

## SISTEM DETEKSI STRUKTUR KALIMAT BAHASA ARAB MENGGUNAKAN ALGORITMA LIGHT STEMMING

Mudafiq Riyan Pratama<sup>1</sup>, Muhammad Yunus<sup>2</sup>

<sup>1,2</sup>Program Studi Rekam Medik, Jurusan Kesehatan, Politeknik Negeri Jember

Jl. Mastrip, Jember, Jawa Timur

Email : [mudafiq.riyan@polije.ac.id](mailto:mudafiq.riyan@polije.ac.id), [m.yunus@polije.ac.id](mailto:m.yunus@polije.ac.id)

### ABSTRAK

Untuk memahami bahasa Arab, perlu mempelajari ilmu nahwu. Ilmu nahwu merupakan ilmu yang mempelajari tata kalimat bahasa Arab. Dengan belajar ilmu nahwu, bisa membedakan subjek, predikat, dan objek dalam kalimat. Salah satu bidang dalam dunia komputer yang mengkaji tentang pengolahan bahasa manusia adalah *Natural Language Processing* (NLP) yaitu pemrosesan bahasa alami manusia melalui analisis sintaksis terhadap struktur kalimat. Salah satu metode dalam menganalisis sintaksis kalimat adalah *stemming*. Salah satu varian dari algoritma *stemming* adalah algoritma *Light Stemming*. Algoritma *Light Stemming* adalah algoritma yang hanya menghilangkan imbuhan depan (*prefix*) dan imbuhan belakang (*suffix*). Berdasarkan pengujian yang dilakukan, algoritma *light stemming* mampu mendeteksi ilmu nahwunya dengan tingkat keberhasilan sebesar 82.22%. Tingkat kegagalan 17.78% terjadi karena pada kata yang tidak memiliki imbuhan akan dideteksi secara otomatis jenisnya *fi'il* (kata kerja), padahal faktanya bisa jadi jenisnya adalah *isim* (kata benda), serta gagalnya hasil deteksi juga dikarenakan tidak mampu melakukan *stemming* pada kata yang imbuhan ditengah, karena memang proses *light stemming* hanya menghilangkan imbuhan depan dan imbuhan belakang.

**Kata kunci:** bahasa arab, ilmu nahwu, natural language processing, light stemming, suffix, prefix

### ABSTRACT

To understand Arabic, it is necessary to study nahwu theory. Nahwu theory is the study of structure of Arabic sentences. By learning nahwu, being able to distinguish subjects, predicates and objects in sentences. One of the fields in the computer that studies about human language processing is NLP (Natural Language Processing), which is natural human language processing through syntactic analysis of sentences structure. One method to analyzing syntactic of sentences is stemming. One variant of the stemming algorithm is Light Stemming algorithm. Light Stemming algorithm is an algorithm that only removes prefix and suffix. Based on the tests conducted, the light stemming algorithm is able to detect nahwu with an accuracy rate of 82.22%. The 17.78% failure rate occurs because words that do not have an affix will be detected automatically the type is *fi'il* (verb), even though the fact may be that the type is *isim* (noun), and the failure of the detection results is also due to not being able to stemming the words that have affix in the middle (infix), because indeed the process of light stemming only eliminates the suffix and prefix.

**Keywords:** arabic, nahwu theory, natural language processing, light stemming, suffix, prefix

Author Korespondensi (Mudafiq Riyan Pratama)

Email : [mudafiq.riyan@polije.ac.id](mailto:mudafiq.riyan@polije.ac.id)

### I. PENDAHULUAN

Bahasa Arab adalah salah satu bahasa tertua di dunia. yang juga termasuk bahasa dengan peringkat keempat yang paling banyak digunakan di dunia, dengan jumlah populasi pengguna sebanyak 315 juta orang di tahun 2018 dan bahasa yang tersebar di lebih dari 25 negara [1]. Bahasa Arab juga menjadi bahasa resminya umat Islam, karena dalam ibadahnya umat Islam

seperti bacaan shalat, doa, dan Al-Qur'an menggunakan bahasa Arab.

Dalam memahami bahasa Arab, perlu mempelajari ilmu nahwu. Ilmu nahwu merupakan ilmu yang mempelajari tata kalimat berbahasa arab yang benar, dengan belajar ilmu nahwu maka kita bisa menguasai bahasa arab dengan baik serta bisa mengerti perubahan-perubahan kata tersebut, walaupun tidak memakai harokat [2]. Dengan memahami

ilmu nahwu, seseorang akan bisa membedakan antara pelaku (*fa'il*) dan objek (*maf'ul bih*). Dengan memahami ilmu nahwu, seseorang akan mengenali keadaan akhir dari suatu kata, apakah ia bisa berubah akhirnya atautkah tetap. Dengan ilmu nahwu pula seseorang akan bisa membaca akhir kata dengan benar, apakah ia harus dibaca *dhammah*, *fat-hah*, atau *kasrah* [3].

Salah satu bidang dalam dunia komputer yang mengkaji tentang pengolahan bahasa manusia adalah *Natural Language Processing* (NLP). NLP merupakan komputasi linguistik yang berkaitan dengan interaksi antara komputer dan manusia melalui bahasa alami [4]. Dalam memproses bahasa alami manusia, perlu adanya proses yang dapat melakukan analisis sintaksis terhadap struktur kalimat. Salah satu metode dalam menganalisis sintaksis kalimat adalah *stemming*.

Algoritma *Stemming* adalah proses melakukan normalisasi pada kata seperti penghapusan prefiks (awalan), sufiks (akhiran), dan standarisasi tanda-tanda lainnya, sehingga proses ini mereduksi semua kata menjadi bentuk umum atau bentuk dasarnya [5]. Salah satu varian dari algoritma *stemming* adalah algoritma *Light Stemming*. Algoritma *Light Stemming* adalah algoritma yang hanya menghilangkan imbuhan depan (*prefix*) dan imbuhan belakang (*suffix*) [6].

## II. METODOLOGI

### 1) Kamus Ilmu Nahwu

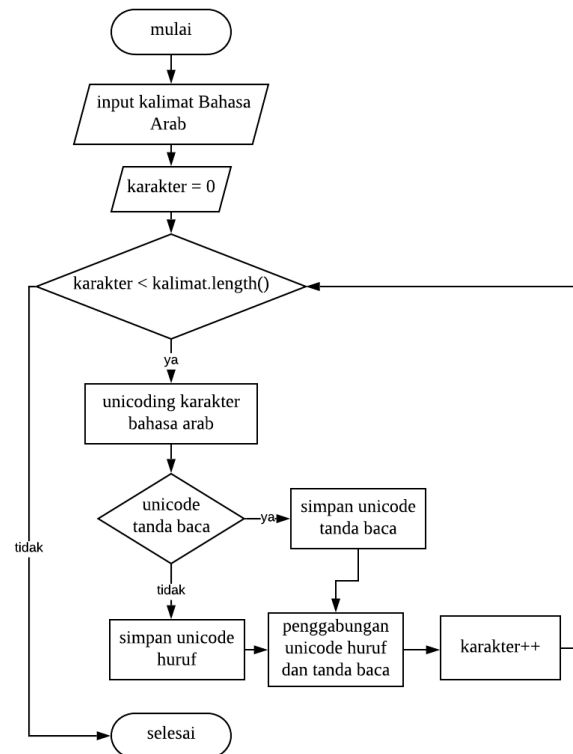
Langkah pertama dalam penelitian ini adalah membuat KamusIlmuNahwu yang berisi kombinasi *unicode* per kata. Contoh isi kamus kata ilmu nahwu tersebut adalah sebagai berikut:

```
1573;0;1604;0;1609;0;=Harfun Al-Jarr  
1601;0;1610;0;1607;0;=Harfun Al - A'tof  
1581;0;1578;0;1609;0;=Isim  
1571;0;1606;0;1575;0;=Isim Dhomir  
1608;1575;1604;=Isim Ma'rifah  
1576;1575;1604;=Isim Ma'rifah  
1601;1575;1604;=Isim Ma'rifah  
1573;0;1584;0;=Harfun Tanbih  
1575;0;1606;0;=Harfun Syartin  
1573;0;1604;0;1575;0;=Harfun Istisna'  
1603;1614;1610;1618;=Harfun An-Nasab
```

Kamus kata ilmu nahwu tersebut disimpan dalam file txt agar proses pencariannya lebih cepat. Semakin banyak isi kamus kata tersebut, dapat menentukan tingkat akurasi pendeteksian struktur kalimat bahasa Arab.

### 2) Unicoding

Dalam memproses kalimat berbahasa Arab, diperlukan proses yang harus dilakukan sebelum melakukan *stemming*, yaitu proses *unicoding*. Pada proses *unicoding*, kalimat bahasa Arab dikonversi menjadi kode tertentu sesuai dengan kode standar komputer agar dapat diproses pengolahan katanya. Alur proses *unicoding* dijelaskan dalam Gambar 1.



Gambar 1. Flowchart Unicoding

Pada proses *unicoding*, kalimat yang diinput diubah menjadi *unicode* terlebih dahulu, kemudian dipisahkan antara *unicode* tanda baca seperti  $\text{ﷲ}$ ,  $\text{ﷻ}$  dan lain-lain dengan *unicode* huruf, kemudian hasil masing-masing *unicode* ditampung dalam sebuah variabel dengan ditambahkan tanda titik-koma (;) di akhir huruf sehingga tidak menghilangkan keaslian *unicode* per karakter tersebut. Tabel *unicode* tanda baca dan huruf bahasa Arab dapat dilihat pada Tabel 1 dan Tabel 2 dibawah ini.

**Tabel 1.** Unicode Tanda Baca Arab

Tanda Baca	Unicode
◌ْ	1617
◌َ	1614
◌ِ	1611
◌ُ	1615
◌ُ	1612
◌ِ	1616
◌ِ	1613
◌ِ	1618

membedakannya dengan jenis kata lain. Cirinya antara lain:

- Diawali dengan al ( ال ).  
Contoh : الهدى، الكتاب
- Diakhiri dengan tanwin.  
Contoh : كِتَابٌ
- Diawali salah satu huruf jar.  
Huruf jar tersebut adalah :  
على، إلى، ك، ل، ب، في، رب، عن، من
- Diawali diawali oleh huruf nida.  
Contoh : يا ادم = يا + ادم
- Disandarkan (diidlofatkan) kepada kata benda lainnya, kebanyakan untuk menunjukkan pemilihan.

**Tabel 2.** Unicode Huruf Arab

Unicode	Huruf	Unicode	Huruf
1575	ا	1591	ط
1576	ب	1592	ظ
1577	ة	1593	ع
1578	ت	1594	غ
1579	ث	1600	-
1580	ج	1601	ف
1581	ح	1602	ق
1582	خ	1603	ك
1583	د	1604	ل
1584	ذ	1605	م
1585	ر	1606	ن
1586	ز	1607	ه
1587	س	1608	و
1588	ش	1609	ى
1589	ص	1610	ي
1590	ض	1591	ظ

b) Fi'il

Fi'il (kata kerja) adalah kata yang menunjukkan suatu pekerjaan yang berkaitan dengan waktu, baik lampau, sekarang, maupun yang akan datang. Oleh karena itu, kata kerja (fi'il) dalam bahasa arab juga di bedakan menjadi tiga macam yaitu: *fi'il madi* (lampau), *fi'il mudari'* (sekarang dan akan datang) dan *fi'il amr* (perintah). Ciri khas dari fi'il sebagai berikut :

- Diawali قَدْ .  
Contoh : قَدْ أَفْلَ .
- Diawali salah satu huruf mudlori' ( ن، أ، ت، ي ).
- Diawali س atau سوف .  
Contoh : تعلمون، سوف سيقول : .
- Diakhiri dengan Ta' taknis (ت).  
Contoh : ربحت
- Diakhiri dengan Ta' fa'il (ت، تم، ت، تما، تن).
- Diakhiri dengan Naa fa'il (نا).  
Contoh : انزلنا:

c) Harf

Harf adalah setiap kata yang tidak akan memiliki makna kecuali menyertainya dengan kata yang lain. Harf atau huruf dipandang sebagai kata tugas atau kata penghubung [7].

**3) Rule Ilmu Nahwu**

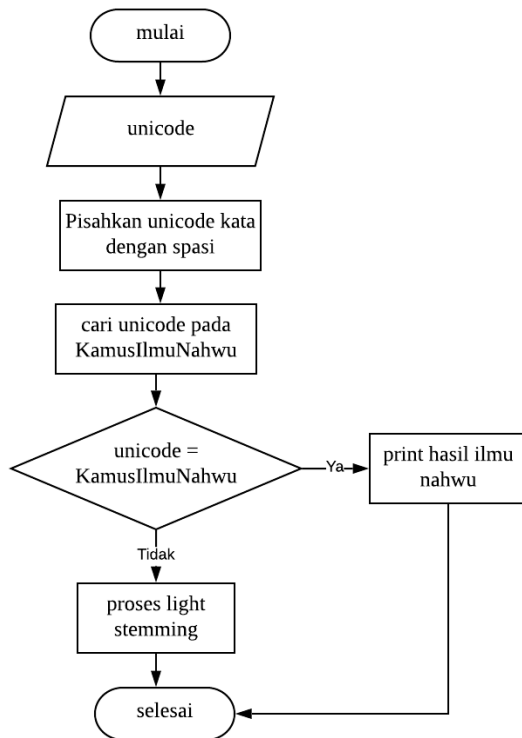
Setelah proses unicoding dan *light stemming*, maka langkah selanjutnya adalah membaca hasil *stemming* sesuai dengan rule ilmu nahwu. Berikut penjelasan mengenai rule ilmu nahwu:

a) Isim

Isim (kata benda) adalah setiap kata yang mempunyai arti benda, baik konkret maupun abstrak, tanpa ada unsur waktu di dalamnya. Isim memiliki ciri khusus yang

#### 4) Pemeriksaan pada Kamus Kata

*Stemming* yang dilakukan dengan tambahan proses dalam menentukan kata yang harus *distemming* yaitu pemilihan kata baku, agar kata yang *distemming* merupakan kata yang memang harus dicari kata dasarnya, bukan sebaliknya. Dengan proses pemilihan kata ini, proses *stemming* lebih efektif dan lebih cepat. Alur pemilihan kata dijelaskan melalui *flowchart* pada Gambar 2 di bawah ini.



**Gambar 2.** Flowchart Pemeriksaan pada Kamus Kata

Alur dari proses *filtering* kata dijelaskan di bawah ini:

- Input diambil dari hasil proses *unicoding*.
- Sistem melakukan proses pemisahan per kata dengan dipisahkan spasi, kemudian unicode kata yang sudah dipisahkan, disimpan dalam variabel *array* sesuai dengan urutan kata tersebut.
- Selanjutnya sistem mencari kombinasi *unicode* kata dasar berdasarkan KamusIlmuNahwu yang telah dibuat sebelumnya. Jika terdapat *unicode* kata di dalam kamus kata, maka sistem akan langsung menampilkan hasil jenis katanya,

jika tidak sesuai maka diproses *stemming* menggunakan algoritma *light stemming*.

#### 5) Light Stemming

Proses *Light Stemming* adalah proses *Stemming* yang lebih difokuskan pada imbuhan depan (*prefix*) dan imbuhan belakang (*suffix*), jadi *Light Stemming* ditekankan pada proses penghapusan imbuhan depan dan belakang secara *simultan* sesuai dengan rule dalam operasi *Light Stemming* berikut:

- Jika kata tersebut jumlah hurufnya 4 atau lebih, hapus 3 karakter pertama pada kata jika karakternya adalah:  
وال , بال , فال , كال , وائل , مال , ال , سال , لال
- Jika kata tersebut jumlah hurufnya 4 atau lebih, hapus 2 karakter pertama jika karakternya adalah:  
ال , وا , يا , لل , وم , وت , وب , لا , سي , وس , وي , ول , كما , فا
- Jika kata tersebut jumlah hurufnya 4 atau lebih, dan diawali dengan huruf و (waw), hapus و pada karakter pertama.
- Jika kata tersebut jumlah hurufnya 4 atau lebih, dan diawali dengan huruf ب (ba') atau ل (lam), maka hapus ب (ba') atau ل (lam) pada kata tersebut.
- Menghapus secara *recursive* 2 huruf pada belakang kata tersebut, jika kata tersebut jumlah hurufnya 4 atau lebih dan 2 huruf belakang tersebut termasuk dalam daftar ini:  
ون , ات , ان , ين , تن , تم , كن , كم , هن , يا , ني , وا , ما , نا , هم , ية , ها
- Menghapus secara *recursive* 1 huruf pada belakang kata tersebut, jika kata tersebut jumlah hurufnya 3 atau lebih dan 1 huruf belakang tersebut adalah: ت , ي , ه , ة

Dari rule di atas akan dihasilkan kata dasar berdasarkan imbuhan depan dan imbuhan belakang, tetapi tidak menutup kemungkinan juga bahwa ada perbedaan arti pada sebuah kata setelah di *stemming*. Sebagai contoh adalah kata الملكة yang artinya adalah seorang ratu, setelah ada penghilangan menjadi ملك yang artinya adalah seorang raja.

Dari rule di atas, maka menghasilkan daftar imbuhan depan dan imbuhan belakang dalam ilmu nahwu sesuai Tabel 3 dan Tabel 4 di bawah ini:

**Tabel 3.** Unicode Imbuhan Depan (*prefix*)

Imbuhan	Unicode
وال	1608;1575;1604;
بال	1576;1575;1604;
فال	1601;1575;1604;
كال	1603;1575;1604;
ولل	1608;1604;1604;
مال	1605;1575;1604;
ال	1575;1575;1604;
سال	1587;1575;1604;
لال	1604;1575;1604;
ال	1575;1604;
وا	1608;1575;
لل	1604;1604;
وم	1608;1605;

هن	1607;1606;
ان	1575;1606;
با	1576;1575;
ما	1605;1575;
ت	1578;
وا	1608;1575;
ني	1606;1610;
يا	1610;1575;
ية	1610;1577;
ي	1610;
ه	1607;
ب	1576;
ة	1577;
تم	1578;1605;
تن	1578;1606;

**Tabel 3 Lanjutan.** Unicode Imbuhan Depan (*prefix*)

Imbuhan	Unicode
وت	1608;1578;
وب	1608;1576;
لا	1604;1575;
سى	1587;1610;
وس	1608;1587;
وى	1608;1610;
ول	1608;1604;
كا	1603;1575;
فا	1601;1575;

**Tabel 4.** Unicode Imbuhan Belakang (*suffix*)

Imbuhan	Unicode
يا	1610;1575;
ون	1608;1606;
ين	1610;1606;
ات	1575;1578;
نا	1606;1575;
كم	1603;1605;
كن	1603;1606;
هم	1607;1605;
ها	1607;1575;

Dari identifikasi imbuhan depan dan imbuhan belakang pada proses *Light Stemming* di atas didapatkan sebuah rule yang digunakan untuk menentukan ilmu nahwu yang dapat dilihat pada Tabel 5, Tabel 6, dan Tabel 7:

**Tabel 5.** Rule Imbuhan pada *Fi'il Madhi*

Imbuhan Depan	Imbuhan Belakang
kosong	تن
kosong	تم
kosong	نا
kosong	تما
kosong	ت
kosong	ا
kosong	يا
kosong	وا
kosong	ما
kosong	ها

**Tabel 6.** Rule Imbuhan pada *Fi'il Mudhari'*

Imbuhan Depan	Imbuhan Belakang
وت	kosong
لا	ksoong
سى	kosong
وس	kosong
وى	kosong

سال	kosong
وت	ان
لا	ان
سى	ان
وس	ان
وى	ان
سال	ان
وت	ون
لا	ون
سى	ون
وس	ون
وى	ون
سال	ون
وت	ا
لا	ا
سى	ا
وس	ا
وى	ا
سال	ا
وت	ن
وى	ن

Tabel 7. Rule Imbuhan pada *Isim*

Imbuhan Depan	Imbuhan Belakang
ال	Kosong
وال	Kosong
بال	Kosong
فال	Kosong
كال	Kosong
مال	Kosong
ال	Kosong
لال	Kosong
لل	Kosong
ولل	Kosong
ول	Kosong
كا	Kosong
فا	Kosong
با	Kosong
وب	Kosong
ب	Kosong

Dengan teridentifikasinya rule tersebut maka sistem akan lebih mudah mendeteksi ilmu nahwu pada setiap kata, rule dibuat berdasarkan proses light Stemming, pada beberapa imbuhan sudah di gabungkan dengan tambahan harfun yang melekat pada kata nya, jadi sudah termasuk mengidentifikasi kata sebelumnya.

Pada harfun, identifikasi langsung berdasarkan kecocokan kata, karena harfun merupakan kata yang paten jika berada di tengah kalimat tanpa menggandeng pada kata tersebut, jadi langsung diberikan hasil. Tidak di berikan rule yang signifikan.

### III. HASIL DAN PEMBAHASAN

Untuk melakukan uji coba proses light stemming pada kalimat bahasa Arab, maka diperlukan sebuah perangkat lunak. Perangkat lunak yang digunakan untuk pengujian dibangun menggunakan bahasa pemrograman Java. Kemudian dilakukan 3 kali pengujian pada kalimat yang berbeda.

#### 1) Pengujian 1

Pengujian 1 dilakukan terhadap kalimat yang terdiri dari 50 kata, sebagai berikut:

لم يكن الصبي يذعن إذعانا. كان تحرشها به كزيت صبت فوق صفيح ساخن. يزداد هياجها، ويرشح صدره بالحقن، مقاوما تسلطها ما أمكنه. بعضها، وقد يدمي شفتها بلكمة تشفي بعض غليله. فتنتهقر وهي تلعن هذا الصبي، الوغد. حينئذ، ينساب إلى كيانه شعور أقرب إلى الغبطة، يحيله على حالة من الدعة والصفاء

Detail hasil pendeteksian paragraf di atas ditampilkan dalam Tabel 8 di bawah ini.

Tabel 8. Hasil Pengujian 1

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
1	لم	لم	Harfun Al-Jazm	
2	يكن	يكن	Fiil Mudhari'	
3	الصبي	صبي	Isim	
4	يذعن	يذعن	Fiil Mudhari'	
5	إذعانا	اذعا	Fiil Madhi	
6	كان	كان	Fiil	
7	تحرشها	تحرشها	Fiil Mudhari'	
8	به	به	Fiil	Harf Jar
9	كزيت	كز	Isim Maushul	
10	صبت	صبت	Fiil	

11	فوق	فوق	Fiil	Isim
12	صفيح	صفيح	Fiil	Isim
13	ساخن	ساخن	Fiil	
14	يزداد	يزداد	Fiil Mudhari'	
15	هياجه،	هياجه،	Fiil	
16	يرشح	يرشح	Fiil Mudhari'	
17	صدره	صدر	Isim Mufrod Tunggal	
18	بالحنق،	حنق،	Isim	
19	مقاوما	مقاو	Fiil Madhi	
20	تسلطها	تسلطها	Fiil Mudhari'	
21	ما	ما	Harfun Nafyi	
22	أمكنه	امكن	Isim Mufrod Tunggal	
23	بعضها،	بعضها،	Fiil Mudhari'	
24	قد	قد	Harfun Al-Jazm	
25	يدمي	يدمي	Fiil Mudhari'	
26	شفتها	شفت	Fiil Madhi	
27	بلكمة	لكم	Isim	
28	تشفي	تشفي	Fiil Mudhari'	
29	بعض	بعض	Fiil	Isim
30	غليله	غليل	Isim Mufrod Tunggal	
31	فتتقهقر	فتتقهقر	Fiil	
32	هي	هي	Isim Mufrad	
33	تلعن	تلعن	Fiil Mudhari'	
34	هذا	هذا	Fiil	Isim Isyarah
35	الصبي،	صبي،	Isim	
36	الوعد	وعد	Isim	
37	حينئذ،	حينئذ،	Fiil	Isim

Tabel 8 Lanjutan. Hasil Pengujian 1

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
38	ينساب	ينساب	Fiil Mudhari'	
39	إلى	الى	Harfun Al-Jarr	
40	كيانه	كيان	Isim Mufrod Tunggal	
41	شعور	شعور	Fiil	
42	أقرب	اقرب	Fiil	
43	إلى	الى	Harfun Al-Jarr	
44	الغبطة،	غبطة،	Isim	
45	يحبيله	يحبيله	Fiil Mudhari'	
46	على	على	Harfun Al-Jarr	
47	حالة	حال	Isim Mu'annas	
48	من	من	Harfun Al-	

			Jarr	
49	الدعة	دعة	Isim	
50	الصفاء	صفاء	Isim	

Hasil percobaan 1 terdapat 6 kesalahan dari 50 kata yang diinputkan. Sehingga kata yang benar adalah 44 kata.

## 2) Pengujian 2

Pengujian 2 dilakukan terhadap kalimat yang terdiri dari 42 kata sebagai berikut:

إحدى قدميه كعجلة مطاطية فارغة كلياً من الهواء، تعلق بجسده كطرف غريب، لا تنفعه في شيء، بل تعيق حركته ولهوه باستمرار. لطالما خامرته فكرة الفكك منها، فحاول ذلك عبثاً، بيديه لأول مرة، ثم بأظفاره وأسنانه فيما بعد، بلا فائدة، ودائماً بلا فائدة

Hasil deteksi kalimat, dituangkan dalam Tabel 9 di bawah ini.

Tabel 9. Hasil Pengujian 2

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
1	إحدى	احدى	Fiil	Isim
2	قدميه	قدمي	Isim mufrod tunggal	
3	كعجلة	كعجل	Isim Mu'annas	

Tabel 9 Lanjutan. Hasil Pengujian 2

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
4	مطاطية	مطاط	Isim Mu'annas	
5	فارغة	رغة	Isim	
6	كلياً	كلياً	Fiil	Isim
7	من	من	Harfun Al-Jarr	
8	الهواء	هواء	Isim	
9	تعلق	تعلق	Fiil Mudhari'	
10	بجسده	جسد	Isim	
11	كطرف	كطرف	Fiil	
12	غريب	غريب	Fiil	
13	لا	لا	Harfun Al-Jawab	

14	تتفعه	تتفعه	Fiil Mudhari'	
15	في	في	Harfun Al-Jarr	
16	شيء	شيء	Fiil	Isim
17	بل	بل	Harfun Al-'Ataf	
18	تعيق	تعيق	Fiil Mudhari'	
19	حركته	حركت	Isim Mufrod Tunggal	
20	لهوه	لهو	Isim Mufrod Tunggal	
21	باستمرار	ستمرار	Isim	
22	لطالما	طال	Fiil	Isim
23	خامرته	خامرت	Isim Mufrod Tunggal	
24	فكرة	فكر	Isim Mu'annas	
25	الفكاك	فكاك	Isim	
26	منها	منها	Fiil	Harfun
27	فحاول	فحاول	Fiil	Isim
28	ذلك	ذلك	Fiil	Isim
29	عبثا	عبثا	Fiil	Isim
30	بيديه	يدي	Isim	
31	لأول	لاول	Fiil	Isim
32	مرة	مرة	Fiil	Isim
33	ثم	ثم	Harfun Al-'Ataf	
34	بأظافره	اظافر	Isim	
35	أسنانه	اسنان	Isim Mufrod Tunggal	
36	فيما	فيما	Fiil	
37	بعد	بعد	Fiil	Isim

Tabel 9 Lanjutan. Hasil Pengujian 2

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
38	بلا	بلا	Fiil	Harfun Al-Jawab
39	فائدة	ئدة	Isim	
40	دائما	دائ	Fiil Madhi	
41	بلا	بلا	Fiil	Harfun Al-Jawab
42	فائدة	ئدة	Isim	

Hasil pengujian 2 terdapat 13 kesalahan dari 42 kata yang diinputkan. Sehingga kata yang benar adalah 29 kata.

### 3) Pengujian 3

Pengujian 3 dilakukan terhadap kalimat yang terdiri dari 43 kata, sebagai berikut:

تراجعت للوراء وصدورها يعلو وينخفض،  
والشرر يقدح من عينيها. لا يمكنها أن تأمن  
على الغرفة بعد اليوم وهذا الشقي بها، كما لا  
يمكنها أن تأمن على حياته إن تركته يلهو  
خارجا خوفا من أن تلدغه حية من الحيات  
الكثر التي تستوطن الدغل القريب

Hasil deteksi kalimat, dituangkan dalam Tabel 10 di bawah ini.

Tabel 10. Hasil Pengujian 3

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
1	تراجعت	تراجعت	Fiil Mudhari'	
2	للوراء	وراء	Isim	
3	صدورها	صدر	Fiil Madhi	
4	يعلو	يعلو	Fiil Mudhari'	
5	ينخفض،	ينخفض،	Fiil Mudhari'	
6	الشرر	شرر	Isim	
7	يقدح	يقدح	Fiil Mudhari'	
8	من	من	Harfun Al-Jarr	
9	عينيها	عين	Isim Maushul	
10	لا	لا	Harfun Al-Jawab	

Tabel 10 Lanjutan. Hasil Pengujian 3

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
11	يمكنها	يمكنها	Fiil Mudhari'	
12	أن	ان	Harfun Nahyi	
13	تأمن	تأمن	Fiil Mudhari'	



14	على	على	Harfun Al-Jarr	
15	الغرفة	غرف	Isim Ma'rifah	
16	بعد	بعد	Fiil	Isim
17	اليوم	يوم	Isim	
18	هذا	هذا	Fiil	Isim
19	الشقي	شقي	Isim	
20	بها،	بها،	Fiil	Harfun Al - Jarr
21	كما	كما	Fiil	Isim
22	لا	لا	Harfun Al-Jawab	
23	يمكنها	يمكنها	Fiil Mudhari'	
24	أن	ان	Harfun Nahyi	
25	تأمن	تامن	Fiil Mudhari'	
26	على	على	Harfun Al-Jarr	
27	حياته	حيات	Isim MUfrod Tunggal	
28	إن	ان	Harfun Syartin	
29	تركته	تركته	Fiil Mudhari'	
30	يلهو	يلهو	Fiil Mudhari'	
31	خارجا	خارجا	Fiil	
32	خوفا	خوفا	Fiil	
33	من	من	Harfun Al-Jarr	
34	أن	ان	Harfun Nahyi	
35	تلدغه	تلدغه	Fiil Mudhari'	
36	حية	حية	Fiil	Isim
37	من	من	Harfun Al-Jarr	
38	الحيات	حيا	Isim Ma'rifah	
39	الكثير	كثير	Isim	
40	التي	الت	Isim Maushul	

Tabel 10 Lanjutan. Hasil Pengujian 3

No.	Kata Asli	Kata Dasar	Hasil Deteksi	Koreksi
41	تستوطن	تستوطن	Fiil Mudhari'	
42	الدغل	دغل	Isim	

43	القريب	قريب	Isim	
----	--------	------	------	--

Hasil pengujian 3 terdapat 5 kesalahan dari 43 kata yang diinputkan. Sehingga kata yang benar adalah 38 kata.

#### 4) Evaluasi Hasil Pengujian

Dari 3 kali pengujian di atas, dapat dihitung tingkat akurasi dengan rumus:

$$\frac{\text{jumlah kata yang benar}}{\text{jumlah keseluruhan kata}} \times 100\%$$

Maka hasil perhitungan tingkat akurasi adalah sebagai berikut:

$$\frac{44 + 29 + 38 \text{ kata}}{50 + 42 + 43 \text{ kata}} \times 100\% = 82.22\%$$

Dari hasil pengujian, tingkat akurasi algoritma *Light Stemming* terhadap deteksi struktur kalimat bahasa Arab adalah sebesar **82.22%**. Maka tingkat *error* atau kesalahan dalam deteksinya bernilai **17.78%**.

Kesalahan-kesalahan yang terjadi pada hasil deteksi, terjadi karena beberapa hal, yaitu:

- Jika kata tidak memiliki imbuhan, baik imbuhan depan maupun imbuhan belakang, maka sistem secara otomatis mendeteksi jenis kata tersebut merupakan *fi'il*, padahal pada kenyataannya tidak selalu *fi'il*, karena bisa jadi kata tersebut *isim*. Sebagai contoh pada kata: **صرخ** pada peletakan katanya **صرخ** bisa bermakna *fi'il madhi* karena bisa berakhiran fathah, tetapi pada peletakan kata yang berbeda, kata ini bisa berakhiran tanwin, pada saat kata ini menjadi tanwin, maka kata ini termasuk *isim*.
- Gagalnya hasil deteksi juga dikarenakan tidak mampu melakukan *stemming* pada kata yang imbuhan ditengah, karena memang proses *light stemming* hanya menghilangkan imbuhan depan dan imbuhan belakang.

#### IV. SIMPULAN DAN SARAN

Dalam memproses bahasa Arab dengan komputer, maka diperlukan proses *unicoding*

yaitu mengkonversi atau mengubah karakter bahasa Arab menjadi kode tertentu sesuai dengan kode standar komputer yang disebut sebagai *unicode* agar dapat diproses lebih lanjut. Proses *unicoding* ini dilakukan pada huruf maupun pada tanda bacanya.

Algoritma *Light Stemming* dapat mendeteksi struktur kalimat bahasa Arab dengan tingkat akurasi yang baik yaitu sebesar 82,22%. Kesalahan sebesar 17.78% dalam deteksi terjadi karena 2 hal, yaitu: (1) kata yang tidak memiliki imbuhan secara otomatis terdeteksi sebagai *fi'il*, padahal kata tanpa imbuhan tidak selalu *fi'il*, bisa jadi kata tersebut berjenis *isim*, (2) tidak mampu melakukan *stemming* pada kata yang imbuhan ditengah. Hasil deteksi struktur kalimat kata bahasa Arab terdiri dari 3 jenis, yaitu: *isim*, *fi'il*, dan *harf*, meskipun ada sub-sub jenisnya seperti *isim ma'rifah*, *fi'il mudhari'*, *harfun al-jar*, dan lain-lain.

#### REFERENSI

- [1] Gary F. Simons and Charles D. Fenning, "Ethnologue: Languages of the World, Twenty-first edition". Dallas, Texas: SIL International, 2018. Online version: <https://www.ethnologue.com/21>.
- [2] Mukhlis, "Dasar-Dasar Bahasa Arab". Yogyakarta: Pustaka Insan Madani, 2007.
- [3] Ari Wahyudi, "Langkah-Langkah Untuk Bisa Membaca Kitab Arab Gundul", Muslim.Or.Id, 13 April 2013. [Online]. Tersedia: <https://muslim.or.id/13164-langkah-langkah-untuk-bisa-membaca-kitab-arab-gundul.html> [Diakses: 26 September 2019].
- [4] R. Raman Nair and L. Sulochana Devi, "Sankrit Informatics: Informatics for Sankrit Studies and Research", Kerala: Centre for Informatics Research and Development, 2011.
- [5] Maqbool R. Al-Maimani, Ahmed Al-Naamany, Ahmed Zaki Abu Bakar, "Searching For Arabic-based Contents: An Approach Towards Using Arabic Derivatives", in *International Conference on Communication, Computer and Power (ICCCP)*, February 2009.
- [6] Aito Chen and Fredric Gey, "Building Arabic Stemmer for Information Retrieval". Berkeley: University of California. 2003.
- [7] Muh. Haris Zubaidillah, "Pengantar Ilmu Nahwu Belajar Bahasa Arab Sampai Bisa". Hulu Sungai Utara: Amuntai Hemat.