query

Query yang kemungkinan ambiguitasnya besar, seperti A16 yang memungkinkan ambiguitas penandaan *hamzah*, secara umum memiliki kinerja yang rendah dibandingkan dengan query yang kemungkinan ambiguitasnya kecil seperti A2. Hal ini terkait juga dengan panjang query, karena kesalahan pada query yang panjang dapat lebih ditoleransi oleh pencarian trigram dibandingkan dengan kesalahan pada query yang pendek.

Perbandingan Pencarian V dan N

Nilai AVP untuk pencarian dengan vokal (V) sebesar 0.651, sedangkan untuk pencarian tanpa vokal (N) sebesar 0.641. Dengan demikian, dapat dilihat bahwa pencarian dengan memperhitungkan vokal memiliki kinerja yang lebih baik daripada pencarian tanpa memperhitungkan vokal meskipun tidak signifikan. Dari hasil ini dapat diketahui bahwa pengguna dapat membedakan huruf vokal dengan baik untuk pencarian dalam Al-Quran. Hal ini juga sesuai dengan Hammo *et al.* (2007) bahwa pencarian lafal dalam Al-Quran tanpa memperhitungkan vokal hasilnya tidak begitu baik.

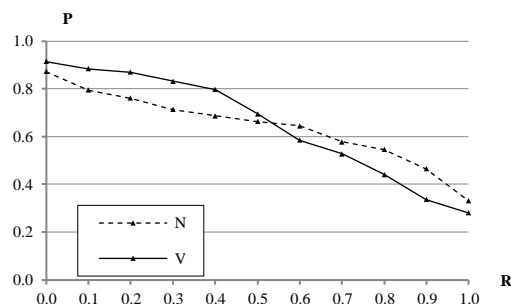
Dengan melihat grafik *precision-recall* seperti tercantum pada Gambar 8, terlihat bahwa pencarian dengan vokal memiliki kinerja lebih baik pada titik *recall* < 0.5, sedangkan pada titik *recall* > 0.5 pencarian tanpa vokal yang memiliki kinerja lebih baik. Hal ini terjadi karena pencarian dengan vokal tidak dapat menemukan ayat relevan yang memiliki huruf vokal yang sedikit berbeda dengan yang ada pada query. Sebagai contoh, ayat yang

mengandung “ULIL ALBAB” (أُولِي الْأَبَابِ) dapat ditemukan pada posisi teratas dengan query “ULUL ALBAB” (query A1) jika dihilangkan huruf vokalnya, sedangkan jika tetap dipertahankan huruf vokalnya maka akan mengurangi skor jumlah trigram yang cocok dan tidak mendapat tempat teratas pada hasil pencarian.

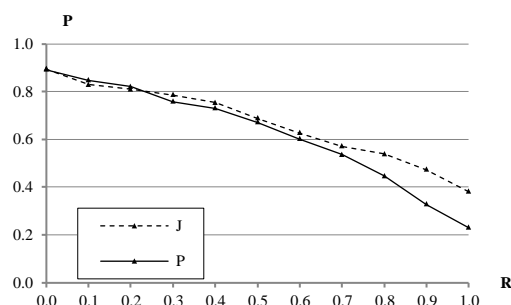
Perbandingan Pemeringkatan J dan P

Pemeringkatan jumlah (J) memiliki nilai AVP sebesar 0.668, sedangkan pemeringkatan posisi (P) sebesar 0.624. Dengan demikian, pemeringkatan hasil cari dengan peringkat jumlah trigram memiliki kinerja yang lebih baik dibandingkan dengan pemeringkatan dengan peringkat posisi kemunculan trigram. Hasil ini tidak sesuai dengan perkiraan bahwa pemeringkatan P dapat memperbaiki kekurangan pada pemeringkatan J.

Bila dilihat grafik *precision-recall*-nya pada Gambar 9, pemeringkatan P lebih baik pada titik *recall* rendah, kemudian jatuh pada titik *recall* selanjutnya. Hal ini telah dijelaskan sebelumnya bahwa penyebabnya terkait pengaruh letak kemunculan query serta



Gambar 8 Grafik perbandingan pencarian dengan vokal dan tanpa vokal.



Gambar 9 Grafik perbandingan pemeringkatan jumlah dan posisi.

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
 2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

keunikan trigram yang ada pada *query*. Dengan demikian, pemeringkatan P dapat meningkatkan *precision* hanya pada titik *recall* rendah.

Contoh perbaikan *precision* yang dapat dilakukan oleh pemeringkatan P ialah pada *query* A3 dan A11 karena kedua *query* tersebut biasanya berada pada bagian awal ayat, serta pada *query* A12, A13, A15, dan B5 karena dapat menemukan ayat relevan yang kemunculan trigramnya dari *query*-nya unik.

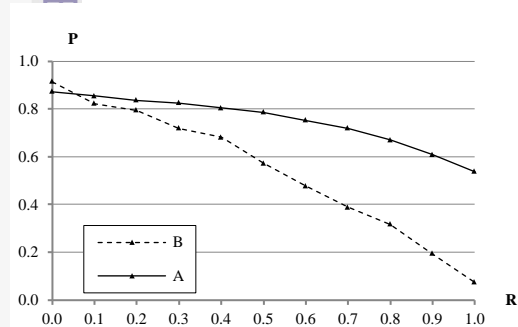
Perbandingan Query A dan B

Nilai AVP untuk pengujian dengan kelompok *query* A sebesar 0.751, sedangkan kelompok *query* B sebesar 0.541. Terlihat jelas bahwa pengujian untuk *query* A lebih baik daripada pengujian untuk *query* B dengan selisih yang signifikan.

Dari grafik *precision-recall* pada Gambar 10, terlihat perbedaan yang jelas antara *query* A dan *query* B. *Query* A memiliki nilai *precision* yang lebih stabil dibandingkan dengan *query* B. Hal ini menunjukkan bahwa sistem yang dibangun lebih sesuai untuk pencarian lafal tanpa memerhatikan makna alih-alih pencarian suatu informasi tertentu yang memiliki makna. Selain itu, kriteria dokumen relevan untuk *query* B sebagian mengandung kata-kata yang tidak mungkin ditemukan oleh *query* yang dimasukkan, misalnya "TAQWA" yang tidak mungkin ditemukan oleh *query* "MUTTAQIN" (B4) dengan metode trigram. Pada kasus demikian, *precision* bernilai 0, atau dengan kata lain tidak semua dokumen relevan dapat ditemukan.

Perbandingan Keempat Unit Percobaan

Pada pengujian kelompok *query* A, nilai AVP paling besar dihasilkan oleh VJ sebesar 0.792, selanjutnya NJ sebesar 0.762, selanjutnya VP sebesar 0.753, dan terakhir NP sebesar 0.698. Pada pengujian kelompok *query*

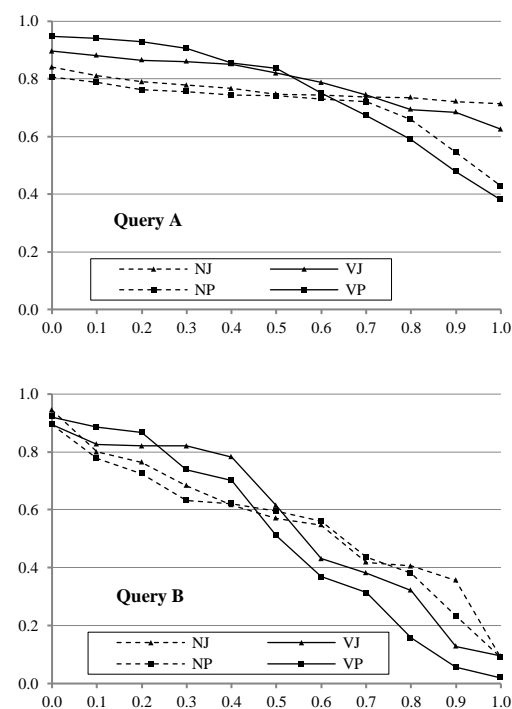


Gambar 10 Grafik perbandingan pengujian kelompok *query* A dan B.

B, nilai AVP paling besar dihasilkan oleh NJ sebesar 0.563, selanjutnya VJ sebesar 0.556, selanjutnya NP sebesar 0.540, dan terakhir VP sebesar 0.503.

Dari hasil tersebut, dapat dilihat bahwa untuk pencarian lafal, skema yang menghasilkan kinerja paling baik ialah VJ (pencarian atau pengindeksan dengan vokal dan pemeringkatan jumlah). Namun, dari grafik *precision-recall* yang ditunjukkan pada Gambar 11, terlihat bahwa VP (pencarian dengan vokal dan pemeringkatan posisi) lebih mendominasi pada titik *recall* < 0.4. Hasil seperti ini secara umum lebih menarik bagi pengguna karena berarti ayat yang relevan ditempatkan pada urutan teratas hasil pencarian, berbeda dengan NJ dan NP yang stabil namun *precision*-nya lebih rendah.

Sementara itu, untuk pencarian topik, kinerja paling baik dihasilkan oleh NJ. Hal ini terkait dengan perbandingan pencarian V dan N yang telah disebutkan karena kriteria ayat relevan yang lebih longgar dari sisi pemberian vokal dibandingkan dengan kriteria ayat relevan pada pencarian tanpa memperhitungkan makna. Dari grafik dapat dilihat bahwa VJ dan VP lebih baik daripada NJ dan NP pada titik *recall* rendah, namun NP dan VP tidak lebih baik daripada NJ dan VJ. Pada kasus ini,



Gambar 11 Grafik perbandingan keempat unit percobaan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

pemeringkatan posisi tidak dapat membantu memperbaiki *precision* karena kelompok *query* B sebagian besar adalah kata yang sering terdapat pada bagian akhir ayat.

Perbandingan Waktu Komputasi

Statistik waktu komputasi (waktu yang dibutuhkan sistem untuk memproses *query* dan mengembalikan hasil pencarian) dari masing-masing unit percobaan untuk seluruh *query* tercantum pada Tabel 7.

Tabel 7 Waktu komputasi setiap unit percobaan (dalam detik)

	NJ	NP	VJ	VP
Minimum	0.117	0.119	0.350	0.355
Maksimum	2.828	3.035	8.289	9.103

Dari perhitungan rata-rata dari rasio antara waktu pencarian dengan vokal dan tanpa vokal, diketahui bahwa secara umum waktu komputasi untuk pencarian tanpa vokal sekitar 8 kali lebih cepat dibandingkan dengan pencarian dengan vokal. Hal ini terkait algoritme pencarian yang digunakan yang bergantung pada jumlah trigram dari *query* serta panjang *posting list*. Iterasi paling banyak terjadi pada proses penggabungan *posting list*, sehingga bila *posting list* semakin panjang, waktu yang dibutuhkan juga semakin besar. Hasil pengindeksan pada Tabel 5 menunjukkan panjang rata-rata *posting list* untuk indeks dengan vokal 3 kali lebih besar daripada indeks tanpa vokal, sehingga jumlah dokumen hasil pencarian menjadi lebih banyak. Selain itu, kode fonetis dari *query* jika dihilangkan vokalnya akan menghasilkan jumlah trigram yang lebih sedikit, sehingga akan mengurangi waktu pencarian. Pembobotan tambahan untuk posisi yang berakhir pada karakter spasi juga memperlambat waktu komputasi karena harus memproses dan mengecek seluruh hasil pencarian.

Dari sisi metode pemeringkatan, pemeringkatan jumlah diproses lebih cepat dibandingkan dengan pemeringkatan posisi dengan perbedaan yang tidak terlalu signifikan. Hal ini terkait pekerjaan tambahan yang harus dilakukan untuk memberi skor pada dokumen, yaitu menghitung LIS serta skor kerapatan.

Secara keseluruhan, sistem mampu memproses *query* dalam waktu yang relatif cepat, kecuali untuk pencarian dengan vokal. Untuk memperbaiki hal tersebut, dapat diterapkan pemotongan hasil pencarian untuk

skor yang rendah, misalnya jika jumlah trigram yang cocok antara *query* dan suatu dokumen kurang dari 75% dari jumlah trigram pada *query*, maka dokumen tersebut tidak diikutsertakan sebagai hasil pencarian (dianggap tidak relevan).

KESIMPULAN DAN SARAN

Kesimpulan

Hasil penelitian ini menunjukkan bahwa:

1. Telah dihasilkan sistem pencarian ayat Al-Quran berbasis kemiripan fonetis dan metode pengodean fonetis untuk teks Al-Quran beserta penulisan lafalnya yang sesuai untuk pembicara bahasa Indonesia.
2. Kinerja sistem paling baik pada pencarian dengan menggunakan vokal, menggunakan pemeringkatan jumlah, serta untuk keperluan pencarian lafal tanpa memerhatikan makna.
3. Pemeringkatan berdasarkan posisi kemunculan hanya dapat memperbaiki *precision* pada titik *recall* rendah.

Saran

Saran untuk penelitian selanjutnya yaitu:

1. Penggunaan kamus atau daftar kata bahasa Arab untuk memperbaiki ambiguitas pada proses pengodean fonetis dari teks Latin.
2. Perbaikan metode pengindeksan sehingga dapat mencatat seluruh posisi kemunculan trigram alih-alih hanya posisi kemunculan pertama untuk meningkatkan kinerja pemeringkatan posisi.
3. Kuesioner pencarian variasi penulisan lafal menggunakan media suara berupa potongan lafal Al-Quran untuk kemudian dituliskan bunyinya oleh responden.

DAFTAR PUSTAKA

- Ahmad JIJ. 1988. A retrieval and display system for the Holy Quran [tesis]. Dhahran: Faculty of the College of Graduate Studies, King Fahd University of Petroleum and Minerals, Arab Saudi.
- Aqeel SU, Beitzel S, Jensen E, Grossman D, Frieder O. 2006. On the development of name search techniques for Arabic. *JASIST* 57:728-739.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

- Baeza-Yates R, Ribeiro-Neto B. 1999. *Modern Information Retrieval*. New York: Addison Wesley.
- Hammo B, Sleit A, El-Haj M. 2007. Effectiveness of query expansion in searching the Holy Quran. Di dalam: *Proceedings of the Second International Conference on Arabic Language Processing, CITALA '07*; Maroko, 18-19 Jun 2007. Rabat: IERA. hlm 1-10.
- [IC] IslamiCity. 2001. Phonetic Search. <http://islamicity.com/ps> [26 Nov 2011].
- Ismail AM, Nawawi MU. 1995. *Pedoman Ilmu Tajwid*. Surabaya: Karya Abditama.
- Junaedi. 2009. Kesalahan pelafalan para imam/khatib Jum'at dalam membaca Al-Qur'an: analisis studi kasus di Masjid Al-Muqadimah Kelurahan Ledeng Bandung [skripsi]. Bandung: Fakultas Pendidikan Bahasa dan Sastra, Universitas Pendidikan Indonesia.
- Kim MS, Whang KY, Lee JG, Lee MJ. 2005. N-gram/2L: a space and time efficient two-level n-gram inverted index structure. Di dalam: *Proceedings of the 31st VLDB Conference, VLDB '05*; Trondheim, 30 Agu - 2 Sep 2005. Trondheim: VLDB Endowment. hlm 325-336.
- Liben-Nowell D, Vee E. 2006. Finding longest increasing and common subsequences in streaming data. *Journal of Combinatorial Optimization* 11(2):155-175.
- Manning CD, Raghavan P, Schütze H. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- Nurhanifah S. 2001. Pencarian informasi dengan metode trigram [skripsi]. Bogor: Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Nwesri AFA. 2008. Effective retrieval techniques for Arabic text [disertasi]. Melbourne: School of Computer Science and Information Technology, RMIT University.
- [TANZ] Tanzil Contributors. 2011. Tanzil - Quran Navigator. <http://tanzil.net> [26 Nov 2011].
- [ZEKR] Zekr Contributors. 2010. Zekr - Quran Study Software for Linux. <http://zekr.org> [26 Nov 2011].
- Zobel J, Dart P. 1996. Phonetic string matching: lessons from information retrieval. Di dalam: *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '96*; Zurich, 18-22 Agu 1996. New York: ACM. hlm 166-172.



LAMPIRAN

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 1 Daftar *fawatihush shuwar* dan bacaannya

No.	Rangkaian Huruf	Bacaan	Nomor Surat	Nomor Ayat
1	الم	أَلِفٌ لَامٌ مِيمٌ	2	1
2	الم	أَلِفٌ لَامٌ مِيمٌ	3	1
3	المص	أَلِفٌ لَامٌ مِيمٌ صَادٌ	7	1
4	الر	أَلِفٌ لَامٌ رَا	10	1
5	الر	أَلِفٌ لَامٌ رَا	11	1
6	الر	أَلِفٌ لَامٌ رَا	12	1
7	الر	أَلِفٌ لَامٌ مِيمٌ رَا	13	1
8	الر	أَلِفٌ لَامٌ رَا	14	1
9	الر	أَلِفٌ لَامٌ رَا	15	1
10	كهيعص	كَافٌ هَا يَاءٌ عَيْنٌ صَادٌ	19	1
11	طه	طَا هَا	20	1
12	طسم	طَا سَيْنٌ مِيمٌ	26	1
13	طس	طَا سَيْنٌ	27	1
14	طسم	طَا سَيْنٌ مِيمٌ	28	1
15	الم	أَلِفٌ لَامٌ مِيمٌ	29	1
16	الم	أَلِفٌ لَامٌ مِيمٌ	30	1
17	الم	أَلِفٌ لَامٌ مِيمٌ	31	1
18	الم	أَلِفٌ لَامٌ مِيمٌ	32	1
19	يس	يَاءٌ سَيْنٌ	36	1
20	ص	صَادٌ	38	1
21	حم	حَا مِيمٌ	40	1
22	حم	حَا مِيمٌ	41	1
23	حم	حَا مِيمٌ	42	1
24	عسق	عَيْنٌ سَيْنٌ قَافٌ	42	2
25	حم	حَا مِيمٌ	43	1
26	حم	حَا مِيمٌ	44	1
27	حم	حَا مِيمٌ	45	1
28	حم	حَا مِيمٌ	46	1
29	ق	قَافٌ	50	1
30	ن	نُونٌ	68	1

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 2 Contoh isi dari file penyimpanan indeks

Vocabulary

(trigram | offset)

Posting list

(id dokumen : jumlah kemunculan trigram : posisi kemunculan trigram)

ABA	0	24:1:60,26:1:42,27:1:98,28:1:19,32:1:2,33:1:35,...
ABD	14232	30:1:35,68:1:144,185:2:67,228:1:122,665:1:29,12...
ABF	14687	1712:1:10,2092:1:7,2445:1:9
ABG	14715	376:1:19,392:1:53,527:1:220,719:1:22,953:1:16,9...
ABH	14910	289:1:200,473:1:21,530:1:10,700:1:20,1039:1:127...
ABI	15057	2:1:16,10:1:17,11:2:17,12:1:20,15:2:24,23:2:29,...
ABK	34319	1317:1:21,1612:1:21,1977:1:39,2138:1:22,2419:1:...
ABL	34511	11:1:51,28:1:51,32:1:133,96:1:61,98:1:154,115:1:...
ABM	37256	1515:1:75
ABN	37266	56:1:63,94:1:65,153:1:52,184:1:180,222:1:94,253...
ABR	38110	52:1:13,160:1:31,257:1:55,342:1:155,486:1:130,4...
ABS	38433	14:1:45,27:2:21,252:1:82,306:1:169,463:1:41,464...
ABT	39119	72:1:38,94:1:177,141:1:45,148:1:45,194:1:165,20...
ABU	39887	9:1:11,14:1:72,17:1:48,37:1:11,44:1:58,46:2:20,...
ABW	46353	196:1:130,272:1:93,833:1:39,994:1:58,1608:1:31,...
ABX	46528	36:1:66,136:1:8,203:1:306,268:1:68,528:1:28,693...
ABY	47072	194:1:229,399:1:7,400:1:12,1680:1:40
ABZ	47109	636:1:4,2055:1:54
ADA	47127	7:1:53,17:2:21,24:2:27,27:1:37,29:1:125,30:1:71...
ADD	56395	125:1:120,254:1:30,333:1:30,345:1:99,404:1:47,4...
ADD	56715	115:1:85,609:1:81,629:1:170,660:1:37,681:1:293,...
ADF	56848	244:1:43,478:1:104,499:1:60,886:1:66,887:1:61,9...
ADG	56987	1640:1:6,2488:1:9,4574:1:49,4582:1:41
ADH	57025	65:2:10,118:1:10,121:1:91,141:1:14,148:1:14,221...
ADI	58053	16:1:4,23:1:70,27:1:146,30:1:99,37:1:131,38:1:9...
ADK	64137	82:1:31,306:1:2,436:1:6,477:1:17,657:1:11,686:1:...
ADL	64601	55:1:79,71:1:36,97:1:76,112:1:126,130:1:55,205:...
ADM	65541	433:1:21,457:1:4,939:1:74,964:1:6,1049:1:51,119...
ADN	65728	58:1:9,68:1:161,94:1:90,151:1:2,260:1:115,289:1...
ADR	66623	33:1:22,67:1:32,80:1:7,372:1:156,416:1:25,436:1...
ADS	67428	445:1:6,474:1:4,515:1:45,516:1:348,646:1:60,771...
ADT	67640	133:1:144,205:1:54,240:1:300,242:1:26,244:2:48,...
ADU	68103	17:1:14,27:1:4,41:1:42,43:1:79,60:1:47,68:2:213...
ADW	73072	593:1:128,624:1:38,878:1:76,897:1:48,998:1:36,1...
ADX	73185	30:1:64,67:1:85,68:1:45,72:1:6,94:1:6,106:1:6,1...
ADY	75485	203:3:62,671:1:63,764:1:136,766:1:62,1039:1:8,1...
ADZ	75716	99:1:6,238:1:135,342:1:32,466:1:32,476:1:94,663...
AFA	76289	13:1:12,23:1:36,30:1:40,33:3:48,41:1:36,42:1:95...
AFD	87932	514:1:25,1870:1:27,2050:1:74,2099:1:110,2335:1:...
AFH	87982	1887:1:83,2251:1:24,2529:1:16,3992:1:43,4330:1:...
AFI	88042	9:1:20,26:1:119,29:1:24,31:1:75,34:1:88,36:1:20...
AFK	98065	571:1:196,814:1:60,854:1:153,887:1:85,1133:1:62...
AFL	98277	2289:1:48,2412:1:35,2484:1:53,2580:1:78,2674:1:...
AFM	98384	2417:1:24,3309:1:35
AFN	98404	1089:1:10,1376:1:67,1462:1:74,2070:1:11,2118:1:...
AFR	98642	76:1:69,243:1:54,257:1:42,413:1:52,481:1:19,500...
...
ZYT	3732028	1462:1:93,1539:1:71,4234:1:68
ZYU	3732058	92:1:251,121:1:125,337:1:55,601:1:56,702:1:161,...
ZZA	3732181	66:1:73,132:1:5,557:1:59,616:1:57,689:1:56,832:...
ZZI	3732683	55:1:16,130:1:16,437:1:141,438:1:129,779:1:330,...
ZZU	3733166	1088:1:24,1581:1:96

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 3 Algoritme untuk mencari LIS (Liben-Nowell & Vee 2006)

Input : sekuens S
Output : subsekuens menaik σ dari S dengan panjang maksimum

- 1 Inisialisasi
$$\begin{aligned}k' &= 0 \\ A[0] &= -\infty \\ A[1] &= \infty \\ \sigma^0 &= \{ \}\end{aligned}$$
- 2 Selama masih terdapat elemen pada S :
 - 1 Baca elemen berikutnya x_i dari S
 - 2 Cari nilai l sedemikian sehingga $A[l] \leq x_i \leq A[l+1]$
 - 3 Set nilai $A[l+1] = x_i$
 - 4 Set nilai $\sigma^{l+1} =$ penggabungan antara $(\sigma^l, \{x_i\})$
 - 5 Jika $l+1 > k'$, maka naikkan nilai k' dan set nilai $A[k'+1] = \infty$
- 3 Output $\sigma^{k'}$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 4 Daftar *query* (aksara Arab) untuk pengujian pencarian lafal

Kode	Query	Dokumen Relevan (Surat : Ayat)	Jumlah Relevan	Cakupan
A1	أُولُو الْأَرْثَابِ	38:29, 14:52, 13:19, 2:269, 39:18, 39:9, 3:7, 40:54, 2:179, 38:43, 3:190, 65:10, 5:100, 12:111, 39:21, 2:197	16	Huruf ال ب ء
A2	الْحَمْدُ لِلَّهِ	37:182, 1:2, 6:45, 35:34, 40:65, 27:59, 27:93, 14:39, 31:25, 23:28, 18:1, 10:10, 27:15, 39:29, 39:74, 17:111, 39:75, 29:63, 34:1, 6:1, 16:75, 35:1, 7:43	23	Huruf ح د
A3	يَا أَيُّهَا النَّبِيُّ	33:45, 8:64, 66:9, 9:73, 33:1, 33:28, 66:1, 8:70, 33:59, 8:65, 60:12, 65:1, 33:50	13	Huruf ي ه ن
A4	رَسُولَ اللَّهِ	91:13, 63:5, 63:1, 61:5, 61:6, 7:158, 48:29, 4:171, 33:21, 33:40, 49:3, 63:7, 49:7, 9:81, 9:61, 4:157, 9:120, 33:53, 6:124	19	Huruf ر س
A5	خَالِدِينَ فِيهَا أَبَدًا	9:22, 33:65, 4:169, 72:23, 98:8, 4:122, 4:57, 5:119, 9:100, 64:9, 65:11	11	Huruf خ ف
A6	الصَّابِرِينَ	21:85, 47:31, 16:126, 3:17, 2:153, 3:142, 22:35, 8:46, 2:155, 3:146, 37:102, 8:66, 33:35, 2:177, 2:249	15	Huruf ص
A7	وَلَقَدْ يُؤْمِنُ لَلْمُكَذِّبِينَ	83:10, 77:49, 77:47, 77:45, 77:40, 77:37, 77:34, 77:28, 77:24, 77:19, 77:15, 52:11	12	Huruf ذ م و , <i>idgham</i> , diftong “ai” dan “au”
A8	عَلَى كُلِّ شَيْءٍ قَدِيرٌ	11:4, 3:189, 5:120, 57:2, 22:6, 48:21, 67:1, 33:27, 6:17, 42:9, 2:106, 9:39, 29:20, 5:40, 30:50, 16:77, 2:148, 3:165, 3:29, 64:1, 46:33, 41:39, 59:6, 3:26, 65:12, 2:284, 2:20, 5:19, 24:45, 35:1, 2:109, 8:41, 5:17, 66:8, 2:259	35	Huruf ع ش ق ك , <i>tanwin</i>
A9	يَسْتَلْبِغُونَ	26:211, 36:75, 7:192, 36:50, 68:42, 25:9, 17:48, 21:40, 18:101, 7:197, 21:43, 4:98, 16:73, 11:20, 2:273	15	Huruf ت ط
A10	إِنَّ اللَّهَ غَفُورٌ رَحِيمٌ	2:192, 24:5, 3:89, 8:69, 2:199, 5:39, 2:226, 2:182, 9:102, 64:14, 16:115, 58:12, 2:173, 49:14, 9:99, 9:5, 60:12, 24:62, 5:3, 73:20	20	Huruf غ
A11	مَثَلُ الَّذِي	2:171, 59:15, 2:17, 29:41, 14:18, 2:261, 62:5, 2:265, 2:214, 2:228	10	Huruf ث
A12	قَوْمِ الظَّالِمِينَ	6:47, 23:94, 26:10, 10:85, 28:21, 23:41, 7:47, 23:28, 61:7, 11:44, 3:86, 46:10, 28:50, 66:11, 6:68, 62:5, 9:109, 5:51, 9:19, 28:25, 2:258, 7:150, 6:144	23	Huruf ظ
A13	ضَالِّينَ	68:26, 26:86, 26:20, 37:69, 56:92, 56:51, 83:32, 15:56, 1:7, 23:106, 3:90, 6:77, 2:198	13	Huruf ض
A14	مِنْ بَعْدِ مَا جَاءَ	98:4, 2:209, 3:105, 2:211, 45:17, 3:19, 42:14, 3:61, 2:145, 4:153, 2:253, 2:213	12	Huruf ج , <i>iqlab</i>
A15	تَنْزِيلَ	36:5, 69:43, 56:80, 41:2, 26:192, 76:23, 46:2, 45:2, 40:2, 20:4, 32:2, 25:25, 17:106, 39:1, 41:42	15	Huruf ز , <i>ikhfa</i>
A16	فَأُولَٰئِكَ هُمْ	23:102, 70:31, 23:7, 3:82, 7:178, 3:94, 24:52, 7:8, 5:47, 2:121, 63:9, 64:16, 30:39, 9:23, 60:9, 59:9, 5:45, 49:11, 24:55, 5:44, 2:229	21	“Ai” dan “au” yang bukan diftong

Hak Cipta Dilindungi Undang-Undang

Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 5 Daftar *query* (aksara Arab) untuk pengujian pencarian topik

Kode	Query	Dokumen Relevan (Surat : Ayat)	Jumlah Relevan	Keterangan
B1	سُبْحَانَ اللَّهِ	37:159, 69:52, 56:96, 56:74, 37:180, 87:1, 52:43, 17:108, 15:98, 43:82, 30:17, 68:29, 21:22, 52:48, 16:57, 10:10, 110:3, 57:1, 61:1, 59:1, 40:55, 27:8, 39:75, 28:68, 16:1, 62:1, 39:4, 48:9, 12:108, 64:1, 50:39, 24:36, 7:206, 19:35, 24:41, 20:130, 41:38, 3:41, 2:116, 59:23, 23:91, 10:18, 42:5, 6:100, 10:68, 32:15, 3:191, 9:31, 59:24, 4:171, 13:13, 39:67, 30:40, 40:7, 43:13, 7:143, 5:116, 17:93, 17:1, 2:30	60	Ayat tentang menyebut kesucian Allah
B2	فَاسِقُونَ	3:82, 51:46, 43:54, 7:102, 2:99, 59:19, 9:53, 10:33, 9:96, 5:47, 9:84, 21:74, 57:26, 27:12, 5:59, 7:145, 24:4, 5:108, 61:5, 18:50, 9:67, 28:32, 7:163, 2:26, 9:24, 24:55	26	Ayat tentang ciri-ciri orang fasik
B3	مُفْلِحُونَ	91:9, 87:14, 23:102, 23:1, 31:5, 2:5, 28:67, 20:64, 7:8, 3:200, 8:45, 22:77, 3:130, 5:35, 62:10, 3:104, 5:100, 30:38, 9:88, 64:16, 5:90, 24:51, 7:69, 2:189, 59:9, 7:157, 58:22, 24:31	28	Ayat tentang ciri-ciri orang yang beruntung
B4	مُتَّقِينَ	78:31, 26:90, 44:51, 77:41, 54:54, 52:17, 51:15, 15:45, 50:31, 38:49, 68:34, 43:67, 3:76, 25:15, 43:35, 3:133, 28:83, 45:19, 16:31, 7:128, 9:123, 11:49, 13:35, 16:30, 9:4, 9:7, 5:27, 2:194, 9:36, 47:15, 41:18, 27:53, 16:128, 12:57, 19:72, 39:61, 2:103, 7:201, 3:200, 3:123, 20:132, 47:36, 39:20, 2:212, 39:73, 8:29, 3:120, 3:15, 3:186, 7:156	50	Ayat tentang balasan bagi orang yang bertakwa
B5	أَكْثَرُ نَاسٍ	12:103, 40:57, 34:28, 17:89, 27:73, 40:59, 25:50, 30:6, 34:36, 40:61, 30:30, 45:26, 12:38, 2:243, 16:38, 10:60, 13:1, 12:68, 12:40, 12:21, 11:17, 7:187	22	Ayat tentang sifat kebanyakan manusia

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 6 Daftar variasi penulisan *query* untuk pengujian[illegible]

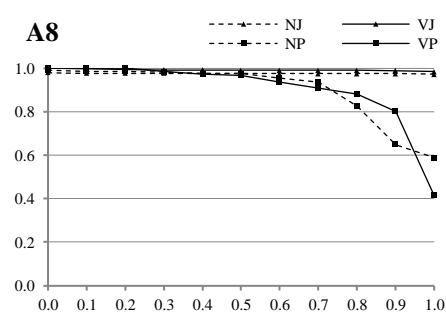
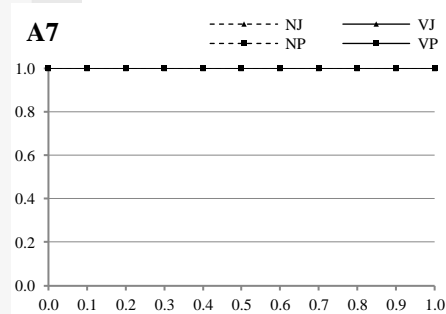
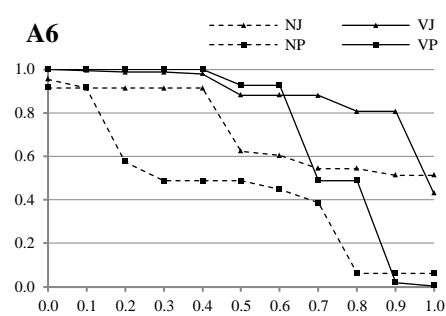
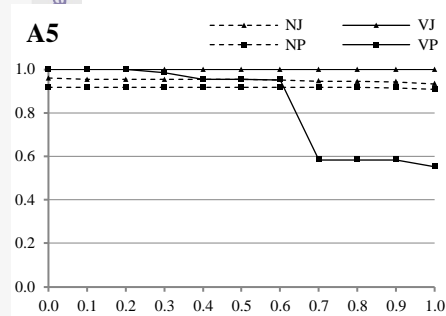
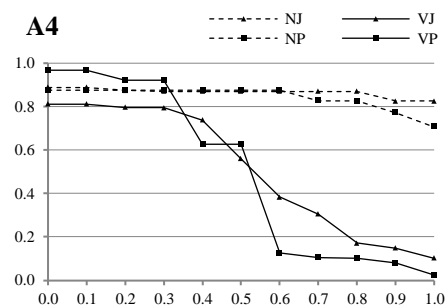
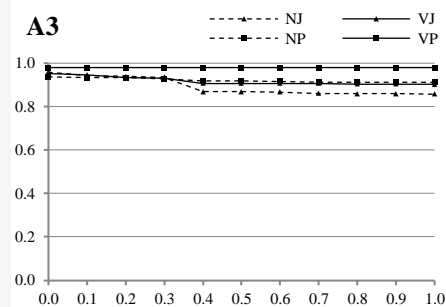
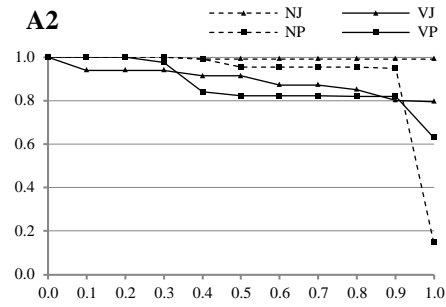
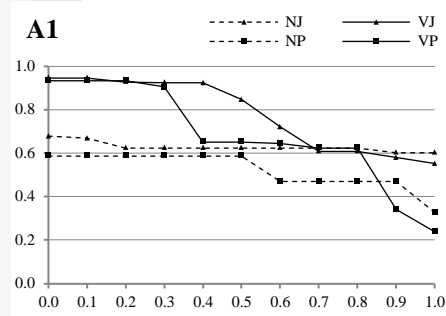
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 6 Lanjutan

A9	yastathii'uun, yastathi'un, yastathi'uun, yastatii uun, yastatii'uun, yastathiy'uun, yastati'awn, yastathiiun, yastathii'uuna, yasta thii'uun, yastii uun, yas tatii uun, yastathiy'awn, yastathiiu'uun, yastathii'un, yastathingun, yastatwii'uun, yastat'ii 'uun, yastatiuun, yas tathii'uun	20
A10	innalloha ghofururrohiim, innallaha ghofururrohiim, innallaha ghofuurrurrahiim, innallaaha ghafuurur rahiim, innallaaha ghofuurur rohiim, innallaaha ghafuururrohiim, innallaha ghafururrahim, innalloha ghofuururrohiim, innalloha ghofuururrohiim, innallaha ghofururrahiim, innalloha ghafuurrurrahiim, innallaha ghofururrohiim, innallaha ghofurur rohiim, innallaha ghafururrahiim, inna allaha ghafuurur rahiim, innallaha ghafururrohim, innalloha ghofuurur rohiim, innallahaghfururrahim, innaalloha ghafuurrurrahiim, innallaha ghofuurur rohiim, innallaaha ghafuurrurrahiim, innalloha ghofuurrurrohiim, innallaha ghafuurun rahiim, innallaha ghofuururrohim, innalloha ghopuu rurohiim, innalloha ghofurur rhoiim, innallaha ghafuurur rahiim, innallaha ghafuurrurrahiim, innallahaghofuurrurrahiim, innaallaha ghafuurrurochiym, innallaha ghofururrohiim, innaalloha ghofururrohim, innallaha ghafururrahiim, innalloha ghofurur rohim, innallaha ghofururrohim, innallaha ghafuurur-rahiim, inna allaaha ghafuurrurrahiim, innallaha ghofururrahiim, innallaha ghafuru rahim, innallaha ghofuurur rohiim, innallaha qhofururrohiim	42
A11	masyallullazii, masyallulladzii, matsalulladzii, matsalul ladzii, matsalul lazi, matsalullazi, matsaluladzii, matsalulladzi, matsalulladziy, matsalulladzi, masaluladzi, matsalul ladzi, masyallulladzi, matsalullazii, masalul ladzi, matsaludzin, mathallullazii, mastalul ladziy, masalullazii, masalullazi, matsalu al ladzii, matsalul ladziy, masalulladzi	23
A12	qaumizzolimiin, qoumizh zhooliiman, qow mizhoolimiin, qawmizhzhoolimiin, qowmidz dzoolimiin, qaumidzhoolimiin, qaumidhalimin, qoumizhzhoolimiin, qowmidzhoolimiin, qoumidzhdzoolimiin, qaumidh dhoolimiin, qoumitzalimiin, qoumizzholimiin, qoumizhoolimiin, qoumidz dzoolimiin, qoumizzoolimiin, qaumidh dholimiin, qaumidzh dzhaliimin, qaumidzdzoolimiin, qaumidzoolimiin, qawmizh zhaaliimin, qawmidzdzoolimiin, qowmidzoolimiin, qoumizh-zholimiin, khumizhzhoolimiin, qoumadzdzoolimiin, qoumi dzaaliimin, qoumizh zhoolimiin, qaumizhzhoolimiin, kouwimizzoolimiin, qowmizh zhoolimiin, qaumidhhaaliimin, qowmidh dholimiin, qowmizhoolimiin, qaumizh zhaaliimin, qoumidzoolimiin, qaumidzoolimiin, qowmi adhdhoolumiin, kowmittholimiin, qoumizhzhoolimin, qoumidhoolimiin, qoumizzhoolimiin, qaumizzhoolimiin, qoumidh dholimin, kouwimizzholimin, qoumi dzaaliimin, qoumidhholimiin, qoumi al dzaaliimin, khomizoolimiin, qoumidzhoolimiin, qoumdzoolimiin, qawmi zholimin, qowmidz dzoolimiin, qoumizzholimin, qaumizzoolimiin, qaumizzhaaliimin, qaumizh zhaaliimin, qoumidzoolimiin	58
A13	dhooliin, dhoooliin, dhaaliin, dhoooliin, dhallin, dhoooliin, dlooliin, dhool liin, dolliin, dhallin, dhaaliin, dhollin, zooliin, dhaaaliin, dholliin, dholliin, dlaaliin, dhdhalliin, dho lin, dhaaaaaliin, dlooliin, dzooliin, dhool liin, dhaal liin	24
A14	mimba'di maa jaa a, mimba'dimaajaa, mim ba'di maa jaa'a, mim ba'di maa jaa a, mim ba'di ma ja a, mimba'di maa jaaaaa'a, mimba'di maajaa', min ba'di maa jaa'a, mim ba'di maa jaa, mimba'dimaajaa a, mim ba'di maa jaa a, mimba'di maajaa a, min ba'di maa jaa'a, min ba'di maa jaa a, min ba'di maa jaa a, mim ba'di maa jaa-a, minba'kimja', mimmba'di maa jaa a, mimba'di maa jaa, min ba'di maa jaa-a, min ba'di maa ja a, minba'dimaaja, mimm ba'di maa jaa a, mim ba'dimaajaaa', min ba'di maja-a, mimba'dimaajaa a, mimba'di majaa a, mimba'di ma ja'a, minba'di ma ja a, mimmba'kdimaajaa, min ba'di maa jaa a', minba'di maajaa-a, mamba'di maa jaa a, mimba'dimaajaa', min ba'di majaa, minba'dimaajaa a, mimba'dimaajaa'a, min' ba'di maa jaa a, min ba'di maajaa a	39
A15	tanziiil, tangziil, tanzil, tan ziil, tanziyl, tandziyl, tannziil, tan nziil, tangzil, tanjiil, tanzilyl, tanzhil, tangziil, tandziil	14
A16	fa-ula-ikahum, faulaaikaikum, fa `ulaaika hum, fa ulaa ikahum, faulaaika hum, fa ulaika hum, fa ulaaaaaikaikum, faulaikaikum, fa uulaikaikum, fa uu la ikahum, fauulaikaikum, faula ika hum, fa`ulaikaikum, faulaaika, fauulaikaikum, fauulaika hum, fa ulaa ika hum, faulaa ika hum, fa ula ikahum, fa uu la ika hum, fa-uulaikaikum, fa ulaaika hum, fa ula ika hum, fau laika hum, fa u laika hum, fa'ula'ika hum, fa-ula-ika hum, fa uulaika hum, faulaikakhum, fa uulaikahum, faulaika hum, fauu laika hum	32
B1	subhanallah, subhaanallaah, subhaanallah, subhaanalloh, subhanalloh, subhanallaah, subhaanallooh	7
B2	faasiqun, faasiqun, fasiqun, faa siqun, fasiqun, fasiqun, faa sikuun, faa si quun, faaasikuun, fasquun, faasiqun, faasiqun	12
B3	muflihun, muflihuun, muflihuwn, muflihuun, muflihun	5
B4	muttaqiin, muttaqin, mutaqiin, muttaqiyn, muttakin, muttaqqiin, muttakiin, mutaqin, muttaqiin, muttaqien	10
B5	aksyaronas, aktsaronaas, aktsara naas, aktsaranas, aktsaro naas, ikstaranaas, aktsaronas, aktsaranaas, aksaronaas, aktsaro nas, aktsanas, ak tsaranaas, aktsaran naas, aktsaronas, aktsaro naas, aktharanaas, aktsaranas, aktsaranas	18

- Hak Cipta Dilindungi Undang-Undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 7 Grafik *precision-recall* untuk masing-masing *query*


Hak Cipta Dilindungi Undang-Undang

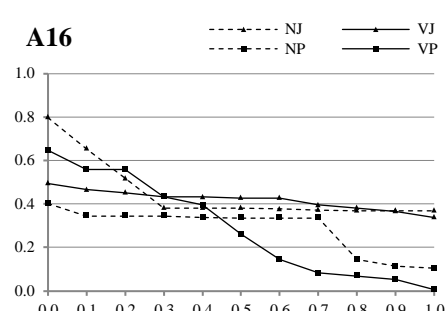
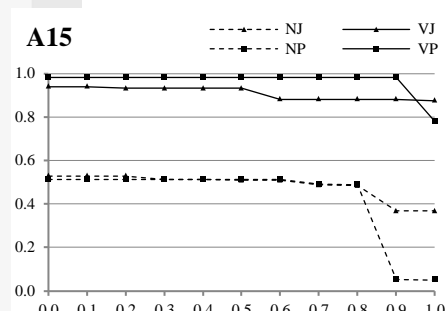
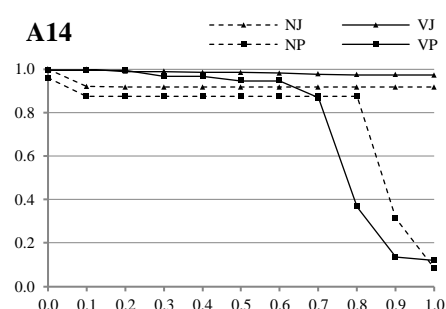
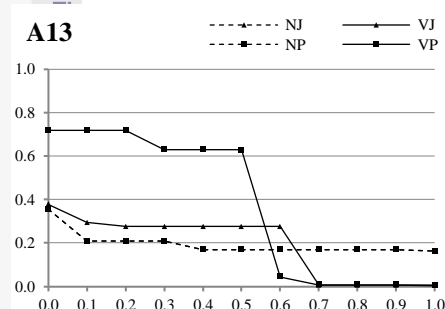
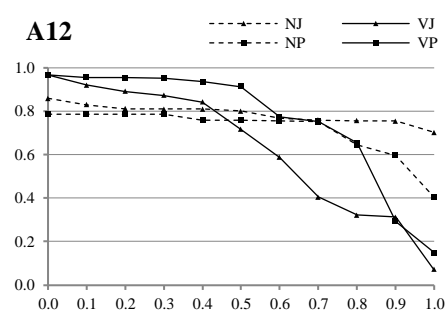
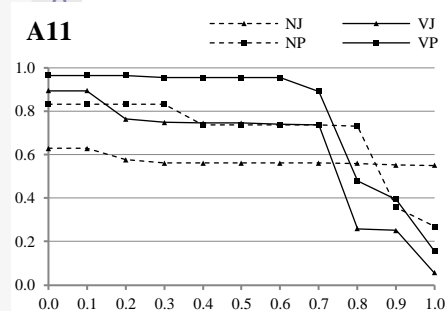
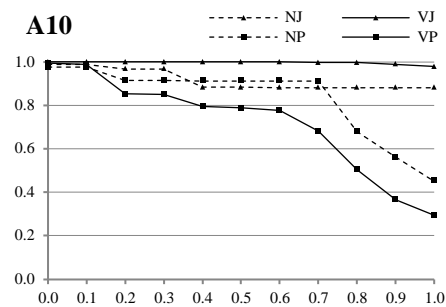
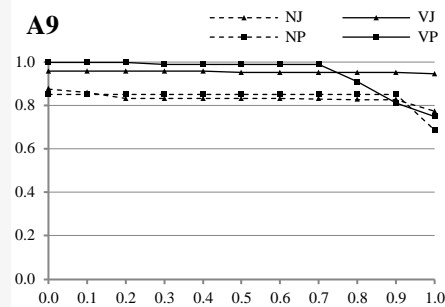
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

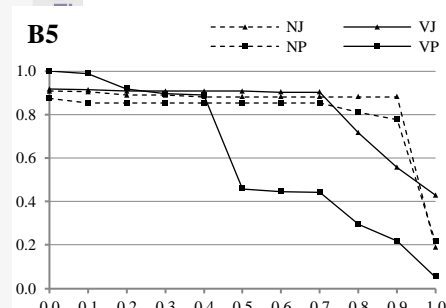
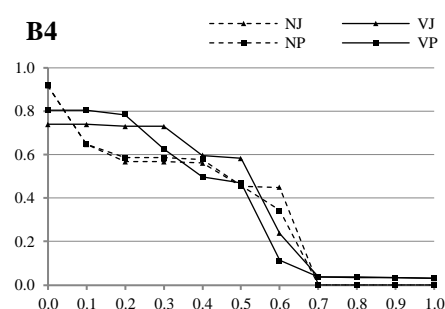
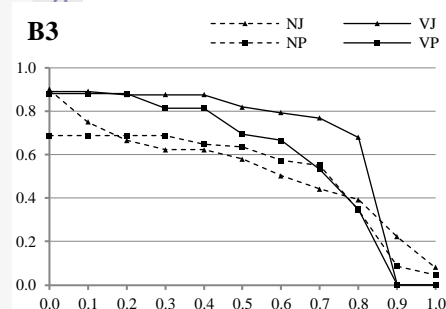
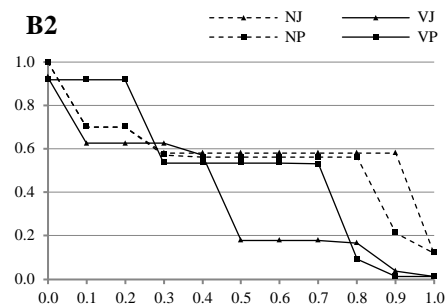
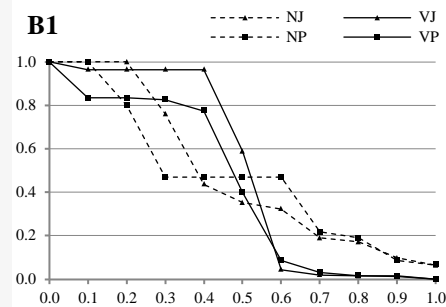
Lampiran 7 Lanjutan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Lampiran 7 Lanjutan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.