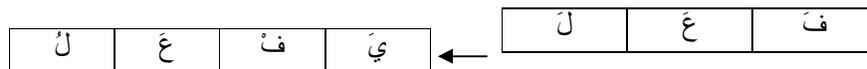


# 1. Pendahuluan

## 1.1 Latar Belakang

Indonesia merupakan negara dengan penduduk muslim terbanyak di dunia. Hasil sensus pada tahun 2000 menunjukkan kira-kira 86,1% dari 240.271.522 penduduk Indonesia adalah pemeluk Agama Islam [6]. Dalam Islam, kita diperintahkan untuk memahami Al-Qur'an untuk dapat dijadikan pedoman dalam hidup kita. Al-Qur'an merupakan *kalamullah* (perkataan Allah) yang diturunkan dalam Bahasa Arab. Sehingga bagi banyak penduduk Indonesia yang dalam keseharian menggunakan bahasa Indonesia, terkadang sulit untuk memahami arti dari kata-kata yang terdapat dalam Al-Qur'an. Salah satu upaya untuk dapat memahami arti dari kata-kata berbahasa Arab yaitu dengan mempelajari Bahasa Arab terlebih dahulu. Adapun upaya untuk dapat memahami dengan benar makna-makna yang terkandung dalam Al-Qur'an, dapat melalui buku-buku tafsir *'ulama*.

Ilmu *nahwu* merupakan ilmu dasar dalam mempelajari Bahasa Arab. Dalam ilmu *nahwu* akan dipelajari tentang jabatan kata dalam kalimat dan harakat akhirnya. Ilmu *nahwu* mengkaji tiga hal yaitu huruf, kata, dan kalimat. Adapun jenis kata/kelas kata secara umum yaitu huruf, kata benda, dan kata kerja [16]. Buah dari mempelajari ilmu nahwu adalah menjaga lisan dari kesalahan berbahasa Arab dan memahami Al-Qur'an dan hadits Nabi dengan pemahaman yang benar [10]. Ilmu *nahwu* memiliki keterkaitan dengan ilmu *sharaf*, yaitu ilmu yang mempelajari tentang bentuk kata dan perubahannya dengan penambahan maupun pengurangan huruf [3]. Perubahan makna dalam Bahasa Arab dapat terjadi jika kata tersebut diberikan imbuhan pada awal katanya. Sehingga, akan lebih memudahkan bagi kita untuk memahami arti dari kata-kata berbahasa Arab apabila kita mengetahui kata dasar, imbuhan, dan kelas kata dari suatu kata tersebut. Sebagai contoh : jika kita menemui kata يَذْهَبُ (yadzhabu – dia (laki-laki) sedang atau akan pergi – kata kerja), maka kita dapat mengetahui bahwa kata dasar dari kata يَذْهَبُ adalah kata ذَهَبَ (dzahaba – dia (laki-laki) telah pergi) berdasarkan pola :



Gambar 1-1: Contoh pola perubahan kata dalam Bahasa Arab

*Stemming* merupakan suatu proses untuk menemukan kata dasar dari sebuah kata dengan cara menghilangkan semua imbuhan baik yang terdiri dari awalan, sisipan, akhiran, maupun kombinasi dari awalan dan akhiran [8]. *Stemming* digunakan untuk mengganti bentuk suatu kata menjadi kata dasar sesuai dengan morfologi yang baik dan benar [11]. Terdapat beberapa teknik yang dapat digunakan dalam proses *stemming* teks berbahasa Arab, misalnya *dictionary technique*, *affix removal technique*, dan *morphological analysis technique*. *Dictionary technique* mengacu kepada pencarian kata-kata ke dalam kamus yang menyimpan banyak perbendaharaan kata (baik kata dasar maupun imbuhan). *Affix removal technique* merupakan suatu teknik mencari akar kata dengan

membuang imbuhan-imbuhan yang melekat pada kata tersebut. Sedangkan *morphological analysis technique* merupakan suatu teknik mencari akar kata dengan memperhatikan susunan dari pembentukan kata tersebut.

Pada tugas akhir ini, penulis akan mengimplementasikan Algoritma *Pattern-based Stemmer* yang merupakan penerapan dari *morphological analysis technique* dan *affix removal technique* untuk menentukan kata dasar dari suatu kata dalam dokumen berbahasa Arab. Selain itu, sistem juga akan melakukan penentuan kelas kata dari masing-masing kata yang berada dalam dokumen. Alasan dipilihnya metode ini karena Bahasa Arab merupakan bahasa yang kaya akan pola. Baik itu pola pembentukan kata kerja, kata benda, maupun pola imbuhan lain yang menyatakan kelas kata dari suatu kata. Sehingga akan memudahkan mencari akar kata dengan memerhatikan pola-pola atau susunan dari suatu kata.

Setelah dilakukan stemming, maka akan dilihat performansi dari stemmer, yaitu tingkat akurasi (*accuracy*) dan ICF (*Index Compression Factor*). Akurasi (*accuracy*) adalah perbandingan antara banyaknya *stem* benar dengan seluruh *term* yang ada di seluruh dokumen. Semakin tinggi akurasi, maka *stemmer* yang digunakan akan semakin baik pula. Sedangkan ICF (*Index Compression Factor*) adalah nilai dari kumpulan *stem* yang berhasil dipadatkan (direduksi) dengan *stemming*. Semakin tinggi nilai ICF yang didapatkan, maka makin bagus performansi *stemmer* yang digunakan. Selain itu, juga akan dilihat performansi dari *rules* penentu kelas kata, yang berkaitan dengan *stem* yang telah dihasilkan pada proses sebelumnya. Performansi dari *rules* ini dilihat dari akurasi, *precision*, dan *recall* dari *rules* penentu kelas kata.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang permasalahan, maka masalah yang akan dirumuskan adalah :

1. bagaimana membangun sebuah sistem yang dapat digunakan untuk mengetahui kata dasar dan kelas kata dari suatu kata dalam Bahasa Arab dengan mengimplementasikan algoritma *Pattern-based Stemmer*
2. bagaimana mengevaluasi proses stemming dan penentuan kelas kata

### Hipotesa:

Semakin banyak pola dan rule penentu kelas kata yang disimpan dalam sistem, maka performansi stemmer dan kelas kata dengan mengimplementasikan algoritma *Pattern-based Stemmer* akan semakin baik.

## 1.3 Batasan Masalah

Dalam implementasi tugas akhir ini dibatasi oleh beberapa hal, sebagai berikut :

1. Kata dalam Bahasa Arab yang akan diketahui kata dasar dan kelas katanya hanya diambil dari beberapa Surah yang berada dalam Juz 'Amma (Juz 30)
2. Kata dasar yang akan dicocokkan dengan *stem* yang dihasilkan sistem berasal dari Kamus Digital Al-Munawwir dan Kamus Al Mufid

3. Kelas kata merupakan jenis kata dalam Bahasa Arab secara umum yaitu *huruf* (kata tugas/kata bantu), *isim* (kata benda), dan *fi'il* (kata kerja)
4. Sistem hanya menentukan satu kelas kata yang lebih spesifik pada dua buah kata yang bergabung menjadi satu. Jika kata benda bergabung dengan *huruf* (kata tugas atau kata bantu), maka akan dikenali sebagai kata benda. Jika kata kerja bergabung dengan huruf, maka akan dikenali sebagai kata kerja. Apabila huruf bergabung dengan huruf, maka akan dikenali sebagai huruf
5. Sistem tidak bisa mengenali kelas kata dari suatu kata khusus yang merupakan gabungan dari huruf-huruf tertentu dengan benar. Contoh : kelas kata dari kata الم (alif lam mim) tidak akan dikenali sebagai huruf.

#### 1.4 Tujuan

Tujuan dari Tugas Akhir ini adalah :

1. Mengimplementasikan *Pattern-based Stemmer* dalam menentukan kata dasar dari suatu kata didalam teks berbahasa Arab
2. Menerapkan beberapa *rule* yang dapat menentukan kelas kata
3. Menganalisis tingkat akurasi sistem berdasarkan *output* yang dikeluarkan oleh sistem dengan menghitung jumlah *stem* dan kelas kata yang benar dibandingkan jumlah *term* dalam dokumen
4. Menganalisis nilai ICF yang dihasilkan dari algoritma *Pattern-based Stemmer* dalam melakukan *stemming* berdasarkan representasi tingkat dari kumpulan stem yang unik yang berhasil direduksi
5. Menganalisis nilai *precision* dan *recall* berdasarkan banyaknya jumlah kelas kata yang berhasil ditentukan benar, jumlah kelas kata yang salah, dan jumlah kata yang tidak dapat diketahui kelas katanya.

#### 1.5 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan digunakan adalah :

- a. Studi literatur  
Mempelajari pembentukan kata dalam Bahasa Arab. Mempelajari algoritma *stemming* yang akan digunakan dalam tahap implementasi, yaitu algoritma *Pattern-based Stemmer*.
- b. Implementasi sistem  
Perancangan perangkat lunak *stemming* untuk teks berbahasa Arab. Sistem ini akan dibangun dengan menggunakan PHP sebagai bahasa pemrograman.
- c. Testing dan Analisis Hasil  
Melakukan pengujian terhadap sistem yang telah dibuat, kemudian menganalisis performansi sistem penentu kata dasar dan kelas kata, berdasarkan *output* yang dikeluarkan oleh sistem, .
- d. Pembuatan Laporan  
Mendokumentasikan langkah-langkah yang dilakukan dalam proses pembuatan Tugas Akhir ini dalam bentuk laporan berupa Buku Tugas Akhir.

## **1.6 Sistematika Penulisan**

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut :

1. **Pendahuluan**  
Bab ini berisi uraian tugas akhir ini secara umum, meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan, dan metode penyelesaian masalah.
2. **Landasan Teori**  
Bab ini membahas mengenai teori-teori penunjang yang digunakan dalam penyelesaian tugas akhir yang berkaitan dengan proses stemming dan proses penentuan kelas kata pada suatu teks berbahasa Arab.
3. **Analisis Kebutuhan dan Perancangan Sistem**  
Bab ini berisi analisis kebutuhan sistem yang dibuat, pemodelan dari sistem dan metode yang digunakan, serta masalah-masalah yang terkait di dalamnya. Dari tahap analisis kemudian dilanjutkan ke tahap implementasi.
4. **Pengujian Dan Analisis**  
Bab ini membahas mengenai pengujian hasil implementasi yang telah dilakukan pada bab sebelumnya. Tahap ini dilanjutkan dengan tahap analisis hasil pengujian.
5. **Kesimpulan Dan Saran**  
Bab ini berisi kesimpulan dari Tugas Akhir ini dan saran-saran yang diperlukan untuk pengembangan lebih lanjut.