

**STUDI PERBANDINGAN AKURASI LIGHT STEMMING DAN  
KHOJA STEMMING PADA FI'IL MADHI DAN MASHDAR**

**SKRIPSI**

Oleh

**ALDITA RAHMA MEI M.L**  
**NIM. 10650113**



**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2015**

**HALAMAN PENGAJUAN**

**STUDI PERBANDINGAN AKURASI LIGHT STEMMING DAN  
KHOJA STEMMING PADA FI'IL MADHI DAN MASHDAR**

**SKRIPSI**

**Diajukan Kepada :  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk Menempuh Salah Satu Persyaratan Dalam  
Memperoleh Gelar Sarjana Komputer**

**oleh :  
ALDITA RAHMA MEI M.L  
NIM. 10650113 / S-1**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2015**

**HALAMAN PERSETUJUAN**

**STUDI PERBANDINGAN AKURASI LIGHT STEMMING DAN  
KHOJA STEMMING PADA FI'IL MADHI DAN MASHDAR**

**SKRIPSI**

Oleh :

Nama : Aldita Rahma Mei M.L  
NIM : 10650113  
Jurusan : Teknik Informatika  
Fakultas : Sains Dan Teknologi

Telah Disetujui, 6 Januari 2015

**Dosen Pembimbing I,**

**Dosen Pembimbing II,**

Zainal Abidin, M.Kom  
NIP. 19760613 200501 1 004

Umaiatus Syarifah, M.A  
NIP. 19820925 200901 2 005

Mengetahui,  
**Ketua Jurusan Teknik Informatika**

Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008

**HALAMAN PENGESAHAN****STUDI PERBANDINGAN AKURASI LIGHT STEMMING DAN  
KHOJA STEMMING PADA FI'IL MADHI DAN MASHDAR****SKRIPSI**

Oleh :  
**Aldita Rahma Mei M.L**  
**NIM. 10650113**

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal : 12 Januari 2015

<b>Susunan Dewan Penguji:</b>	<b>Tanda Tangan</b>
1. Penguji Utama : <b><u>Totok Chamidy, M.Kom</u></b> NIP. 19691222 200604 1 001	( )
2. Ketua Penguji : <b><u>Ririen Kusumawati, M.Kom</u></b> NIP. 19720309 200501 2 002	( )
3. Sekretaris : <b><u>Zainal Abidin, M.Kom</u></b> NIP. 19760613 200501 1 004	( )
4. Anggota Penguji : <b><u>Umayatus Svarifah, M.A</u></b> NIP. 19820925 200901 2 005	( )

Mengetahui,  
**Ketua Jurusan Teknik Informatika**

**Dr. Cahyo Crysdiان**  
NIP. 19740424 200901 1 008

## HALAMAN PERSEMBAHAN

Alhamdulillah.. Puji syukur yang sebesar-besarnya tercurahkan kepada Allah SWT atas selesainya tugas akhir ini. Skripsi ini kupersembahkan untuk kedua orangtuaku :

**Abah Drs. H. Masykur Lukman, S.H M.Si**

**Ibu Hj. Badiyah Nur Afifah**

Yang tak pernah bosan memberikan semangat, motivasi, dukungan selama menyelesaikan studi di UIN Maliki Malang. Yang tak pernah mengeluh ketika aku merengek. Dan yang tak pernah lupa, absen dalam mendoakan putrinya ini untuk sukses di dunia maupun akhirat, dan diberikan segala kemudahan. Terimakasih abah, ibukk yang mbak sayaaang ..

Dan ucapan terimakasih yang sebanyak-banyaknya kepada :

- ☞ 2 adekku tercinta, yang selalu ada untuk mbak.. Adek **Aldiki Zaki Zamani M.L** dan adek **M. Wildan Dewanta M.L** terimakasih sudah memberi mbak semangat, dukungan dan doa untuk menyelesaikan skripsi ini.
- ☞ Para master, Taufan , Afif yang telah membantu proses pengerjaan program dan tak bosan aku tanya setiap waktu. Tanpa kalian skripsi ini tak kan selesai. Terimakasih banyak.
- ☞ Teman-teman Infinity 2010 seperjuangan menyelesaikan skripsi ini sampai tidak mengenal makan dan waktu, Vivid, Dzikri, Agus Cahyono, Musfiroh, Naufal, Kholid, Taufan. Semoga kita sukses selalu.
- ☞ Special thank's for my roommate Ade Durotun Nisa' sahabatkuu, terimakasih semangat, doa dan dukungannya setiap waktu, terimakasih sudah selalu mendengarkan apa keluhanku sehari-hari.
- ☞ Unni Puspa Safitri , terimakasih selalu menemaniku di saat penat datang melanda.
- ☞ Sahabat-sahabatku Mbak Itta, dewi, Ade, Puspa, mbak vivid, terimakasih dukungan dan semangat kalian. Sukses buat kita semua.
- ☞ Teman- teman MMM , Ade, Puspa, Itta , Dewi, Vivid, Syafei, Dian, Naufal, Agus, Haris, Nopi, Zaenal. Terimakasih semangat , doa, canda dan tawa kalian, serta dukungan kalian semua.
- ☞ Terimakasih yang sudah memberi semangat , dukungan dan selalu mendoakanku ☺
- ☞ Teman-teman INFINITY (TI'10) khususnya 4GTI yang telah memberikan keceriaan selama menempuh studii.
- ☞ Teman – Teman INFINITY 2010, terimakasih semangat dan dukungannya. Sukses buat kita semua.

## HALAMAN MOTTO

*Jadikanlah kami hamba yang selalu bersyukur  
atas nikmat-Mu,*

*Yang selalu berikhtiar atas uji-Mu*

*Yang selalu berprasangka baik terhadap-Mu*

*Yang selalu berusaha mendekatkan diri pada-  
Mu*



## HALAMAN PERNYATAAN ORISINALITAS PENELITIAN

Saya yang bertandatangan di bawah ini:

Nama : Aldita Rahma Mei M.L  
NIM : 10650113  
Fakultas : Sains Dan Teknologi / Teknik Informatika  
Judul Penelitian : Studi Perbandingan Akurasi Light Stemming dan Khoja Stemming pada Fi'il Madhi dan Mashdar.

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Januari 2015  
Yang Membuat Pernyataan,

Aldita Rahma Mei M.L  
10650113

## KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

*Assalamu'alaikum Wr. Wb.*

Segala puji bagi Allah SWT, karena atas rahmat, taufik dan hidayahnya, penulis dapat menyelesaikan studi di Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Sekaligus menyelesaikan skripsi ini dengan baik. Shalawat serta salam tetap tercurahkan kepada junjungan Nabi Muhammad SAW., yang telah membimbing umatnya menuju jalan yang *diridhoi* oleh Allah SWT.

Selanjutnya, penulis haturkan ucapan terima kasih seiring do'a dan harapan kepada semua pihak yang telah membantu terselesaikannya skripsi ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang, dan para pembantu Rektor, atas segala motivasi dan layanan fasilitas yang telah diberikan selama penulis menempuh studi.
2. Dr. drh. Hj. Bayyinatul Muchtaromah, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Cahyo Crys dian selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Zainal Abidin, M.Kom. selaku Pembimbing dalam penulisan skripsi ini yang telah memberikan memotivasi, membantu dan memberikan penulis arahan yang baik dan benar dalam menyelesaikan penulisan skripsi ini.

5. Umaiatus Syarifah, M.A selaku Dosen Pembimbing integrasi yang bersedia meluangkan waktu untuk memberikan masukan dan arahan mengenai laporan dan permasalahan integrasi Al-Quran.
6. Dr. Muhammad Faishal, M.T selaku Dosen Wali, terima kasih atas bimbingan, masukan dan saran ketika penulis mengalami kesulitan selama proses perkuliahan dari semester awal sampai semester akhir.
7. Seluruh Dosen Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang, khususnya Dosen Teknik Informatika dan staf yang telah memberikan ilmu kepada penulis serta dukungan dalam menyelesaikan penulisan Skripsi ini.
8. Abah Drs. H. Masykur Lukman, S.H, M.Si dan ibuk Hj. Badiyah Nur Afifah yang selalu tiada hentinya mendukung dan mendoakan aku.
9. Semua pihak yang ikut membantu dalam menyelesaikan skripsi ini baik berupa materiil maupun moriil.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap semoga skripsi ini bisa memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amiin Yaa Robbal Alamin.*

*Wassalamu 'alaikum Wr. Wb.*

Malang, 12 Januari 2015  
Penulis

Aldita Rahma Mei M.L

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>HALAMAN PENGANTAR</b> .....	<b>ii</b>
<b>HALAMAN PERSETUJUAN</b> .....	<b>iii</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>iv</b>
<b>HALAMAN PERSEMBAHAN</b> .....	<b>v</b>
<b>HALAMAN MOTTO</b> .....	<b>vi</b>
<b>HALAMAN PERNYATAAN</b> .....	<b>vii</b>
<b>KATA PENGANTAR</b> .....	<b>viii</b>
<b>DAFTAR ISI</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>DAFTAR GAMBAR</b> .....	<b>xv</b>
<b>ABSTRAK</b> .....	<b>xvi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Penelitian .....	4
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	5
1.6 Metodologi Penelitian .....	5
1.7 Sistematika Penulisan .....	6
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>10</b>
2.1 Penelitian Terkait .....	10
2.2 Pengolahan bahasa alami .....	15
2.3 Bidang Pengetahuan Dalam Bahasa Natural .....	16

2.4 Penelusuran Informasi .....	17
2.5 Text minning .....	19
2.5.1 Text Preprocessing .....	19
2.5.2 Text transformation .....	20
2.5.3 Pattern discovery/Analysis .....	21
2.6 Bahasa Arab .....	22
2.6.1 Sejarah Bahasa Arab .....	22
2.6.2 Pembagian Bahasa Arab.....	27
2.6.3 Jenis-Jenis Kata Kerja .....	29
2.7 Ilmu Sharaf.....	30
2.7.1 Pengertian Ilmu Sharaf.....	31
2.7.2 Istilah dalam Ilmu Sharaf.....	31
2.8 Al Quran .....	31
2.8.1. Pengertian Al Quran.....	31
2.6 Stemming .....	34
<b>BAB III ANALISIS DAN PERANCANGAN SISTEM .....</b>	<b>40</b>
3.1 Desain Aplikasi .....	40
3.2 Desain Proses .....	41
3.3 Desain Database .....	48
3.4 Teknik Stemming Imbuhan .....	49
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>51</b>
4.1 Implementasi Sistem .....	51
4.2 Implementasi Interface .....	52
4.3 Uji Coba .....	54
4.3 Pembahasan .....	63
4.4 Integrasi Aplikasi Penerjemah Dan Islam .....	65

<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>67</b>
5.1 Kesimpulan .....	67
5.2 Saran .....	68
<b>DAFTAR PUSTAKA .....</b>	<b>70</b>



## DAFTAR TABEL

Tabel 4.1 Uji coba menggunakan Light Stemming dengan output fi'il madhi.....	54
Tabel 4.2 Uji coba menggunakan Khoja dengan output fi'il madhi.....	56
Tabel 4.3 Uji coba menggunakan Light Stemming dengan output mashdar.....	58
Tabel 4.4 Uji coba menggunakan Khoja dengan output mashdar.....	60



## DAFTAR GAMBAR

Gambar 2.1 Contoh <i>stemming</i> .....	22
Gambar 3.1 Blok diagram .....	41
Gambar 3.2 Flowchart <i>stemming fi'il madhi</i> .....	43
Gambar 3.3 Flowchart <i>stemming mashdar</i> .....	44
Gambar 3.4 Contoh <i>stemming</i> .....	45
Gambar 3.5 Flowchart <i>stemming</i> dengan <i>khoja</i> .....	46
Gambar 3.6 flowchart Light <i>stemming</i> dengan output <i>fi'il madhi</i> .....	47
Gambar 3.7 flowchart <i>khoja</i> dengan output <i>mashdar</i> .....	34
Gambar 3.8 Desain database .....	48
Gambar 3.9 Potongan tabel <i>fi'il mudhori'</i> .....	48
Gambar 3.10 Jalannya aplikasi <i>stemming</i> .....	49
Gambar 4.1 Tampilan halaman utama .....	52
Gambar 4.2 Contoh penggunaan aplikasi .....	53
Gambar 4.3 Aplikasi pemotongan imbuhan.....	53

## ABSTRAK

Rahma, Aldita, 2015. **Studi Perbandingan Akurasi Light Stemming dan Khoja Stemming pada Fi'il Madhi dan Mashdar**. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing : (I) Zainal Abidin, M.Kom, (II) Umaiatus Syarifah, M.A

---

**Kata Kunci** : *stemming*, bahasa Arab, kata dasar

Salah satu bahasa yang sangat populer di kalangan umat Islam adalah bahasa Arab karena bahasa ini adalah bahasa Quraniah yang dipakai dalam ibadah sehari-hari seperti tilawah Quran, Sholat, dan juga berdoa pada sang khalik Allah SWT. Selain itu, bahasa Arab merupakan pedoman bahasa pada dua pedoman umat Islam (al-Quran dan al-Hadits) bahasa para nabi dan bahasa para penghuni surga. Maka sudah selayaknya umat Islam mempelajari bahasa Arab sejak dini agar memudahkan dirinya dalam memahami ajaran agamanya. Pada dasarnya, dalam belajar bahasa Arab, harus mengetahui kata dasar dari suatu kalimat bahasa Arab tersebut, karena dari kata dasar tersebut, kita dapat mengetahui kedudukan dan arti dari kalimat Arab. Pada *sharaf* terdapat jenis-jenis kata kerja, diantaranya adalah yang menjadi cakupan dalam skripsi ini *fi'il mudhori'*, *fi'il madhi* dan *mashdar*. Dan untuk mencari kata dasar dari kata Arab tersebut, membutuhkan keakurasian. Tidak sedikit umat Islam di dunia ini tidak mengetahui kata dasar dari kata Arab yang dibacanya sehari-hari, sehingga menimbulkan kesalahan dalam arti yang sesungguhnya. Huruf per huruh hijaiyah dalam kata Arab, mempunyai arti, sehingga salah sedikit menentukan kata dasar, kata dasar tersebut salah dan tidak dapat diartikan. Maka peneliti termotifasi membuat aplikasi studi perbandingan akurasi *light stemming* dan *khoja stemming* pada *fi'il madhi* dan *mashdar* sehingga memudahkan pengguna dalam mengetahui ketepatan perubahan bentuk dari *fi'il mudori'* menjadi *fi'il madhi* dan *mashdar*. Berdasarkan pengujian dari 44 kalimat Arab dapat disimpulkan bahwa aplikasi *stemming* dengan menerapkan algirotma *light stemming* dan *khoja* mampu membentuk kata dasar. Dari hasil uji coba dapat diambil prosentase keakurasian dengan *lighth stemming* 93% sedangkan dengan *khoja* 20%. Adapun kesalahan yang terjadi disebabkan karena terdapat kata yang mengalami *overstemming* atau *understemming* dan kata dasar tidak terdeteksi dan tidak terdapat kamus bahasa Arab Al-Munawwir. Untuk pengembangan lebih lanjut, perlu ditambahkan kata kerja (*fi'il*) agar dapat mengetahui kata dasar dari berbagai *fi'il*.

## ABSTRACT

Rahma, Aldita, 2015. **Comparative Study of Accuracy Light Stemming and Stemming Khoja on fi'il Madhi and mashdar**. Department of Informatics Engineering. Faculty of Science and Technology. Islamic State University of Maulana Malik Ibrahim Malang. Advisors : (I) Zainal Abidin, M.Kom, (II) Umaiatus Syarifah, M.A

---

**Keywords** : *stemming*, Arabic Language, basic words

A very popular language among Muslims is Arabic, because this is a "Quraniah" language that used everyday as worship like recitations of the Quran, praying, and also pray to Allah the creator of universe. As addition, Arabic used as guideline for Muslims (al-Quran and al-Hadith), the language of the prophets and the language of the inhabitants of heaven. By that reasons, Muslims are suppose to learn the Arabic language from an early age, so that allows him to understand religion teaching. Basically, in learning Arabic, we should know the basic words of an Arabic sentence, because of the basic words, we can know the position and the meaning of the sentences. In Arabic there is a word called "sharaf" are the types of verbs, such as being in the scope of this thesis, "fi'il Mudhori", "fi'il madhi" and "mashdar". And to find the root of the basic word, requires accuracy. Its only few Muslims in this world does know the basic words from an Arabic they reads everyday, so it giving errors in the sense from the original. Letters of hijaiyah in Arabic, has a meaning, so if one bit of determining the base is wrong, it will ruined the whole words and can not be interpreted. So it motivated the researchers to make a comparative study of the application of light stemming accuracy and stemming Khoja on ficil madhi and mashdar making it easier for users to determine the accuracy of deformation of fi'il mudori 'into fi'il madhi and mashdar. Based on testing of 44 Arabic sentence it can be concluded that the applications stemming by applying light stemming and Khoja algoitma able to form basic words. From the test results can be taken percent accuracy with lighth stemming Khoja 93%, while the 20%. The error is caused because there is a word that is experiencing overstemming or understemming and basic word is not detected and there is no dictionary Arabic Al-Munawwir. For further development, need to be .added verb (fi'il) that can also learn basic words from various fi'il

## الملخص

رحمة الديتا، 2015. دراسة مقارنة لالدقة الخفيفة وانطلاقا ووقف خوجة على فعل ماضي و مسدر. أطروحة. قسم المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج. المشرف: زينل أبدین الماجستير أميا نُشرفة الماجستير

الكلمات البحث: النابعة، اللغة العربية، كلمة الأساسية

لغة بشعبية كبيرة بين المسلمين هي اللغة العربية، لأن هذا هو "قرآنية" اللغة التي تستخدم كل يوم كما عبادة مثل تلاوة القرآن، والصلاة، والدعاء أيضا إلى الله خالق الكون. كما بالإضافة إلى ذلك، تستخدم اللغة العربية بوصفها المبدأ التوجيهي للمسلمين (سورة القرآن، و آل الحديث)، لغة الأنبياء ولغة سكان السماء. قبل أن الأسباب، من المفترض المسلمين على تعلم اللغة العربية منذ سن مبكرة، بحيث يسمح له لفهم الدين تدريس. أساسيا، في تعلم اللغة العربية، ونحن يجب أن يعرف الكلمات الأساسية للجملة العربية، لأن من الكلمات الأساسية، ونحن يمكن معرفة موقف ومعنى الجمل. في اللغة العربية هناك كلمة واحدة تسمى "شرف" هي أنواع الأفعال، مثل أن تكون في نطاق هذه الأطروحة، "فعل مضارع"، "فعل ماضي" و "مسدر". وإيجاد جذر الكلمة الأساسية، ويتطلب دقة. في المسلمين قلة الوحيد في هذا العالم لا تعرف الكلمات الأساسية من اللغة العربية التي يقرأ كل يوم، لذلك إعطاء أخطاء بمعنى من الأصل. خطابات الحروف العربية باللغة العربية، له معنى، حتى إذا بت واحد من تحديد قاعدة خاطئ، وسوف دمر كلمات كاملة، ولا يمكن أن تفسر. لذلك بدافع الباحثون إلى إجراء دراسة مقارنة لتطبيق ضوء النابعة دقة ووقف خوجة على ماضي فعل ومسدر مما يسهل على المستخدمين لتحديد دقة تشوه فعل مضارع "إلى ماضي فعل و مسدر. وبناء على اختبار من 44 الجملة العربية يمكن الاستنتاج أن التطبيقات المنبثقة عن طريق تطبيق ضوء وقف و حوارزمية خوجة قادرة على تشكيل الكلمات الأساسية. من نتائج الاختبار يمكن أن تؤخذ دقة في المئة مع ضوء النابعة خوجة 93٪، في حين أن 20٪ في المائة. ويتسبب الخطأ لأن هناك الكلمة التي تشهد ويتسبب الخطأ لأن هناك الكلمة التي تتوافر لديها الخبرة فوق أو تحت الناجمة الناجمة. وكلمة الأساسية لم يتم الكشف وليس هناك القاموس آل منور. لمزيد من التطوير، تحتاج إلى أن يكون الفعل المضافة (فعل) التي يمكن أيضا تعلم الكلمات الأساسية من مختلف فعل.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Menurut Mujtahid (2006:1) , bahasa adalah alat komunikasi yang sangat vital bagi manusia. Bahasa dipakai untuk menghubungkan perbedaan, persamaan serta berbagai dialektika peradaban dari zaman kuno hingga zaman sekarang. Tanpa bahasa seolah-olah dunia ini terasa gelap gulita. Begitu pentingnya bahasa, Wittgenstein menyebutnya sebagai paradigma.

Salah satu bahasa yang sangat populer di kalangan umat Islam adalah bahasa Arab karena bahasa ini adalah bahasa Quraniah yang dipakai dalam ibadah sehari-hari seperti tilawah Quran, Sholat, dan juga berdoa pada sang khalik Allah SWT. Selain itu, bahasa Arab merupakan pedoman bahasa pada dua pedoman umat Islam (al-Quran dan al-Hadits) bahasa para nabi dan bahasa para penghuni surga (Abu Hamzah Yusuf al-Atsary, 2007). Maka sudah selayaknya umat Islam mempelajari bahasa Arab sejak dini agar memudahkan dirinya dalam memahami ajaran agamanya (Nurlailah, 2006).

Allah SWT memberikan sifat kepada al-Quran sebagai kitab yang mulia, yang di dalam al-Quran tersebut tidak ada kebatilan. Hal ini sebagaimana yang dicantumkan dalam ayat al-Quran, surat Fushshilat(41):41-42, :

إِنَّ الَّذِينَ كَفَرُوا بِالذِّكْرِ لَمَّا جَاءَهُمْ وَإِنَّهُ لَكِتَابٌ عَزِيزٌ ﴿٤١﴾ لَا يَأْتِيهِ الْبَطْلُ مِنْ بَيْنِ يَدَيْهِ

وَلَا مِنْ خَلْفِهِ ۖ تَنْزِيلٌ مِّنْ حَكِيمٍ حَمِيدٍ ﴿٤٢﴾

“Dan sesungguhnya Al-Quran itu adalah kitab yang mulia. Yang tidak datang kepadanya (Al-Quran) kebatilan baik dari depan maupun dari belakangnya, yang diturunkan dari Tuhan Yang Maha Bijaksana lagi Maha Terpuji”. (Fushshilat: 41-42)

Bahasa Arab adalah bahasa Agama Islam, al-Quran sebagai sumber pengetahuan dan doa dalam islam menggunakan bahasa Arab. Namun, kegiatan agama yang kita lakukan kurang begitu sempurna karena kita tidak memahami arti dari apa yang kita baca. Oleh karena itu, harus ada sebuah pemahaman yang ditanamkan dalam diri untuk menyempurnakan ibadah umat muslim.

Hal ini diperjelas dengan firman Allah SWT dalam surah Yusuf yang berbunyi :

إِنَّا أَنْزَلْنَاهُ قُرْءَانًا عَرَبِيًّا لَعَلَّكُمْ تَعْقِلُونَ ﴿٢﴾

Artinya :

“Sesungguhnya Kami menurunkannya berupa Al Quran dengan berbahasa Arab, agar kamu memahaminya.” (QS Yusuf : 2).

Dalam belajar bahasa, kita tidak harus mengetahui arti kata bahasa tersebut, tetapi ada juga tata bahasa yang kita perhatikan. Dalam bahasa Inggris dikenal dengan istilah *Grammar* sedangkan dalam bahasa Arab dikenal dengan istilah *Nahwu Sharaf*.

Ilmu *Sharaf* merupakan salah satu dari ilmu *grammar* bahasa Arab *adabiyatul arabiyah*. Secara bahasa ilmu *Sharaf* diartikan sebagai

“Perubahan”. Mengetahui *Sharaf* atau mengkaji bentuk perkataan sangat penting dalam mempelajari bahasa Arab karena dengan perubahan bentuk suatu perkataan, maka perkataan itu akan berubah. Dengan *Sharaf* tersebut, kita mampu mengetahui kata dasar dari sebuah kata.

Pada dasarnya, dalam belajar bahasa Arab, harus mengetahui kata dasar dari suatu kalimat bahasa Arab tersebut, karena dari kata dasar tersebut, kita dapat mengetahui kedudukan dan arti dari kalimat Arab. Pada *sharaf* terdapat jenis-jenis kata kerja, diantaranya adalah yang menjadi cakupan dalam skripsi ini *fi'il mudhori'*, *fi'il madhi* dan *mashdar*. Dan untuk mencari kata dasar dari kata Arab tersebut, membutuhkan keakurasian. Tidak sedikit umat Islam di dunia ini tidak mengetahui kata dasar dari kata Arab yang dibacanya sehari-hari, sehingga menimbulkan kesalahan dalam arti yang sesungguhnya. Huruf per huruh hijaiyah dalam kata Arab, mempunyai arti, sehingga salah sedikit menentukan kata dasar, kata dasar tersebut salah dan tidak dapat diartikan.

Berdasarkan permasalahan tersebut maka peneliti termotifasi membuat aplikasi studi perbandingan akurasi *light stemming* dan *khoja stemming* pada *fi'il madhi* dan *mashdar* sehingga memudahkan pengguna dalam mengetahui ketepatan perubahan bentuk dari *fi'il mudori'* menjadi *fi'il madhi* dan *mashdar*.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah bagaimana studi perbandingan akurasi *light stemming* dan *khoja stemming* pada *fi'il madhi* dan *mashdar* dengan inputan *fi'il mudhori'* yang ada pada Al Quran Juz 30?

## 1.3. Batasan Masalah

Dalam penyusunan skripsi ini, penyusun perlu untuk membatasi masalah yang akan dibahas. Adapun masalah yang dibatasi oleh penyusun adalah sebagai berikut :

1. Aplikasi yang dibangun adalah aplikasi studi perbandingan akurasi *light stemming* dan *khoja stemming* pada *fi'il madhi* dan *mashdar*.
2. Bahasa pemrograman yang digunakan peneliti adalah *Netbeans* dengan database *MySQL*.
3. *Fi'il mudhori'* yang terdapat di Al Quran juz 30 .

## 1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk:

- a. Mengukur keakurasian stemming menggunakan *light stemming* dan *khoja stemming* pada *fi'il madhi* dan *mashdar*.

## 1.5. Manfaat Penelitian

Manfaat yang dapat diambil dengan adanya penelitian ini antara lain sebagai berikut :

- a. Menambah referensi tentang penggunaan bahasa Arab.
- b. Mendukung pembelajaran masyarakat luas tentang penggunaan kata bahasa Arab.
- c. Dengan adanya aplikasi ini dapat membantu orang awam untuk mengetahui kata dasar dalam bahasa Arab.

## 1.6. Metodologi Penelitian

Dalam penelitian ini digunakan metode sebagai berikut :

### 1. Pengumpulan data dan studi literature

Pada tahap ini dilakukan pencarian dan pemahaman literatur serta pengumpulan informasi dengan cara membaca buku referensi yang dapat dijadikan pedoman pembahasan dalam masalah ini.

### 2. Perumusan Masalah dan penyelesaiannya

Tahap ini meliputi perumusan masalah, batasan-batasan masalah dan penyelesaiannya serta penentuan parameter untuk mengukur hasilnya.

### 3. Perancangan dan desain aplikasi

Pada tahap ini akan dilakukan perancangan proses-proses utama desain mengenai aplikasi stemming kata Arab berbasis *fi'il madhi* dan *mashdar* yang akan dibangun berdasarkan teori yang telah dipahami.

### 4. Implementasi sistem

Pada tahap ini akan dilakukan pembangunan aplikasi stemming kata Arab, yang mana dalam pembangunan aplikasi tersebut akan

diterapkan teori/algorithm yang telah dipelajari yaitu *light stemming* dan *khoja*. Perancangan dan desain aplikasi ini diimplementasikan dengan bahasa pemrograman Java dan database *MySQL*. Aplikasi dibangun dengan IDE NetBeans 7.2 untuk mempermudah desain antarmuka dan database yang digunakan adalah *MySQL*.

5. Pengujian aplikasi

Uji coba dilakukan sampai sistem benar – benar *ready to use*, kekurangan yang terjadi diperbaiki dalam lingkup batasan masalah. Evaluasi dilakukan untuk mengetahui apakah sistem yang dibangun sudah sesuai yang diharapkan.

6. Penyusunan laporan

Penyusunan laporan akhir merupakan dokumentasi dari keseluruhan pelaksanaan penelitian dan diharapkan bermanfaat bagi penelitian lebih lanjut.

## 1.7. Sistematika Penulisan

### **BAB I Pendahuluan**

Bab ini berisi latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi dan sistematika penyusunan laporan tugas akhir . Uraian di bab ini pendahuluan ini memberikan gambaran kepada pembaca terkait maksud dan tujuan dalam penelitian ini.

### **BAB II Landasan Teori**

Bab ini menjelaskan konsep dan teori dasar yang mendukung penulisan tugas akhir ini. Adapun yang dibahas dalam bab ini adalah dasar teori yang berkaitan dengan pembahasan terkait judul penelitian, yaitu perbandingan stemming kata arab berbasis fiil madhi dan *mashdar* menggunakan algoritma *light stemming* dan *khoja*. Mulai dari penjabaran mengenai bahasa Arab, yang dalam penelitian ini digunakan sebagai objek penelitian, dan tahapan-tahapan dalam text minning.

### **BAB III Analisis dan Perancangan Aplikasi**

Bab ini menjelaskan mengenai analisis dan perancangan aplikasi stemming kata Arab berbasis *fi'il madhi* dan *mashdar* menggunakan algoritma *light stemming* dan *khoja*. Perancangan aplikasi terdiri atas perancangan proses-proses utama dan desain aplikasi.

### **BAB IV Hasil dan Pembahasan**

Bab ini berisi tentang implementasi dari aplikasi *stemming* kata arab yang mampu *stemming* dengan menggunakan metode *light stemming* dan *khoja*. Perancangan dan desain aplikasi diimplementasikan dengan bahasa pemrograman Java dan database *mySQL*. Aplikasi ini dibangun dengan *IDE NetBeans 7.0.1* untuk mempermudah desain antarmuka dan database yang digunakan adalah *MYySQL*. Disini juga dipaparkan pengujian terhadap aplikasi yang dibuat.

## **BAB V**

Bab ini berisi kesimpulan dan saran-saran perbaikan yang berhubungan dengan masalah yang dibahas untuk pengembangan lebih lanjut.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Penelitian Terkait

Berikut ini adalah beberapa penelitian terdahulu yang berkaitan dengan penelitian yang akan dilakukan :

*Aplikasi Stemmer Fi'il dalam Jus 'Amma Menggunakan Porter Stemmer* oleh Nurlailah (2009). Hasil dari penelitian tersebut adalah bagaimana suatu aplikasi dapat memotong kata berimbuhan menjadi kata dasar dalam Bahasa Arab sehingga dapat memudahkan user dalam pencarian arti suatu kata Bahasa Arab.

*Analisis Dan Implementasi Stemming Teks berbahasa Indonesia Dengan Menggunakan Porter Stemmer* oleh Ni Nyoman Budiasih(2009). Penelitian tersebut menghasilkan setelah dilakukan modifikasi, dilakukan perbandingan nilai keakuratan dari stem yang dihasilkan oleh Porter stemmer dan gabungan porter stemmer (stemmer hasil modifikasi Porter stemmer). Hasil dari pengujian menunjukkan bahwa *stemming* yang menggunakan gabungan porter stemmer menghasilkan keakuratan stem yang lebih baik dibandingkan dengan menggunakan Porter stemmer. Sehingga dapat disimpulkan bahwa modifikasi yang dilakukan dapat meningkatkan keakuratan stem.

*Analisis Stemming pada Information Retrieval System dengan Algoritma Porter dan Krovetz* oleh Rini Riandha Asri (2010). Dari penelitian tersebut

menghasilkan *stemming* dalam sistem temu kembali informasi digunakan untuk membatasi varian bentuk kata yang berbeda menjadi bentuk dasarnya, sehingga nantinya dapat meningkatkan kemampuan sistem dalam menemukan dokumen relevan sesuai *query* yang ada. Dalam tugas akhir ini, dibuat sebuah sistem temu kembali informasi yang mengimplementasikan teknik *stemming* dengan menggunakan algoritma *Porter dan Krovetz*).

*Pencarian ayat-ayat Al-Quran Berdasarkan Konten Menggunakan Text mining Berbasis Aplikasi Dekstop* oleh *Aditya Herdianto* (2012). Proses pencarian ayat - ayat Al Quran dengan cara konvensional ataupun dengan Al-Quran digital yang selama ini ada di internet tidak cukup membantu, jika hasil yang kita inginkan adalah ayat - ayat tertentu yang sesuai dengan masalah yang kita hadapi. Dengan begitu dibutuhkan sebuah sistem untuk mengenali, mencari dan mengelompokkan masalah yang dibutuhkan oleh *user*. Sehingga sistem tersebut dapat menampilkan ayat - ayat Al-Quran sebagai referensi dan solusi. Dengan masalah diatas, maka dapat digunakan sebuah proses pengenalan teks yang disebut *Text mining*. Dengan proses tersebut maka masalah yang dibutuhkan *user* dilakukan dengan beberapa metode yaitu, *parsing*, *stemming*, dan *morphing*. Sehingga dapat mengenali masalah yang dibutuhkan oleh *user* dan ayat - ayat yang berhubungan dengan masalah *user* dapat langsung ditampilkan.

Penelitian tentang penerjemah kalimat tunggal Bahasa Indonesia ke dalam Bahasa Jawa menggunakan context-free grammer. Tujuan dari penelitian ini untuk membantu masyarakat luas mempelajari bahasa Jawa lebih mudah serta memberikan referensi tambahan tentang penggunaan bahasa Jawa. Proses yang

terjadi dalam penelitian yakni diawali dengan inputan kalimat oleh user kemudian dilanjutkan dalam tahap preprocessing, analisis sintaks menggunakan aturan context free grammar, stemming kata dengan algoritma porter, analisa semantiks sesuai dengan aturan struktur kalimat yang ada dan yang terakhir proses penerjemahan dalam kalimat Bahasa Jawa. Penelitian ini menerapkan context free grammar sebagai aturan dalam menganalisa kalimat yang masuk. Dari hasil uji coba ini dapat diambil persentase keberhasilan analisis sebesar 76% dengan jumlah struktur benar sebanyak 38 kalimat. Program juga dapat menerjemahkan kalimat bahasa Indonesia dari hasil analisa sintaks, dengan persentase keberhasilan sebesar 55%. Persentase ini didapat dari 21 kalimat yang dapat diterjemahkan sesuai hasil dari sumber data pada buku praktis Jawa-Indonesia disertai contoh, dibagi dengan jumlah kalimat yang strukturnya benar. Kesalahan yang sering terjadi dalam proses penerjemah disebabkan kurangnya kosakata yang ada di dalam database sistem. Selain itu kurangnya aturan atau imbuhan pada kata terjemahan bahasa Jawa (Choiroh, 2011).

Aplikasi ini hanya menerjemahkan kalimat yang sesuai dengan aturan struktur kalimat yang ada. Jika kalimat tersebut tidak sesuai dengan aturan struktur kalimat maka kalimat tersebut akan muncul peringatan atau kalimat tersebut tidak dapat diterjemahkan ke dalam kalimat bahasa Jawa. Untuk aturan struktur kalimat dapat dilihat pada Gambar 2.1.



Gambar 2.1. Aturan struktur kalimat

Dan pada penelitian yang meneliti tentang kamus elektronik kalimat Bahasa Indonesia. Terdapat 2 mode kamus, yaitu kamus kata kata dan kamus kalimat. Penelitian ini menggunakan Markov Model. Proses penerjemahan kalimat dimulai dari penguraian kalimat dengan parsing tree, proses markov, pencarian kata dengan binary search, dan penggabungan kata menjadi kalimat kembali. Dari hasil pengujian penggunaan markov dan binarysearch dengan data kamus 775 kata, pada data 775 pada kamus kata mempercepat waktu pencarian, diperlukan waktu rata-rata 56,67 $\mu$ s, dan jika tanpa markov memerlukan waktu 102,36  $\mu$ s. Sedangkan penggunaan markov pada binary search pada kamus kalimat (4 kata) diperlukan waktu rata-rata 264.16 ms, dan jika tanpa markov memerlukan waktu 53.96 ms. Sedangkan dengan data kamus 1530 kata (kamus Jawa- Indonesia) penggunaan markov dan binary search pada kamus kata juga mempercepat waktu

pencarian, diperlukan waktu rata-rata 69.08  $\mu$ s, dan jika tanpa markov memerlukan waktu 154.98  $\mu$ s. Sedangkan penggunaan markov pada binary search pada kamus kalimat (4 kata) diperlukan waktu rata-rata 402.48 ms, dan jika tanpa markov memerlukan waktu 155.84 ms. Pada kamus kalimat penggunaan markov tidak mempengaruhi kecepatan proses searching, hal ini dipengaruhi beberapa faktor diantaranya adalah jumlah kata pada kamus terlalu sedikit sehingga fungsi markov kurang terlihat, dan pada kamus kalimat diperlukan lebih dari 1 kali proses markov untuk tiap kata, sehingga memerlukan waktu lebih banyak daripada tidak menggunakan markov (Afifah, 2012).

Penelitian yang meneliti tentang stemming dokumen teks bahasa Indonesia. Penelitian ini dilakukan dengan membandingkan antara algoritma Porter dengan algoritma Nazief dan Adriani, menghitung presisi dan waktu proses serta menguji menggunakan 30 sampel. Proses perbandingan algoritma Porter dengan algoritma Nazief & Adriani dilakukan dengan membuat program sederhana yang memproses dokumen teks inputan sehingga diketahui stem, waktu proses, presisi dari hasil stemming dokumen. Hasil dari penelitian yaitu proses stemming dokumen teks bahasa Indonesia menggunakan algoritma porter membutuhkan waktu yang lebih singkat dibandingkan dengan stemming menggunakan algoritma Nazief dan Adriani. Selain itu stemming menggunakan algoritma porter memiliki prosentase keakuratan (presisi) lebih kecil dibandingkan dengan stemming algoritma Nazief dan Adriani (Agusta, 2009).

Selanjutnya penelitian yang meneliti tentang stemming sinonim bahasa Indonesia. Penelitian ini menggunakan metode pendekatan teknik stemming yang

dibuat oleh Bobby Nazief dan Mirna Adriani. Data yang diambil untuk penelitian ini adalah kumpulan kata kerja dan kata benda yang berhubungan dengan teknologi. Peneliti mengambil 100 kata dasar yang akan dipasangkan dengan jenis imbuhan. Jenis imbuhan tersebut terdiri dari : awalan, sisipan, akhiran, dan Awalan-Akhiran. Lalu menghitung berapa tingkat keberhasilan dari program yang dibuat oleh peneliti dalam mengubah kata berimbuhan menjadi bentuk dasar (kata dasar) (Mario, 2011).

Penelitian tentang bentuk-bentuk afiks verba pada bahasa Jawa dan bahasa Indonesia. Penelitian ini bertujuan untuk mengetahui makna yang terkandung di dalam verba bahasa Jawa dan bahasa Indonesia. Penelitian ini menganalisis bahasa untuk menunjukkan persamaan dan perbedaan dari bentuk-bentuk afiks verba bahasa Jawa dan bahasa Indonesia. Persamaan dan perbedaan itu meliputi dua aspek, yaitu adanya kesejajaran bentuk afiks verba dan bentuk dasar yang dilekati oleh afiks tersebut. Dalam penelitian ini, bahasa Jawa akan menjadi kajian peneliti merupakan bahasa yang serumpun dengan bahasa Melayu sehingga bahasa ini memiliki persamaandengan unsur-unsur yang terdapat di dalam bahasa Indonesia. Di samping persamaan juga ada perbedaannya (Krishandini, 2011).

Selain itu penelitian tentang sistem morfemis nomina bahasa Jawa dan bahasa Indonesia. Bentuk dasar yang dilekati oleh bahasa Jawa dan bahasa Indonesia mempunyai bentuk yang sama. Jika yang dilekati oleh afiks bahasa Jawa berupa kelas verba, dalam bahasa Indonesia pun kelas verba. Bentuk dasar yang dapat dilekati oleh nomina itu adalah verba, nomina, adjektiva, adverbial, dan pokok kata (praktegorial). Di dalam proses morfofonemik terdapat adanya kekhasan

sistem penggabungan bunyi vokal yang dalam bahasa Indonesia hal itu tidak akan terjadi secara morfemis. Selain itu, terjadi kekontrasan pula adanya proses penambahan bunyi, penghilangan bunyi (Sukarto, 2010).

## 2.2 Pengolahan Bahasa Alami

Bahasa alami atau bahasa natural adalah suatu bahasa yang diucapkan, ditulis, atau diisyaratkan (secara visual atau isyarat lain) oleh manusia untuk komunikasi umum. Natural language adalah bahasa yang dapat dipahami dan dimengerti oleh individu pada lingkungan tertentu.

*Natural Language Processing* atau pemrosesan bahasa alami merupakan salah satu cabang dari *Artificial Intelligence* (kecerdasan buatan) yang memiliki kemampuan untuk memahami bahasa manusia. Pada prinsipnya bahasa alami adalah suatu bentuk representasi dari suatu pesan yang ingin dikomunikasikan antar manusia. Bentuk utama representasinya adalah berupa suara/ucapan, tetapi sering pula dinyatakan dalam bentuk tulisan. Inti dari pemrosesan bahasa alami adalah penguraian kalimat yang berfungsi untuk membaca kalimat atau kata serta menentukan jenis kata apa saja yang boleh mengikuti kata tersebut.

Pengolahan bahasa alami (*Natural Language Processing*) merupakan kemampuan suatu komputer untuk memproses bahasa baik lisan maupun tulisan yang digunakan oleh manusia dalam percakapan sehari-hari. Kesulitan yang sering terjadi dalam pemrosesan bahasa alami yakni adanya makna ganda (*ambiguitas*) atau jumlah kosa kata dalam bahasa alami mengalami perkembangan dari waktu ke waktu.

### 2.3 Bidang Pengetahuan Dalam Bahasa Natural

Secara singkat pengolahan bahasa alami (natural language processing) mengenal beberapa tingkatan pengolahan yaitu (Savitri, 1996) :

1. Fonetik dan fonologi, berhubungan dengan suara yang menghasilkan kata yang dapat dikenali. Bidang ini menjadi penting dalam proses aplikasi yang memakai metode *speech based system*.
2. Morfologi, yaitu pengetahuan tentang kata dan bentuknya dimanfaatkan untuk membedakan satu kata dengan lainnya. Pada tingkat ini juga dapat dipisahkan antara kata dan elemen lain seperti tanda baca.
3. Sintaktis, yaitu pemahaman tentang urutan kata dalam pembentukan kalimat dan hubungan antar kata tersebut dalam proses perubahan bentuk dari kalimat menjadi bentuk yang sistematis. Meliputi proses pengaturan tata letak suatu kata dalam kalimat akan membentuk kalimat yang dapat dikenali. Selain itu dapat pula dikenali bagian-bagian kalimat dalam suatu kalimat yang lebih besar.
4. Semantik, yaitu pemetaan bentuk struktur sintaksi dengan memanfaatkan tiap kata ke dalam bentuk yang lebih mendasar dan tidak tergantung struktur kalimat. Semantik mempelajari arti suatu kata dan bagaimana arti dari kata tersebut membentuk suatu arti dari kalimat yang utuh. Dalam tingkatan ini belum tercakup konteks dari kalimat tersebut.

5. Pragmatik, pengetahuan pada tingkatan ini berkaitan dengan masing-masing konteks yang berbeda tergantung pada situasi dan tujuan pembuatan sistem
6. *Discourse knowledge*, melakukan pengenalan apakah suatu kalimat yang sudah dibaca dan dikenali sebelumnya akan mempengaruhi arti dari kalimat selanjutnya. Informasi ini penting diketahui untuk melakukan pengolahan arti terhadap kata ganti orang dan untuk mengartikan aspek sementara dari informasi.
7. *World knowledge*, mencakup arti sebuah kata secara umum dan apakah ada arti khusus bagi suatu kata dalam suatu percakapan dengan konteks tertentu.

Definisi ini tidaklah bersifat baku, dan untuk setiap bentuk bahasa alami yang ada biasanya ada pendefinisian lagi yang lebih spesifik sesuai dengan karakter bahasa tersebut. Pada beberapa masalah mungkin hanya mengambil beberapa dari pendekatan tersebut bahkan mungkin ada yang melakukan tambahan proses sesuai dengan karakter dari bahasa yang digunakan dan sistem yang dibentuk.

#### **2.4 Penelusuran Informasi (*Informasi Retrieval*)**

Pada prinsipnya penelusuran informasi merupakan sebuah proses pengidentifikasian, pencarian, penyediaan dan pemberian informasi atas kebutuhan atau permintaan pemakai unit informasi. Penelusuran informasi merupakan kegiatan yang bertujuan untuk menyediakan dan memasok informasi bagi pemakai sebagai jawaban atas permintaan atau berdasarkan kebutuhan pemakai.

*Information Retrieval* merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. *Information retrieval* merupakan suatu pencarian informasi (biasanya berupa dokumen) yang didasarkan pada suatu *query* (*inputan user*) yang diharapkan dapat memenuhi keinginan *user* dari kumpulan dokumen yang ada. Sedangkan, definisi *query* merupakan sebuah formula yang digunakan untuk mencari informasi yang dibutuhkan oleh *user*, dalam bentuk yang paling sederhana, sebuah *query* merupakan suatu *keywords* (kata kunci) dan dokumen yang mengandung *keywords* merupakan dokumen yang dicari dalam *IRS* (Mustaqim, 2009 : 1).

Menurut Lancaster dalam Ratu Siti Zaenab (2002 : 41), Sistem temu kembali informasi dapat didefinisikan sebagai suatu proses pencarian dokumen dengan menggunakan istilah – istilah pencarian dokumen dengan menggunakan istilah – istilah pencarian untuk mendefinisikan dokumen sesuai dengan subjek yang diinginkan.

Menurut Zainal A. Hasibuan (1997 : 2), *Indexing* merupakan sebuah proses untuk melakukan pengindeksan terhadap kumpulan dokumen yang akan disediakan sebagai informasi kepada pemakai. Adapun tahapan dari pengideksan adalah sebagai berikut :

a. *Parsing* dokumen yaitu proses pembuangan kata – kata dari kumpulan dokumen.

- b. *Stoplist* yaitu proses pembuangan kata buang seperti: tetapi, yaitu, sedangkan, dan sebagainya.
- c. *Stemming* yaitu proses penghilangan/pemotongan dari suatu kata menjadi bentuk dasar. Kata “beradaptasi” menjadi kata “adaptasi”
- d. *Term weighting* dan *Inverted File* yaitu proses pemberian bobot pada istilah.

## 2.5 Text mining

Text mining memiliki definisi menambang data yang berupa text dimana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen (Raymond, 2006). Dengan text mining tugas-tugas yang berhubungan dengan penganalisaan teks dengan jumlah yang besar, penemuan pola serta penggalian informasi yang mungkin berguna dari suatu teks dapat dilakukan. Proses text mining dibagi menjadi 3 tahap utama, yaitu proses awal terhadap teks (text preprocessing), transformasi teks ke dalam bentuk antara (text transformation/feature generation), dan penemuan pola (pattern discovery) (Even dan Zohar: 2002).

### 2.5.1 Text Preprocessing

Tahapan awal dari text mining adalah text preprocessing yang bertujuan untuk mempersiapkan teks menjadi data yang akan diproses pada tahapan berikutnya. Terdapat beberapa hal yang dilakukan dalam tahapan ini, baik itu berupa tindakan yang bersifat kompleks seperti part-of-speech (pos), tagging,

parse tree, maupun tindakan sederhana seperti tokenization, yaitu tugas memisahkan deretan kata di dalam kalimat, paragraf atau halaman menjadi token atau potongan kata tunggal atau termmed word. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua token ke bentuk huruf kecil (case folding) (Husni).

### 2.5.2 Text transformation

Pada tahap ini dilakukan penyaringan (filtration). Penyaringan dilakukan dengan menentukan term mana yang akan digunakan untuk merepresentasikan dokumen sehingga dapat mendiskripsikan isi dokumen dan membedakan dokumen tersebut dengan dokumen lain dalam koleksi. Term yang sering dipakai tidak dapat digunakan untuk tujuan ini, setidaknya karena dua hal. Pertama, jumlah dokumen yang relevan terhadap suatu query kemungkinan besar merupakan bagian kecil dari koleksi. Term yang efektif dalam pemisahan dokumen yang relevan dari dokumen tidak relevan kemungkinan besar adalah term yang muncul pada sedikit dokumen. Ini berarti bahwa term dengan frekuensi kemunculan tinggi bersifat poor discriminator. Kedua, term yang muncul dalam banyak dokumen tidak mencerminkan definisi dan topik atau sub-topik dokumen. Karena itu, term yang sering digunakan dianggap sebagai stop-word dan dihapus. (Husni)

Stop-word didefinisikan sebagai term yang tidak berhubungan (irrelevant) dengan subjek utama dari database meskipun kata tersebut sering kali hadir di dalam dokumen (Cios, 2007). Stopword merupakan kata-kata yang bukan

merupakan ciri (kata unik) sehingga dengan menghilangkannya dari suatu teks maka sistem hanya akan memperhitungkan kata-kata yang dianggap penting. Penghapusan stop-word dari dalam suatu koleksi dokumen pada satu waktu membutuhkan banyak waktu. Solusinya adalah dengan menyusun suatu pustaka stop-word atau stop-list dari term yang akan dihapus. (Husni)

Konversi term ke bentuk akar (stemming) juga merupakan tindakan yang dapat dilakukan pada tahap ini. Menurut Tala (2003) stemming merupakan proses untuk mereduksi kata ke bentuk dasarnya. Kata-kata yang muncul di dalam dokumen sering mempunyai banyak varian morfologik. Karena itu, setiap kata yang bukan stop-words direduksi ke bentuk stemmed word yang cocok. Dengan cara ini, diperoleh kelompok kata yang mempunyai makna serupa tetapi berbeda wujud sintaksis dari dengan lainnya. Kelompok tersebut dapat direpresentasikan oleh satu kata tertentu.

### **2.5.3 Pattern discovery/Analysis**

Tahap ini merupakan tahap terpenting dari seluruh proses text mining. Ada beberapa macam operasi yang dapat dilakukan pada tahap ini, diantaranya:

- a. Categorization/classification/supervised learning
- b. Clustering/unsupervised learning
- c. Visualization
- d. Summarization

#### e. Trends analysis

Teknik yang digunakan pada tahap ini adalah dengan melakukan pembobotan (weighting) terhadap term dari hasil tahap text transformation. Setiap term diberikan bobot sesuai dengan skema pembobotan yang dipilih, baik itu pembobotan lokal, global atau kombinasi keduanya. Dan dalam kemiripan antar dokumen didefinisikan berdasarkan representasi bag-of-words dan dikonversi ke suatu model ruang vektor (vector space model, VSM). Model ini diperkenalkan oleh Salton (Salton, 1983) dan telah digunakan secara luas.

## **2.6 Bahasa Arab**

### **2.6.1 Sejarah Bahasa Arab**

Menurut pendapat ahli bahasa, bahasa Arab adalah merupakan salah satu rumpun bahasa Semit Selatan. Sedangkan bahasa Semit adalah bahasa yang berakar dari bahasa yang dipakai oleh keturunan Nabi Nuh. Lebih dari itu, mereka juga berpendapat, bahwa untuk mengkaji tentang sejarah bahasa Arab, dimana sebelum datangnya agama Masehi (Abad ke-1 Masehi), para ahli belum memperoleh gambaran apapun mengenai bahasa Arab, karena pada saat itu belum adanya prasasti atau peninggalan yang dapat diperoleh sebagai bukti munculnya bahasa Arab, dengan kata lain masih misteri.

Prasasti tentang bahasa Arab baru terungkap dengan ditemukannya ukiran-ukiran tulisan yang bernama “al-Nimarah” di dekat kota Damaskus yang bertanda tahun 328 M. Walaupun ditemukan prasasti yang mengungkap misteri sejarah tentang asal-usul dan kapan dimulainya adanya bahasa Arab. Tetapi, masih ada diantara para ahli yang masih bersikap skeptis dengan pemikiran

tersebut untuk dijadikan kepastian telah adanya bahasa Arab pada masa itu, karena sebagian besar kata-kata yang terukir dalam prasasti tersebut hanyalah nama-nama orang saja. Menurut pendapat mereka, sebaiknya untuk menentukan asal-usul dan kapan dimulainya ada bahasa Arab, kiranya cukup berpedoman kepada teks-teks dari sastra Jahiliyah yang tidak diragukan kebenarannya untuk menjelaskan keadaan-keadaan bahasa Arab sebelum datangnya agama Islam.

Dari kedua pendapat tersebut di atas, tentang asal usul bahasa Arab tersebut sebenarnya tidak terlalu berbeda. Adapun pendapat yang menjadikan ukiran al-Nimarah yang bertanda tahun 328 M. dan pendapat yang menjadikan teks-teks Jahiliyah sebagai patokan awal adanya atau dimulainya bahasa Arab yang seperti kita kenal adalah sama atau berdekatan tahunnya. Eksistensi ukiran al-Nimarah bertanda tahun 328 M, sedangkan Nabi Muhammad SAW dilahirkan tahun 571 M. sementara itu syair-syair Jahiliyah yang dikenal dalam Nushush al-Adab al-‘Arabi, pengarangnya hidup antara abad ke-4 dan ke-6. Dengan demikian, perbedaan tahun tidak perlu lagi dipermasalahkan. Yang tinggal hanya perbedaan tentang keberadaan ukiran al-Nimarah sebagai prasasti yang menunjukkan telah adanya bahasa Arab pada saat itu, sementara prasasti al-Nimarah hanya memuat nama-nama orang saja. Oleh karena itu, orang-orang yang meragukan telah adanya bahasa Arab saat-saat itupun tidak dapat disalahkan, dan juga pendapat yang menjadikan al-Nimarah sebagai indikasi telah adanya bahasa Arab pada saat itu, juga tidak dapat disangkal kebenarannya.

Masyarakat Semenanjung Arab sejak dahulu kala sudah terbagi ke dalam beberapa kabilah, suku bani atau garis keturunan, dan lain-lain. Masing-masing

dari mereka itu sangat menonjolkan tradisi, membanggakan suku, fanatik ‘ashabiyah dan tidak ketinggalan pula bersaing dalam logat dan dialek antara kabilah-kabilah yang ada.

Situasi dan kondisi seperti itu berlangsung cukup lama sampai menjelang datangnya agama Islam. Namun sejak mereka berkepentingan untuk lebih banyak berkomunikasi di musim-musim haji dan suatu kepentingan untuk menyelenggarakan pekan raya di Ukaz dan Zulmajaz, maka mereka mulai merasakan kebutuhan adanya alat untuk saling mengerti bagi semua kabilah. Maka dalam pekan raya tersebut, mereka harus menjauhkan dari ciri-ciri lokal yang berkenaan dengan dialek. Mereka berusaha untuk menggunakan bahasa yang dapat dipahami oleh semua pihak.

Karena dalam pekan raya, justru diadakan perlombaan bersyair yang diikuti oleh berbagai suku yang beraneka-ragam dialek adalah merupakan kegiatan yang paling populer dan menonjol. Dengan keadaan yang demikian, maka semakin kokoh dan kuatlah keinginan untuk mencari solusi dari perbedaan bahasa dan dialek tersebut. Pada akhirnya, terbentuklah suatu bahasa Arab kesusasteraan yang menjadi bahasa Arab standar.

Bahasa Arab standar tersebut berasal dari dialek Quraisy yang disempurnakan oleh dialek suku-suku lain. Kontribusi dialek suku Quraisy menjadi bahasa standar yang berkaitan dengan perkembangan kota Makkah yang menjadi pusat kegiatan ibadah haji, dan di kota Makkah inilah kebanyakan suku Quraisy menetap.

Dalam perkembangan selanjutnya, bahasa standar tersebut menjadi bahasa sastra jahili sekaligus menjadi bahasa yang digunakan oleh hampir seluruh lapisan masyarakat dalam waktu yang singkat. Pesatnya perkembangan bahasa standar tersebut disebabkan oleh munculnya para ahli dan para genius dari tiap-tiap kabilah yang mampu berbicara dengan fasih, serta mampu menyusun syair-syair dalam sistematika bahasa yang sangat bermutu.

Adapun pengertian bahasa Arab standar adalah bahasa Arab yang digunakan oleh umat Islam dalam bentuk teks-teks klasik sastra jahili bahasa Arab yang dipakai oleh Al-Quran dan Hadis serta dipakai untuk menulis cabang-cabang ilmu agama, seperti fiqih, tafsir, ilmu kalam dan lain-lain.

Akhirnya asal-usul bahasa Arabpun ditemukan, ini berdasarkan penelitian dari ditemukannya sebuah prasasti. Prasasti ini diperkirakan dibuat antara tahun 469-470 Masehi. Ini merupakan tahun tertua dari tulisan Arab yang ditemukan, sehingga disebut “bagian yang hilang” antara Nabatean dan tulisan Arab.

Tim arkeolog dari Museum Louvre, Perancis, mengumumkan telah menemukan Epigrafi atau prasasti yang menggunakan alfabet Nabatean atau akar dari alfabet Arab. Dilansir dari Emirates 247, Juru Bicara Museum Louvre, Christelle Guyader mengatakan Prasasti Nabatean Arab ini ditemukan sekitar 100 km di Utara Najran dekat perbatasan Yaman.

“Para arkeolog yakin bahwa artefak yang ditemukan tersebut merupakan prasasti berbahasa Arab tertua,” kata Guyader.

Guyader juga menambahkan bahwa:

1. Hal pertama yang menjadikan penemuan ini penting adalah prasasti ini merupakan teks campuran, dikenal dengan Nabatean Arab, tahap pertama tulisan Arab. Prasasti seperti ini sebelumnya hanya ditemukan di utara Hejaz, Sinai, dan Levant.

2. Hal kedua adalah prasasti ini diperkirakan dibuat antara tahun 469-470 Masehi. Ini merupakan tahun tertua dari tulisan Arab yang ditemukan, sehingga disebut sebagai “bagian yang hilang” antara Nabatean dan tulisan Arab.

Prasasti yang diperkirakan dibuat antara tahun 469-470 Masehi ini diyakini sebagai missing link antara alphabet Yunani dan alphabet Arab.

Ahli prasasti yang juga seorang profesor dari Universitas Aix-Marseille, Frederic Imbert menilai banyaknya contoh tulisan dan prasasti, yang sebagian besar merupakan nama dan salam, merupakan bukti luasnya penyebaran keaksaraan dalam budaya Nabatean, yang tersebar dari ujung utara Arab Saudi hingga ujung utara Laut Mati.

Penemuan ini, dianggap penting untuk sejarah Semenanjung dan Bahasa Arab, dan merupakan sebuah keberhasilan misi arkeologi yang dibiayai oleh Kementerian Luar Negeri Prancis, yang melakukan 148 proyek di 60 negara pada tahun 2014.

Misi Arkeologi ini merupakan strategi yang dipakai Perancis untuk mengembangkan hubungan kerjasamanya dengan negara lain. Perancis melihat kesempatan untuk mentransfer pengetahuannya dalam kegiatan-kegiatan ini. Seperti matematika, arkeologi sebenarnya merupakan salah satu bintang bidang penelitian Perancis dan banyak negara mengajukan permohonan kerjasama

dengan Perancis. Dewan Penasihat Riset Arkeologi Luar Negeri Perancis, terdiri dari arkeolog dan para ahli dari Institut de France, CNRS, lembaga-lembaga budaya maupun riset besar Perancis (Louvre, Museum Sejarah Alam Nasional, universitas-universitas) dan luar negeri (Institut Arkeologi Jerman, Museum Nasional Mali), bertugas menangani proyek-proyek terbaik.

### 2.6.2. Pembagian Bahasa Arab

Semua bahasa manusia tersusun dari tiga komponen dasar yaitu:

1. Satuan bunyi yang disebut huruf atau abjad.

Contoh : م-س-ج-د

2. Susunan huruf yang memiliki arti tertentu yang disebut kata.

Contoh : مسجد (masjid)

3. Rangkaian kata yang mengandung pikiran yang lengkap yang disebut kalimat.

Contoh : أَصَلِّي فِي الْمَسْجِدِ (Saya sholat di masjid)

Pembagian kata dalam Bahasa Arab dibagi menjadi tiga golongan besar, yaitu *isim* (اسم), *fi'il* (فعل), dan *harf* (حرف).

#### 1. *Isim*(اسم)

*Isim* secara bahasa memiliki arti yang dinamakan atau nama atau kata benda. Sedangkan menurut ulama *nahwu*, *isim* adalah kata yang menunjukkan suatu makna yang ada pada zatnya akan tetapi tidak berkaitan dengan waktu. *Isim* itu terbagi-bagi menjadi beberapa jenis yang dapat dikelompokkan sesuai dengan kelompoknya (Al-Batawy, 2008:2).

Contoh: زيد artinya Zaid (Isim ‘Alam = nama orang), فيل artinya gajah (isim ‘Alam = nama hewan), جاكرتا artinya Jakarta (nama tempat), هذا artinya ini (kata tunjuk), انا artinya saya (kata ganti) dan contoh yang lain.

## 2. *Fi'il* (فعل) atau kata kerja

*Fi'il* adalah suatu kata yang menunjukkan suatu makna sempurna dengan sendirinya dan menggambarkan latar belakang waktu kejadian, dalam bahasa Indonesia *fi'il* dikenal dengan istilah ‘kata kerja’. Namun, terdapat sedikit perbedaan yaitu dalam Bahasa Arab bentuk kata kerja itu berubah sesuai dengan latar belakang waktu kejadiannya, tidak sebagaimana pada bahasa Indonesia (Mushlih, 2010 : 1).

Contoh : اصلي (saya shalat)

## 3. *Harf*(حرف)

*Harf* secara bahasa memiliki arti huruf seperti yang kita kenal dalam bahasa Indonesia ada 26 huruf. Sedangkan dalam Bahasa Arab kita mengenal ada 28 huruf yang kita kenal dengan Huruf Hijaiyah. Akan tetapi, huruf yang dimaksud disini bukan setiap Huruf Hijaiyah melainkan Huruf Hijaiyah yang memiliki arti seperti و (dan), ف (maka), ب (dengan), ل (untuk), س (akan), ك (seperti), في (di dalam).

Adapun huruf-huruf seperti *Alif, Ta, Tsa* dan yang lain yang tidak memiliki arti maka tidak dapat menyusun suatu kalimat, melainkan hanya menyusun suatu kata saja (Al-Batawy, 2008:3).

### 2.6.3. Jenis-jenis Kata Kerja (*Fi'il*)

*Fi'il* merupakan salah satu jenis kata yang mengandung *morfem* rangkap dalam bahasa Arab/al-Quran. Letak *fi'il* dalam kalimat dapat menentukan jenis kalimat itu sendiri. Apabila diletakkan di awal kalimat atau mendahului *isim*, maka kalimat itu dinamakan kalimat verbal (*jumlah fi'liyah*). Sebaliknya, apabila *fi'il* terletak sesudah *isim*, maka kalimat itu disebut kalimat nominal (*jumlah ismiyah*) (Masor, 2009 : 1).

*Al-fi'lu* atau *fi'il* secara bahasa memiliki makna perbuatan atau kata kerja. Sedangkan menurut istilah dalam ilmu *nahwu*, *fi'il* adalah kata yang menunjukkan suatu makna yang menunjukkan suatu makna yang ada pada zatnya serta terkait dengan waktu. *Fi'il* terbagi menjadi tiga (Al-Batawy, 2008 :1):

1. *Fi'il Madhi* adalah kata kerja untuk masa lampau atau dalam istilah bahasa Inggrisnya adalah *past tense* yang memiliki arti telah melakukan sesuatu. Contoh : **قام** (telah berdiri) atau **جلس** (telah duduk).
2. *Fi'il Mudhori'* adalah kata kerja yang memiliki arti seang melakukan sesuatu atau dalam istilah bahasa Inggrisnya *present continues tense*. Contoh : **يقوم** (sedang berdiri) atau **يجلس** (sedang duduk).
3. *Fi'il Amr* adalah kata kerja untuk perintah. Contoh : **قم** (bangunlah!) atau **اجلس** (duduklah!).

## 2.7. Ilmu *Sharaf*

### 2.7.1. Pengertian *Sharaf*

*Sharaf* atau dibaca *Shorof* adalah satu nama cabang ilmu dalam pelajaran Bahasa Arab yang khusus membahas tentang perubahan bentuk kata (dalam Bahasa Arab disebut dengan kalimat). Perubahan bentuk kata ini dalam prakteknya disebut *Tashrif*. Oleh karena itu, dinamakan Ilmu *Sharaf* (perubahan: berubah), karena ilmu ini khusus mengenai pembahasan *Tashrif* (pengubahan; mengubah) (Toha, 2010 : 1).

### 2.7.2. Istilah – istilah Dasar dalam Ilmu *Sharaf*

#### a. *Tashrif*

Pengertian dari *tashrif* adalah perubahan asal suatu kata kepada kata-kata yang berbeda untuk mencapai arti yang dikehendaki. Seperti perubahan dari bentuk *Mufrad* (satu) kepada *Tatsniyah* (dua) dan *Jamak* (banyak) dan atau bentuk *Masdar* (kata benda) kepada *fi'il* (kata kerja) atau *wasf* (kata sifat) atau dengan maksud *lafadz* yakni meringankan ucapan seperti perubahan kata *qawala* (قَوْل) dan *gazawa* (غَزَوْ) menjadi *qaala* (قَالَ) dan *gaaza* (غَاَزَ) (Al-Maraghi 2010 :4).

Secara umum, suatu kata kerja berubah menjadi jenis perubahan kata sebagai berikut:

#### 1. *Fi'il Madhi* (kata kerja lampau, *past tense*)

*Fi'il Madhi* adalah kata kerja (*fi'il*) yang menunjukkan terhadap suatu kejadian/peristiwa sebelum masa pembicaraan (lampau, telah berlalu)

## 2. *Fi'il Mudhori'* (kata kerja sekarang, *present continuous tense*)

*Fi'il Mudhori'* adalah kata kerja (*fi'il*) yang menunjukkan terhadap suatu peristiwa/kejadian yang berlangsung pada saat masa pembicaraan (sekarang) atau sesudahnya (akan datang)

## 3. *Mashdar* (kata benda)

Isim yang menunjukkan kejadian (*huduts*) yang sepi dari zaman dan mencukupi atas huruf-huruf *Fi'ilnya* atau melebihinya.

## 2.8. Al Quran

### 2.8.1. Pengertian Al Quran

Al Quran adalah Kalam Allah SWT yang bernilai mukjizat, yang diturunkan kepada Rasulullah Muhammad SAW, dengan perantara malaikat Jibril a.s. yang tertulis pada mashahif. Secara mutawattir, dan bernilai ibadah bagi pembacanya. Al-Quran diawali surat al-Fatihah dan ditutup dengan surat an-Naas. (Ash-Shobuni, 2001)

Definisi tersebut telah disepakati para ulama' (Ash-Shobuni, 2001). Kemudian, Allah SWT menurunkan al-Quran sebagai tata kehidupan umat dan petunjuk bagi makhluk. Al Quran merupakan tanda kebesaran Rasulullah SAW, disamping bukti yang jelas atas kenabian dan kerasulannya. Selain itu, Al Quran juga sebagai *hujjah* yang akan tetap tegak sampai hari kiamat.

### 2.8.2 Kedudukan Al Quran

Banyak ayat dan hasits Nabi yang menerangkan keutamaan-keutamaan Al-Qur'an dan ilmu-ilmunya, baik yang berhubungan dengan belajar mengajar

maupun membaca dan menghafalkannya. Banyak pula ayat yang mengajak kaum mukminin untuk memikirkan dan menggunakan hukum-hukum Kitabullah. Keutamaan yang Al Quran yang paling besar ialah bahwa ia adalah kalam Allah (Ash-Shobuni, 2001), yang pujian terhadapnya telah difirmankan Allah di beberapa ayat, diantaranya sebagaimana yang disebutkan dalam surat al-An'am ayat 92 dan surat al-Isra' ayat 9.

وَهَذَا كِتَابٌ أَنْزَلْنَاهُ مُبَارَكٌ مُصَدِّقُ الَّذِي بَيْنَ يَدَيْهِ  
وَلِتُنذِرَ أُمَّ الْقُرَىٰ وَمَنْ حَوْلَهَا وَالَّذِينَ يُؤْمِنُونَ بِالْآخِرَةِ يُؤْمِنُونَ بِهِ ۗ  
وَهُمْ عَلَىٰ صَلَاتِهِمْ يُحَافِظُونَ ﴿٩٢﴾

92. Dan ini (Al Quran) adalah kitab yang telah Kami turunkan yang diberkahi; membenarkan kitab-kitab yang (diturunkan) sebelumnya<sup>[492]</sup> dan agar kamu memberi peringatan kepada (penduduk) Ummul Qura (Mekah) dan orang-orang yang di luar lingkungannya. Orang-orang yang beriman kepada adanya kehidupan akhirat tentu beriman kepadanya (Al Quran) dan mereka selalu memelihara sembahyangnya.

(Al-An'am: 92 diambil dari Software Quran in Word)

*Al Quran petunjuk ke jalan yang benar*

إِنَّ هَذَا الْقُرْآنَ يَهْدِي لِلَّتِي هِيَ أَقْوَمُ وَيُبَشِّرُ الْمُؤْمِنِينَ الَّذِينَ يَعْمَلُونَ  
الصَّالِحَاتِ أَنَّ لَهُمْ أَجْرًا كَبِيرًا ﴿٩﴾

9. Sesungguhnya Al Quran ini memberikan petunjuk kepada (jalan) yang lebih lurus dan memberi khabar gembira kepada orang-orang Mu'min yang mengerjakan amal saleh bahwa bagi mereka ada pahala yang besar,

(Al-Isra': 9 diambil dari Software Quran in Word).

Sebagai kalam Allah, Al-Quran sangat mulia yang diturunkan untuk memberi peringatan serta sebagai petunjuk jalan yang harus ditempuh oleh orang-orang mu'min sehingga memperoleh keselamatan di dunia dan di akherat. Begitu

utamanya kedudukan Al Quran hingga Rasul pun mengistimewakan orang-orang yang mempelajari dan mengajarkan Al Quran hingga menjadikan mereka sebagai orang-orang yang terbaik. Hal ini disebutkan dalam salah satu hadits Rasul yang diriwayatkan oleh Imam Bukhori.

Tidak pernah terjadi dalam sejarah umat manusia adanya umat yang sangat menaruh perhatian terhadap kitabnya sebagaimana perhatian umat Muhammad terhadap Al-Quran. Bahkan belum pernah didengar adanya kitab suci yang mendapatkan penjagaan, pemeliharaan, penghormatan dan penghargaan yang telah didapatkan oleh kitab suci al-Quran al-karim. Mukjizat Nabi Muhammad saw yang abadi dan hujjahnya yang amat paripurna. Serta seruannya yang universal untuk seluruh umat manusia. Al-Quran di atas segala kitab yang dibawa para nabi dan Rasul dalam memberikan petunjuk dan kedamaian, pendidikan dan pelajaran, ketinggian derajat dan tasyri'. (Ash-Shobuni, 2001: 131)

### **2.8.3 Kemukjizatan Al Quran**

Al-Quranul Azhim adalah Kalamullah yang bernilai mukjizat bagi makhluk, baik uslub dan nadzamnya, keindahan penjelasannya, ilmi-ilmu dan hukum yang terkandung di dalamnya, pengaruh petunjuknya, dan cermatnya menyibak hal-hal ghaib yang sudah lampau atau yang akan datang. Para ulama telah datang membuktikan rahasia-rahasia dari penjelasan dari berbagai segi kemukjizatan Al-Quran setelah mereka mempunyai petunjuk dan bukti. Seluruh orang arab serta ahli bahasa dan bayan telah bersepakat bahwa Al-Quran itu sendiri telah merupakan mukjizat. Yaitu bahwa kemukjizatan Al-Quran terletak

pada kefasihan lafalnya, keindahan keterangannya, dan uslubnya yang mempunyai gaya tersendiri yang sama sekali tidak bisa ditiru baik dari bentuk prosa atau syair. Bahkan lebih dari itu, Al-Quran menyajikan lafal yang memikat yang nampak jelas pada aturan suara, bagusya bahasa, dan keindahan yang menghanyutkan. (Ash-Shobuni, 2001: 152-153)

## 2.9. Stemming

Menurut Hopper dan Paice (2005 : 1), banyak kata – kata yang mempunyai penafsiran *semantic* serupa dan dapat diperlakukan sama untuk kepentingan *IR (Information Retrieval)* aplikasi. Karena alasan ini, *algoritma stemming* atau *stemmer* telah dikembangkan.

*Stemming* merupakan suatu proses untuk menemukan kata dasar dari sebuah kata. Dengan menghilangkan semua imbuhan (*affixes*) baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan. *Stemming* digunakan untuk mengganti bentuk dari suatu kata menjadi kata dasar dari kata tersebut yang sesuai dengan struktur morfologi Bahasa Indonesia yang baik dan benar.

Imbuhan (*affixes*) pada Bahasa Indonesia lebih kompleks bila dibandingkan dengan imbuhan (*affixes*) pada Bahasa Inggris. Karena seperti yang telah disebutkan di atas bahwa imbuhan (*affixes*) pada Bahasa Indonesia terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*), bentuk perulangan (*repeated forms*) dan *confixes* (kombinasi dari awalan dan akhiran). Imbuhan-

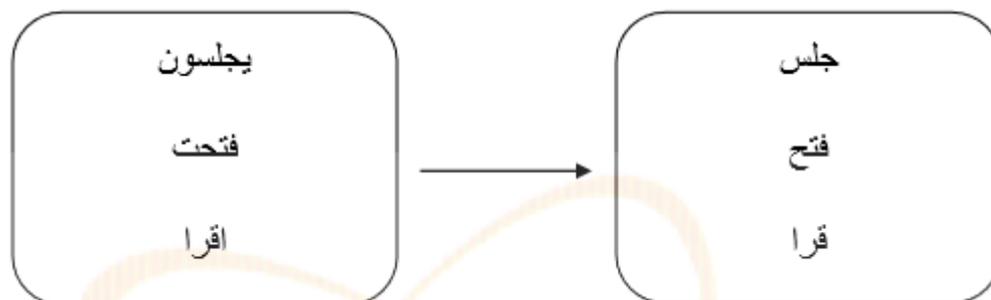
imbunan yang melekat pada suatu kata harus dihilangkan untuk mengubah bentuk kata tersebut menjadi bentuk kata dasarnya.

*Stemming* teks berbahasa Indonesia memiliki beberapa masalah yang sangat khusus terhadap bahasa. Salah satu masalah tersebut adalah perbedaan tipe dari imbunan-imbunan (*affixes*), yang lain adalah bahwa awalan (*prefixes*) dapat berubah tergantung dari huruf pertama pada kata dasar. Sebagai contoh "me-" dapat berubah menjadi "mem-" ketika huruf pertama dari kata dasar tersebut adalah "b", misalnya "membuat" (to make), tetapi "me-" juga dapat berubah menjadi "meny-" ketika huruf pertama dari kata dasar melekat adalah "s", misalnya "menyapu" (to sweep). Selanjutnya ketika ada lebih dari satu imbunan (*affixes*) yang melekat pada suatu kata, maka urutan untuk menghilangkan imbunan-imbunan (*affixes*) pada kata tersebut menjadi sangat penting. Jika dalam proses menghilangkan imbunan-imbunan (*affixes*) tersebut kita tidak memperhatikan urutan penghilangan imbunan-imbunan (*affixes*) tersebut, maka kata dasar yang benar dari kata tersebut tidak akan ditemukan. Sebagai contoh pada kata "di-beri-kan" (to be given) yang diturunkan dari kata dasar "beri" (to give). Jika kita menghilangkan akhiran (*suffixes*) "kan" terlebih dahulu sebelum menghilangkan awalan (*prefix*) "di-" maka pada proses *stemming* ini kita mendapatkan kata dasar yang benar yaitu "beri" (to give), akan tetapi jika algoritma *stemming* mencoba untuk menghilangkan awalan (*prefixes*) terlebih dahulu sebelum akhiran (*suffixes*) maka hasil kata dasar yang dihasilkan dari proses *stemming* dengan menggunakan algoritma tersebut adalah "ikan" (fish) (setelah menghilangkan awalan "di" dan "ber") dimana "ikan" merupakan kata

dasaryang valid yang terdapat dalam kamus tetapi ”ikan” bukan merupakan kata dasar yang benar untuk kata turunan ”diberikan”.

Contoh produk yang menggunakan *algoritma stemming* adalah *search engine* seperti *Lycos* dan *Google* dan juga kams serta produk laik menggunakan NLP untuk kepentingan IR (Haryati, 2007 : 17).

Penelitian terhadap *stemming* untuk *text retrieval*, *machine translation*, *document summarization* dan *text classification* sudah pernah dilakukan sebelumnya. Untuk *stemming* yang dilakukan pada *Text Retrieval*, *stemming* ini meningkatkan kesensitivan *retrival* dengan meningkatkan kemampuan untuk menemukan document yang relevan, tetapi hal itu terkait dengan pengurangan pada pemilihan dimana pengelompokkan menjadi kata dasar menyebabkan penghilangan makna kata. Pada *Text Retrival stemming* diharapkan dapat meningkatkan recall, tetapi memungkinkan untuk menurunkan precision. Metode *stemming* memerlukan input yang berupa *term* yang terdapat dalam suatu dokumen dan outputnya berupa stem (bagian dari kata yang tersisa setelah dihilangkan imbuhan) (Nurlailah, 2009).



Gambar 2.1 Contoh stemming

Beberapa contoh stemmer adalah sebagai berikut :

- a. *Porter Stemmer*. Algoritma ini terkenal digunakan sebagai stemmer untuk bahasa Inggris. Porter Stemmer dalam bahasa Indonesia akan menghasilkan keambiguan karena aturan morfologi bahasa Indonesia (Tala, 2003).
- b. *Light Stemmer*. Algoritma ini adalah algoritma *Stemming* yang hanya menghilangkan imbuhan depan (*prefix*) dan imbuhan belakang (*suffiks*). Sesuai dengan rule yang telah di tentukan dari karakter yang paling banyak di gunakan, adapun karakter yang paling banyak di gunakan tersebut di bagi menjadi 2 yaitu karakter imbuhan depan dan imbuhan belakang yang sering di pakai. Yang di maksudkan disini adalah di *light Stemming* bisa di inisialisasikan menjadi 3 inisial karakter yang paling sering di gunakan dalam document –dokumen berbahasa arab yaitu karakter depan (imbuhan depan) dan karakter belakang (imbuhan belakang) dengan mendaftarkan imbuhan – imbuhan tersebut akan mendapatkan sebuah rule / aturan dalam menjadikan kata tersebut ke bentuk dasarnya.
- c. *Paice/Huck Stemmer*. Algoritma ini hanya menggunakan satu tabel rule, dimana setiap rule dapat menspesifikasikan penghapusan atau penggantian akhiran dari autu kata. Rule-rule ini dikelompokkan menurut huruf akhiran dari suffix, sehingga pengaksesan tabel rule dilakukan dengan melakukan

look-up terhadap huruf akhir dari kata yang sedang diolah (Yogatama, 2008 : 12).

- d. *Lovins Stemmer*. Algoritma ini banyak dipengaruhi oleh technical vocabulary. Lovins menyatakan bahwa bentuk yang paling diharapkan dari *context sensitive rule* adalah bentuk yang dapat digeneralisasi untuk diterapkan dalam berbagai situasi. Dalam kenyataannya, sedikit sekali rule-rule yang memenuhi syarat di atas. Untuk setiap endings, terdapat beberapa kasus khusus yang dapat menyebabkan kesalahan *stem* yang dihasilkan. *Lovins Stemmer* mengatasi hal ini dengan menangani *exception-exception* yang lebih sering muncul dengan harapan dapat membatasi jumlah kesalahan dalam daftar *exception*. (Yogatama, 2008:13). Algoritma *Lovins* terdiri dari dua langkah utama, yaitu fase *stemming* dan fase *recording*.
- e. *Nazien & Adriani Stemmer*. Algoritma ini paling sering dibicarakan dalam *stemming* bahasa Indonesia. Algoritma ini merupakan hasil penelitian internal UI (Universitas Indonesia) dan tidak dipublish secara umum (Nazif, 1996). Algoritma ini merupakan gabungan antara algoritma menghilangkan imbuhan dan *brute force stemming*. Namun algoritma ini mempunyai dua masalah, yang pertama kemampuannya tergantung dari besarnya database kata dasar, dan yang kedua, hasil *stemming* tidak selalu optimal untuk aplikasi information retrieval (Tala, 2013).
- f. *Khoja*. Algoritma *khoja* adalah algoritma *stemming* yang menghapus semua *afiks* untuk menghasilkan root. Dalam bahasa arab ini berarti

penghapusan *prefiks, sufiks dan infiks*. Digunakan untuk membantu menentukan tag kata. Dalam algoritma ini proses *stemmingnya* dengan cara memisahkan per huruf dalam kata arab yang akan di *stemming*.



## BAB III

### ANALISIS DAN PERANCANGAN SISTEM

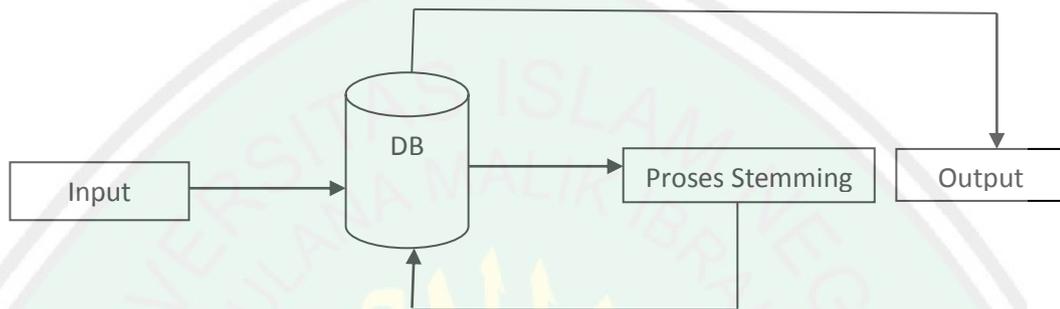
#### 3.1 Desain Aplikasi

Aplikasi yang dibangun adalah sebuah aplikasi *stemming* kata Arab yang digunakan untuk menentukan kata dasar dari kata Arab. Aplikasi ini dibuat dengan menggunakan perbandingan akurasi *Light Stemming* dan *Khoja*. Aplikasi ini berbasis java. Dalam melakukan proses *stemming* ini, input kata ke textarea yang sudah disediakan, cek imbuhan, lalu *stemming*.

Tahapan-tahapan yang harus dilakukan dalam membangun aplikasi *stemming* kata Arab ini adalah masukan sistem berupa input kata, input kata akan dicek dalam database, apabila inputan dikenali dalam database maka akan menghasilkan output, tetapi apabila tidak dikenali di dalam database, maka akan dilakukan pemotongan imbuhan (*stemming*). Kemudian hasil pemotongan (stem) akan dicek kembali di dalam database. Jika ada, maka akan menghasilkan output.

Pada penelitian ini akan dibangun sebuah aplikasi *stemming* dengan konsep perbandingan. Dengan aplikasi ini, pengguna tidak hanya mengetahui satu *fi'il* saja, akan tetapi pengguna juga dikenalkan dengan bentuk *fi'il* lainnya. Dalam aplikasi ini pengguna memasukkan kata dari database yang nanti akan dimasukkan secara acak. Pengguna bisa memasukkan kata dasar atau kata dalam bentuk yang lain berdasarkan huruf yang tersedia dalam database. Metode stemmer dalam aplikasi ini digunakan saat pengguna memasukkan kata berimbuhan, kata tersebut akan di *stemming* yang kemudian akan dilakukan pengecekan dalam database.

Blok diagram proses dapat dilihat pada Gambar 3.1.



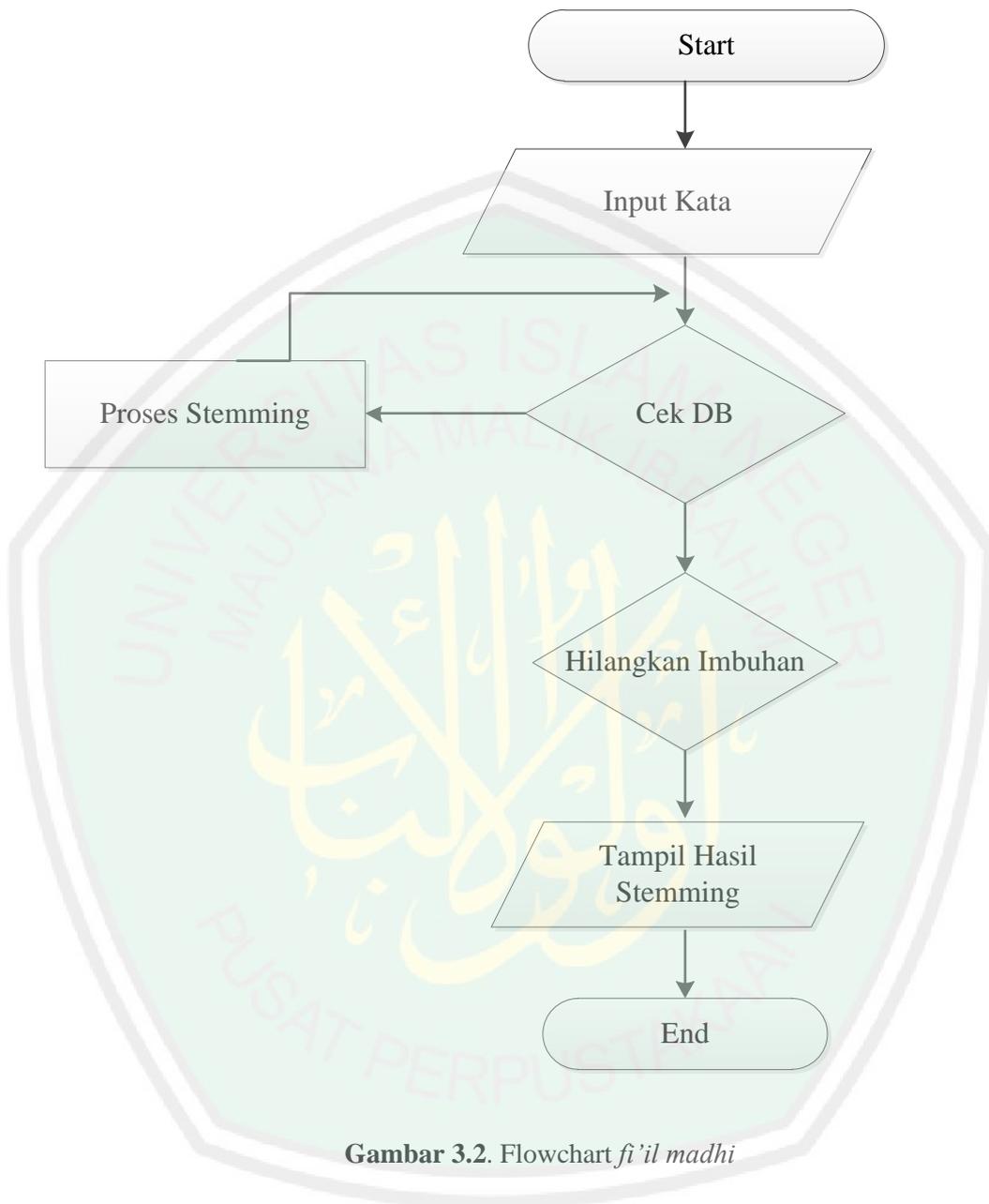
Gambar 3.1. Blok Diagram

### 3.2 Desain proses

Desain proses keseluruhan dari aplikasi penelitian ini digambarkan dalam sebuah flowchart untuk menjelaskan alur jalannya program yang akan peneliti buat. Dalam pembuatan aplikasi ini terdapat beberapa flowchart, yaitu :

#### a. Flowchart *fi'il madhi* menggunakan *light stemmer*

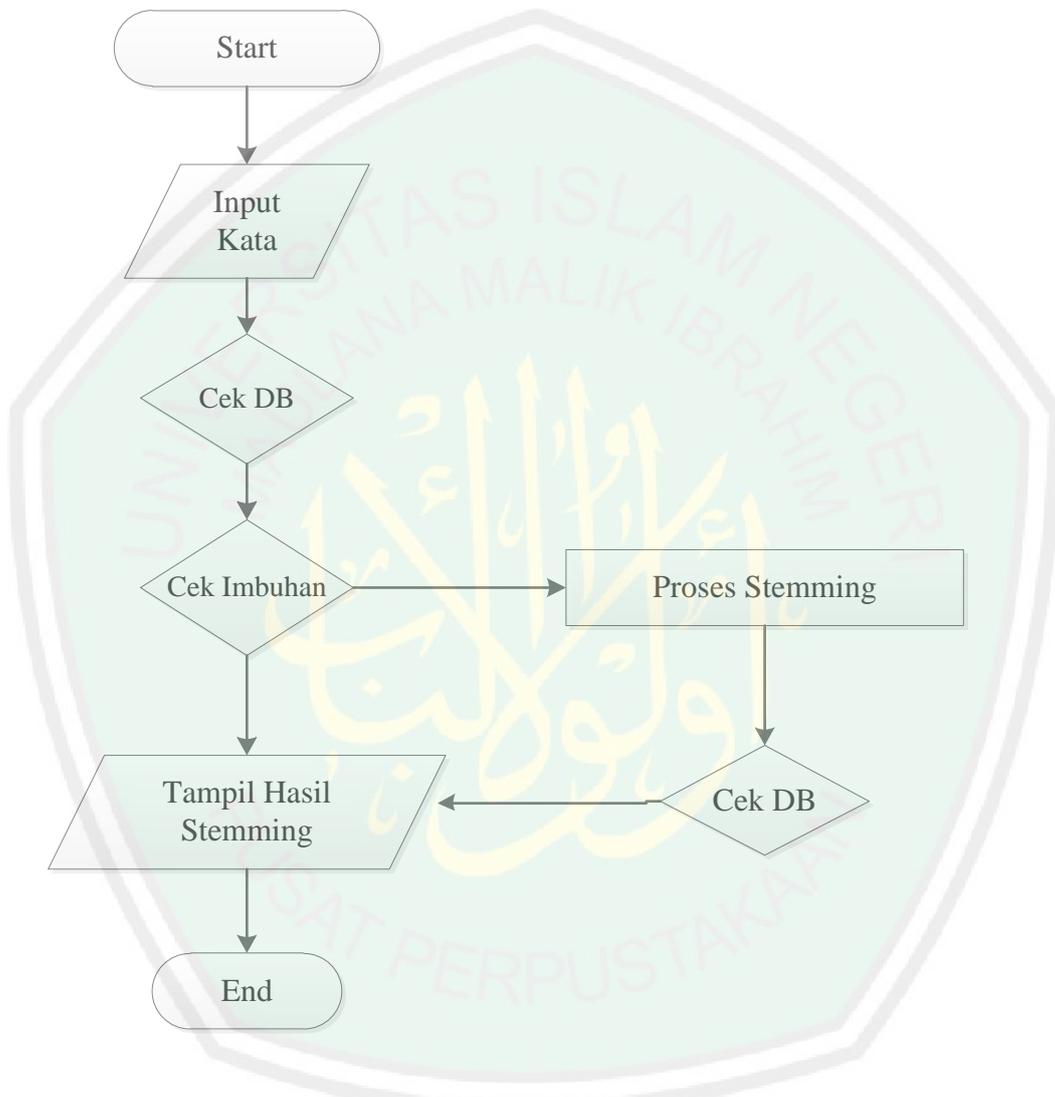
Dari flowchart pada gambar 3.2 menunjukkan flowchart untuk *stemming* fill madhi menggunakan *light stemmer*. Langkah-langkah untuk *stemming* ini adalah dimulai pada titik mulai, kemudian menginputkan kata. Kemudian cek di database kata Arab tersebut setelah itu melakukan proses *stemming*, *stemming* ini suatu proses untuk menentukan kata dasar dari sebuah kata, setelah itu cek di database kemudian berakhir pada titik end.



#### b. Flowchart mashdar menggunakan *Light Stemming*

Pada gambar 3.3 menunjukkan flowchart untuk *stemming* mashdar menggunakan *light stemmer*. Langkah-langkah untuk *stemming* ini adalah dimulai pada titik mulai, kemudian menginputkan kata Arab. Setelah itu cek di database kata tersebut, kemudian cek imbuhan yang ada, kalau beserta dengan imbuhan

masuk ke dalam proses *stemming* , cek database kembali kemudian tampil hasil *stemming*, jika tidak ada imbuhan maka langsung pada proses titik end.

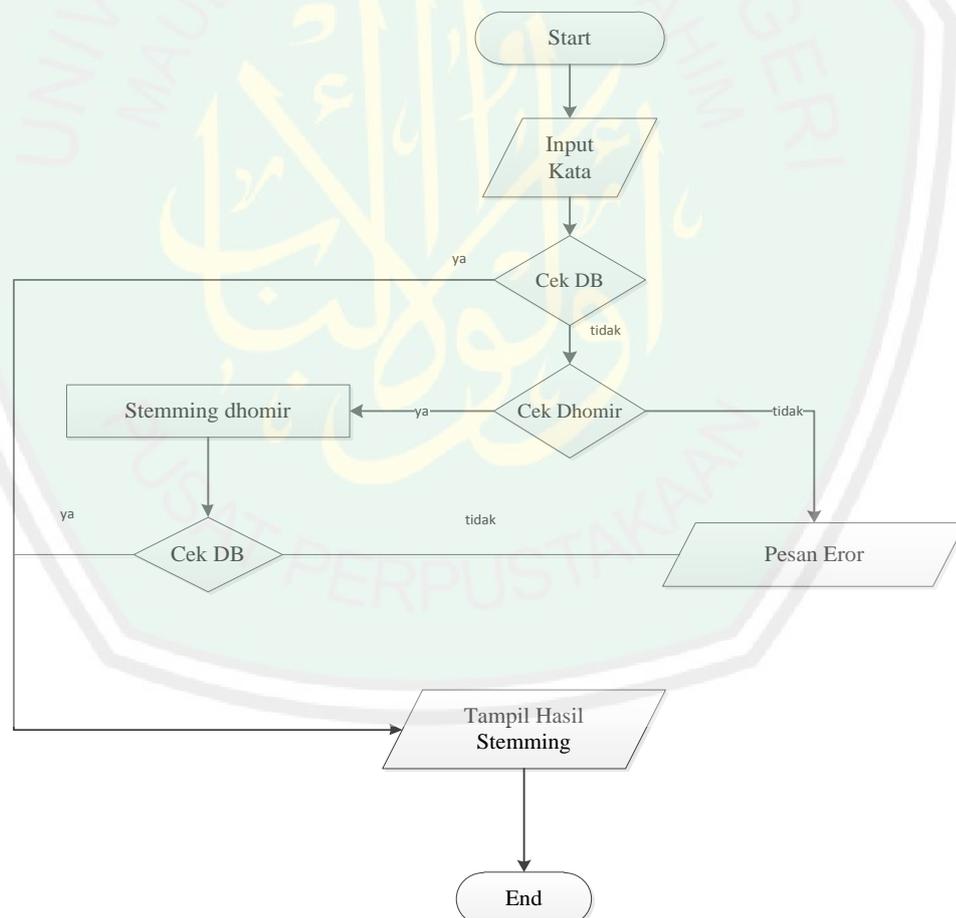


**Gambar 3.3.** Flowchart mashdar

### c. Flowchart *Stemming*

Pada gambar 3.4 menunjukkan flowchart untuk contoh *stemming* pada umumnya, yang menggunakan *light stemmer*. Langkah-langkah untuk *stemming* ini adalah dimulai pada titik mulai, kemudian menginputkan kata Arab. Setelah itu

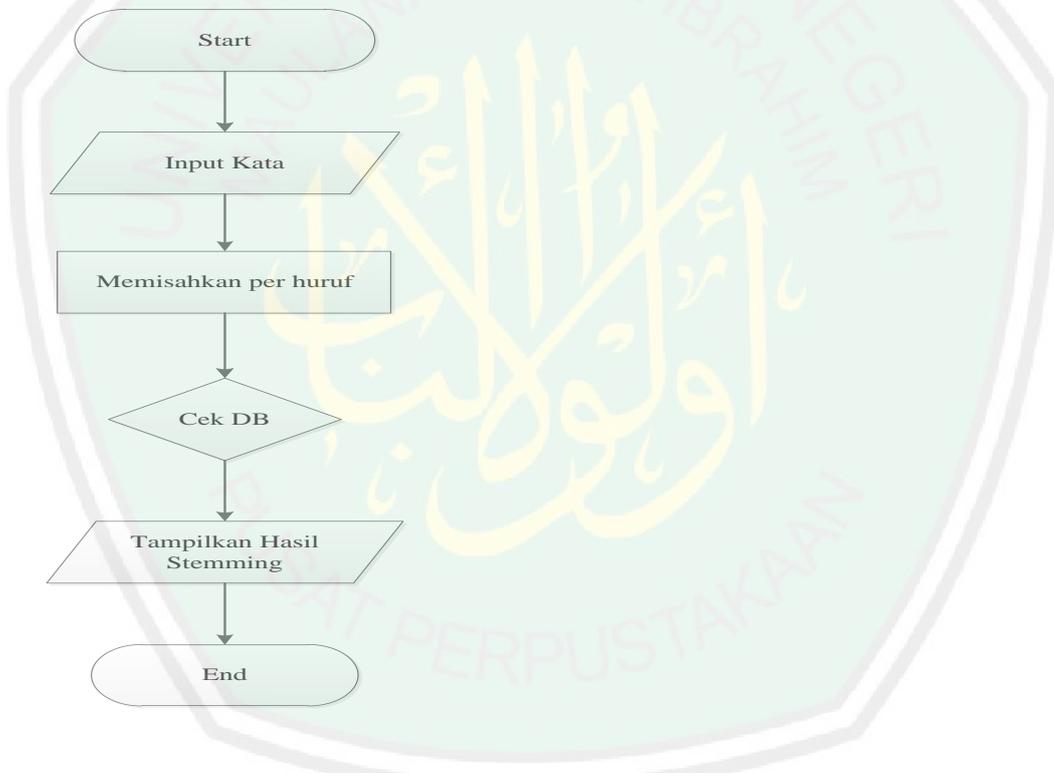
cek di database kata tersebut, jika ya atau sudah tidak memerlukan *stemming* maka langsung tampil dan berakhir pada titik end. Apabila tidak ada atau belum menemukan kata dasar maka cek dhommir dalam kata tersebut kemudian melakukan *stemming* dhommir tersebut setelah itu cek database jika iya sudah menemukan kata dasar maka tampil hasil *stemming* dan berakhir pada titik end. Jika tidak maka tampil pesan error.



**Gambar 3.4.** Flowchart *stemming*

#### d. flowchart stemming dengan *khoja*

Pada gambar 3.5 menunjukkan flowchart untuk stemming menggunakan *khoja*. Langkah-langkah untuk stemming ini adalah dimulai pada titik mulai, kemudian menginputkan kata Arab. Setelah itu memisahkan per huruf dari kata tersebut untuk lebih memudahkan proses kemudian cek dalam database dan tampilkan hasil stemming dan berakhir pada titik end.

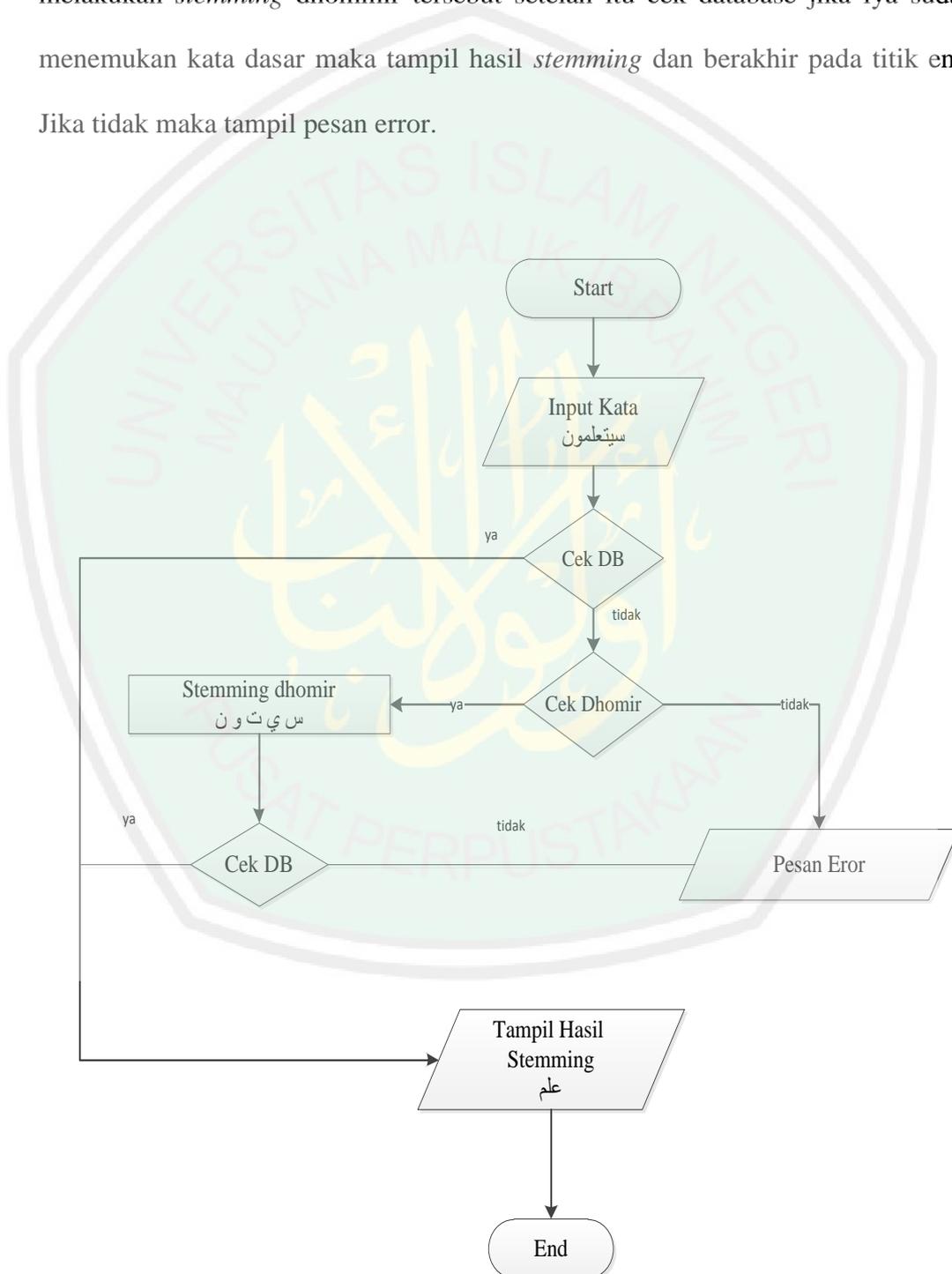


Gambar 3.5. Flowchart *stemming khoja*

#### e. flowchart Light stemming dengan output fi'il madhi

Pada gambar 3.6 menunjukkan flowchart untuk contoh *stemming* pada umumnya, yang menggunakan *light stemmer*. Langkah-langkah untuk *stemming* ini adalah dimulai pada titik mulai, kemudian menginputkan kata Arab. Setelah itu cek di database kata tersebut, jika ya atau sudah tidak memerlukan *stemming*

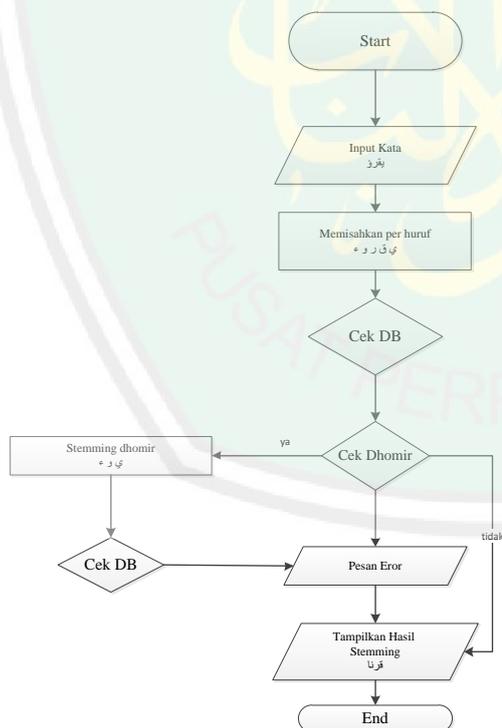
maka langsung tampil dan berakhir pada titik end. Apabila tidak ada atau belum menemukan kata dasar maka cek dhommir dalam kata tersebut kemudian melakukan *stemming* dhommir tersebut setelah itu cek database jika iya sudah menemukan kata dasar maka tampil hasil *stemming* dan berakhir pada titik end. Jika tidak maka tampil pesan error.



**Gambar 3.6.** Flowchart *light stemming* dengan output fi'il madhi

### f. flowchart khoja dengan output mashdar

Pada gambar 3.7 menunjukkan flowchart untuk contoh *stemming* pada umumnya, yang menggunakan *khoja stemming*. Langkah-langkah untuk *stemming* ini adalah dimulai pada titik mulai, kemudian menginputkan kata Arab. Setelah itu memisahkan per huruf, jika ya atau sudah tidak memerlukan *stemming* maka langsung tampil dan berakhir pada titik end. Apabila tidak ada atau belum menemukan kata dasar maka cek dhommir dalam kata tersebut kemudian melakukan *stemming* dhommir tersebut setelah itu cek database jika iya sudah menemukan kata dasar maka tampil hasil *stemming* dan berakhir pada titik end. Jika tidak maka tampil pesan error.



**Gambar 3.7.** Flowchart *khoja stemming* dengan output *mashdar*

### 3.3 Desain database

hasil  
imbuhan

**Gambar 3.8** desain database

Desain database untuk aplikasi *stemming* ini menggunakan satu database. Penelitian ini menggunakan database *SQL* karena memerlukan data yang jumlahnya tidak sedikit. Ada dua tabel yang digunakan pada database ini yaitu tabel hasil, dan tabel imbuhan. Tabel hasil ini digunakan untuk menyimpan kata Arab asli, hasil light stemming dan khoja. Tabel imbuhan berfungsi untuk menyimpan *suffix*, *prefix*, dan *infix* yang ada pada kata Arab.

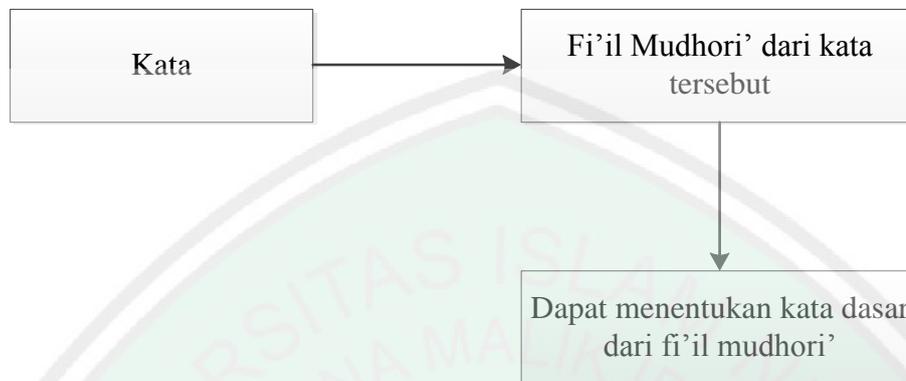
#### a. tabel fi'il mudhori'

Tabel fi'il mudhori' berisi kumpulan fi'il mudhori' yang ada pada Al Quran Juz 30. Dapat dilihat beberapa contoh *fi'il mudhori'* dari gambar potongan isi tabel *fi'il mudhori'* pada **gambar 3.9** Terdapat satu field dalam tabel ini, yaitu kolom fi'il mudhori' yang akan *distemming*. sehingga dapat dilakukan pengecekan terhadap tiap kata dari input kata.

Fi'il Mudhori'
يتعلمون
يقول
ينظرون
تسمع
فليحيدي
يتدبر

**Gambar 3.9** Potongan tabel fi'il mudhori'

### 3.4 Teknik *Stemming* Imbuhan



Gambar 3.10 Skema Jalannya Aplikasi *Stemmer*

Dalam *Information Retrieval*, untuk mempercepat pencarian informasi dari sekumpulan data, biasanya orang menyusun data berdasarkan *key* tersebut data akan lebih mudah dicari kembali. Misalkan dalam kamus, data disusun urut berdasarkan kata secara *alphabetic*, contoh lain adalah informasi buku pada perpustakaan dapat dilihat berdasarkan kelompok indeks tertentu.

Dalam upaya pengkomputerisasian, cara yang termudah untuk menyusun urutan kata adalah dengan mengindeks seluruh kata berikut urutannya dalam satu berkas indeks yang semakin besar setiap kali terjadi penambahan kata atau subkata baru, akibatnya dapat mempengaruhi kinerja pencarian kata, seiring dengan perkembangan bahasa itu sendiri dari waktu ke waktu (Nurlailah, 2009 : 68-69).

Salah satu cara untuk meningkatkan kinerja sistem pencarian kata adalah memperkecil ukuran berkas indeks dengan menyimpan urutan katanya saja.

Pencarian variasi turunannya dilakukan dengan menyimpan urutan katanya saja. Pencarian variasi turunannya dilakukan dengan cara menentukan kata yang membentuk turunan tersebut terlebih dahulu, kemudian berdasarkan kata tersebut, dan turunannya dilakukan dengan menggunakan algoritma *stemming*.

Keuntungan penggunaan algoritma ini adalah dimungkinkannya penyimpanan indeks berdasarkan kata, agar dapat memperkecil ukuran berkas indeks dan meningkatkan kinerja pencarian data.

Terdapat dua kesalahan fatal yang dimungkinkan terjadi dari pemotongan imbuhan : pemotongan yang terlalu berlebihan (*overstemming*) dan pemotongan yang terlalu sedikit (*understemming*). Pemotongan imbuhan yang berlebihan akan mengakibatkan gagalnya pencarian kata di *database*. Misalnya pemotongan *prefix* ن pada kata نظر proses *stemm* akan memotong kata tersebut menjadi ظر dan kata seperti itu tidak mungkin dijumpai dalam *database*, karena huruf ن dan kata ظر adalah satu kesatuan kata yang tidak bisa dipisahkan. Demikian juga pemotongan imbuhan yang terlalu sedikit, seperti pemotongan kata ستهذب. Pemotongan *prefix* س akan menghasilkan kata هذب. Hasil pemotongan kata ini juga tidak mungkin akan ditemui di *database* karena kata tersebut bukan merupakan kata dasar, ت هذب masih mempunyai *prefix* ت yang belum dipisahkan. Dan bisa dikatakan sistem akan gagal dalam mencari data. Aplikasi ini akan memperkecil tingkat kesalahan dalam *overstemming* (pemotongan terlalu banyak) maupun *understemming* (pemotongan terlalu sedikit).

## BAB IV

### HASIL DAN PEMBAHASAN

Dalam bab ini dibahas mengenai rangkaian uji coba dan evaluasi yang telah dilakukan. Uji coba bertujuan untuk mengetahui tingkat keakurasian dari aplikasi dalam *stemming* kata Arab. Sedangkan evaluasi dilakukan bertujuan untuk analisa hasil uji coba agar mendapatkan kesimpulan dan saran untuk pengembangan aplikasi selanjutnya.

#### 4.1. Implementasi Sistem

Pada bab ini membahas perancangan aplikasi yang sudah dibuat, dan melakukan uji coba aplikasi untuk mengetahui apakah aplikasi yang telah dibuat sesuai dengan yang diinginkan. Untuk mengimplementasikan sistem yang dibuat ada beberapa komponen yang digunakan yaitu komponen perangkat lunak (*software*) dan perangkat keras (*hardware*). Spesifikasi perangkat keras dan perangkat lunak yang digunakan di dalam penelitian ini sebagai berikut :

##### 1. Perangkat Keras (*hardware*)

Perangkat keras (*hardware*) yang digunakan dalam pembuatan aplikasi ini adalah sebagai berikut :

###### a. Toshiba Netbook PC.

Processor : Intel® Core™ i3-2350M Processor (2.30 GHz, 3MB Cache)

Chipset : Intel® HM65

Memory : 2 GB DDR3 PC-10600

Video : Intel® Graphics Media Accelerator HD 729MB (shared)

Tampilan : 13.3" WXGA LED, Display Max. Resolution 1366 x 768

Hard Drive : 640 GB Serial ATA 5400 RPM

## 2. Perangkat Lunak (*software*)

Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

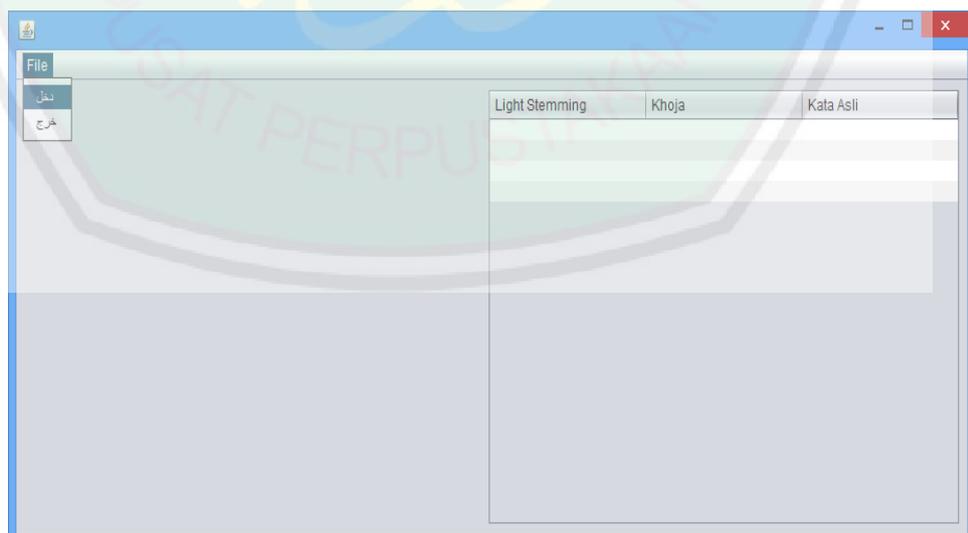
- a. NetBeans IDE 7.2
- b. Database Mysql. Untuk media penyimpanan data.

### 4.2. Implementasi Interface

Pada implementasi *interface* akan dijelaskan mengenai komponen-komponen dari aplikasi *stemming* yang diimplementasikan pada java . Berikut beberapa tampilan java yang terdapat dalam aplikasi.

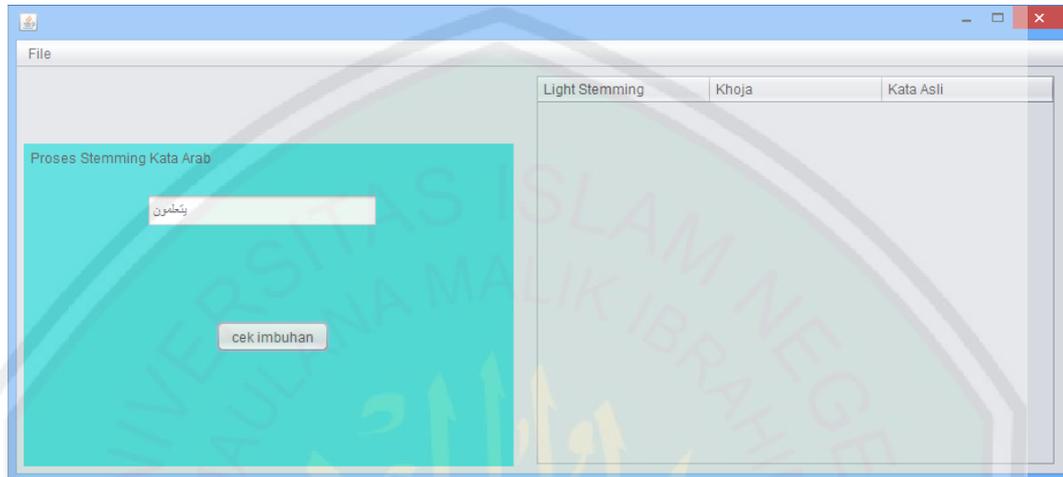
#### 4.2.1. Tampilan Halaman Utama

Halaman utama ini muncul ketika program pertama kali dijalankan. Halaman utama ini memuat form untuk menginput kalimat Arab, cek imbuhan, dan proses stemming. Halaman utama ditunjukkan oleh Gambar 4.1.



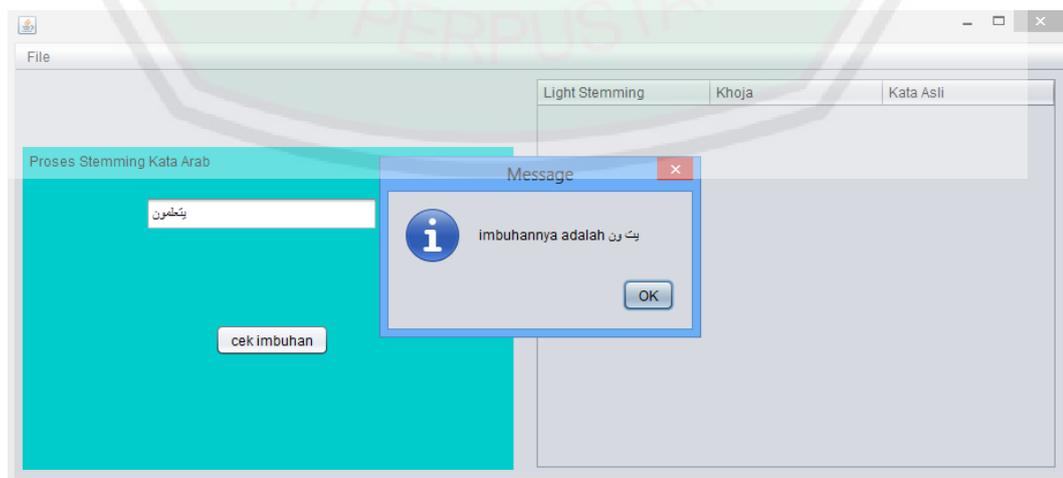
**Gambar 4.1.** Halaman Utama

Selanjutnya *user* menginputkan sebuah kata arab. Contoh penggunaan aplikasi terdapat pada Gambar 4.2.



**Gambar 4.2.** Contoh Penggunaan Aplikasi

Setelah menginputkan kata arab, kemudian *user* bisa klik pada button sek imbuhan untuk mengetahui imbuhan yang terdapat pada kata arab yang telah diinputkan. Contoh penggunaan aplikasi terdapat pada gambar 4.3



**Gambar 4.3.** Contoh Penggunaan Aplikasi

### 4.3 Uji Coba

Uji coba Uji coba *stemming* dilakukan dengan menginputkan kalimat bahasa Arab. Data pengujian yang digunakan dalam penelitian ini adalah beberapa *fi'il mudhori'* yang diambil dari Al Quran juz 30.

Untuk data yang diuji terdiri dari 44 kalimat tunggal bahasa Indonesia dengan inputan yang berbeda. Dan kata dasar asli diambil dari kamus Arab Al-Munawwir. Data lengkapnya dapat dilihat pada tabel dibawah. Adapun prosentase keberhasilan tersebut diperoleh dengan rincian sebagai berikut.

$$\frac{\text{Jumlah kalimat yang distemming}}{\text{Jumlah data yang di ujicoba}} \times 100\%$$

#### 4.3.1 Uji Coba menggunakan *Light Stemming* dengan output *fi'il madhi*

Uji coba menggunakan *Light Stemming* dengan output *fi'il madhi* pada tabel dibawah ini, dengan menginputkan kata *fi'il mudhori'* yang ada pada Al Quran Juz 30. Hasil dari uji coba dapat dilihat pada tabel 4.1

Tabel 4.1 Uji coba menggunakan *Light Stemming* dengan output *fi'il madhi*

Data ke-	Fi'il Mudhori	Fi'il Madhi (Light Stemming)	Ketepatan	Kata Dasar (asli)
1.	سيتعلمون	علم	Tepat	علم
2.	يتكلمون	كلم	Tepat	كلم
3.	يتذكر	ذكر	Tepat	ذكر
4.	يعلمون	علمن	Kurang tepat	علم
5	تفعلون	فعل	Tepat	فعل
6	مطهره	-	-	-
7	يؤمنون	أمن	Tepat	أمن
8	نسمع	سمع	Tepat	سمع

(lanjutan tabel 4.1)

Data ke-	Fi'il Mudhori	Fi'il Madhi (Light Stemming)	Ketepatan	Kata Dasar (asli)
9	ينظرون	نظر	Tepat	نظر
10	فيقول	قال	Tepat	قال
11	تكرمون	كرم	Tepat	كرم
12	تحبُّن	حب	Tepat	حب
13	يقديِر	قدر	Tepat	قدر
14	يكذِّبُك	كذَّب	Tepat	كذَّب
15	يعلمون	علم	Tepat	علم
16	فليعبدو	عبد	Tepat	عبد
17	مأكل	-	-	-
18	وستغفر	غفر	Tepat	غفر
19	يمنعون	منع	Tepat	منع
20	يقوم	قما	Tepat	قما
21	يقول	قال	Tepat	قال
22	يحسب	حسب	Tepat	حسب
23	يخرج	خرج	Tepat	خرج
24	يدخل	دخل	Tepat	دخل
25	يُعلِّمُ	علم	Tepat	علم
26	يحفظون	حفظ	Tepat	حفظ
27	يعملون	عمل	Tepat	عمل
28	يقرؤُ	قرأ	Tepat	قرأ
29	تخبرون	خبر	Tepat	خبر
30	تكذبون	كذب	Tepat	كذب
31	يصفون	صف	Tepat	صف
32	يتفكرون	فكر	Tepat	فكر
33	يتطهرون	طهر	Tepat	طهر
34	تجهلون	جهل	Tepat	جهل
35	يعدلون	عدل	Tepat	عدل
36	يظلمون	ظلم	Tepat	ظلم
37	يبصرون	صبر	Tepat	صبر
38	يسجدون	سجد	Tepat	سجد
39	يستغفرون	غفر	Tepat	غفر
40	يعجزون	عجز	Tepat	عجز
41	يحذرون	حذر	Tepat	حذر
42	يتفكرون	فكر	Tepat	فكر
43	يبصرون	بصر	Tepat	بصر
44	يجمعون	جمع	Tepat	جمع

Dari tabel 4.1 tersebut dapat disimpulkan bahwa *stemming fi'il mudhori'* dengan output *fi'il madhi* menggunakan *light stemming* lebih mendominasi keakurasian sama dengan kata dasar asli yang ada pada kamus bahasa Arab Al-Munawwir

$$\begin{aligned} \text{Sehingga prosentase yang didapat} &= \frac{41}{44} \times 100\% \\ &= 93\% \end{aligned}$$

#### 4.3.2 Uji Coba menggunakan *Khoja Stemming* dengan output *fi'il madhi*

Uji coba menggunakan *Khoja Stemming* dengan output *fi'il madhi* pada tabel dibawah ini, dengan menginputkan kata *fi'il mudhori'* yang ada pada Al Quran Juz 30. Hasil dari uji coba dapat dilihat pada tabel 4.2

Tabel 4.2 Uji coba menggunakan *Khoja* dengan output *fi'il madhi*

Data ke-	Fi'il Mudhori	Fi'il Madhi (Khoja)	Ketepatan	Kata dasar (asli)
1.	سيتعلمون	علم	Tepat	علم
2.	يتكلمون	كلم	Tepat	كلم
3.	يتذکر	يذكر	Kurang tepat	ذكر
4.	يعلمون	علمو	Kurang tepat	علم
5	تفعلون	تفعل	Kurang tepat	فعل
6	مطهره	-	-	-
7	يؤمنون	أمن	Kurang tepat	من
8	نسمع	سمع	Tepat	سمع
9	ينظرون	نظرو	Kurang tepat	نظر
10	فيقول	فقل	Kurang tepat	قل
11	تكرمون	كم	Kurang tepat	كرم
12	تحبن	حبي	Kurang tepat	حب
13	يقدير	قدير	Kurang tepat	قدر
14	يكذبك	كذب	Tepat	كذب
15	يعلمون	علمن	Kurang tepat	علم

(lanjutan tabel 4.2)

Data ke-	Fi'il Mudhori	Fi'il Madhi (Khoja)	Ketepatan	Kata Dasar (asli)
16	فليعبدو	عبدو	Kurang tepat	عبد
17	مأكل	-	-	-
18	وستغفر	غفرا	Kurang tepat	غفر
19	يمنعون	منعن	Kurang tepat	منع
20	يقوم	قم ا	Kurang tepat	قم
21	يقول	قال	Kurang tepat	قل
22	يخسب	حسبي	Kurang tepat	حسب
23	يخرج	خارج	Kurang tepat	خرج
24	يدخل	دخل	Kurang tepat	دخل
25	يُعلمُ	أعلم	Kurang tepat	علم
26	يحفظون	حيفظ	Kurang tepat	حفظ
27	يعملون	عمل	Tepat	عمل
28	يقرؤ	قرأ	Tepat	قرأ
29	تخبرون	خاب ر	Kurang tepat	خبر
30	تكذبون	كذبو	Kurang tepat	كذب
31	يصفون	صاف	Kurang tepat	صف
32	يتفكرون	يفك ر	Kurang tepat	فكر
33	يتطهرون	طهرو	Kurang tepat	طهر
34	تجهلون	جهلي	Kurang tepat	جهل
35	يعدلون	عدل	Tepat	عدل
36	يظلمون	ظلمن	Kurang tepat	ظلم
37	يبصرون	صبر	Kurang tepat	بصر
38	يسجدون	سجود	Kurang tepat	سجد
39	يستغفرون	إستغفرون	Kurang tepat	غفر
40	يعجزون	عجزن	Kurang tepat	عجز
41	يحذرون	حذر	Tepat	حذر
42	يتفكرون	تفك ر	Kurang tepat	فكر
43	يبصرون	بصور	Kurang tepat	بصر
44	يجمعون	جموع	Kurang tepat	جمع
16	فليعبدو	عبدو	Kurang tepat	عبد
17	مأكل	-	-	-
18	وستغفر	غفرا	Kurang tepat	غفر
19	يمنعون	منعن	Kurang tepat	منع
20	يقوم	قم ا	Kurang tepat	قم
21	يقول	قال	Kurang tepat	قل

Dari tabel 4.2 di atas, dapat disimpulkan bahwa *stemming fi'il mudhori'* dengan output *fi'il madhi* menggunakan *khoja stemming* kurang tepat dengan kata dasar asli yang ada pada kamus bahasa Arab Al-Munawwir. Dan dari tabel hasil uji coba di atas, diperoleh keakurasian sebagai berikut

$$\begin{aligned} \text{Akurasi yang didapat} &= \frac{9}{44} \times 100\% \\ &= 20\% \end{aligned}$$

#### 4.3.3 Uji Coba menggunakan *Light Stemming* dengan output *Mashdar*

Uji coba menggunakan *Light Stemming* dengan output *mashdar* pada tabel dibawah ini, dengan menginputkan kata *fi'il mudhori'* yang ada pada Al Quran Juz 30. Hasil dari uji coba dapat dilihat pada tabel 4.3

Tabel 4.3 Uji coba menggunakan *Light Stemming* dengan output *mashdar*

Data ke-	Fi'il Mudhori'	Mashdar (Light Stemming)	Ketepatan	Kata dasar (asli)
1.	سيتعلمون	علماً	Tepat	علماً
2.	يتكلمون	كلماً	Kurang tepat	كلماً
3.	يتذکر	ذکرًا	Tepat	ذکرًا
4.	يعلمون	علماً	Tepat	علماً
5	تفعلون	فعلًا	Tepat	فعلًا
6	مطهره	-	-	-
7	يؤمنون	أمنًا	Kurang tepat	منًا
8	نسمع	سمعاً	Tepat	سمعاً
9	ينظرون	نظرًا	Kurang tepat	نظرًا
10	فيقول	-	-	-
11	تكرمون	كرماً	Tepat	كرماً
12	تحبُّن	حبَّ	Tepat	حبَّ
13	يقدير	قدرًا	Tepat	قدرًا

(lanjutan tabel 4.3)

Data ke-	Fi'il Mudhori'	Mashdar (Light Stemming)	Ketepatan	Kata dasar (asli)
14	يَكْذِبُكَ	كُذِّبًا	Tepat	كُذِّبًا
15	يَعْلَمُونَ	عِلْمًا	Tepat	عِلْمًا
16	فَلْيَعْبُدُوا	عِبَادًا	Tepat	عِبَادًا
17	مَأْكُلٌ	-	-	-
18	وَسْتَغْفِرُ	غَفْرًا	Tepat	غَفْرًا
19	يَمْنَعُونَ	-	-	-
20	يَقُومُ	قِمًا	Tepat	قِمًا
21	يَقُولُ	-	-	-
22	يَحْسِبُ	حِسْبًا	Tepat	حِسْبًا
23	يَخْرُجُ	خُرْجًا	Tepat	خُرْجًا
24	يَدْخُلُ	دُخْلًا	Tepat	دُخْلًا
25	يُعَلِّمُ	عِلْمًا	Tepat	عِلْمًا
26	يَحْفَظُونَ	حَفْظًا	Tepat	حَفْظًا
27	يَعْمَلُونَ	-	-	-
28	يَقْرَأُ	قِرَاءً	Tepat	قِرَاءً
29	تَخْبِرُونَ	خَبْرًا	Tepat	خَبْرًا
30	تَكْذِبُونَ	كُذِّبًا	Tepat	كُذِّبًا
31	يَصِفُونَ	صِفًا	Tepat	صِفًا
32	يَتَفَكَّرُونَ	فِكْرًا	Tepat	فِكْرًا
33	يَتَطَهَّرُونَ	طَهْرًا	Tepat	طَهْرًا
34	تَجْهَلُونَ	جَهْلًا	Tepat	جَهْلًا
35	يَعْدِلُونَ	عَدْلًا	Tepat	عَدْلًا
36	يَظْلِمُونَ	ظِلْمًا	Tepat	ظِلْمًا
37	يَبْصُرُونَ	بَصْرًا	Tepat	بَصْرًا
38	يَسْجُدُونَ	سُجْدًا	Tepat	سُجْدًا
39	يَسْتَغْفِرُونَ	غَفْرًا	Tepat	غَفْرًا
40	يَعْجِزُونَ	عَجْزًا	Tepat	عَجْزًا
41	يَحْذَرُونَ	حَذْرًا	Tepat	حَذْرًا
42	يَتَفَكَّرُونَ	فِكْرًا	Tepat	فِكْرًا
43	يَبْصُرُونَ	بَصْرًا	Tepat	بَصْرًا
44	يَجْمَعُونَ	جَمْعًا	Tepat	جَمْعًا

Dari tabel 4.3 di atas, dapat disimpulkan bahwa *stemming fi'il mudhori'* dengan output *mashdar* menggunakan *light stemming* lebih mendominasi keakurasian sama dengan kata dasar asli yang ada pada kamus bahasa Arab Al-Munawwir. Tetapi pada stemming dengan output *mashdar* terdapat 6 kata yang tidak dapat di *stemming* karena menghasilkan kata yang tidak baku dan tidak terdapat dalam kamus Arab Al-Munawwir.

$$\begin{aligned} \text{Sehingga akurasi yang didapat} &= \frac{35}{44} \times 100\% \\ &= 79\% \end{aligned}$$

#### 4.3.4 Uji Coba menggunakan *Khoja Stemming* dengan output *mashdar*

Uji coba menggunakan *Khoja Stemming* dengan output *fi'il mashdar* pada tabel dibawah ini, dengan menginputkan kata *fi'il mudhori'* yang ada pada Al Quran Juz 30. Hasil dari uji coba dapat dilihat pada tabel 4.4

Tabel 4.4 Uji coba menggunakan *Khoja* dengan output *mashdar*

Data ke-	Fi'il Mudhori	Mashdar (Khoja)	Ketepatan	Kata dasar (asli)
1.	سيتعلمون	علماً	Tepat	علماً
2.	يتكلمون	كلماً	Kurang tepat	كلماً
3.	يتذکر	ذکرًا	Tepat	ذکرًا
4.	يعلمون	علماً	Tepat	علماً
5.	تفعلون	فعلًا	Tepat	فعلًا
6.	مطهره	-	-	-
7.	يؤمنون	أمنً	Kurang tepat	أمنًا
8.	نسمع	سمعاً	Tepat	سمعاً

(lanjutan tabel 4.4)

Data ke-	Fi'il Mudhori	Mashdar (Khoja)	Ketepatan	Kata dasar (asli)
9	ينظرون	نظراً	Kurang tepat	نظاراً
10	فيقول	-	-	-
11	تكرمون	كرماً	Tepat	كرماً
12	تحبن	حباً	Tepat	حباً
13	يقدير	قدرأ	Tepat	قدرأ
14	يكذبك	كذبأ	Tepat	كذبأ
15	يعلمون	علمأ	Tepat	علمأ
16	فليعبدو	عبداً	Tepat	عبداً
17	ماكل	-	-	-
18	وستغفر	غفرأ	Tepat	غفرأ
19	يمنعون	-	-	-
20	يقوم	قما	Tepat	قما
21	يقول	-	-	-
22	يحسب	حسبأ	Tepat	حسبأ
23	يخرج	خرجأ	Tepat	خرجأ
24	يدخل	دخلأ	Tepat	دخلأ
25	يُعلم	علمأ	Tepat	علمأ
26	يحفظون	حفظأ	Tepat	حفظأ
27	يعملون	-	-	-
28	يقرؤ	قرناً	Kurang tepat	قرأ
29	تخبرون	خبرأ	Tepat	خبرأ
30	تكذبون	كذبأ	Tepat	كذبأ
31	يصفون	صفأ	Tepat	صفأ
32	يتفكرون	فكرأ	Tepat	فكرأ
33	يتطهرون	طهراً	Tepat	طهراً
34	تجهلون	جهلاً	Tepat	جهلاً
35	يعدلون	عدلاً	Tepat	عدلاً
36	يظلمون	ظلمأ	Tepat	ظلمأ
37	يبصرون	صبرأ	Tepat	صبرأ
38	يسجدون	سجدأ	Tepat	سجدأ
39	يستغفرون	غفرأ	Tepat	غفرأ
40	يعجزون	عجزأ	Tepat	عجزأ
41	يحذرون	حذرأ	Tepat	حذرأ
42	يتفكرون	فكرأ	Tepat	فكرأ
43	يبصرون	بصرأ	Tepat	بصرأ
44	يجمعون	جمعأ	Tepat	جمعأ

Dari tabel 4.4 d atas, dapat disimpulkan bahwa *stemming fi'il mudhori'* dengan output *mashdar* menggunakan *khoja*. Hasilnya sama dengan *stemming* menggunakan *light stemming*. Dan pada *stemming* ini juga terdapat 6 kata yang tidak dapat di *stemming* karena menghasilkan kata yang tidak baku dan tidak terdapat dalam kamus Arab Al-Munawwir

$$\begin{aligned} \text{Sehingga akurasi yang didapat} &= \frac{35}{44} \times 100\% \\ &= 79\% \end{aligned}$$

Dari tabel-tabel uji coba diatas, dapat disimpulkan bahwa, *Stemming* menggunakan *light stemming* dengan output *fi'il madhi* memperoleh prosentase 93% , *stemming* menggunakan *khoja* dengan output *fi'il madhi* memperoleh prosentase 20%. Sedangkan *stemming* menggunakan *light stemming* dan *khoja* dengan output *mashdar* mempunyai prosentase yang sama yaitu 79%.

#### 4.4 Tabel kebenaran

Tabel 4.5 Data *stemming* dengan *light stemming* dan *khoja* pada *fi'il madhi*

	Data Sesuai	Data tidak sesuai
Data sesuai	44 (tp)	9 (fp)
Tidak sesuai	41 (fn)	35 (tn)

Tabel 4.6 Data *stemming* dengan *light stemming* dan *khoja* pada *mashdar*

	Data Sesuai	Data tidak sesuai
Data sesuai	44	41
Tidak sesuai	35	9

Rumus akurasi diperoleh dari :

$$A = \frac{(tp + tn)}{N}$$

$$N = tp + fp + tn + fn$$

$$N = 44 + 41 + 35 + 9 = 129$$

$$A = (44 + 35) / 129 = 44,2\%$$

$$A = (44 + 9) / 129 = 41\%$$

#### 4.5 Pembahasan

Berdasarkan pengujian, akurasi yang didapatkan dari hasil pengujian aplikasi *stemming* menggunakan *light stemming* dengan output *fi'il madhi* keakurasiannya mencapai 93%, sedangkan *stemming* menggunakan *khoja* dengan output *fi'il madhi* keakurasiannya 20%. *Stemming* menggunakan *light stemming* dengan output *mashdar* keakurasiannya 79% dan *stemming* menggunakan *khoja* dengan output *mashdar* keakurasiannya sama dengan menggunakan *light stemming* yaitu 79%.

Dari prosentase keberhasilan uji coba penerjemah tersebut belum mencapai 100%, karena terdapat kalimat yang tidak dapat di stemm dengan baik oleh sistem. Ketidakberhasilan aplikasi *stemming* disebabkan karena sebagai berikut.

- a. Adanya proses *overstemming* dan *understemming* pada algoritma *Light Stemming*. Sehingga menyebabkan hasil *stemming* yang diperoleh tidak sesuai dengan yang diharapkan. *Overstemming* adalah kondisi dimana pemenggalan suatu imbuhan yang melebihi dari seharusnya. Kesalahan ini terjadi karena kata dasar tersebut menyerupai kata imbuhan. Misalnya pemotongan *prefix* ن pada kata نظر proses *stemm* akan memotong kata tersebut menjadi ظر dan kata seperti itu tidak mungkin dijumpai dalam *database*, karena huruf ن dan kata ظر adalah satu kesatuan kata yang tidak bisa dipisahkan. Demikian juga pemotongan imbuhan yang terlalu sedikit, seperti pemotongan kata ستهب. Pemotongan *prefix* س akan menghasilkan kata ت ذهب. Hasil pemotongan kata ini juga tidak mungkin akan ditemui di *database* karena kata tersebut bukan merupakan kata dasar, ت ذهب masih mempunyai *prefix* ت yang belum dipisahkan.
- b. Selain itu terdapat kata yang memiliki makna yang berbeda tetapi ejaan sama.
- c. Kendala lain yang menjadi hambatan aplikasi ini adalah kata yang bermakna ganda . Satu kata dapat memiliki makna lebih dari satu.

Di luar keterbatasan tersebut aplikasi ini dapat berjalan baik. Secara keseluruhan dari 44 kata yang di uji cobakan terdapat beberapa yang tidak sesuai dengan yang diharapkan.

#### 4.6 Integrasi *Stemming* dalam Islam

Bahasa Arab yang merupakan salah satu bahasa yang banyak digunakan di Timur Tengah. Bahasa Arab juga merupakan bahasa yang digunakan Al-Qur'an. Hal ini diperjelas dalam firman Allah SWT sebagai berikut :

وَكَذَلِكَ أَنْزَلْنَاهُ حُكْمًا عَرَبِيًّا ۗ وَلَئِنِ اتَّبَعْتَ أَهْوَاءَهُمْ بَعْدَ مَا جَاءَكَ مِنَ الْعِلْمِ مَا لَكَ مِنَ اللَّهِ مِنْ وَلِيٍّ وَلَا وَاقٍ ﴿٣٧﴾

Artinya :

*“Dan Demikianlah, Kami telah menurunkan Al Quran itu sebagai peraturan (yang benar) dalam bahasa Arab. Dan seandainya kamu mengikuti hawa nafsu mereka setelah datang pengetahuan kepadamu, Maka sekali-kali tidak ada pelindung dan pemelihara bagimu terhadap (siksa) Allah.” (QS Ar Ra’d : 37)*

Penggunaan bahasa Arab dalam Al-Qur'an karena keistimewaan bahasa Arab. Diantara keistimewaan bahasa Arab adalah sejak dahulu kala hingga sekarang bahasa Arab itu merupakan bahasa yang hidup. Bahasa Arab adalah bahasa yang lengkap dan luas untuk menjelaskan tentang ketuhanan dan keakhiratan. Dan bentuk-bentuk kata dalam bahasa Arab mempunyai tasrif (konjugasi) yang amat luas sehingga dapat mencapai 3000 bentuk perubahan.

Huruf per huruf hijaiyah dalam kata Arab, mempunyai arti sendiri. Jika salah menentukan kata dasar dalam suatu kata Arab, maka salah juga arti dari kata tersebut.

Hal ini diperjelas dengan firman Allah SWT dalam surah Yusuf yang berbunyi :

إِنَّا أَنْزَلْنَاهُ قُرْآنًا عَرَبِيًّا لَعَلَّكُمْ تَعْقِلُونَ ﴿٢١﴾

Artinya :

*“Sesungguhnya Kami menurunkannya berupa Al Quran dengan berbahasa Arab, agar kamu memahaminya.” (QS Yusuf : 2).*

Dalam surat diatas, al Quran diturunkan dalam bahasa Arab agar umat Islam memahami setiap kata dalam al Quran tersebut. Menentukan sebuah kata dasar dalam kata Arab, sangat penting karena dari kata dasar tersebut kita dapat mengetahui arti dari kata Arab yang kita baca.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 KESIMPULAN

1. *Stemming* dengan menggunakan *light stemming* lebih akurat daripada menggunakan *khoja*. Tetapi, *stemming* dengan output *mashdar* menggunakan *light stemming* maupun *khoja*, mempunyai hasil prosentase yang sama, karena *stemming* pada *mashdar* baik menggunakan *light stemming* maupun *khoja* menghasilkan kata dasar yang sama. Dan terdapat dua *fi'il mudhori'* yang tidak dapat di *stemming* baik menggunakan *light stemming* maupun *khoja stemming*. Sedangkan pada tabel uji coba *mashdar*, terdapat 6 *fi'il mudhori'* yang tidak dapat di *stemming*, karena kata tersebut menghasilkan kata Arab yang tidak baku.

2. *Stemming* menggunakan *light stemming* dengan output *fi'il madhi* memperoleh prosentase 93% , *stemming* menggunakan *khoja* dengan output *fi'il madhi* memperoleh prosentase 20%. Sedangkan *stemming* menggunakan *light stemming* dan *khoja* dengan output *mashdar* mempunyai prosentase yang sama yaitu 79%.

#### 5.2 SARAN

Berikut ini adalah beberapa saran yang dimaksudkan dengan tujuan agar pembuatan dan pembangunan aplikasi ini nantinya menjadi lebih baik lagi, yaitu :

1. Melengkapi imbuhan dan perubahan yang terjadi pada kata stemming bahasa Arabnya.
2. Diharapkan juga dapat menyertakan terjemahan agar memudahkan pengguna belajar Al-Quran.
3. Aplikasi ini hendaknya dapat dilakukan tidak hanya menggunakan fi'il madhi dan mashdar saja.
4. Pada penelitian ini, peneliti menggunakan *NetBeans*. Untuk penelitian selanjutnya dapat dikembangkan dengan menggunakan berbasis android atau berbasis web.



## DAFTAR PUSTAKA

Al-Quran al-Karim

Achmad Solichin, Gatot Wicaksono, 2013, Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta.

Al Ameer, Hayder. 2011. Arabic Light Stemmer: A New Enhanced Approach. UAE Univesity

Ash-Shobuni, Ali. 2001. At-Tibyan Fi Ulumil Qur'an.

B. Nazief and M. Adriani. 1996. Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. Technical report, Faculty of Computer Science, University of Indonesia, Depok, 1996

Nurlailah. 2009, Aplikasi Stemmer Fi'il dalam Jus 'Amma Menggunakan Porter Stemmer. Skripsi Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta.

Cios, Krzysztof J. Etc. 2007. *Data Mining A Knowledge Discovery Approach*, Springer.

Even.Yahir dan Zohar. 2002. *Introduction to Text mining. Automated Learning Group National Center For Supercomputing Aplications*. University of Illionis.

Haryatai, Tuti. 2007. *Aplikasi Kamus Bahasa Perancis – Indonesia menggunakan natural language processing dan stemming*. Skripsi Fakultas Sains dan Teknologi, Universitas Islam Negeri Starif Hidayatullah Jakarta.

Khoja, Shereen. 2012. *APT : Arabic Part-of-speech Tagger*. Lancaster University.

Ledy Agusta. 2009. Perbandingan *Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia*. Konferensi Nasional Sistem dan Informatika 2009: Bali, 14 November 2009.

Muhammad Ma'sum bin Ali. 1965. *AlamsilatutTasrifiyah*. Maktabah Syaih Salim bin Said Annabhani, Surabaya.

Mustafa, Suleiman H. 2012. Word Stemming for Arabic Information Retrieval: The Case for Simple Light Stemming.

Sutanta, Edhy. 2014. Algoritma Teknik Penyelesaian Permasalahan Untuk Komputasi. Graha Ilmu. Yogyakarta.

Trunojoyo, Husni. Buku Ajar. Information Retrieval dan Klasifikasi. <http://husni.trunojoyo.ac.id/wp-content/uploads/2010/03/Husni-IR-dan-Klasifikasi.pdf>. Diakses pada 20 Desember 2014.

