

ARABIC LANGUAGE PROCESSING FOR RECOGNITION HOLY QURAN TAJWEED RULES USING FUZZY LOGIC AND PRODUCTION RULES

Dr. Kairoddin Omar*, UKM, Malaysia, email: ko@ftsm.ukm.my

Dr. Nazliah Omar*, UKM, Malaysia, email: no@ftsm.ukm.my

Mr. Majdi Salameh, UKM, Malaysia, majdisalameh@yahoo.com, majdi_salameh@aabu.edu.jo.
Faculty of information Science and technology, University of Kapangsaam. Malaysia, Bangi, 43600,
Selangor, Malaysia

Abstract

Arabic language is one of the natural languages which represents good domain for automatic processing, QURAN TAJWEED rules recognition is also one the interesting subjects in the past few years there were many attempts for processing Arabic language and just few which focused on Tajweed rules as result of the (Arabic Language Processing) [1].

this paper discuss new system that uses new techniques for computerize TAJWEED rules , the system uses QURAN text as an array of characters and vowels , which each have unique ASCII code representing it , then the array elements is analyzed to classify and recognize all TAJWEED rules that are in the selected text , all TAJWEED rules are stored in table , then the rules will be stored in a set with specific number of elements , after that each rule will be represent by a vector that contain ordered elements (characters ,vowels and spaces) ,then the text will be scanned for TAJWEED rules , and these rules will be converted into vectors with specific format , by calculating the distance between the new rule vector and all TAJWEED rules vectors - the rule with the lowest distance- the correct Tajweed rule is recognized .

Keywords: Fuzzy Logic, production rules, AI, Tajweed Rules, Holy Quran, Arabic Language Processing, Language Processing.

Back ground:

there were many attempts for creating software that deal with QURAN TAJWEED rules one of them is “ computerize holy QURAN TAJWEED rules “[1] the research applies program that computerize TAJWEED rules using production rules (if-then) and BNF rules for TAJWEED rules , however this study and the software “العاصم” have some errors in recognizing some of TAJWEED rules – not 100% correct , which caused system error , however the suggested system will correct those errors by using fuzzy computations to find the best TAJWEED rule –the lowest distance –with threshold for accepting these TAJWEED rules , also the system uses extra features for reducing the amount of

data to be searched by searching for specific conditions which will ignore the text and discard it from the searching process ,which will fast up and enhance the system.

Tajweed rules:

Tajweed rules they are the rules for pronouncing during reading QURAN words , the way of reading QURAN must be according to special rules specified with characters and vowels , unlike traditional arabic text which is not . all Muslims must know how to read the holy QURAN using tajweed rules , however there are many arabic and non arabic Muslims who don't know these rules , the system help them to know , learn

,and read the QURAN according to these rules.

the system also have a features that define the rules , explain them , color the characters that represent TAJWEED rules in the text with different color for each TAJWEED rule , to be easy recognized and classified .

The system cover most of QURAN TAJWEED rules , such as :

1- silent noon with sokon (سکون) and tanween rules (النون الساكنة و التنوين) includes :

- a- ethhar halqi (اظهر حلقى)
- b- edgham (with ghonnah and without ghonnah) (ادغام بغنة و بلا غنة)
- c- eqlab (اقلاب)
- d- ekhafa hakiki (real hiding) (اخفاء حقيقي)

2- silent meem (with sokon) (ميم ساكنة) includes :

- a- ekhafa shafawi (اخفاء شفوي)
- b- edgham shafawi (ادغام شفوي)
- c- ethhar shafawi (اظهر شفوي) .

3- meem and noon with shadah (ميم و نون مشددين)

- a- ghonnah (غنة)
- b- qalqalah (قفلة)
- c- b.1- grate qalqalah (قفلة عظمى)
- d- b.2- medium qalqalah (قفلة وسطى)
- e- b.3- small qalqalah (قفلة صغرى)

4- ra'a rules (احكام الراء) includes:

- a- tafkheem (تفخيم)
- b- tarqeeq (ترقيق) .

shared rules:

there are some rules which are shared by the same word , characters and vowels, which means that it can represent more than one tajweed rule in the same time such as :

- 1- edgham with ghonnah (ادغام بغنة) and ghonnah (غنة)
- 2- edgham without ghonnah (ادغام بلا غنة) with both tafkheem (تفخيم) and tarqeeq (ترقيق)

3- ethhar shafawi (اظهر شفوي) with both tafkheem (تفخيم) and tarqeeq (ترقيق)

4- edgham shafawi (ادغام شفوي) and ghonnah (غنة)

Tajweed rules uses all arabic alphabetical characters , however it uses some of arabic vowels which are (shadah (شدة), tanween fath (تنوين فتح), tanween dam (تنوين ضم), tanween kaser (تنوين كسر), sokon (سکون), mad (مد), fatha (فتحة), damah (ضمه), and kasrah (كسره)

Quran Tajweed Rule System:

the system is divided into many steps , at the beginning the text for the QURAN “ayah (آية) is chosen either by written it directly in rich text box or by import it from the database where all QURAN ayah (آية) is already stored , ordered by the sorah (سورة) it belongs to , and number that present the order of that ayah (آية) in the selected sorah (سورة), the ayah is stored in one diminution array including the characters, the vowels , and even the spaces inside the ayah with the same order as they were in the text , then the text will be scanned searching for a match with any Tajweed rule ,by using production rules with fuzzy logic the best suitable rule that match the text is recognized as the TAJWEED rule for that text , and for all rules the text colored , defined and explained according to specified tajweed rule that been recognized .

CONDITIONS FOR QURAN TAJWEED RULES:

in this step there will be determination for each TAJWEED rule that is in the same word , in two different words or in two different sentence in QURAN text (آية), preparing for tajweed rule recognition.

TAJWEED rules:

1- Allah : the word Allah (الله) (حكم) (لفظ الجلالة) has two rules

a – tafkheem (تفخيم) “Allah”: Allah (الله)

* If the vowel on the letter before Allah (الله) is damah (ضمة) (ُ)

* If the vowel on the letter before Allah (الله) is fatha (فتحة) (َ)

* if the vowel on the letters before Allah (الله) is sokon (سكون) () and the vowel before sokon (سكون) is either fatha (فتحة) (َ) or damah (ضمة) (ُ، ُ)

b- Tarqeeq (ترقيق) “Allah”: Allah (الله)

* If the vowel on the letter before Allah (الله) is kasrah (كسرة) (ِ)

* If the vowel on the letter before Allah (الله) is sokon (سكون) and the Letter before sokon is kasrah (كسرة)

* If the vowel on the letter before Allah (الله) is tanween (all Kinds of tanween)

1. tanween fath (تنوين فتح) ()
2. tanween dam (تنوين ضم) ()
3. tanween kaser (تنوين كسر) ()
4. shadah with tanween (تنوين مشدد) ()

2- lam rules (silent lam)

أحكام اللام الساكنة التي لا تحوي حركة

a- lam “al shmesih” اللام الشمسية

it must be (compound) (إدغام) (edgham) with out ghonah (بدون غنة) with the character after it (اللام) if is one of 14 letters :

1	2	3	4	5	6	7
{tt,	tha`a,	sad,	raaa,	ta,,	daa,	tha
ط،	ث،	ص،	ر،	ت،	ض،	ذ

8	9	10	11	12	13	14
non,	dal,	sen,	tha`a,	az ,	shen ,	lam}
ن	د،	س،	ظ،	ز،	ش،	لام،

b- lam “qmareiah” اللام القمرية
it must be (pronounced) (ethhar lam), إظهار
if it is before one of the following characters

1	2	3	4	5	6	7
{alef,	ba,	gen,	hha,	jem,	kaf,	waw,
و،	ك،	ج،	ح،	غ،	ب،	أ

kha,	fa,	ein,	qaf,	ya,	men,	ha }
خ	هـ،	م،	ي،	ق،	ع،	ف،

3- al qalqalah القلقلة

Al qalqalah is rule of Tajweed, which is done just for 5 arabic characters (characters) Which are

1	2	3	4	5
(qaf,	tta,	ba,	jem,	dal)
د،	ج،	ب،	ط،	ق،

al qalqalah القلقلة have 3 kinds

1- Small qalqalah: if the above (5 characters) one of them “is in the middle of the word and the vowel for it is sokon (سكون).

1- Middle qalqalah: if one of the 5 characters is in the end of the word- last character - and the vowel for it is sokon(سكون).

2- Grate qalqalah : if one of the 5 characters is in the end of the word – last character -and the vowel for it is shadah (شدة) .

4- Noon (نون) with sokon (سكون) and al tanween
there are many rules for the noon(نون) and tanween(تنوين)

4- a – hiding “ekhafa”

إخفاء

the “non ن” must be hided and composed with the followed character if the non vowel is sokon and the character after it is one of the following 15 characters :

1	2	3	4	5	6	7	8	9	10
{sad, thl, th, kaf, jem, shen, qf, sen, dal, tia,									
ط، د، س، ق، ش، ج، ك، ث، ذ، ص									
11	12	13	14	15					
zen, fa , taa dhad, tha }									
ز	ف،	ت،	د،	ظ					

the non and the tanween may come in three different places.

4-a-1 non char is in one word and one of 15 character is the same word- both in the same word-.

4-a-2 non is word and the character (one of the 15 characters)is in another word – each one of the character in different followed words .

4-a-3 tanween in one word and one of (15 character) in another word.

4-b- eqlab إقلاب

it is to convert the non and tanween into mem before (ba ب)

b-1 (non) نوناً

if it is followed by (ba ب) it must converted into (mem م)

b-2- (tanween)

if it is (tanween) followed by (ba) it must be converted into mem.

4-c- edgham إدغام

it is to compose the non with vowel sokon and tanween with the character after it if the character is one of the following {ya, non, mem, waw, lam, ra}

{	ر	ل	و	م	ن	ي	}
	1	2	3	4	5	6	

the edgham (إدغام) is two kinds :

4-c-1 with ghonnah (غنة) if the letter (character) follows the non or tanween is one of {ya, non, mem, waw }

4-c-2 with out ghonnah (غنة): if the character follows the non or tanween is either

{	lam	,	or	ra	}
	{	ل	,	ر	}

4-d- ethhar إظهار

it is to appear the non with sokon and tanween. if the non or tanween came after following 6 characters :

1	2	3	4	5	6		
{alef, ha, ein, ha, gein, kha}							
{	أ	هـ	ع	ح	غ	خ	}

4-d-1 - non sokon and one of the 6 character follows in the same word.

4-d-2 - non with sokon vowel on it and one of the 6 character follows in two words .

4-d-3 - tanween followed with on of 6 character in two different words .

5- mem rules

إحكام الميم الساكنة

if the mem(الميم) vowel is sokon (سكون)is followed by one of alphabetical Arabia character it has many rules

5-a- “ekhafa”

إخفاء الميم

if the mem vowel is sokon and the follow character is ba (ب)

5-b edgham إدغام

if the mem vowel is sokon and followed by character mem (another mem)

5-c- ethhar إظهار

if the mem vowel is sokon and followed by all other character except for (mem م , ba ب) = the character set is :

{	alef, ta, th, jem, ha, kha, dal, thal, ra, zai,
{	1 2 3 4 5 6 7 8
	9 10
	sen, shen, sad, daws, tan, thah, qaf, kaf, fah,

11 12 13 14 15 16
 17 18 19
 non, hah, waw, ein, giev, ya, lam}
 20 21 22 23 24 25 26}

6- Hamza wasel أحكام همزة الوصل

- 6- a -if it is at the beginning of the word, the hamza appeared and it must be specked تتنطق
- 6- b- if the hamza is not at the beginning it is not specked but must appear.
- 6-c- if it is t the middle and if it is with “edgham”
- 6-d - convert the “hamza” with “mad” if the hamza vowel is fatha after hamza for question .

1- al-edgham “general” الاضغام العام

It is to compose more than one character and pronounce anew character, if they followed each other, there are many kinds of edgham :

7-a - general heterogeneous edgham (الاضغام العام المتجانس)

7-a-1- if the characters followed to others at any order

{ tta ط , ta`a ت , dal د } are
 { ط ، ت ، د }
 tta → ta`a ط → ت
 ta`a → tta ت → ط
 ta`a → dl د → ت
 dal → ta`a ت → د
 tta → dal د → ط
 {dal → tta } {ط → د }

7-a -2 - if the following characters follow each other in any position (order)

{tha` , thal, ttha` } { ث ، ذ ، ظ }

7-a-3 - if the following characters follows each other in any order

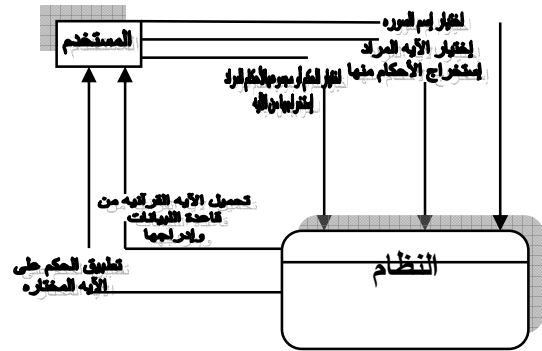
(mem, ba) (م ، ب)

7- b -general edgham Simi similar إدغام عام متقاربين

7-b-1 - if the following characters follows each other in any position (order)

(lam, ra) (ل ، ر)

7-b-2 - if the following characters follow each other in any order
 (qaf, kaf) (ق ، ك)



شكل 1-5 السور القرآنية المعطاة (DFD)

System stages:

1- read and load the QURAN text

in this step there will be two options , first is chosen the QURAN text or ayah (آية) from database that store all QURAN sorah (سورة) , and from specific sorah one ayah or more will be chosen, performed using lists and combo boxes , or by display the QURAN text page for the specific sorah in al Othmani copy of Quran , the text may be written in different font type, font size , and font style , also the text may be written with vowel and in multi lines different than it is in the holy QURAN, also the text may be copied or imported from external source.

2- QURAN text process and analyst

in this step all characters , vowels an spaces in the ayah will be stored in array of characters , after a copy of that array will be stored in one another array of integer representing the ASCII , then the index for all characters vowels and spaces is stored , the number of words in the ayah is counted , the position for each word – the index of the first character in the word – is located and stored ,and the number of

characters in each word - excluding the vowels- is counted . all these stored data and information is going to be used later in next steps.

3- exclusive words discarding

this step will discard some of the words in the ayah (آية) because it is impossible to have rule that match any of Tajweed rules [exclusive rules], the main purpose of this step is to reduce the source text which will be processed to recognize TAJWEED rule and speed up the recognition process to enhance the performance of the system, the following exceptions represents the words which will be discard :

- 1- If the word consists of less than three characters, by checking the length of each word, which have been calculated from the previous step.
- 2- if the word were one of the following words
{Waw (واو), etha (اذا), kathalek (كذلك), allaty (التي), allaaty (اللاتي), allaawaaty (اللواتي), allthy (الذي), ayoha(ايها), thalek(ذلك), telk (تلك)}
- a. if the first and the last character in the word is noon (ن)
- 3- often the words that begin with characters some characters don't mach any TAJWEED rules , characters such as :
{ خ ، ث ، د، ز ، ص ، ض ، غ ، ط ، ظ ، و }
- 4- if the word is (Allah) the rule is tafkheem lam
- 5- if the word was one of the following
{ يس ، طس ، كهيعص ، طسم ، الر ، المص ، الم ، الحاقة ، ن ، هؤلاء ، الضالين ، ق ، ص ، حم }
the rule is mad with 6 interval .
- 6- if the text is (من راق) then the rule is silent after noon(ن).
- 7- if the text is (بل ران) then the rule is silent after noon(ن).
- 8- if the word was one of the following (الدنيا ، بنيان ، صنوان ، قنوان)
(اظهار) ethhar (اظهار)

4- tajweed rules recognition using production rules :

The recognition in this step done by using the production rules (IF – THEN), the production rules are used with the BNF (Backus Naur Form) [1].

5- TAJWEED rules Fuzzification and recognition:

in this step fuzzy logic principles are used , with soft computing to find the best TAJWEED rule for specific text (recognition) if it is not found by the production rules , first let v be a set of all vowels used in TAJWEED rules , $v = \{ \text{shadah (شدة) , tanween fath (تنوين فتح) , tanween dam (تنوين ضم) , tanween kaser (تنوين كسر) , sokon (سكون) , and mad (مد) } \}$ and presented by ASCII code set , where each vowel is presented by its ASCII code, and let C be a set of all characters used in TAJWEED rules , $C = \{ \text{all Arabic alphabetical} \}$, also presented by there ASCII code values . let each TAJWEED rule be presented by vector of ordered elements of both sets (v ,c) , let each rule be denoted with r with different sub number ,

- $r_1 = \text{ethhar halqi (اظهار حلقي)}$
- $r_2 = \text{edgham (with and without ghonnah)}$
- $r_3 = \text{eqlab}$
- $r_4 = \text{ekhafa hakiki (real hiding)}$
- $r_5 = \text{ekhafa shafawi (اخفاء شفوي)}$
- $r_6 = \text{edgham shafawi (اضغام شفوي)}$
- $r_7 = \text{ethhar shafawi (اظهار شفوي)}$
- $r_8 = \text{ghonnah (غنه)}$
- $r_9 = \text{grate qalqalah (قلقله عظمى)}$
- $r_{10} = \text{medium qalqalah (قلقله وسطى)}$
- $r_{11} = \text{small qalqalah (قلقله صغرى)}$
- $r_{12} = \text{tafkheem (تفخيم)}$
- $r_{13} = \text{tarqeeq (ترقيق)}$

after that any checked text will be converted into vector of characters and vowels using the one diminution array the text (ayah)with many conditions, these conditions are compounded of the number of parameters such as : word characters, the position of the word, the first elements value , the last element value , counter for spaces ...etc.

6- TAJWEED rule recognition:

in this step the each text , paragraph , sentence , and word is scanned , and analyzed searching for specific vowel , spaces and characters in proper order to match with one of TAJWEED rules that been stored before , from previous step all rules are stored as vectors of elements (vowel, character , and space) in specific order , the vector from the QURAN text will be compared with all vowels using similarity or distance . the smallest distance value from all vectors is going to be the most suitable rule that match with the vector from QURAN

$$\text{distance}1(\text{input}, R1) = \sum |M_{\text{input}}(x) - M_R(x)|$$
where $x \in X, R=1 \text{ to } m$

Distance (input vector, r_1) = | input vector (element1) – R1(element 1) + | R input vector(element2) – 1(element 2) | +.. | R input vector (element n) – R1(element n) | For all R $1 \text{ to } m$

7- Tests:

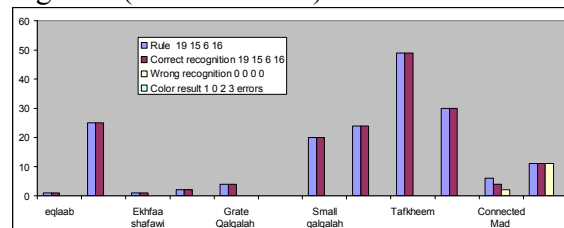
in this system ten sorah(سورة) were tested, and the results of recognition for the selected sorah(سورة) shown in table 1.(العصر ، الغاشية ، الزلزلة ، الهمزة ، الماعون ، الناس ، الجمعة ، الضحى الليل)

Table 1 recognition 10 sorah
Table 1

TAJWEED rule	Rule recognition	Wrong recognition	Color result
Real ekhafa	19	0	1 error
Real ethhar	15	0	No error

Edgham without ghonnah	6	0	2 errors
Edgham with ghonnah	16	0	3 errors
Eqlab	1	0	No error
Ethhar shafawi	25	0	No error
Ekhafa shafawi	1	0	No error
Edgham shafawi	2	0	No error
Grate Qalqalah	4	0	No error
Medium qalqalah	0	0	No error
Small qalqalah	20	0	No error
Tarqeeq	24	0	No error
Tafkheem	49	0	No error
Ghonnah	30	0	No error
Connected Mad	4	2 missing	No error
Separated Mad	11	11	No error

Figure 1 (chart of table1)

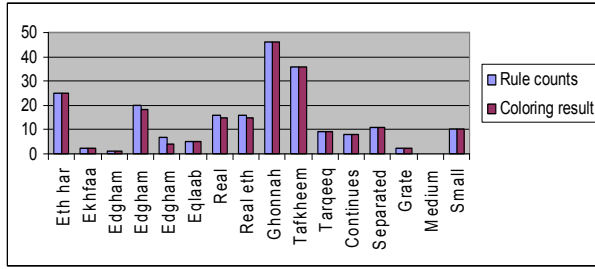


also another test on all TAJWEED rules were done separately , and table two shows the results of recognition for all TAJWEED rules .

Table 2 table of tajweed recognition results

TAJWEED rule	Rule counts	Coloring result
Ethhar shafawi	25	25 correct
Ekhafa shafawi	2	2 correct
Edgham shafawi	1	1 correct
Edgham with ghonnah	20	18 correct
Edgham without ghonnah	7	4 correct
Eqlab	5	5 correct
Real ekhafa	16	15 correct
Real ethhar	16	15 correct
Ghonnah	46	46 correct
Tafkheem	36	36 correct
Tarqeeq	9	9 correct
Continues Mad	8	8 correct
Separated mad	11	11 correct
Grate qalqalah	2	2 correct
Medium qalqalah	0	0
Small qalqalah	10	10 correct

Figure 2 (chart of table 2)



8- Results:

The results were very good for the first test as shown two errors of 227 recognized rule with (99.992 %)correct percentage for recognition .

The results of the second test were also very good with seven errors of 214 recognized rule (99.968 %)correct percentage for recognition .

9- Conclusion:

it is possible to use production rules or crisp – hard computing - to recognize holy QURAN TAJWEED rules , but there will be some errors not because of the production rules it self but because of the text it process , and because of some special vowels and characters in holy QURAN , this study used both crisp and soft computing operations by using production rules with fuzzy logic principles to recognize TAJWEED rules , and the results were ((99.98 % %)) , which are better than the previous studies which use production rules only, because there will be no missing for any of TAJWEED rule if it was checking and it will take the closest rule that may match the correct TAJWEED rule ,with suitable threshold and correct condition .

10- Problems:

Some rules are recognized in correct because it takes the nearest rule that may match, but those rules in other studies are missing and not recognized, but still have to solve this problem.

the program used in this research coloring characters that catch any of the TAJWEED rules, some colors covers the colors which was before it, then the character will have the latest used TAJWEED rule color, this case appears in shared TAJWEED rules as been mentioned before in this paper and in some examples part of the characters representing one TAJWEED rule is colored differently.

11- Future work:

- Making speaking TAJWEED rules for holy QURAN.
- Enhancing this research to have 100% correctness percentage.

Special thanks: special thanks for the students who help in developing the program of tajweed rules using production rules and the results of the program. (Ms. Noor and yasmin).

References:

- [1] Computerized TAJWEED rules for the holy QURAN by maen al horani, Ahmed tawfeeq al ta'an 2005.
- [2] <http://www.islamicgoodsdirect.co.uk>
- [3] الواضح في احكام التجويد تاليف الدكتور محمد عصام مفلح القضاة 1998
- [4] المنير في احكام التجويد اعداد د. احمد شكري ، د. احمد القضاة ، د. محمد سليمان ، د. محمد المجالي ، د. محمد القضاة، عبد الرحمن ابو غليون علي الجيوسي، عمر حماد ، مامون الشمالي ، 2005
- [5] arabic character recognition using structural approach by Mustafa ramadneh
- [6] مختصر فتح الرحمن في تجويد احكام القران الكريم ، د.احمد حسين علي ، 2000 .
- [7] holy quran in othmani font , web site “ www.dar-al-maarifah.com

- [8] Question answering system to support the arabic language , by hammo bassam , abu salem hani and lytinen steven, JASIS.50(6),pp524-529,2001.
- [9]statistics classifier of the holy quran verses , by al kabi mohammed , ghassan kannan , riyad al shalabi , kahlid M.O . and basel bani-ismail , journal of applied sciences 5(3):580-583,2005.
- [10] lexical analysis of inflected arabic words using exhaustive search of augmented transition network,by rafea A. and shaalan K.1993,software practice and experience , Vol .123(6),pp567-588,john wiley & sons U.K.9.
- [11] A computational morphology system for arabic databases , by dichy joseph , university lumiere-lyon, france ,<http://www.elda.fr>.
- [12] a computational morphology system for arabic , by al-shalabi riyad , and marth evens , <http://www.humanities.uchicago.edu>
- [13] expert system for searching in arabic text by zeyab , ali , master research , al albayte university , 2000.
- [14] <http://www.quranexplorer.com/quran>.
- [15] نموذج محوسب لمحلل نحوي للجمل الاسمية غير المشكولة في اللغة العربية ، سليمان ، معتصم فتحي ، جامعة ال البيت ، 2002 .
- [16] AI and Islam , by professor mohammed zaki khedher , 2004.
- [17] A rule –based framework for GIS system applications , by abid al ajeeli, Bahrain university .