

Smart I'rab: Smart Aplication for Arabic Grammar Learning

**Syd. Ali Zein Farmadi, Ali Ridho Barakbah,
Entin Martiana Kusumaningtyas**

Department of Information and Computer Engineering
Electronic Engineering Polytechnic Institute of Surabaya
Email: farmady@student.eepis-its.edu, ridho@eepis-its.edu , entin@eepis-its.edu

Abstract

Arabic grammar, known as nahwu, is necessary to comprehend the Holy Qur'an that is completely written in Arabic. However, many people get trouble to study this skill because there are various kinds of word formation and sentences that may be created from a single verb, noun, adjective, subject, predicate, object, adverb or another formation. This research proposes a new approach to identify the position and word function in Arabic sentence. The approach creates smart process that employs Natural Language Processing (NLP) and expert system with modeling based on knowledge and inference engine in determining the word position. The knowledge base determines the part of speech while the inference engine shows the word function in the sentence. On processing, the system uses 82 templates consisting of 34 verb templates, 34 subject pronouns, 14 pronouns for object or possessive word. All the templates are in the form of char array for harakat (vowel) and letters which become the comparators for determining the part of speech from input word sentence. Output from the system is an i'rab (the explanation of word function in sentence) written in Arabic. The system has been tested for 159 times to examine word and sentence. The examination for word that is done 117 times has not made any error except for the word that is really like another word. While the detection for word function in sentence that is done 42 times experiment, there is no error too. An error happens when the part of speech from the word being examined is not included in the system yet, influencing the following word function detection.

Keywords: I'rab, Arabic grammar, NLP, expert system, knowledge base, inference engine.

1. INTRODUCTION

Nowdays, there are many students from religious high school and houses of Koranic studies who want to study Islam. One of the skills is nahwu, the principles of arabic language. Nahwu studies about word function

in a sentence. Nahwu makes the arabic translation easier to understand the meaning and to avoid misinterpretation on the bias meaning. Principelly, nahwu is used to comprehend the holy qur'an of musleem book that is written in arabic. Nahwu has a branch of knowledge called as i'rab meaning the transformation of suffix to give a word distinctive use in sentence such as lafdhz, (spoken word) and taqdir, (predicted word) [4]. So the i'rab can show the function word in sentence which gives readers clear explanation for easy understanding.

2. RELATED WORKS

There were some researchers such as Harmain, El-Khatib dan Lakas[1]who have made softwares about arabic language. They use NLP (Natural Language Processing) modul and database access to determine the part of speech. Besides, El-Kourdi et all [2]made a system by using algoritm Naive Bayes to clarify an arabic document . Furthermore, El Halees [3]used a method with an association rule to clarify arabic language document. But, these studies do not highlight the arabic grammar.

3. ORIGINALITY

This research proposes a new approach for identifying the position and part of speech in arabic sentence by developing smart process involving NLP (Natural Language Processing) and expert system with a modelling on knowledge base and inference engine to determine the part of speech. The knowlgele base together with char array and string database creates the part of speech, while the inference engine shows the word function in sentence.

4. SYSTEM DESIGN

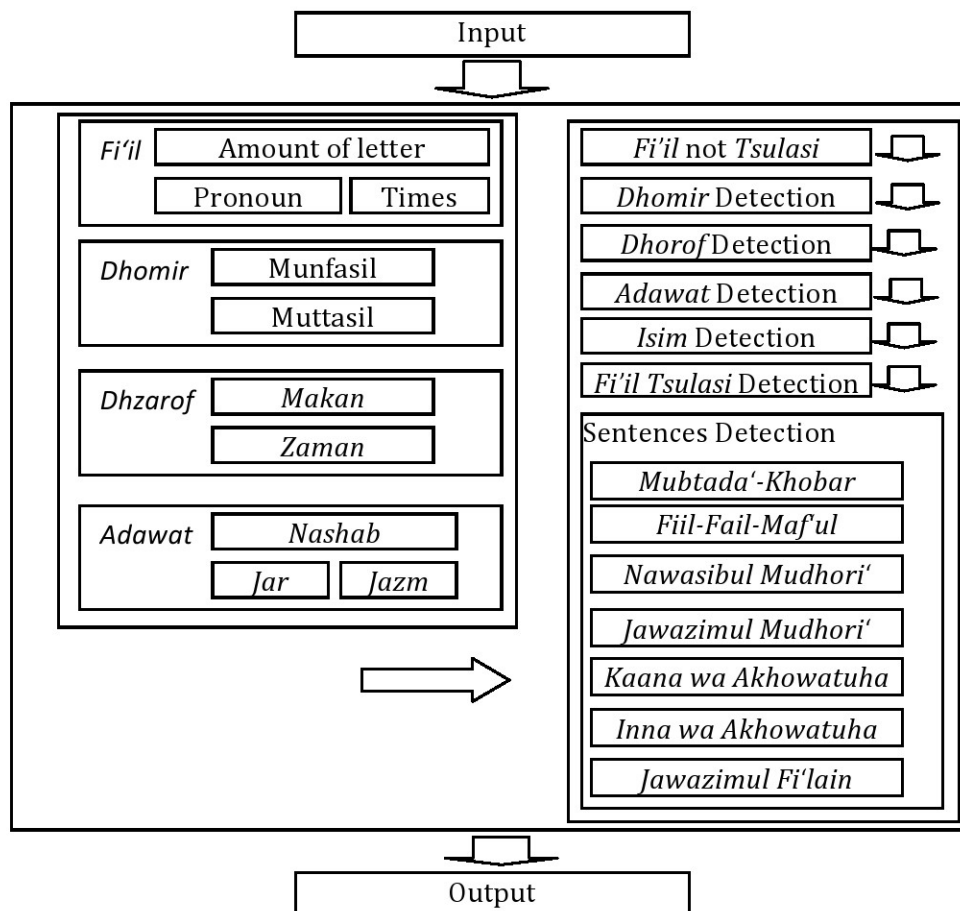
Arabic grammar has several branches of knowledge. One of them is nahwu that is useful to determine the word function in an arabic sentence. The nahwu itself has another branch that is called as i'rab. I'rab means to change the suffic of word to distinct the use in sentence. Two of the i'rab are lafdhz (spoken word) and taqdir (predicted word) [4].

This research proposes a new approach to identify the types and parts of speech in the sentence in arabic language by developing smart process that involves NLP (Natural Language Processing), and expert system with modelling based on knowledge base and inference engine. The knowledge base creates the types of word, while inference engine creates the function word in the sentence.

NLP is used to form an expert system., the branches of NLP that are applied in this research are morphology and syntaxes. Morphology is the knowledge about word and word formation. The morphology gives clear distinction among the derivatives [5]. Syntaxes are knowledge about the way how arrange words in making sentences [5]. In this research, morphology is used to form knowledge base for detecting a word. However, the syntaxes are used by the inference engine to detect a sentence.

Meanwhile the expert system is a tool with software developed to provide flexible system for knowledge designer to solve specific problems [6].

This research proposes a smart process involving both Knowledge Base and Inference Engine. The knowledge base consists of templates functioning as comparators for all kinds of word. Does the word belong to fi'il (verb), dhomir (pronoun), dhzaraf (adverb) or adawat (word that changes the part of speech). The Inference Engine contains the process for detecting part of speech which is then determined to find out the function of the word in a sentence. The Architecture System proposed can be seen on Figure 1.



Gambar 1. Our proposed system architecture

4.1 Knowledge Base

In this research, facts are classified into four kinds of word, namely fi'il, dhomir, dhzaraf, and adawat.

4.1.1 فعل (Fi'il) means verb.

The verb is limited only to *fi'il shahih* that is classified into three parts:

1. The time when an event happens is classified into:
 - a. ماض (Madhi): past.

- b. مضارع (*Mudhori*): now and future.
 c. امر (*Amr*): imperative.
2. There are many letters that can change word into verb which then becomes wazan (norm). These norms are:
- a. ثلاثي (*Tsulatsi*): a word consisting of three letters and the pattern only uses vowel /a/. In this study the three types of *wazan tsulatsi* are introduced.
- b. رباعي (*Ruba'i*): a word consisting four letters including the addition of vowel /i/ and another letter. In this study the three types of *wazan ruba'I* are introduced.
- c. خماسي (*Khumasi*): a word consisting of five letters. In this study the five types of *wazan khumasi* are introduced.
- d. سداسي (*Sudasi*): a word consisting of six letters. In this study the six types of *wazan sudasi* are introduced.
3. The pronoun of subject is implied in it.

The *template* for verb division can be seen on Table 1 and Table 2.

Tabel 1. Template of letter for *fi'il madhi*

Wazan	Huruf				
	0	1	2	3	4
Fa'ala					
Fa'ila					
Fa'ula					
Fa'ala (with tasydid)					
Af'ala	ا				
Faa'ala		ا			
If'alla	ا				
Infa'ala	ا	ن			
Ifta'ala	ا		ت		
Tafa'ala	ت				
Tafaa'ala	ت		ا		
Istaf'ala	ا	س	ت		

Tabel 2. Template *harakat* for *fi'il madhi*

Wazan	Harakat				
	0	1	2	3	4
Fa'ala	اَ	اَ	اَ		
Fa'ila	اِ	اِ	اِ		
Fa'ula	اُ	اُ	اُ		
Fa'ala (dengan tasydid)	اَ	اُ	اَ		
Af'ala	اَ	اُ	اَ		
Faa'ala	اَ	اَ			
If'alla	اِ	اِ	اِ	اِ	
Infa'ala	اِ	اِ	اِ	اِ	
Ifta'ala	اِ	اِ	اِ	اِ	
Tafa'ala	اَ	اَ	اَ	اَ	
Tafaa'ala	اَ	اَ	اَ		
Istaf'ala	اِ	اِ	اِ	اِ	اِ

4.1.2 ضمير (*Dhomir*) means the pronoun for (first, second, and third person), in which it implies:

1. *Munfashil* (pronoun that can stand alone without being linked to the following word). *DhomirMunfashil* can easily be recognized by the string of letters.
2. *Muttashil* (that is linked to the following word). This includes:
 - a. When the pronoun position is before *fi'il*, it becomes the pronoun of the object.
 - b. When the pronoun position is before *isim* (noun), it becomes possessive pronoun.

The template for *dhomir muttashil* can be seen on Table 3.

Table 3. Template of *dhomir muttashil*

JenisDhomir	Huruf (Length - x)					Harakat (Length - x)				
	L-1	L-2	L-3	L-4	L-5	L-1	L-2	L-3	L-4	L-5
Huwa 1	هـ					وُ				
Huwa 2	هـ					وِ				
Humaa (L) 1	ا	م	هـ			وُ	وُ			
Humaa (L) 2	ا	م	هـ			وِ	وِ			
Hum 1	م	هـ				وُ	وُ			
Hum 2	م	هـ				وِ	وِ			
Hiya	ا	هـ				وُ				
Humaa (P) 1	ا	م	هـ			وُ	وُ			
Humaa (P) 2	ا	م	هـ			وِ	وِ			
Hunna 1	ن	هـ				وُ	وُ	وُ		
Hunna 2	ن	هـ				وِ	وِ	وِ		
Anta	ك					وُ				
Antumaa (L)	ا	م	ك			وُ	وُ			
Antum	م	ك				وُ	وُ			
Anti	ك					وِ				
Antumaa (P)	ا	م	ك			وُ	وُ			
Antunna	ن	ك				وُ	وُ	وُ		
Ana 1	ي					وُ				
Ana 2	ي	ن				وُ	وِ			
Ana 3	ي					وِ	وِ			
Nahnu	ا	ن				وُ				

4.1.3 ظرف (*Dhzaraf*) has functions that determine the time and place of which it implies:

1. *DhzarafMakan* (determining the place)
2. *DhzarafZaman* (determining the time)

4.1.4 ادوات (*Adawat*) has functions to change the position in a sentence. The functions are limited to the instrument that can stand alone (separated by space). The *adawat* contains:

1. *Nashab* (the instrument that changes the part of speech into word ending with *fathah*)
2. *Jar* (the instrument that changes the part of speech into word ending with *kasrah*)
3. *Jazm* (the instrument that changes the part of speech into word ending with *sukun*)

4.2 Inference Engine

Executes the detection of word function according to the following chronological steps:

1. Detection is intended to verb that is *nottsulatsi* (including *rubā'i*, *khumasi*, *orsudasi* detected from its pattern).
2. Detection for *dhomir* (pattern detection).
3. Detection for *dhzaraf* (*string database* detection).
4. Detection for *adawat* (*string database* and pattern detection).
5. Detection for *isim*.
6. Detection for a verb *tsulatsi*.
7. Detection for syntax. In this research, the detection is intended to the formation of 7 types of sentence in Arabic language, that is:
 - a. *Mubtada'-Khabar*. It is concluded when a sentence is begun with *isim* of which the *isim* has *dhommahordhommatain* as ending *harakat*, the word becomes *mubtada'*. Then, when the *mubtada'* is followed by another *isim* having ending *dhommahordhommatain* as final *harakat*, the word becomes *khabar mubtada'*.
 - b. *Fi'il-Fa'il-Maf'ul*. It is concluded when a sentence is begun with *fi'il*, and followed by *isim* with final *harakat* in *dhommahordhommatain* that becomes *fa'il* (subject), then followed by *isim* having final *harakat* *fathah* or *fathatain* becoming *maf'ul bih* (object).
 - c. *Nawaashibul Mudhori'*. It is concluded when a sentence is begun with *adawatnashab*, the sentence is followed by *fi'il mudhori'* having ending sound vowel /a/, *fathah*.
 - d. *Jawaazimul Mudhori'*. It is concluded when a sentence is begun with *adawat jazm*, the sentence is followed by *fi'il mudhori'* having ending sound vowel /u/, *sukun*.
 - e. *Kaana wa Akhawatuha*. It is concluded when a sentence is begun with the word *kaana* or word having function as relatives of *kana*, the word is followed by *isim* having ending sound as vowel /u/, *dhommah* atau *dhommatain*, called as *isim kaana* which is then followed by another word having ending sound vowel /a/, *fathah* or *fathatain* later called as *khabar kaana*.
 - f. *Inna wa Akhawatuha*. It is concluded when a sentence begins with word *inna* or word having function as *inna* relatives, the word is followed by *isim* having ending sound vowel /a/, *fathah* or *fathatain* later called as *isiminna*, then it is followed by word having ending

sound /u/, *dhommahordhommatain* which is then called as *khobar inna*.

- g. *Jawaazimul Fi'lain*. It is concluded when a sentence is begun with *adawatjazm*, the word will be followed by *fi'il mudhori'* having ending sound as vowel /u/ *sukun*. It is then followed by the *fi'il mudhori'* of which it has similar ending sound as *sukun*.

5. EXPERIMENTS AND ANALYSIS

The research has already done 159 experiments consisting of 117 times for words and 42 for sentences.

5.1 Experiment for words

The word experiment is divided into 26 experiments for *fi'ilmadhi* (12 wazan, 14 dhomir), 26 experiments for *fi'ilmudhori'* (12 wazan, 14 dhomir, 17 experiments for *fi'il 'amr* (11 wazan, 6 dhomir), 14 experiment for *dhomirmusfashil*, 28 experiments for *dhomirmuttashil* (14 for pronouns of object, 14 for possessive pronouns), 2 experiments for *dhzaraf* (eating and period), 3 experiments for *harf* (*hasb*, *jar*, *jazm*) and 1 experiment for *isim* (noun). However, the error of experiment refers to the following:

1. The subject pronoun for *fi'ilmadhi* and *fi'il 'amr*, *dhomirantumamuannats* has the same form with *dhomirantumamudzakkar*.
2. The subject pronoun for *fi'ilmudhori'* has made two errors such as:
 - a. *Dhomir anta* has similar form to *dhomir hiya*.
 - b. *Dhomir antuma mudzakkar* dan *dhomir antuma muannats* has similar form to *dhomir huma muannats*.
3. The *dhomir muttashil* dan *dhomir munfashil* has made two errors such as:
 - a. *Dhomirhumamuannats* has similar form to *dhomir hum mudzakkar*.
 - b. *Dhomirantumamuannats* has similar form to *dhomirantumamudzakkar*.

5.2 Experiment for sentences

The sentence experiment is done for 12 times using common sentence according to the criteria tested, and 30 times using sentences taken from the Holy Qur'an.

Errors happen because the detection word is not added yet in the system and this can disturb the detection function in the following word. The result of experiment can be seen in Table 4.

Surah inAl-Qur'an	Trial	Mistake	Arguments
An- Nas:4	<p>مِنْ = حرف الجر نَسْرٌ = اسم مجرور بِمِنْ بالكسرة الظاهرة وهو مضاف الجار والمجرور متعلق بمحذوف تقديره كائن في محل رفع المبتدأ الْوَسْوَسُ = مضاف اليه مجرور بالكسرة الظاهرة اِحْتِسَابٌ = صفة لالْوَسْوَسِ مجرورة بمثله بالكسرة الظاهرة</p>	Are not found	الخناس become the characters of الوسوس because they are isimma'rifah having position as majrur
Al-'Aadiyat:6	<p>اِنَّ = حرف مشبه بالفعل للتوكيد تنصب الاسم وترفع الخبر الْإِنْسَانَ = اسم إنَّ منصوب بالفتحة الظاهرة لِيُرِيَهُ لَكُنْتُ = خبر إنَّ مرفوع بالضممة الظاهرة</p>	cannot be detected	Because there is aharfjar connecting to isim. While theharf is linked yet error cannot be detected
Al-'Aadiyat:7	<p>وَ = حرف العطف أَلَيْهِ = حرف مشبه بالفعل للتوكيد تنصب الاسم وترفع الخبر هُ = ضمير متصل مبني على الضمة في محل نصب اسم إنَّ عَلَى = حرف الجر ذَلِكَ = اسم مجرور بعلى بالكسرة الظاهرة وهو مضاف كَ = ضمير متصل مبني على الفتحة في محل جر مضاف اليه لَتَسْتَهْتِئُ = خبر إنَّ مرفوع بالضممة الظاهرة</p>	ذلك detected as isimmajrur andmudhof	Because ذلك is not added yet to the system, the system searches its own system in different condition.
Al-Kautsar:1	<p>إِنَّ = حرف مشبه بالفعل للتوكيد تنصب الاسم وترفع الخبر نا الدالة على المتكلمين = ضمير متصل مبني على السكون في محل نصب اسم إنَّ أَعْطَيْنَاكَ = فعل ماض مبني على السكون لإتصاله بنا الدالة على المتكلمين ونا الدالة على المتكلمين = ضمير متصل مبني على السكون في محل رفع فاعله كَ = ضمير متصل مبني على الفتحة في محل نصب مفعول به الْكُوتِرُ = مفعول به ثان منصوب بالفتحة الظاهرة</p>	Errors cannot be found	الكوثر becomes second maf'ulbih because it has been preceded by maf'ulbih
Al-Humazah:9	<p>فِي = حرف الجر عَمْدٍ = اسم مجرور بفِي بالكسرة الظاهرة الجار والمجرور متعلق بمحذوف تقديره كائن في محل رفع المبتدأ مُذَدَّةٌ = صفة لعَمْدٍ مجرورة بمثله بالكسرة الظاهرة</p>	Errors cannot be found	ممددة detected as character of عمد having similar condition that isisimnakirahwith h positionmajrur.

An- Nashr:2	<p>وَ = حرف العطف رَأَيْتَ = فعل ماض مبني على السكون لإتصاله ببناء المخاطب وتاء المخاطب = ضمير متصل مبني على الفتححة في محل رفع فاعله التَّاسِمَ = مفعول به منصوب بالفتحة الظاهرة يَدْحَلُونَ = فعل مضارع مرفوع بثبوت النون لأنه من الأفعال الخمسة وواو الجماعة = ضمير متصل مبني على السكون في محل رفع فاعله في = حرف الجر بِئْسَ = اسم مجرور بفي بالكسرة الظاهرة وهو مضاف الله = لفظ الجلالة مضاف إليه مجرور بالكسرة الظاهرة أَفْوَاجًا = مفعول به منصوب بالفتحة الظاهرة</p>	أفواجا detected as maf'ulbih	The word should become haal(adjective). Because the condition that determines the haalis not addedyet, the word is detected as an object .
-------------	--	------------------------------------	---

5.3 Work Evaluation by Users

The questionnaires given to 5 persons, the users having teachers' profession at religious high school and Koranic school show the following responses:

1. The application is interesting because the users never see similar application. So the application is really new for the users.
2. The application is very complicated since the accuracy for storing the letters and the vowels is not easy. When *harakat*(vowel) is not properly entered, the result is not maximal. And when the harakat is not entered as the input, the system cannot make detection to the word.
3. As a whole the application is correct. When an error happens, it is caused by the system that is not added with suitable input or caused by the word which has similar word order and similar harakat with another word.
4. The application, as a whole, does not help much because the focus is on desktop. The user who does not have a desktop will never use the application.
5. From the perspective of users, the application should be used by students under the guidance of their teachers, so they can get assistance to learn the Arabic grammar.

6. CONCLUSION

Based on the survey, some conclusions can be proposed as follows:

1. Arabic grammar, especially *nahwu*, is really needed to comprehend the Holy Qur'an that is completely written in Arabic.
2. There are so many kinds of word and sentence forms which make the learners difficult to learn the nahwu knowledge.
3. This research employs NLP and expert systems to determine the word position in a sentence.

4. The expert system employs knowledge base to determine the types of word and the inference engine is used to determine the function and position of word in a sentence.
5. When determining, the system involves 82 templates for every letter and harakat. The templates are then divided into 34 templates for verb, 34 for subject pronouns, and 14 object pronoun or possessive.
6. The output from this system is an i'rab (description of word function in sentence) that is written in Arabic language.
7. From the experiment, some conclusions are drawn as follows:
 - a. To detect the parts of speech from 117 experiments, the error happens because the condition of word that is equally the same with the other word.
 - b. To detect the word function in sentence, 42 experiments are done. The error happens because the system is not supplied with the word relatives yet. This makes the following word function detection in trouble.
8. From the questionnaires given to 5 persons, the users having profession as religious high school and Kuranic school teachers, it is concluded that the application needs more features to complete and improve the detection result. Besides the application is good when it is intended to help high school students, who still need the guidance from their teacher. The application can help them to study the grammar of Arabic language.

REFERENCES

- [1] H. M. Harmain, H. El Khatib, A. Lakas, **Arabic Text Mining**, *IADIS International Conference Applied Computing*, 2004.
- [2] M. Elkourdi, A. Bensaid, T. Rachidi, **Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm**, *School of Science & Engineering Alakhawayn University*, 2004.
- [3] M. A. El Halees, **Arabic Text Classification Using Maximum Entropy**, *The Islamic University Journal*, Vol. 15, No.1, pp. 157-167, 2007.
- [4] Khairul Umam, Abu Abdin Nafi', **Terjemahan Matan al Ajrumiyah**, *Sinar Baru Algensindo*, Bandung, 2010.
- [5] Ali Ridho Barakbah, **Natural Language Processing**, *Lecture Notes*, Departments of Information and Computer Engineering, Electronic Engineering Polytechnic Institute of Surabaya.
- [6] Ali Ridho Barakbah, **Expert System**, *Lecture Notes*, Departments of Information and Computer Engineering, Electronic Engineering Polytechnic Institute of Surabaya.