

# SISTEM DETEKSI JENIS KATA DALAM BAHASA ARAB DENGAN MENGUNAKAN ALGORITMA LIGHT STEMMING

<sup>1</sup> Nurul Maliya (131 065 1113)

<sup>2</sup> Mudafiq Riyan Pratama, S.Kom, M.Kom. (12 03 720)

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember  
Jln. Karimata No. 49, Telp (0331) 336728, Jember  
E-mail : [nurul.maliya69@gmail.com](mailto:nurul.maliya69@gmail.com)

## ABSTRAK

Bahasa Arab menjadi sumber utama pengetahuan tentang Islam atau sarana pokok untuk memahami isi dan kandungan kitab suci umat Islam. Karena itu, bagian-bagian yang tersusun didalamnya memiliki makna yang mendalam, termasuk susunan kalimat yang terbentuk dari *al-asmâ'* (kata benda, nama) dan *alaf'âl* (kata kerja, verba). Pada dasarnya, dalam belajar bahasa Arab harus mengetahui kata dasar dari suatu kalimat bahasa Arab tersebut. Dengan kata dasar tersebut kita dapat mengetahui jenis dan arti dari kalimat Arab. *Stemming* adalah salah satu cara yang digunakan untuk mentransformasi kata-kata dalam sebuah kalimat teks ke kata dasarnya. *Arabic Unicode* adalah kode yang dipergunakan untuk memunculkan/membangkitkan sebuah karakter Arab sehingga dikenali oleh komputer. Karakter Arab ini ada dalam bentuk desimal ataupun heksa desimal. Untuk representasi Heksa decimal *Arabic Unicode* menempati 0600 – 06FF. Dan untuk Desimal menempati 1536 – 1791. Dilakukan pengujian dengan memasukkan kalimat-kalimat bahasa arab untuk mengetahui tingkat akurasi sistem dalam mendeteksi jenis kata dalam bahasa arab. Dari percobaan yang telah dilakukan menghasilkan tingkat akurasi sitem mencapai 97,54%.

**Kata kunci :** *bahasa arab, unicode, stemming*

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Bahasa adalah alat komunikasi yang paling penting dalam berinteraksi dengan siapapun di dunia ini, banyak sekali bahasa yang tercipta, semua itu untuk mempermudah dalam berkomunikasi dengan yang lainnya. Bahasa juga merupakan alat komunikasi yang utama, kreatif, dan cepat bagi manusia untuk menyampaikan ide, pikiran dan perasaannya. Bahasa tidak mungkin terpisahkan dari kehidupan manusia, karena manusialah yang menggunakan bahasa itu sendiri untuk berinteraksi. (Hidayat, 2012)

Bahasa yang digunakan oleh Al-Qur'an adalah bahasa Arab sehingga bahasa Arab

menjadi sumber utama pengetahuan tentang Islam atau sarana pokok untuk memahami isi dan kandungan kitab suci umat Islam. Karena itu, bagian-bagian yang tersusun didalamnya memiliki makna yang mendalam, termasuk susunan kalimat yang terbentuk dari *al-asmâ'* (kata benda, nama) dan *alaf'âl* (kata kerja, verba). Masing-masing *isim* dan *fi'il* yang terdapat didalam Al-Qur'an, dengan demikian, mengandung makna tersendiri yang spesifik. (Wahidi, 2014).

Problema dalam pembelajaran bahasa arab merupakan suatu faktor yang bisa menghalangi dan memperlambat pelaksanaan proses belajar mengajar dalam bidang studi bahasa Arab. Problema tersebut muncul dari dalam bahasa Arab itu sendiri (problematika linguistik) dan non linguistik atau dikalangan

pengajar (guru) dan peserta didik itu sendiri. (Hizbullah & Mardiah, 2014)

Mempelajari bahasa Arab tidak harus melalui suatu lembaga pembelajaran bahasa Arab, tetapi juga dapat dilakukan melalui belajar kepada orang yang sudah ahli bahasa Arab, melalui buku, dan melalui sistem atau aplikasi yang bisa membantu dalam memahami dan mempelajari jenis-jenis kata dalam bahasa Arab tersebut. Aplikasi pengolahan kata dalam bahasa Arab menggunakan *unicode*.

Dalam perjalanan sejarah *unicode* merupakan standar industri yang dirancang untuk mengizinkan suatu teks atau simbol agar bisa ditampilkan kedalam tampilan yang sesuai dengan penulisan semua huruf dan simbol yang ada di dunia seperti bahasa Latin, Jepang, Arab dan lain sebagainya. *Unicode* ini diciptakan oleh organisasi bernama *Unicode Consortium* dengan misi mengkodekan semua alphabet di dunia menjadi sebuah kode sehingga kode tersebut sesuai dengan huruf atau kalimat dalam berbagai bahasa seperti bahasa Arab sambung, Latin, dan lain sebagainya. Sistem ini sanggup untuk menentukan setiap kalimat tersebut dengan rinci. (Lovins, 1968).

Untuk mempermudah dalam mengenali jenis kata dalam bahasa Arab maka digunakan metode Algoritma *Light Stemming*, *Stemming* adalah salah satu cara yang digunakan untuk mentransformasi kata-kata dalam sebuah kalimat teks ke kata dasarnya. Dengan menggunakan metode ini akan didapatkan kalimat dasar dan imbuhan yang terdapat dalam kalimat tersebut, sehingga bisa ditentukan jenis dari kalimat tersebut. Berdasarkan uraian diatas, maka penulis mengusulkan “**Sistem Deteksi Jenis Kata dalam Bahasa Arab dengan Menggunakan Algoritma Light Stemming**”.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada tugas akhir ini adalah:

1. Bagaimana menentukan jenis kata dalam Bahasa Arab dengan

menggunakan Algoritma *Light Stemming*?

2. Berapa nilai dari akurasi Algoritma *Light Stemming* dalam penentuan jenis kata Bahasa Arab?
3. Apa kelebihan dan kekurangan Algoritma *Light Stemming* dalam penentuan jenis kata Bahasa Arab?

## 1.3. Batasan Masalah

Agar pembahasan tidak menyimpang dari topik permasalahan yang ada, maka diperlukan batasan-batasan sebagai berikut:

1. Mendeteksi *isim* (kata benda, kata sifat, kata keterangan) dan *fi'il* (kata kerja).
2. Tidak mendeteksi teks Arab berupa inputan gambar.
3. Tidak mendeteksi kesalahan kata atau kalimat.

## 1.4. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah :

1. Menentukan jenis kata dalam Bahasa Arab dengan menggunakan Algoritma *Light Stemming*.
2. Menghitung tingkat akurasi dari Algoritma *Light Stemming* dalam menentukan jenis kata didalam Bahasa Arab.
3. Menganalisa kelebihan dan kekurangan dalam mengimplementasi Algoritma *Light Stemming* untuk menentukan jenis kata didalam Bahasa Arab.

## 1.5. Manfaat Penelitian

Dengan mengimplementasikan deteksi jenis kata dalam Bahasa Arab diharapkan dapat memperoleh manfaat sebagai berikut :

1. Memberi kemudahan kepada pemula untuk mengetahui kata *isim* dan kata *fi'il* dalam Bahasa Arab.
2. Menambah pengetahuan dan wawasan peneliti tentang pentingnya mempelajari Bahasa Arab dan bagaimana cara mudah untuk mengetahui kata *isim* dan kata *fi'il* dalam Bahasa Arab sehingga dapat

mengembangkan peneliti yang lebih luas dimasa yang akan datang.

## 2. TINJAUAN PUSTAKA

### 2.1. Bahasa Arab

Bahasa Arab sebagai bahasa agama, bahasa integrasi dunia Arab (dan Islam), dan bahasa resmi PBB, tetap penting, tidak hanya dalam pengembangan kajian keislaman tetapi juga dalam pengembangan ilmu pengetahuan dan peradaban. Bahasa Arab dan peradaban Islam sepanjang sejarahnya tidak dapat dipisahkan, bagaikan dua sisi dari mata uang. Di satu sisi bahasa Arab bisa berkembang maju karena Al-Qur'an, dan di sisi lain, bahasa Arab perlu dikembangkan sebagai ilmu karena dibutuhkan untuk melayani kajian Al-Qur'an. Bahkan para *qurrâ'* adalah juga para ahli bahasa. Ilmu-ilmu bahasa Arab dan keislaman bisa berkembang, antara lain, karena adanya inspirasi dan motivasi dari Al-Qur'an yang berbahasa Arab. (Wahab, 2014)

Bahasa Arab dan Al-Qur'an merupakan kesatuan yang tidak bisa dipisahkan satu dengan yang lainnya. Dalam belajar Al-Qur'an bahasa Arab adalah syarat mutlak yang harus dikuasai, demikian halnya dengan belajar bahasa Al-Qur'an berarti belajar bahasa Arab. Bahasa Arab termasuk salah satu diantara bahasa yang banyak digunakan di dunia, karena banyak yang menggunakannya, maka bahasa Arab ini menjadi bahasa Internasional dan diakui oleh dunia. (Hidayat, 2012)

### 2.2. Konsep Bahasa Arab

Kata dalam bahasa arab pada dasarnya : Menurut jumhur ulama ilmu nahwu telah membagi kata kepada tiga, yaitu *isim*, *fi'il* dan *harf*. (Sudirman, 2015). Berikut ini pembagian tiga jenis kata tersebut:

#### 2.2.1 *Isim* (اسم)

Kata benda (*isim*) yaitu suatu kata menunjukkan makna tidak terikat dengan waktu (lampau, sekarang, akan datang).

1) Tanda-tanda kata benda (اسم):

- Didahului oleh *alif-lam* ( ا ل )
- Memiliki harokat *tanwin* ( ً ٍ ٌ )
- Didahului oleh huruf *jar*

مِنْ ، إِلَى ، عَنْ ، عَلَى ، فِي ، رَبِّ ، .. )  
بِ .. ، كِ .. ، لِ

• *Mudhof/Mudhofilaih* ( مضاف/مضاف إليه )

2) Pembagian Kata Benda (اسم)

Kata benda apabila dipandang dari jumlahnya, maka terbagi tiga yaitu: benda jumlahnya satu dinamakan *mufrad* ( مفرد ), benda jumlahnya dua dinamakan *tastniyah* ( مثني ), benda jumlahnya tiga ke atas dinamai *jamak* ( جمع ).

a. Cara pembentukan kata benda *isim* yang jumlahnya satu ( مفرد ) menjadi dua ( مثني ) ada dua cara :

- 1.) Menambah *alif* ( ا ) dan *Nun* ( ن )
- 2.) Menambah *Ya* ( ي ) dan *Nun* ( ن )

b. Pembagian benda yang jumlahnya tiga ke atas ( جمع ). Kata benda jumlahnya tiga ke atas disebut ( اسم جمع ), namun ( اسم جمع ) ini terbagi tiga yaitu :

1.) Benda yang jumlahnya lebih dari tiga jenisnya laki-laki dinamakan *jamak mudzakkar* ( جمع مذكر ) dengan ditandai dengan ada tambahan:

- *Waw* ( و ) dan *Nun* ( ن ), contoh مُسَلِّمُونَ
- *Ya* ( ي ) dan *Nun* ( ن ), contoh مُسَلِّمِينَ

2.) Benda yang jumlahnya lebih dari tiga jenisnya perempuan dinamakan *jamak mu'annats* ( جمع مؤنث ) dengan ditandai:

- *Alif* ( ا ) dan *Ta* ( ت ), contoh مُسَلِّمَاتُ

3.) Benda yang jumlahnya lebih dari tiga jenisnya bisa laki-laki dan perempuan dan bentuk ini dinamai, kata tak beraturan atau dinamakan *jamak taksir* ( جمع تكسير ).

Contoh :

- *كُتُبٌ* jamaknya menjadi كُتُبٌ
- *أَقْلَامٌ* jamaknya menjadi أَقْلَامٌ
- *بُيُوتٌ* jamaknya menjadi بُيُوتٌ

c. Kata benda dipandang dari maknanya. Kata benda dipandang dari segi maknanya terbagi dua :

- 1.) Kata benda yang memberi pengertian umum dinamakan *isim nakiroh* (اسم الكثرة)
- 2.) Kata benda yang memberi peringatan khusus dinamakan *isim ma'rifah* (اسم المعرفة). Kata benda yang termasuk kata benda khusus, adalah kata benda yang tergolong atau disebut dengan:
  - Kata ganti  
Artinya: Kata ganti yang digunakan sipembicara ketika pada posisi orang; pertama, kedua dan orang ketiga.
  - Kata Benda Alam  
Artinya : Suatu lafaz yang memberi petunjuk langsung tanpa tujuan keluar dari lafad yang sama, apakah tanda bagi manusia, hewan dan kota. *Isim* alam, terbagi tiga :
    1. Nama
    2. Panggilan
    3. Gelar
  - Kata Tunjuk  
Artinya: Suatu lafaz memberi makna yang jelas. Kata tunjuk ada tiga tingkatan yaitu : Dekat, Sedang, Jauh
  - Kata Benda Penghubung  
Artinya: Sesuatu pemberian makna tertentu dengan Perantara satu kalimat yang disebutkan sesudahnya kata yang dinamai kata penghubung.
  - Kata Benda Yang *Beralif Lam*  
Artinya: Suatu kata benda yang dimasuki alif lam memberi makna khusus.
  - Kata benda yang disandarkan  
Dua benda dikaitkan maknanya yang pertama memiliki yang kedua.
  - Kata benda dari panggilan  
Artinya: Kata benda yang berada sesudah huruf *nida*.

### 2.2.2. *Fi'il* (الفعل)

Kata kerja (*Fi'il*) yaitu kata yang menunjukkan terjadinya suatu peristiwa pada

waktu tertentu (lampau, sekarang dan akan datang). Kata kerja bila diperhatikan dari segi waktunya maka terbagi 3 yaitu:

- 1) Kata Kerja (الفعل) yang menunjukkan lampau dinamakan *Madhi* (الماضي)
- 2) Kata Kerja (الفعل) yang menunjukkan waktu sedang dan akan datang dinamakan *Madhari'* (المضارع)
- 3) Kata Kerja (الفعل) yang menunjukkan perintah dinamakan *Amr* (الأمر)

*Fi'il* dapat dibedakan dari *isim*, dan *huruf*, yaitu dengan masuknya:

- 1.) *Qad* (قد) artinya sungguh  
Contoh : قَدْ أَفْلَحَ الْمُؤْمِنُونَ (Sesungguhnya beruntunglah orang-orang yang beriman. (Al-Mu'minun:1)
- 2.) *Sin* (س) artinya akan ; berwaktu dekat  
Contoh : سَأَدْخُلُ الْفَصْلَ (Saya akan masuk kelas)
- 3.) *Saufa* (سَوْفَ) artinya akan ; bermakna jauh  
Contoh : سَوْفَ تَعْلَمُونَ (Kamu kelak akan mengetahui)
- 4.) *Ta ta'nis* (ت) : ta' tanis yang berharokat sukun  
Contoh : جَاءَتْ حَلِيمَةٌ (Halimah telah datang)

### 2.2.3. *Harf* (حرف)

*Harf* yaitu suatu kata tidak punya makna kecuali dikaitkan dengan kata yang lainnya.

### 2.3. Unicode

*Arabic Unicode* adalah kode yang dipergunakan untuk memunculkan/membangkitkan sebuah karakter Arab sehingga dikenali oleh komputer. Karakter Arab ini ada dalam bentuk desimal ataupun heksa desimal. Untuk representasi Heksa decimal *Arabic Unicode* menempati 0600 – 06FF. Dan untuk Desimal menempati 1536 – 1791. (Amin, 2006)

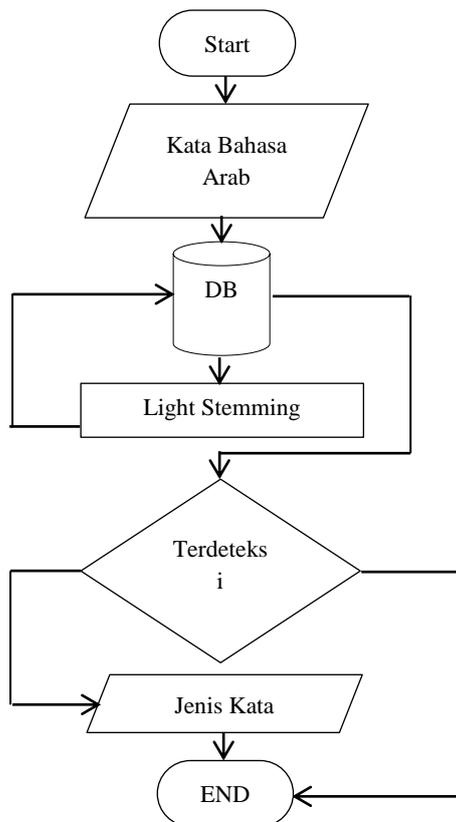
### 2.4. Algoritma Light Stemming

Algoritma *light stemmer* adalah algoritma *Stemming* yang hanya menghilangkan imbuhan depan (*prefix*) dan imbuhan belakang (*suffixs*). Sesuai dengan rule yang telah ditentukan dari karakter yang

paling banyak digunakan, adapun karakter yang paling banyak digunakan tersebut dibagi menjadi 2 yaitu karakter imbuhan depan dan imbuhan belakang yang sering dipakai. *Light Stemming* bisa diinisialisasikan menjadi 3 inisial karakter yang paling sering digunakan 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. (Chen and Gay, 2003).

### 3. Perancangan Sistem

Pada tahap ini dilakukan perancangan sistem menggunakan *flowchart* dan UML (*Unified Modeling Language*) yang meliputi *use case* dan *activity diagram* dengan tujuan untuk menghasilkan rancangan sistem penentuan jenis kata yang dapat digunakan oleh user dalam mempelajari bahasa Arab.



Gambar 3.2 *Flowchart System*

### 4. Pembahasan

#### 4.1. Implementasi Sistem

Implementasi merupakan tahapan dimana aplikasi siap dioperasikan pada keadaan yang sebenarnya sehingga dapat diketahui apakah program yang dibuat benar-benar dapat menghasilkan sebuah keluaran yang sesuai dengan tujuan yang diinginkan.

### 5. Penutup

#### 5.1 Kesimpulan

Berasarkan rancangan, pembangunan dan implementasi aplikasi pengenalan dan pembelajaran bahasa arab dapat ditarik kesimpulan sebagai berikut :

1. Aplikasi yang telah dibuat tergantung pada kata dasar yang diinputkan dalam database.
2. Aplikasi tidak dapat mendeteksi perubahan kata, dari kata dasar ke bentuk jamak yang terdapat imbuhan ditengah kata.
3. Aplikasi ini telah lolos uji akurasi dengan rata-rata tingkat akurasi sebesar 97,54%

#### 5.2 Saran

Dengan segala kelebihan yang terdapat pada tugas akhir ini, tidak terlepas dari kekurangan yang tentunya sangat diharapkan adanya saran-saran yang mendukung proses penyempurnaan. Penulis ingin memberikan beberapa saran yang mungkin dapat membantu dalam pengembangan Tugas Akhir ini sebagai berikut :

1. Pada penelitian ini sistem hanya mendeteksi jenis kata dalam bahasa arab meliputi *isim* dan *fi'il*, dalam pengembangan selanjutnya bisa ditambahkan jenis kata *harf* atau lebih detail dengan mendeteksi macam-macam dari jenis kata *isim* dan *fi'il*.
2. Penelitian ini menggunakan metode algoritma *light stemming* berbasis dekstop, sebaiknya pada penelitian selanjutnya dikembangkan menjadi berbasis android.

## Daftar Pustaka

1. Amin, M. M. 2006. Pembangkitan Karakter Arab Menggunakan Unicode Untuk Pembuatan Aplikasi Kamus Bahasa Arab-Indonesia Berbasis Web. *Jurnal Informatika*, 6(1), 13-25.
2. Chen, A. and Gey, F. 2002. *Building an Arabic Stemmer for Information Retrieval*. USA: University Of California.
3. Hidayat, N. S. 2012. Problematika Pembelajaran Bahasa Arab. Dosen Fakultas Tarbiyah dan Keguruan UIN Suska Riau. *Jurnal Pemikiran Islam*, 37(1), 82-88.
4. Hizbullah, N., & Mardiah, Z. 2015. Masalah Pengajaran Bahasa Arab di Madrasah Aliyah di Jakarta. Program Studi Sastra Arab, Fakultas Sastra, Universitas Al Azhar Indonesia. *Jurnal Al-Azhar Indonesia Seri Humaniora*, 2(3), 189-198.
5. Lovins, J. B. 1968. *Development of a Stemming Algorithm*. Cambridge: Massachusetts Institute of Technology.
6. Sudirman. 2015. *Tata Bahasa Arab Praktis*. Yogyakarta: Azza Grafika. hlm. 3-31.
7. Wahab, M. A. 2014. Peran bahasa arab dalam pengembangan ilmu dan peradaban Islam. Program Pascasarjana FITK UIN Syarif Hidayatullah Jakarta. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 1(1), 1-20.
8. Wahidi, R. 2014. Pola-pola penggunaan kata Isim dan Fi'il dalam Al-Qur'an. Fakultas Ilmu agama Islam, Universitas Islam Indragiri Tembilahan Riau. *Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 1(2), 253-266.
9. Wulandari, A., & Romadhony, A. 2011. *Pattern-Based Stemmer Analysis and Implementation on Arabic Text*. Fakultas Informatika Institut Teknologi Telkom, Bandung. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*.