جامعة السودان للعلوم والتكنولوجيا

كلية علوم الحاسوب و تقانة المعلومات

قسم علوم الحاسوب

نظام رد آلي تفاعلي للقرآن الكريم

Interactive Question Answering System for Holly Quran

مشروع مقدم كأحد متطلبات الحصول على درجة البكالوريوس في علوم الحاسوب

إعداد الطالبات :

آمنة الطيب عبدالرحيم محمد

زينب إدريس محمد إبراهيم

هادية مالك شيخ الدين محمود

إشراف:

د. محمد عدني حمدالسيد

1439هـ - أكتوبر،2017م

بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا

كلية علوم الحاسوب و تقانة المعلومات

قسم علوم الحاسوب

نظام رد آلي تفاعلي للقرآن الكريم

Interactive Question Answering System Holly Quran

مشروع مقدم كأحد متطلبات الحصول على بكالوريوس في علوم الحاسوب

إعداد الطالبات :

آمنة الطيب عبدالرحيم محمد

زينب إدريس محمد إبراهيم

هادية مالك شيخ الدين محمود

إشراف:

د. محمد عدني حمد السيد

1439هـ- أكتوبر،2017

الإستهلال

قال تعالى:

{ وَعَلَّمَ آَدَمَ الْأَسْمَاءَ كُلَّهَا ثُمَّ عَرَضَهُمْ عَلَى الْمَلَائِكَةِ فَقَالَ أَنْبِئُونِي بِأَسْمَاءِ هَؤُلَاءِ إِنْ كُنْتُمْ صَادِقِينَ (31) قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا مَا عَلَّمْتَنَا إِنَّكَ أَنْتَ الْعَلِيمُ الْحَكِيمُ(32) }

صدق الله العظيم
سورة البقرة

الحمد لله

الحمد لله أقصى مبلغ الحمد.. والشكر لله من قبل ومن بعد.. الحمد لله عن سمع وعن بصر.. الحمد لله عن عقل وعن جسد.. الحمد لله عن ساق وعن قدم.. الحمد لله عن كتفي وعن يدي.. الحمد لله عن قلبي وعن رئتي.. الحمد لله عن أمي وعن أبتي.. والحمد لله عن أخوات ذا العبد

الإهداء

يا من أحمل اسمك بكل فخر

يا من كنت له الأمل الذي راوده في حياته فحلم أن يراني في مثل هذا اليوم...

إلى أبي العزيز

إلى شجرةٍ باسقةٍ في وجه أعاصير الحياة ترنو باغصانها إلى السماء متضرعة

دعاءً لي، إلى التي سكنت قلبي، إلى القلب الحنون...

أمي الحبيبة

إلى سندي وقوتي وملاذي بعد الله ..إلى من علموني معنى الحياة

إلى من اظهروا لي ما هو أجمل في الحياة.....

إلى إخوتي

لمن غرس في معنى التميز والإبداع الذي كان عونا لي بعد الله في اكمال هذا البحث..

الدكتور محمد عدني

لمن أضاؤا عقلي بعلمهم فاظهروا بسماحتهم تواضع العلماء...

أساتذتي الأجلاء في كل المراحل التعليمية كم نحن فخورون بكم

إلى أصدقائي و أحبتي و زملائي ومن رافقوني في مسيرتي العلمية، فكانوا نعم الرفقة الصالحة، ونعم العون ...

وطني أُحِبُّكَ لا بديل.. أتريدُ من قولي دليل.. سيظلُّ حُبّك في دمي.. لا لن أحيد ولن أميل.. سيظلُّ ذِكرُكَ في فمي ووصيّتي في كلّ جيل. لا نبخل عنك بما تعلمنا و سنقدم كل ما بوسعنا و سنجعل منك فخوراً بنا...

سنشكركم جميعا تنطقها قلوبنا على ألسنتنا و تخوننا عبرات الشكر على تقديم ما يليق بكم ، فجزاكم الله عنا كل الخير والجزاء... وإلى كل مهتمٌ بلغة الضاد، و مجتهدٌ في تعلمها و ناشد لبلوغها الكمال في عصر البيانات الرقمية نهـــــــــــدي هذا الجهد المتواضع

شكر وتقدير

لابد لنا ونحن نخطو خطواتنا الاخيره في الحياة الجامعيه من وقفة تعود إلى اعوام قضيناها في رحاب الجامعة مع أساتذتنا الكرام الذين قدموا لنا الكثير باذلين جهودا كبيرة في بناء جيل الغد لتبعث الأمة من جديد.

وقبل أن نمضي وانطلاقا من العرفان بالجميل، نقدم أسمى آيات الشكر والأمتنان والتقدير والمحبة إلي الذين حملوا أقدس رسالة في الحياة إلى الّذين مهدوا لنا طريق العلم والمعرفة إلي جميع أساتذتنا الأفاضل....

"كن عالما ..فإن لم تستطع فكن متعلما ،فإن لم تستطع فأحب العلماء ،فإن لم تستطع فلا تبغضهم".

ونخص بالشكر و التقدير الدكتور محمد عدني الذي نقول له بشراك قول رسول الله صلي الله عليه وسلم " إنَّ الله وملائكته وأهل السَّموات والأرض حتَّى النَّملة في جحرها وحتَّى الحوت في البحر ، والطير في السماء ،ليصلون على معلم الناس الخير". ونسأل الله العظيم أن يبارك في علمه و يزيده وأن ينفع به الأمة و يمد في عمره ليبقى نبراساً منيراً في حقل العلم والمعرفة، فجزاه الله عنا كل الخير والجزاء بإرشاده لنا و تمهيده لنا الطريق لمعرفة حقل هذا البحث من حقول علوم الحاسوب.

فإنه ليسرنا و يثلج صدرنا أن نتقدم بجزيل الشكر والإمتنان ’لى كل من جامعتي الجميلة والحبية جامعة السودان للعلوم والتكنولوجيا ممثلة برئيسها، وكلية

علوم الحاسوب و تقانة المعلومات ممثلة بعميدها ،لكل م قدموه لنا للمضئ قدما في المسيرة العلمية. وأسأل الله أن تبقى قلعةً للعلم و المعرفة.

كذالك نشكر كل من ساعد على إتمام هذا البحث وقدم لنا العون ومد يد المساعدة وزودنا بالمعلومه اللازمة لأتمام هذا البحث ونخص بالشكر جامعة افريقيا العالمية متمثلة في بحثها الذي هو بعنوان "تطبيق معالجة النصوص باستخدام لغة بايثون" للمشرف الدكتور البراء أبو عبيدة.

المستخلص

إسترجاع المعلومات إحدى المجالات المهمة في علوم الحاسوب و الذي إرتبط حديثا بمجال الأنترنت، و الذي يهتم بعمليات الفهرسة و البحث و إسترجاع المعلومات و المستندات التي يطلبها المستخدم، من أمثلة نظم إسترجاع المعلومات: محركات البحث ، المكتبات الألكترونية و أنظمة الرد الآلي على السؤال . نظم إسترجاع المعلومات تواجه مجموعة من التحديات في بعض اللغات خصوصا اللغة العربية لأنها لغة إشتقاقية . نظم إسترجاع المعلومات تفشل في التعرف على بعض الكلمات؛ إذا قام المستخدم بكتابة هذا الإستعلام " من هو المسلم " و لكن في المستندات الموجودة لا توجد كلمة "المسلم" بعينها و لكن توجد مسلمين، مسلمات، مسلمون، مسلم ، بما أن عملية البحث تكون بمطابقة الأنماط فكلمة "المسلم" تختلف عن (مسلمين ، مسلمات،...) و هكذا سيقوم النظام بإسترجاع كل المستندات التي يحتوي على كلمة "المسلم" إن وجدت أو لا توجد نتائج، بينما يتم فقد المستندات التي تحتوي على مسلمين، مسلمون، مسلم، مسلمات... و التي ينبغي إسترجاعها.

توسعة الإستعلام بزيادة الأنماط للكلمة الواحدة "من هو المسلم " بعد التوسعة تصبح "من هو المسلم مسلم المسلمون مسلمون المسلمات مسلمات..." لضمان إسترجاع كل الملفات ذات الصلة، عمل هذه المعالجات الآلية لتوسعة الإستعلام واحد من التحديات التي تواجه نظم إسترجاع المعلومات العربية.

علم قواعد الصرف في اللغة العربية علم يهتم بالكلمة و ما يترتب عليها من إعلال أو إبدال أو تغييير في أبنية الكلمة و هو العلم الذي يعطينا أنماط متعددة للكلمة الواحدة، مثال: قال على وزن "فعل" و منها قائل على وزن "فاعل".

هذه الدراسة تهدف لتوضيح كيف أن معالجة الإستعلام و توسعته يمثل تحدي يواجه نظم إسترجاع المعلومات العربية و اقترحت طريقة لتحسين الإسترجاع.

تكونت هذه الدراسة من جزئين: الجزء الأول بناء و برمجة الأنماط حيث تمت معالجة إثنين و ثلاثين نمطاً آلياً وهذه الأنماط هي: اسم الفاعل "فاعل" ، اسما الزمان و المكان "مفعل" ، أفعل التفضيل "افعل"، اسم المرة"فعلة"، اسم الهيئة "فعلة"، صيغ اسم الآلة (" مفعل" " مفعال " "مفعلة " "فعالة" " فعال" "فاعلة" " فاعول")، صيغ و أوزان الصفة المشبهة(" فعل" "فعلة" "افعل" "فعلاء" "فعلان" "فعيل" " فعيلة " "فعال" "فعالة" "فاعل" " فيعل " " فعول " " فعلون " " فعلى" "فاعلة") و صيغ المبالغة (" فعل" "مفعال" " فعال" " فعيل" "فعول") وكذلك بالإضافة إلى الإثنين و الثلاثين نمطاً تمت معالجة إضافة حروف التعريف للكلمة و حروف التعريف هي خمسة حروف(ال بال، كال، فال، وال) توجد في بداية الكلمة ، كما أن هنالك آلية لحذف حروف التعريف. إضافة للمعالجات تمت معالجة إضافة الحروف السابقة واللاحقة الي جذع الكلمة (Stemmer) العربية.

الجزء الثاني من الدراسة هو بناء نظام رد آلي تفاعلي للقرآن الكريم

تطبيق و إختبار تلك المعالجات على النظام، تمت المقارنة بين النتائج قبل المعالجة الآلية للإستعلام(أي من غير معالجة) و بعد عملية المعالجة. بناءً على النتائج هذه الدراسة نجحت في إسترجاع المعلومات و التي هي الآية أو الآيات من القرآن الكريم ذات الصلة بسؤال المستخدم و في حين فشل النظام في استرجاعها من غير معالجات . وبالتالي الطريقة المقترحة قادت إلي تحسين الإسترجاع من حيث الدقة و العدد في النتائج.

Abstract

Information retrieval is one of the most important fields in computer science, which is connected to the Internet. It is concerned with indexing, searching and retrieval of information and documents requested by the user. Examples of information retrieval systems are search engines, electronic libraries, and automated answering systems. Information retrieval systems face a range of challenges in some languages, especially Arabic, because it is a derivative language. Information retrieval systems fail to recognize certain words; if the user type this query " من هو المسلم ", but in the existing documents there is no "المسلم " word, but there are (مسلمين، مسلمات، مسلمون، مسلم), Since the search process matches the patterns, the word " المسلم " is different from مسلمون، مسلم (مسلمين، مسلمات،...) Thus, the system will retrieve all documents containing the word " المسلم " if any or no results, while documents containing (مسلمين، مسلمات، مسلمون، مسلم ...)which should be retrieved. Expanding the query by increasing the patterns of the word "من هو المسلم " after the expansion becomes "من هو المسلم مسلم المسلمون مسلمون المسلمات مسلمات ..."To ensure the retrieval of all relevant files, these processors to expand the query one of the challenges facing the Arabic information retrieval systems and cause In the loss of some documents and information and therefore inaccurate results.

This study aims at clarifying how the processing and expansion of the query represents a challenge facing Arabic information retrieval systems and suggested a method to improve retrieval.

This study was composed of two parts: the first part was the construction and programming of patterns, Thirty-two patterns were processed. The second part of the study is the construction of an Interactive Question Answer System for the Holy Quran application and testing of these processed in first part on the system. Based on the results, this study succeeded in retrieving the information, which is the verse or verses from the Quran related to the question of the user, while the system failed to retrieve them without proposed method. Thus the proposed method led to improved retrieval in terms of accuracy and number in results.

الفهرست

الإستهلال		**ج**
الحمد لله		**د**
الإهداء		**ه**
شكر وتقدير		**ز**
المستخلص		**ط**
ABSTRACT		**ل**
الفهرست		**م**
الأشكال		**س**
الجداول		**ع**
الباب الأول		1
المقدمة العامة		*1*
1- 1	مقدمة عامة	2
1- 2	مشكلة البحث	3
1- 3	أهمية البحث	3
1- 4	أهداف البحث	4
1- 5	الحلول المقترحة	4
1- 6	منهجية البحث	4
1- 7	حدود البحث	5
1- 8	هيكلية البحث	5
الباب الثاني		6
الإطار النظري و الدراسات السابقة		*6*
الفصل الأول		7
الإطار النظري		7
2 - 1	خلفية عن نظم إسترجاع المعلومات	8
2 - 1 - 1	المعالجة الآلية للغة الطبيعية	8
2 - 1 - 2	إسترجاع المعلومات	8
2 - 1 - 3	تعريف إسترجاع المعلومات	8
2 - 1 - 4	مفاهيم أساسية في نظم إسترجاع المعلومات	8
2 - 1 - 5	نظام الإجابة على الأسئلة	10
2 - 1 - 6	نظام الإجابة على الأسئلة التفاعلي	11
2 - 1 - 7	المدونات اللغوية	11
2 – 2	اللغة العربية	11
2 - 2 - 1	خصائص تراكيب اللغة العربية	11
2 - 2 - 2	تحديات تواجه المعالجة الآلية للغة العربية	12
2 - 2 - 3	الفرص التى يمكن اقتناصها عند معالجة اللغة العربية آليا	13
2 - 2 - 4	مدخل إلى تعريف علم الصرف و دراسة المشتقات	13

الفصل الثاني		19
الدراسات السابقة		19
2 - 3	الدراسات السابقة	20
أ.	الدراسة الأولى[9]	20
ب.	الدراسة الثانية**[10]**	20
2 - 4	الخلاصة	22
الباب الثالث		23
منهجية البحث		*23*
3. 1 .	مقدمة	24
3. 2 .	الأدوات والتقنيات	24
3. 2. 1.	لغة جافا (Java JDK_8)	24
3. 2. 2.	محررالنصوص (NetBeanse_8 IDE)	25
3. 2. 3.	المكتبة (ucanaccess)	25
3. 2. 4.	مصمم الأشكال (Diagram designer)	25
3. 3 .	إطار البحث	25
3. 3. 1.	المعالجة الأولية (pre-process)	25
3. 3. 2.	توسعة الإستعلام(query expansion)	26
3. 3. 3.	المطابقة (matching)	26
3. 3. 4.	إسترجاع المعلومات (retrieve verses from holy Quran)	26
3. 4 .	تحليل الأنماط	27
3. 4. 1.	إسم الآلة	27
3. 4. 2.	إسم الفاعل	28
3. 4. 3.	إسما الزمان والمكان	29
3. 4. 4.	صيغة المبالغه	29
3. 4. 5.	الصفة المشبة	31
3. 4. 6.	الكلمات التوقيفية(Stop Words)	32
3. 4. 7.	حذف التشكيل	32
3. 4. 8.	معالجة الحروف الخاصة (Special Characters)	32
3. 4. 9.	التقطيع(Split)	32
3. 5 .	الخلاصة	37
الباب الرابع		38
تطبيق النظام		*38*
4 . 1 .	المقدمة	39
4 . 2 .	الشاشة الرئيسية	39
4 . 3 .	لوحة التحكم	39
4 . 4 .	شاشة النتائج	40
4 . 5 .	تنفيذ شاشة لوحة التحكم جزء بجزء	41
4 . 6 .	نتائج إختبار النظام	44
4 . 7 .	وصف مدونة القرآن الكريم	46
4 . 8 .	الخلاصة	46
الباب الخامس		47
النتائج والتوصيات		*47*
5 . 1 .	مقدمة	48
5 . 2 .	النتائج	48
5 . 3 .	التوصيات	48
5 . 4 .	الخاتمة	49
المراجع		50
الملاحق		51
الملحق أ		52

الأشكال

شكل 1-1: محرك البحث قوقل سكولار	3
شكل 2-1: النموذج العام لنظم إسترجاع المعلومات	9
الشكل 3-1: إطار البحث المقترح	27
الشكل3-2: طريقة معالجة اسم الآلة	30
الشكل3-3: طريقة معالجة اسم الفاعل آليا	31
الشكل3-4: طريقة معالجة اسما الزمان والمكان	34
الشكل 3-5: طريقة معالجة صيغ المبالغة آليا	34
الشكل 3-6: طريقة معالجة صيغ الصفة المشبهة آليا	35
الشكل3-7: طريقة معالجة حذف الكلمات التوقيفية	36
الشكل 3-8: طريقة معالجة حذف التشكيل	36
الشكل 3-9: طريقة معالجة استبدال بعض الحروف الخاصة	37
الشكل 3 -10: طريقة معالجة عملية التقطيع للإستعلام	37
شكل 4-1 :أيقونة البرنامج	39
شكل4-2:الشاشة الرئيسية	39
شكل4-3: الذهاب الي لوحة التحكم	40
شكل 4-4: لوحة التحكم	40
شكل4-5:أعمدة مدونة القرآن الكريم	46

الجداول

جدول 2-1: أمثلة لصيغ المبالغة المشهورة 15

جدول 2-2: أمثلة لصيغ المبالغة غير المشهورة 16

جدول 2-3: أمثلة لصيغ الصفة المشبهة من الفعل الثلاثي على وزن فَعِلَ 16

جدول 2-4: أمثلة لصيغ الصفة المشبهة من الفعل الثلاثي على وزن فَعُلَ 16

جدول 2-4: أمثلة لصيغ إسم الآلة 17

جدول 2-5: مقارنة الدراسات السابقة 22

جدول 4-1: أسئلة الإختبار والإجابات الصحيحة لها. 44

جدول 4-2: نتائج الإختبار بدون معالجة. 45

جدول 4-3: نتائج الإختبار بإختيار جميع المعالجات 45

الباب الأول

المقدمة العامة

1-1 مقدمة عامة

شهد مجال استرجاع المعلومات (Information Retrieval) تطوراً ملحوظاً في العقود الأخيرة وذلك نتيجة النمو المضطرد للشبكة العنكبوتية وحاجة المستخدمين الماسة للبحث في ذلك الكم الهائل من المعلومات الرقمية. ونظراً للنمو المتزايد للمحتوى العربي الرقمي سواءً على الإنترنت أو الوسائط الإلكترونية الأخرى، فإن الحاجة أصبحت أشد إلحاحاً لإيجاد أنظمة استرجاع معلومات ومحركات بحث تعتني بالخصائص الفريدة للغة العربية وتحسن التعامل معها. فالعربية هي لغة الوحيين القرآن والسنة وتعد أكثر اللغات السامية الحية من حيث عدد المتكلمين[1]. فهي في المرتبة الرابعة من لغات العالم من حيث الانتشار نسبة عدد متحدثيها في العالم 6.6%، وهي واحدة من أقدم لغات العالم، ويتحدث بها غالبية سكان الشرق الأوسط وشمال أفريقيا، وبالأخص سكان البلدان العربية كافة. وعلاوة على ذلك، يقبل الملايين على تعلم اللغة العربية لأنها لغة القرآن الكريم، وهناك الملايين من المسلمين في بلدان أخرى يتحدثون العربية أيضاً ،في عام 1974 تم اعتمادها كلغة رسمية سادسة في الأمم المتحدة[2].

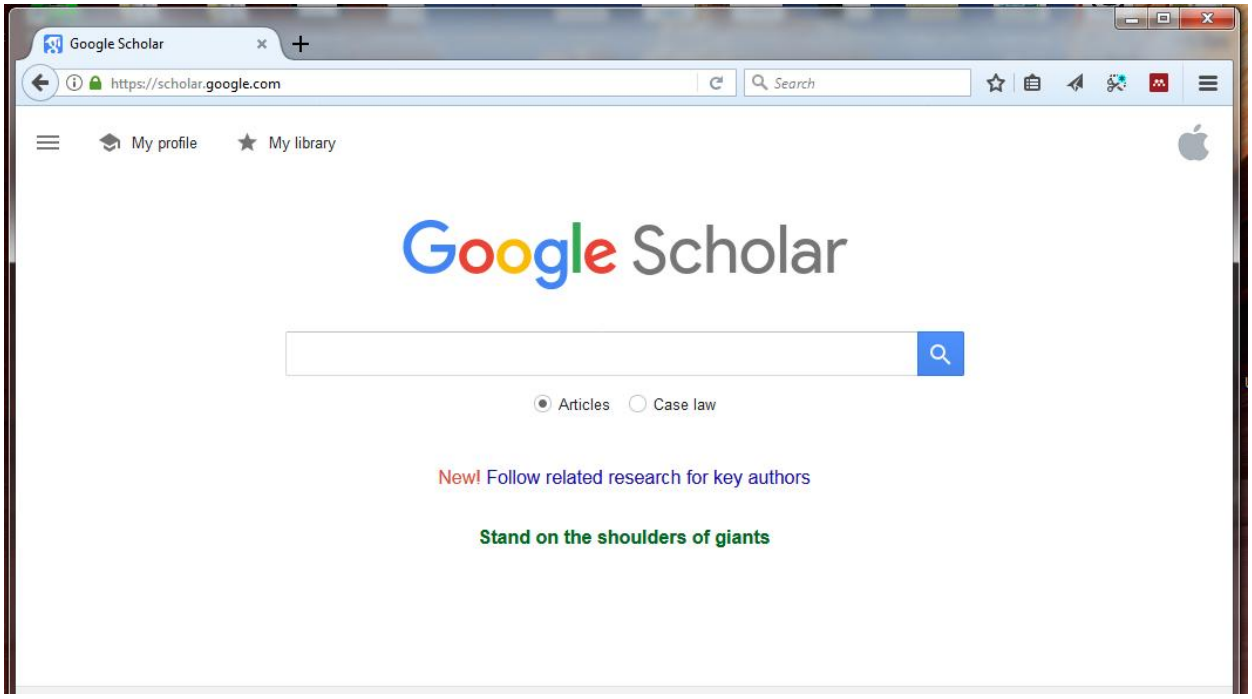
وتتميز العربية بأنها تكتب وتقرأ من اليمين إلى اليسار، كما أن حروفها تكتب بأشكال مختلفة تبعاً لموقعها والحروف المجاورة لها، وتختلف طريقة نطق الحرف وبالتالي معنى الكلمة وموقعها الإعرابي بناءً على حركة التشكيل الموجودة عليه ، بالإضافة إلى أن العربية لغة اشتقاقية وليست إلصاقية، حيث يعد نظامها الصرفي من أكثر النظم الصرفية تقدماً، فهو مبني على تصريف الجذور وفقاً لمجموعة محددة من الأوزان للحصول على كلمات ذات دلالات مختلفة من نفس الجذر. وكل ما سبق ذكره يمثل تحديات لمكننة التحليل الصرفي والإعرابي والدلالي للغة العربية ومن ثم لاسترجاع النصوص العربية[1].

وفي هذا الصدد لا بد من تعريف علم قواعد الصرف للغة العربية وهو التغيير و التحويل لغة ، و اصطلاحا هو العلم الذي يبحث في أبنية الكلمة العربية و صيغتها،وبيان ما في حروفها من أصالة وزيادة أو حذف ، أو صحة أو إعلال أو إبدال الى غير ذلك ، ويسمى التصريف.وقد وضح العلماء- رحمهم الله- على مدى العصور السابقة أهمية هذا العلم حتى وصل إلينا على ما هو عليه الآن ،فجزا الله العلماء عنا كل الخير والجزاء.إذ أنهم بنوا لنا القواعد و صنفوا لنا التصنيفات ؛ كل ذلك للحفاظ على اللغة العربية.

ونحن في هذا البحث سنقوم بشرح و تفصيل القواعد و القوانين التي استخلصناها و استنبطناها من علم قواعد الصرف للغة العربية حيث تم تمثيل و تطبيق تلك القواعد على الحاسب اللآلي لدعم أنظمة إسترجاع المعلومات العربية ، فالحاسب اللآلي عندما يعالج البيانات فإن عمله لا يتعدى إستعمال و تنفيذ سلسلة من القوانين (الخوارزميات) التي نزوده بها عن طريق البرنامج ، تخبر الخوارزمية الحاسب أن يأخذ بعين الإعتبار كل الإحتمالات ماذا يجب أن يفعل ليكمل المهمة المطلوبة و يزودنا بالنتائج. القران الكريم هو كتاب الله الذي أنزل بالحق على رسولنا محمد صلى الله عليه وسلم على لسان العرب الفصيح وهو القائل جل في علاهَ "حم (1) تُنزِيلٌ من الرحمن الرحيم (2) كتاب فصلت ءاياته قرءانا عربيا لقوم يعلمون(3)" سورة فصلت . فبذلك فإن القران الكريم يحتوي على كل الإحتمالات النحوية والصرفية للغة العربية ، وفي هذا الصدد،قرّرنا أن نبدأ العمل مع نصوص سليمة ومختارة بعناية، فكانت البداية من النّص القرآني كأفضل طريقة للبدء بمثل هكذا مجهود.

عندما تستخدم محرك بحث يجب أن تضع بعض الكلمات المفتاحية التي تعرف بشكل واضح ما الذي تبحث عنه على سبيل المثال :إذا أردت أن تبحث عن" معلومات حول العالم " فاستخدم الكلمات المفتاحية التالية "معلومات العالم" . ومع

ذلك فإن بحثًا بسيطا باستخدام كلمات مفتاحية قليلة لايعطيك نتائج مفيدة في الغالب ،يمكنك أن تستعمل ميزات محرك البحث الاخرى لمساعدتك في البحث.الشكل 1-1يوضح محرك البحث(Google Scholar).



شكل 1-1: محرك البحث قوقل سكولار

الشكل 1-1 يوضح محرك البحث Google Scholar و هو واحد من نظم إسترجاع المعلومات (IRS). يجد الباحث العلمي من Google Scholar طريقة بسيطة للبحث بشكل واسع عن الأدب العلمي. من مكان واحد، يمكنك البحث عبر العديد من التخصصات والمصادر: الأوراق العلمية و البحوث والكتب والملخصات والآراء المحكمة، من الناشرين الأكاديميين والجمعيات المهنية والمستودعات على الانترنت والجامعات ومواقع أخرى على شبكة الإنترنت.

1- 2 مشكلة البحث

تتركز مشكلة البحث في أن معظم أنظمة الرد على الأسئلة في اللغة العربية تعطي النتائج مباشرة من غير أي معالجات مما يؤثر في عملية البحث ونتائجه، بزيادة النتائج الغير مناسبة أو الدقة فيها.

يتركز الحل المقترح لهذه المشكلة حول إعطاء النتائج بعد عملية المعالجة، هذه المعالجة تعمل على تحسن السؤال أو الإستعلام و بالتالي تحسين البحث و النتائج المسترجعة من حيث عددها و الدقة فيها.

1- 3 أهمية البحث

تواجه اللغة العربية الكثير من التحديات في عصر البيانات الرقمية تسهم في ظهور العديد من القضايا التي تعوق ظهور المحتوى العربي الرقمي على الإنترنت بشكل يليق بأمة العرب ، حيث تشير الأرقام إلى أن حصة الإنتاج الفكري باللغة العربية المتاح على الإنترنت لا يتجاوز 1% من المعلومات المنشورة على الويب وقد يعود ذلك لبعض الصعوبات اللغوية والفنية لتدعيم هذا المحتوى ، وترجع أهمية هذه الدراسة إلى:

- الإستنهاض باللغة العربية وتطوير قواعدها في انظمة البحث لتحسين أداء نظم إسترجاع المعلومات العربية، و بالتالي إثراء المحتوى العربي الرقمي على الإنترنت أو الوسائط الالكترونية.

نشر الوعي بثقافة الإهتمام بالمعالجة الآلية للغة العربية و بالتالي زيادة عدد الباحثين و البحوث في مجال المعالجة الآلية للغة العربية مما يؤدي الى الإعتمادية على اللغة العربية في محركات البحث.

- خطوة نحو طريق إغناء المحتوى الرقمي العربي.
- واحد من البحوث التي تغطي المعالجة الآلية للغة العربية.
- يستفيد الباحيث القادمين من هذه الدراسة و من النظام المقترح في فهم كيفية المعالجة الآلية للغة العربية و ذلك بمساعدة شاشة النظام الذكية و التفاعلية .

1- 4 أهداف البحث

- توضيح كيف أن اللغة العربية تواجه تحديات في المعالجة الآلية و إيجاد بعض الحلول.
- الإعتناء بالجانب الصرفي للغة العربية في المعالجة الآلية لها و تعريف و برمجة بعض قواعد الصرف للغة العربية مثل:إسم الفاعل،الصفة المشبة بإسم الفاعل،إسم التفضيل،إسما الزمان والمكان،إضافة اللواحق و السوابق للكلمة أو حذفهما.....الخ قدر المستطاع ،لتحسين أداء نظم إسترجاع المعلومات العربية .
- معالجة الإستعلام الذي طرحه المستخدم و توسعته لضمان إسترجاع كل الملفات ذات الصلة في عملية البحث و عدم فقدان الملفات التي تحتوي على الكلمات الأساسية.
- بناء واحد من أنظمة الرد الآلي على السؤال و التي تعمل على معالجة سؤال المستخدم لتحسين نتائج البحث من حيث العدد و الدقة.
- جعل النظام أكثر تفاعل مع المستخدم وسهل الإستخدام بحيث يختار المستخدم المعالجة و يمكن تكرار هذه العملية و أختيار المعالجة من مجموعة المعالجات التي يوفرها النظام إلى أن يرضى المستخدم بالنتائج.

1- 5 الحلول المقترحة

الحل المقترح مكون من جزئين الجزء الأول: برمجة أنماط مأخوذة من علم قواعد الصرف للغة العربية على الحاسب الآلي، و الجزء الثاني هو بناء نظام رد آلي تفاعلي للقرآن الكريم يتم فية تطبيق و إختبار الأنماط .

1- 6 منهجية البحث

لدينا الآيات (91)،(221)،(248)و(278) من سورة البقرة و المذكورة أدناه، كل آية تعتبر معلومة مخزنة يجب إسترجاعها عند مطابقة اي كلمة من سؤال المستخدم مع اي من الكلمات الموجودة في الآية .

" وَإِذَا قِيلَ لَهُمْ آمِنُوا بِمَا أَنْزَلَ اللَّهُ قَالُوا نُؤْمِنُ بِمَا أُنْزِلَ عَلَيْنَا وَيَكْفُرُونَ بِمَا وَرَاءَهُ وَهُوَ الْحَقُّ مُصَدِّقًا لِمَا مَعَهُمْ قُلْ فَلِمَ تَقْتُلُونَ أَنْبِيَاءَ اللَّهِ مِنْ قَبْلُ إِنْ كُنْتُمْ مُؤْمِنِينَ " سورة البقرة الآية(91)

" وَلَا تَنْكِحُوا الْمُشْرِكَاتِ حَتَّى يُؤْمِنَّ وَلَأَمَةٌ مُؤْمِنَةٌ خَيْرٌ مِنْ مُشْرِكَةٍ وَلَوْ أَعْجَبَتْكُمْ وَلَا تُنْكِحُوا الْمُشْرِكِينَ حَتَّى يُؤْمِنُوا وَلَعَبْدٌ مُؤْمِنٌ خَيْرٌ مِنْ مُشْرِكٍ وَلَوْ أَعْجَبَكُمْ أُولَئِكَ يَدْعُونَ إِلَى النَّارِ...الخ" سورة البقرة الآية(221)

" وَقَالَ لَهُمْ نَبِيُّهُمْ إِنَّ آيَةَ مُلْكِهِ أَنْ يَأْتِيَكُمُ التَّابُوتُ فِيهِ سَكِينَةٌ مِنْ رَبِّكُمْ وَبَقِيَّةٌ مِمَّا تَرَكَ آلُ مُوسَى وَآلُ هَارُونَ تَحْمِلُهُ الْمَلَائِكَةُ إِنَّ فِي ذَلِكَ لَآيَةً لَكُمْ إِنْ كُنْتُمْ مُؤْمِنِينَ" سورة البقرة الآية(248)

" يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَذَرُوا مَا بَقِيَ مِنَ الرِّبَا إِنْ كُنْتُمْ مُؤْمِنِينَ" سورة البقرة الآية(278)

إذا قام المستخدم بإدخال إستعلام يحتوي على كلمة "مؤمن" فالنتيجة هي الآية رقم (221) بينما لا يتم إسترجاع الآيات (91)، (248) و (278) و التي من المفترض إسترجاعها، هذا لأن الآية(221) تحتوي على كلمة (مؤمن) و التي تطابق الكلمة في سؤال المستخدم، بينما الآيات(91)، (248) و (278) تحوي على كلمة (مؤمنين) و التي بالتأكيد لا تطابق كلمة (مؤمن). لجعل الكلمة (مؤمن) تطابق (مؤمنين) لابد من عمل معالجة لها، معظم نظم إسرجاع المعلومات العربية بما فيها أنظمة الرد الآلي على السؤال تفتقد هذه المعالجات في كثير من الأنماط هذا يؤدي الى عدم الدقة في النتائج أو المعلومات المسترجعة.

لفهم الطريقة المقترحة لحل المشكلة فيما أعلاه تم إستعراض جانب من مشكلة البحث و يتمثل الحل لها في الآتي:

جمع المذكر السالم واحد من الأنماط التي يغطيها علم الصرف يتم فيه إضافة (ين) في آخر الكلمة، نحو: "مؤمن" مفرد مذكر يتم إضافة اللاحقة(ين) فتصير "مؤمنين" جمع مذكر سالم. وهكذا يتم حل مشكلة جمع المذكر السالم وهي جزء من مشكلة البحث. و هنالك الكثير من الأنماط جمع المؤنث السالم، اسم الفاعل ، اسما الزمان و المكان...الخ و تحل بنفس الطريقة فقط يتم دراسة النمط الذي يقدمه لنا علم قواعد الصرف للغة العربية و ملاحظة التغير في بنية الكلمة يلهمنا برصد القواعد للمعالجة الآلية.

1- 7 حدود البحث

1.7.1. الحدود الزمانية

فترة سبعة أشهر لتنفيذ المشروع بدءاً من أبريل2017م إلى أكتوبر2017م

1.7.2. الحدود المكانية

القرآن الكريم (سورتي الفاتحة و البقرة).

1- 8 هيكلية البحث

يتكون هذا البحث من خمسة أبواب، الباب الأول فيه: المقدمة العامة، الأهداف، الأهمية، المشكلة، الحلول المقترحة، المنهجية المتبعة وحدود البحث. الباب االثاني فيه : الإطار النظري ويتكون من مبحثين، الأول: نظم إسترجاع المعلومات والثاني: اللغة العربية، كما يحتوى الباب الثاني على الدراسات السابقة. الباب الثالث : والذي يمثل منهجية البحث فيه: الأدوات والتقنيات، إطار البحث المقترح وتحليل الأنماط. . الباب الرابع: فيه الجانب العملي للبحث و جانب من الشاشات التوضيحية للنظام. الباب الخامس: وفيه نتائج البحث، توصيات الباحث والخاتمة.

الباب الثاني

الإطار النظري و الدراسات السابقة

الفصل الأول

الإطار النظري

2 - 1 خلفية عن نظم إسترجاع المعلومات

في هذا المبحث يتم توضيح المفاهيم الأساسية و التي تعتبر الكلمات المفتاحية للبحث. و يحتوي على خمسة مقاطع، وهي: المعالجة الآلية للغة الطبيعية(NLP)، خلفية عن إسترجاع المعلومات(Information Retrieval)، نظام الإجابة على الأسئلة(question answering system)، نظام الإجابة على الأسئلة التفاعلي(interactive QAS)، المدونات اللغوية(corpus or data set).

2 - 1 - 1 المعالجة الآلية للغة الطبيعية

تتعدد المصطلحات والتعريفات المعبرة عن قضية المعالجة الآلية للغة الطبيعية (NLP)، فهناك هندسة اللغة واللغويات الحاسوبية، ومن أبرز التعريفات السائدة حول معنى المعالجة الآلية للغة التعريف الذى قدمته جامعة شيفلد البريطانية شرحا آخر لمعنى المعالجة الآلية للغات الإنسانية قالت فيه أن المعالجة الآلية للغة تعنى استخدام أجهزة الحاسب فى معالجة اللغة المكتوبة والمنطوقة من أجل أغراض عملية مفيدة مثل الترجمة الآلية بين اللغات واستخلاص المعلومات من مواقع الويب وقواعد البيانات وبنوك المعلومات المتصلة بالإنترنت للحصول على إجابات للأسئلة أو من أجل إجراء حوار مع الحاسب أو الآلة للحصول على استشارة أو معلومة ما[3].

2 - 1 - 2 إسترجاع المعلومات

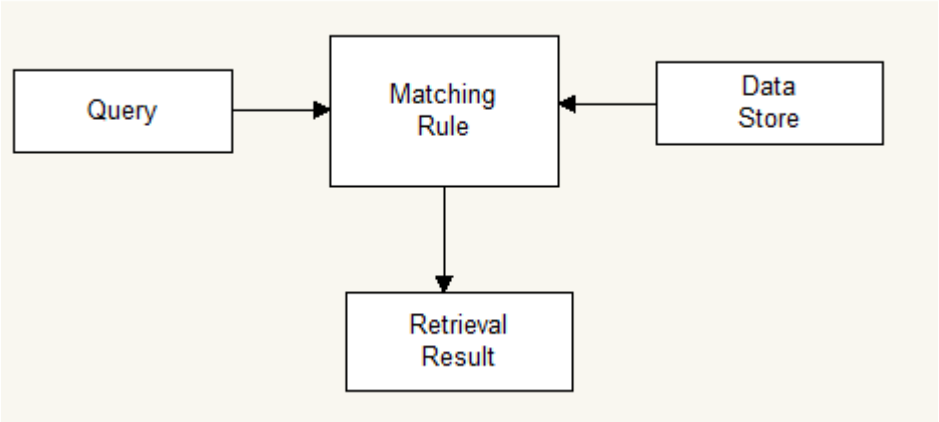
إسترجاع المعلومات (Information Retrieval) هو فرع من علوم الحاسوب يهتم بفهرسة و عملية بحث في الوثائق و الملفات و المواقع الالكترونية لتحسين عملية البحث.أصبح إسترجاع المعلومات واحد من أهم الدراسات و البحوث التي تعمل لتحسين عملية نظم إسترجاع المعلومات[4].

2 - 1 - 3 تعريف إسترجاع المعلومات

هو أحد النظم الفرعية في نظام المعلومات ووظيفته الأساسية هي إختزان واسترجاع المعلومات وفقا لاحتياجات مجتمع المستفيدين. ويمكن تعريفه أيضاً على أنه علم البحث عن الوثائق وعن المعلومات داخل الوثائق وعن البيانات الوصفية التي تصف الوثائق، بالاضافة إلى البحث في قواعد البيانات وشبكة الانترنت[5].

2 - 1 - 4 مفاهيم أساسية في نظم إسترجاع المعلومات

نظم إسترجاع المعلومات تحتوي علي عمليتين :الفهرسة و عملية البحث.الشكل رقم يوضح النموذج العام لنظم إسترجاع المعلومات.



شكل 2-1: النموذج العام لنظم إسترجاع المعلومات

2-1-4-1 الفهرسة

الهدف من الفهرسة (Indexing) زيادة السرعة و الأداء لعملية البحث .من غير عملية الفهرسة ،محرك البحث أو نظام الرد الآلي يجب أن يمسح أو يمر بكل الوثائق أو الملفات حتي يجد الهدف (الملف) الذي يجب إسترجاعه ،وهذا يتطلب زمن و طاقة الكمبيوتر[5].

2-1-4-2 عملية البحث

عملية البحث (Searching) تعني كيف يمكن البحث عن المعلومات في الوثائق مستخدمًا نظم إسترجاع المعلومات(IRS) للحصول علي الأجابة الصحيحة للإستعلام(query)[5].

2-1-4-3 المعالجة الأولية

المعالجة الأولية(Pre-Processing) هي خطوة مهمة للفهرسة و عملية البحث ؛إذا انها تصنع التوافق بين الإستعلام (query statement) و البيانات المفهرسة(indexing data).أهمية هذه الخطوة أننا نسمح للمستخدم كتابة الأستعلام من غير أي تقييد و جعل التوافق بين الأستعلام و بيانات الفهرسة[5].المعالجة الأولية تحتوي على العديد من عمليات كل هذه العمليات كانت نتائج من الدراسات السابقة. و من المعالجات الأولية :

2-1-4-3-1 التقطيع

التقطيع (Tokenization) هي عملية تقطيع الإستعلام الذي طرحه المستخدم والذي هو عبارة عن مجموعة متدفقة من الكلمات يفصل ما بينها مسافات (space) بحيث يتم تقطيعه إلى كلمات مفصولة عن النص. مثال: النص "من هو الرحمن الرحيم ؟" تقطع الي الكلمات الآتية:"من" "هو" "الرحمن" "الرحيم" "؟".

2-1-4-3-2 التحويل الي الحروف الصغيرة

عادة تتم هذه العملية في معالجة اللغة الإنجليزية ،وهي تعني تحويل كل الحروف في الكلمة الواحدة الي حروف صغيرة (Lower case).

2-1-4-3-3 حذف التشكيل

تتميز اللغة العربية عن باقي اللغات الأخرى بالتشكيل (Diacritic) الذي يرسم على حروفها، و التشكيل هو علامات و حركات لضبط الأحرف حتى تنطق بشكل صحيح، و علامات التشكيل هي: (الكسرة "ِ " ، الفتحة "َ "، الضمة "ُ " ، السكون "ْ " و التنوين). فالحاسوب يعتبر الكلمتان (ذَهَبَ) بالتشكيل و (ذهب) من غير تشكيل كلمتان مختلفتان وبالتالي فإن ذلك يؤثر في نتائج البحث إذا أخطأ المستخدم في التشكيل أو لم يشكل الكلمة.

بما أن التشكيل يمثل خطورة في آلية البحث فإن الدراسات السابقة أثبتت أن حذف التشكيل (Remove Diacritic) ضروري في عمليات المعالجة الأولية و يقود إلى فعالية البحث، هذه إحدى مسلمات الدراسات السابقة بحيث يتم أخذ نسخة من إستعلام المستخدم و نسخة من الملفات التي يتم البحث فيها و حذف التشكيل من كلا النسختين و تطبيق آلية البحث فإذا وجدت نتائج يتم إسترجاع النسخ الأصلية.

2-1-4-3-4 حذف الحروف الخاصة بالقرآن الكريم

هنالك رموز أو حروف خاصة تفرد بها النص القرآني تبين هذه الموز بعض أحكام التجويد، ومن هذه الحروف الخاصة مصطلحات ضبط الوقف وهي: (صلى) (قلى) (لا) (النقطة المثلثة) (ج) (مـ). و ايضا هنالك حرف خاص و هو الألف الخنجرية و غيره. كل هذه الحروف الخاصة توجد فوق الكلمة و في عملية المعالجة الأولية شأنها شأن التشكيل يجب حذفها من الملف الموجود و من استعلام المسنخدم إن وجدت و هذا يقود إلي فعالية البحث كما أثبتت الدراسات السابقة.

2-1-4-3-5 إستبدال بعض الحرف

سبعض الكُتّاب يلجأون لكتابة الحروف(إ،أ،آ) بالحرف (ا) تماشيا مع هذا يتم استبدال هذه الحروف من الإستعلام أو الملفات بالحرف(ا). وعلى سبيل المثال الكلمتان (أمة) و (امة) يعتبرهما الحاسوب كلمتان مختلفتان، فإذا كانت الكلمة (أمة) هي التي موجودة في الملف الذي يتم البحث فيه و لكن المستخدم كتب كلمة (امة) للاستعلام و بالتالي لا توجد نتائج لعملية البحث. لجعل الحاسب يفهم أن كلمة (امة) هي نفسها الكلمة (أمة) لا بد في عمليات المعالجة الأولية كل الحروف (أ،إ،آ) بالحرف(ا) من النسختين نسخة الإستعلام و نسخة الملف.

2-1-4-3-6 حذف الكلمات التوقيفية

هنالك كلمات شائعة و تسمى الكلمات التوقيفية (Stop Words) و هي لا تفيد كثيرا في عملية البحث بل قد تؤثر في عملية البحث ويتم حذف هذه الكلمات من الإستعلام .وتختلف هذه الكلمات من لغة الي أخرى وهي if، is ، to ،be،(etc...)في الإنجليزية،أما في اللغة العربية فهي(هي،هو،الذي،على،من،ان،و،عن،....الخ).الملحق(أ) فيه جانب من الكلمات التوقيفية في اللغة العربية.

2 - 1 - 5 نظام الإجابة على الأسئلة

أنظمة الإجابة على الأسئلة (Question Answering System) أو كما تسمى أنظمة الرد الآلي تعتبر من الأنظمة المنتمية لإسترجاع المعلومات، حيث يفترض بالحاسب أن يستطيع الإجابة على أي أسئلة باللغة الطبيعية لدى البشر من خلال البحث فى مجموعة ضخمة من المستندات والوثائق مثل شبكة الويب العالمية أو الوسائط الألكترونية، وتتطلب الإجابة على الأسئلة وسائل معالجة آلية متقدمة للغات الطبيعية، ويعتبرها الكثيرون المرحلة التالية بعد محركات البحث، حيث تقوم بطرح سؤال عن شيء معين، وينبغى على الحاسب أو محرك البحث أن يعرض لك الإجابة فقط أو المستندات التى تحوى الإجابة فقط. ومن الإشكالات في نظم الإجابة على الأسئلة كأن يطرح المستخدم هذا السؤال :

هل الرياض عاصمة السودان

لا الرياض عاصمة السعودية.

لا إنها الخرطوم.

أيهما أفضل الإجابة الأولى أم الثانية،في الواقع أفضل الإجابة الأولى.

هل من الممكن أن تعطني الكوب الذي بجوارك؟ فهذا ليس سؤال أو إستعلام وإنما هو طلب.

2 - 1 - 6 نظام الإجابة على الأسئلة التفاعلي

نظام الإجابة على الأسئلة التفاعلي(Interactive QAS) من أنظمة الإجابة على الأسئلة و لكن فيه جانب من تفاعل المستخدم مع البرنامج وذلك بإعادة صياغة الاستعلام عن طريق التغذية الراجعة من المستخدم الى النظام من خلال واجهة النظام الذكية.حيث يطلب النظام من المستخدم أن يحدد ما إذا كانت الإجابات المسترجعة كرد على استعلامه ذات صلة بالموضوع الذي يريده أم لا، ومن ثم تعاد صياغة الاستعلام وفقاً لذلك عن طريق إزالة كلمات تؤثر على عملية استرجاع الاجابات من حيث الدقة، أو إعادة وزن كلمات البحث. ويدخل هذا الاستعلام الجديد في نظام استرجاع المعلومات لاسترجاع مجموعة أخرى من الوثائق والتي ربما تكون أكثر صلة بالموضوع، وأحياناً تكرر هذه الطريقة حتى يرضى المستخدم عن النتائج التي حققت.

2 - 1 - 7 المدونات اللغوية

المدونات (Corpus or data set) اللغوية أو مهمة جدا لنظم إسترجاع المعلومات إذ أنها تحتوي على مجموعة ضخمة من البيانات المنظمة والمرتبة للأغراض العامة ،فنحن أستخدمنا قاعدة بيانات اكسس للقرآن الكريم تحتوي على سورتي الفاتحة و البقرة ، في الباب الرابع يتم ذكر المدونة المستخدمة .

2 - 2 اللغة العربية

اللغة العربية من اللغات السامية و تعتبر اللغة الأم لأكثر من 330 مليون شخص،اللغة العربية هي لغة مهيكلة و لغة إشتقاقية حيث تعمل المورفولوجيا(علم التشكل) فيها دورا هامًا جدًا. يجب أن تتعامل تطبيقات معالجة اللغة العربية مع العديد من المشاكل المعقدة ذات الصلة بطبيعة و هيكلة اللغة العربية.على سبيل المثال :اللغة العربية تكتب من اليمين الي اليسار،و مثل اللغات الصينية،اليابانية و الكورية لا توجد حروف كبيرة في اللغة العربية،بالإضافة الي ذلك تتغير حروف اللغة العربية وفقا لموقعها من الكلمة،و كذلك مثل اللغة الإيطالية، الإسبانية، الصينية، واليابانية، والعربية هي لغة مؤيدة للإسقاط، فإنه يسمح الضمائر الموضوع لإسقاط[4].

كما سوف يتم توضيح المفاهيم الأساسية لفهم اللغة العربية و قواعد الصرف لتمهيد الطريق لكفية معالجة اللغة العربية آليا، ويحتوي هذا الفصل على: خصائص تراكيب اللغة العربية، تحديات تواجه المعالجة الآلية للغة العربية، الفرص التى يمكن اقتناصها عند معالجة اللغة العربية آليا، مدخل إلى تعريف علم الصرف.

2 - 2 - 1 خصائص تراكيب اللغة العربية

إن اللغة العربية تحتاج إلى إبراز خصائصها ومميزاتها في جميع مستويات تكوينها الصوتية والصرفية والتركيبية والدلالية لندرك الماهية العامة لهذه اللغة المجتباة، التي اختارها الله تعالى لغة لكتابه الكريم المعجز. وهذه المستويات[6] هي:

2-2-1-1 المستوى الصوتي

ويتكون من الأصوات أو ما نسميه الحروف المنطوقة، وهي أصغر وحدة بنائية في اللغة، مثل: بَ، تَ، سَ، طَ وهكذا، ويسمى التكوين الصوتي.

والمستوى الصوتي هو الميدان الذي يبحث فيه علم الأصوات .

ومن الأصوات تتكون المقاطع الصوتية، ومن المقاطع الصوتية تتكون (المورفيمات)، وهي جزء من الكلمة التي تدل على معنى متصل بالكلمة فمثلاً (ال) للتعريف، تعد (مورفيما) يدل على التعريف، ويتكون من مقطع واحد هو (ال)، وهو مرتبط بجزء الكلمة (معلم)، فـ(ال) مقطع ومورفيم، و(معلم) (جزء) من كلمة، تتكون من ثلاثة مقاطع :

(معلـ) و(ون) مورفيم مكون من مقطع واحد يدل على الجمع كل ذلك يكون كلمة متكاملة هي (المعلمون) ذات (مورفيمات) ثلاثة (ال + معلم+ ون).

2-1-2-2 المستوى الصرفي

يتكون المستوى الصرفي من الأصوات التي تتجمع فتتكون منها الوحدات الصرفية، والوحدة الصرفية هي أصغر وحدة لغوية ذات معنى، مثل: (ال) لمعنى التعريف، ونسمي الوحدة الصرفية أيضاً (مورفيمات)، فإذا ضربنا مثلاً كلمة (كتاب) فهي تعد وحدة صرفية واحدة، فإذا أدخلنا عليها (مورفيما) في البداية مثل (ال) أضفنا إليها التعريف فتصبح (الكتاب)، وإذا أضفنا (مورفيما) في النهاية مثل الضمير (الهاء) فنقول (كتابه) وهكذا .

وهكذا يتم التصريف عبر الإضافة، أو الحذف، أو القلب المكاني مثل كتب، كبت، تكب، بكت ... إلخ.

ومحور الدراسة اللغوية في المستوى الصرفي هو الكلمة أو اللفظ .

3-1-2-2 المستوى التركيبي

عندما تتوالي الألفاظ أو الكلمات أو البني اللفظية في تنسيق كلام معين، فإنه يكون الجملة مثل: كتب أحمد الدرس، أو مثل: محمد كتب الدرس، ويكون ذلك بضوابط معينة، هي الضوابط النحوية والدلالية أيضاً .

"وإن محور الدراسة في المستوى التركيبي هو الجملة أو التركيب اللغوي".

4-1-2-2 المستوى الدلالي (علم المعاني)(علم الدلالة)

وهو الغاية من تركيب الألفاظ في جمل منتظمة، فإذا لم تكن للتراكيب أو الجمل دلالات فإنها لا تعني شيئاً في اللغة، وبذلك ندرك أن دلالة هذه الألفاظ أو التراكيب أو الفقرات أو النصوص هي موضوع الدراسة اللغوية في المستوى الدلالي.

يعد أوسع المستويات اللغوية من حيث اشتماله على الموضوعات الكثيرة والمرتبطة بالدلالة في اللغة العربية.والمعالجة الآلية في تطبيقاتها المختلفة تنظر إلى تحقيق هدفين متزامنين: هما دقةُ النتائج، وتعميمُها على الظاهرة اللغوية كلها، وتتفاوت نسبة تحقق هذين الهدفين باختلاف المستوى اللغوي المستهدف من المعالجة الآلية، ففي المستوى الصوتي والصرفي بلغت نسبة عالية.تجاوزت 95%، أما معالجة المستوى الدلالي فلا تزال في بداية الطريق، وذلك لصعوبة الظاهرة اللغوية على المستوى الدلالي.

2 - 2 - 2 تحديات تواجه المعالجة الآلية للغة العربية

تميزت اللغة العربية بميزات و خصائص عديدة مما يجعل حوسبة اللغة العربية عملية صعبة و معقدة ، فهناك تحديات عديدة تواجه اللغة العربية ، ولا شك في ذلك حيث ان لكل لغة خصائص تسهل المعالجة الآلية واخرى اكثر تعقيداً . ولدراسة حالة اللغة العربية هنالك جوانب عديدة منها:

1-2-2-2 الجانب اللغوي

وهو الجانب المتعلق بخصائص اللغة العربية والمقسم الى اربعة مستويات التي ذكرت سابقا مقطع(2.2.2) ، حيث أن اللغة العربية لا تحتوي على الحروف الكبيرة أو الصغيرة فمن الصعوبة الاعتماد على الشكل الطباعي لمعرفة بداية الجملة، كذلك الامر بما يخص الإعتماد على علامات الترقيم والتنقيط لتحديد نهاية الجملة. ففي الكتابات اللاتينية، باستطاعتنا ان نبني قواعد على شكل الحرف الاول من الكلمة لمعرفة بداية الجملة. اما في الكتابات العربية، فلا يوجد اشارات ظاهرة تحدد ذلك. هذا الأمر أيضاً يشكل عائقاً في وجود وادراك اسماء العلم.

ومن الصعوبات الاخرى التي تواجه المعالجة الآلية للغة العربية، مشكلة الإلتباس في حروف الكلمات. فالنصوص العربية غير المشكّلة تؤدي الى الإلتباس والابهام. ونسبة الإلتباس الشكلي تفوق الـ90 بالمئة، مثالاً على ذلك : كلمة"مهم"هل هي الحرف"ميم"+ الضمير"هم"أو كلمة"مهم"؟

كما من ضمن المشاكل المهمة الاخرى، هي مشكلة الادماج أو الإلصاق حيث تتميز اللغة العربية بالإلصاقات(affixes) التي تتعدى الحروف الإلصاقية الى حروف العطف (Conjunction) والجر(Preposition)، فمثلاً حرفي"الواو"و"الفاء"بسبب ادماجهم بالكلمة، من الصعب التمييز ما اذا كان كل حرف من هذين الحرفين هما من اصل الكلمة أم لا، مثالاً على ذلك :كلمة"وهْم"هل"الواو"هي من أصل الكلمة أو هي حرف عطف و في القرآن الكريم "وإسماعيل " وكذالك "وإسرائيل" "ويرسل" ....الخ.

لجوء بعض الكتاب إلى كتابة الحروف (أ ، إ ، آ ، ا) بطريقة متشابهة كـ (ا).

كما أن اتجاه الكتابة العربية هو من اليمين إلى اليسار.

تتعدد الأشكال البصرية للحرف الواحد تبعاً لموقعه من الكلمة(عـ،ـعـ،ع).

هذه المشاكل ليست بالمستعصية. تشهد المختبرات في الجامعات والدول العربية وبعض الشركات الخاصة حلولاً لهذه المشاكل، رغم انه من الصعب في المعالجة الآلية للغة التكلم عن حلول مئة بالمئة: طبيعة اللغة من ناحية فهمها وتحليلها آلياً أو بطريقة غير آلية لا تُحل جذرياً[3].

2-2-2-2 الجانب الفني او التقني

و المتعلق الفني او التقني الذي يعتمد على برمجيات حوسبة اللغة العربية وعدم توفرها والتي خصصت في الغالب لبرمجة اللغة الانجليزية و غالب اللغات اللاتينية ، وقلة المتخصصين في مجال حوسبة اللغويات.

2 - 2 - 3 الفرص التى يمكن اقتناصها عند معالجة اللغة العربية آليا

- استطاعت اللغة العربية الاستمرار كلغة حية لمدة خمسة عشر قرنًا فى حين فشلت اللغات الأخرى فى ذلك.
- فهى اللغة التي يؤدى بها أكثر من مليار مسلم صلواتهم يوميًا.
- وهى لغة ثرية من حيث النظريات اللغوية، وكافية للتعبير عن أفكار واحتياجات الناطقين بها.
- وهى لغة قادرة على تجديد نفسها ذاتيًا.

- من جهة المسائل الايجابية عند تحليل العربية، فبطبيعتها الهيكلية وبتركيبتها الشكلية للمفردات مؤهلة في بعض الحالات للمعالجة الآلية، وخصوصاً فيما يخص حساب الجذر، اذ انه باستطاعتنا الوصول الى جذر الكلمة آلياً عبر معادلات رياضية وقواعد مبنية على تركيبات منظمة وعلى المنهج الكشفي باستعمالنا ادوات الربط والوصل والاحرف الوسطية التي هي منسقة في كثير من حالاتها.

2 - 2 - 4 مدخل إلى تعريف علم الصرف و دراسة المشتقات

1-4-2-2 تعريف علم الصرف

الصرف لغة: التغيير و التحويل،ومنه تصريف الرياح أي تغييرها،وإصطلاحا هو العلم الذي يبحث في أبنية الكلمة العربية وصيغتها، وبيان ما في حروفها من أصالة و زيادة أو حذف،أو صحة أو إعلال أو إبدال إلى غير ذلك.

2-4-2-2 ما يدخله التصريف من أقسام الكلمة

موضع علم التصريف أمران:الأسماء المتمكنة(المعربة) و الأفعال المتصرفة، و من ثم لا تدخل الحروف و لا الأسماء المبنية و الأفعال الجامدة مثل:عسى و ليس و بئس و نعم ، ولا يقبل التصريف ما كان حرفا واحدا أو حرفين إلا إذا كان محزوفا منه مثل(قل من قال) و(ق زيدا) من وقى و يد من "يدي" .

3-4-2-2 الميزان الصرفي

هو معيار لفظي اتفق عليه الصرفيون و اختاروا له ثلاثة أحرف هي الفاء و العين واللام (فعل) ليدلوا بها على هيئة الكلمة و ما يعتريها من تغيير سواء أكان حذفا أم زيادة أم تصفيرا أم تكسيرا أم نسبا مثل : ذَهَبَ بوزن فَعَلَ ،و عَلِمَ بوزن فَعِلَ ،و شَرُفَ بوزن فَعُلَ ،و يُقْبَلُ بوزن يُفْعَلُ،و هكذا.

4-4-2-2 فائدة دراسة الميزان الصرفي

هو بيان حالة الكلمة و هيئتها ،وما في حروفها من أصالة أو زيادة أو حذف أو إعلال أو إبدال أو تصغير أو تكسير،مثل عَلِمَ على وزن فَعِلَ (فحروفها أصلية كلها) ، و(كَافَحَ) على وزن فاعل فالألف مزيدة على حروفها الأصلية، و (قُلْ) على وزن (فُلْ) حذفت الواو المقابلة لعين الميزان، و"قلوب" على وزن فُعُول، و "قِ" على وزن "عِ"، فالميزان هو الذي يرصد هذه التغيرات الصرفية جميعها و يبين الزائد منها و المحزوف و نحوه من التغيرات الحاصلة في الكلمات.

5-4-2-2 كيف توزن الكلمات وزنا صرفيا؟

إذا كانت الكلمات المراد وزنها ثلاثية(اسماً كانت أم فعلاً) قوبل الحرف الأول منها بالفاء، والحرف الثاني بالعين، والحرف الثالث باللام، ثم يضبط الميزان بضبط الموزون، نحو: شَمْس بوزن فَعْل، كُتِبَ بوزن فُعِلَ، و نَصْرٌ بوزن فَعْل، وهكذا.

وزن الكلمات المزيدة:

إذا كانت الكلمة مزيدة بتكرير حرف أصلي(أي تضعيفه)ضعف نظيره في الميزان مثل كُرِّمَ على وزن فُعِّلَ، وإذا كانت الزيادة ليست بتكرير حرف أصلي بأن كانت من حروف(سألتمونيها) نزلت الأحرف المزيدة في أماكنها في الميزان مثل: حامد (فاعل)، مستغفر (مستفعل)، وإذا حصلت في الكلمة زيادتان أعطينا لكل زيادة حكمها نحو: تَعَلَّمَ على وزن (تَفَعَّلَ)، وهكذا.

وزن ما حدث فيه حذف من الكلمة:

إذا حدث في الكلمة حذف حذف نظيره في الميزان، فإذا حذف أول الكلمة حذف أول الميزان نحو: خُذْ (بوزن عُلْ)، وإذا حذف وسط الكلمة حذف وسط الميزان نحو: (قُلْ) بوزن "فُلْ"، وإذا حذف آخر الكلمة حذف آخر الميزان نحو: اُدْعُ بوزن"اُفْعُ"، وكذا إذا حذف أوله و آخره و بقي وسطه نحو: "ع" (أمر من وعى بوزن ع)، وهكذا إلى بقية الإحتمالات.

6-4-2-2 دراسة المشتقات

في هذا البحث تمت دراسة بعض المشتقات[7] و فيما يلي شرحها:

1-6-4-2-2 إسم الفاعل تعريفه و صياغته

تعريفه

هو إسم مشتق من الفعل للدلالة على الحدث و فاعله، مثل محمد كاتبٌ رسالة،(فكاتب) دلت على الكتابة ومن قام بها، و نحو:"إِنِّي جَاعِلٌ فِي الْأَرْضِ خَلِيفَةً".

صياغته

صياغة إسم الفاعل من الفعل الثلاثي

يصاغ إسم الفاعل من الفعل الثلاثي على وزن فاعل، نحو: فهم الطالب الدرس فهو فاهم و نحو:" **ءأَنتُمْ تَخْلُقُونَهُ أَمْ نَحْنُ الْخَالِقُونَ"**.

صياغة إسم الفاعل من الفعل الثلاثي الأجوف

يصاغ إسم الفاعل من الفعل الثلاثي الأجوف(الذي وسطه حرف علة، وحروف العلة هي الالف الواو الياء) على وزن فاعل بعد قلب عين اسم الفاعل همزة سواء أكانت واوية أم يائية نحو:" قَالَ قَائِلٌ مِّنْهُمْ" (فقائل) أصلها (قاول) لأنها من الفعل قال يقول فأصل الألف واو فقلبت همزة لوقوعها في مقابلة عين اسم الفاعل.

صياغة اسم الفاعل من الفعل الثلاثي الناقص

عند صياغة اسم الفاعل من الفعل الثلاثي الناقص(الذي آخره حرف علة) يحذف آخر اسم الفاعل، و يحذف نظيره في الميزان، نحو: تلا القرآن فهو تال ، و دعا فهو داع، قال-تعالى:" كُلُّ مَنْ عَلَيْهَا فَانٍ".

صياغة اسم الفاعل من غير الثلاثي

إذا كان الفعل غير ثلاثي صيغ منه اسم الفاعل على وزن بأن يوتى منه بالمضارع، ثم يبدل حرف المضارعة فيه ميماً مضمومة و كسر ما قبل الآخر نحو:"أقبل يقبل فهو مقبل" و نحو: (فَذَكِّرْ إِنَّمَا أَنْتَ مُذَكِّرٌ (21) لَسْتَ عَلَيْهِمْ بِمُسَيْطِرٍ ((22)).

صياغة اسم الفاعل من الفعل الخماسي الأجوف

يصاغ اسم الفاعل من الفعل الخماسي الأجوف (معتل الوسط) نحو: اختار و اكتال و اختار على وزن مفعال.

وبهذا نكون قد اكتفينا بهذه الصيغ من اسم الفاعل التي تمت معالجتها.

2-2-4-6-2 صيغ المبالغة

تعريف

هي نوع من المشتقات يودي ما يؤديه اسم الفاعل (وهو الدلالة على الحدث و من قام به) و لكن على سبيل المبالغة و التكثير، مثل: ألوفٌ كريمٌ لكنه حذر اي(يألف غيره كثيرا و كثير الكرم و كثير الحذر).

أشهر صيغ المبالغة

صيغ البالغة المشهورة هي خمسة صيغ.

جدول 2-1: أمثلة لصيغ المبالغة المشهورة

الرقم	صيغة المبالغة	الفعل	على وزن صيغة المبالغة
1	فعَّال	جبُرَ،همز،لمز	جبَّار،همَّاز،لمَّاز
2	مِفْعَال	عطر،ضيف	معطار،مضياف
3	فعُول	شكر،غفر،ودد	شكور،غفور،ودود
4	فَعيل	سمع،رحم،بصر	سميع،رحيم،بصير
5	فَعِل	حَذَرَ،نَهَمَ،بَطَرَ	حذِر،نهِم،بطِر

جدول 2-2: أمثلة لصيغ المبالغة غير المشهورة

الرقم	صيغة المبالغة	على وزن صيغة المبالغة
1	فعَال	فساق
2	فُعل	غُدر
3	فُعلة	همزة و لمزة
4	فاعلة	راوية
5	مِفْعيل	معطير
6	**فاعول**	فاروق

هل تصاغ صيغ المبالغة من غير الثلاثي؟

الأصل في صيغ المبالغة أن تصاغ من الثلاثي ولكن سمع مجيئها من غير الثلاثي مثل:معطاء.

2-2-4-6-3 الصفة المشبهة معناها وصياغتها:

تعريف الصفة المشبهة

هي كل اسم صيغ من الفعل اللازم للدلالة على صفة ثابتة ملازمة لصاحبها نحو: طويل،قصير،جميل،جبان...

صياغة الصفة المشبهة من الفعل الثلاثي

تصاغ الصفة المشبهة من الفعل الثلاثي الذي على وزن "فَعِلَ" أو "فَعُلَ".

جدول 2-3: أمثلة لصيغ الصفة المشبهة من الفعل الثلاثي على وزن فَعِلَ

صيغة الصفة المشبهة	الفعل	على وزن الصيغة
فَعِلٌ	حَزِنَ،فَرِحَ	حزِنٌ،فَرِحٌ
أفعل – فعلاء	سود	هو اسود – وهي سوداء
فعلان – فعلى	عطش	هو عطشان – وهي عطشى

جدول 2-4: أمثلة لصيغ الصفة المشبهة من الفعل الثلاثي على وزن فَعُلَ

صيغة الصفة المشبهة	**الفعل**	**على وزن الصيغة**
فعيل	كرم،جمل	كريم،جميل
فَعَلَ	حسُن،بطَل	حسَنَ،بَطَلَ
فُعال	شجُع،همُمَ	شجاع،همام
فَعَال	جَبُن،حصُن	جَبان،حَصان(عفيفة)
فَعْل	ضخُم،صعُبَ	ضَخْم،صَعْب

تم معالجة أوزان للصفة المشبهة بالإضافة الى الأوزان (فعلة،فعيلة،فاعل،فاعلة،فعالة،فيعل،فعول،فعلون).

2-2-4-6-4 أفعل التفضيل:

هو وصف يدل أن شيئين إشتركا في صفة، وزاد أحدهما على الآخر في تلك الصفة نحو:(محمد أكرم من علي) ونحو:" وَالْفِتْنَةُ أَشَدُّ مِنَ الْقَتْلِ " و نحو " إِذْ قَالُوا لَيُوسُفُ وَأَخُوهُ أَحَبُّ إِلَىٰ أَبِينَا مِنَّا...."

شروط صوغ اسم التفضيل

يصاغ اسم التفضيل من كل فعل إستوفى شروطا ثمانية، هي: أن يكون فعلا ثلاثياً متعدياً متصرفاً قابلاً للتفاضل(يعني زيادة و نقصاناً) و ألا يكون الوصف منه على وزن أفعل فعلاء و أن يكون مثبتاً(غير منفي) و أن يكون مبنياً للمعلوم(غير مبني للمجهول) نحو: عليٌ أفضل من خالد، الشمس أكبر من الأرض.

2-2-4-6-5 اسما الزمان والمكان:

- **اسم المكان**

هو اسم مصوغ من الفعل للدلالة على مكان وقوع الفعل(اي: الحدث) نحو: "هنا مجلس العلم ، ذاك موقف الحافلات" و كقوله-تعالى-:" **حَتَّىٰ إِذَا بَلَغَ مَطْلِعَ الشَّمْسِ** " و قوله سبحانه:" **حَتَّىٰ إِذَا بَلَغَ مَغْرِبَ الشَّمْسِ**".

- **اسم الزمان**

هو اسم مصوغ من الفعل للدلالة على زمن وقوعه مثل قوله تعالى:" ..**إِنَّ مَوْعِدَهُمُ الصُّبْحُ ۚ أَلَيْسَ الصُّبْحُ بِقَرِيبٍ** ".

صياغة اسم الزمان و المكان

يصاغ اسم الزمان والمكان على وزن (مَفْعِلْ) او (مَفْعَلْ) من الثلاثي و من غير الثلاثي(مُفَعَّل).

2-2-4-6-6 اسم المرة:

مصدر المرة هو مصدر يدل على وقوع الحدث مرة واحدة، و يصاغ من الفعل الثلاثي على وزن (فَعْلَة) مثل: جلست جلسة و وقفت وقفة.

و هنالك اسم المرة من الأفعال غير الثلاثية هذا ما لم نطرقه في هذا البحث.

2-2-4-6-7 اسم الهيئة:

هو مصدر يصاغ من الفعل الثلاثي على وزن (فِعلة)(بكسر الفاء) ليدل على هيئو و قوع الحدث نحو: وقف وقفة الذاهل و قوله-تعالى-: "فإذا قتلتم فأحسنوا القتلة".

2-2-4-6-8 اسم الآلة:

هو كل اسم مشتق منالفعل للدلالة على الآلة التي يقع بها الفعل، مثل: مِفْتاح،مِقص و قوله-تعلى-:" اللَّهُ نُورُ السَّمَاوَاتِ وَالْأَرْضِ ۚ مَثَلُ نُورِهِ كَمِشْكَاةٍ فِيهَا مِصْبَاحٌ ...".

صياغة اسم الآلة المشتق(اي: له فعل)، وهو معظم اسماء الآلة وهي معلومة في علم الصرف.

جدول 2-4: أمثلة لصيغ إسم الآلة

صياغة اسم الآلة	على وزن اسم الآلة
مِفْعَلْ	مبرد
مِفْعال	منشار
مِفْعلة	مسطرة

اسم الآلة من المضعف الثلاثي: يتم إضافة ميم في أول الكلمة، نحو: قصّ(مقص).

صياغة اسم الآلة الجامد(اي: ليست له أفعال): هذا لا ينضبط تحت قاعدة معينة مثل:رمح،فأس، ولذلك تم تجاوزه في التطبيق و اكتفينا باسم الآلة المشتق بالقياسي.

2-2-4-7 الجذع(Stemmer)

الجذع هو الكلمة من غير السابقة أو اللاحقة التي تعتري الكلمة العربية، مثل: الصلاة (ال التعريف هي سابقة في الكلمة صلاة) و حروف السابقة أو اللاحقة هي كما يلي :

حروف التعريف:

ويمكن أعتبارها خمسة حروف للتعريف في اللغة العربية وهي:(ال،وال،بال،كال،فال) مثل: (الصلاة،والصلاة،بالصلاة،كالصلاة،فالصلاة).

السابقة: السابقة و حروف التعريف كلاهما سابقة، ولكن خصصت حروف التعريف لتعريف الكلمة أما السابقة هي حروف كثيرة مثل:(لل،س،ت،ن،...)[8] مثل :(للصلاة)، (نعبد) فعل مضارع من عبد، (نستعين) فعل مضارع مبني للمعلوم من عون (ن،س،ت) بادئة. الملحق (أ) فيه جانب من السابقةفي اللغة العربية.

اللاحقة: منها الضمير (نا)المتكلمين نحو: (اهدنا) من فعل الأمر(اهد)(مرجع موقع المغربي Corpusالمصحف).ومثل: مسلمات،مسلمون،مسلمان فاللاحقة(ات،ون،ان).في الملحق (أ) فيه جانب من اللاحقة في اللغة العربية.

تم استخدام استخدام ملف يحتوي على كل السابقةفي اللغة العربية والآخر يحتوي على اللاحقة من إعداد الناشطة في اللغة العربية "شيرين خوجة"[8].

8-4-2-2 كيف تمت برمجة و تطبيق كل هذه القواعد على الحاسب الآلي

لنبين كيفية برمجة هذه القواعد لا بد من أن تأخذ بعض الترتيبات التي وضعناها في عين الإعتبار وهي:

المعالجة الأولية للكلمة بما فيها حذف التشكيل سواء أكان من الكلمة أو من الوزن لها، مثال: (فُعال و فَعَال)من صيغ الصفة المشبهة، يعتبران وزناً واحداً وهو(فعال) بغض النظر عن التشكيل. وكذا، (فَعِلٌ و فُعْلٌ) من صيغ المبالغة، يعتبران وزناً واحداً وهو(فعل)، وهكذا...

بما أنه يتم حذف التشكيل في البدء، يقود هذا الى عدم معرفتنا إن كانت الكلمة مضعفة أم لا، مثال: مدّ، هذه الكلمة مكونة من ثلاثة أحرف وهي(الميم،الدال،الدال)، ولكن عند حذف التشكيل تصبح من حرفين هما(الميم،الدال) هذا يؤثر في تطبيق بعض القواعد التي تعمل على الفعل الثلاثي . كيف تمت معالجة هذا الإستثناء، وبأي جانب تم الإهتمام، هل بالثلاثي المضعف، أم الرباعي المضعف هذا ما سوف يأتي شرحه في الباب التالي.

من تطبيق هذه الأوزان على الكلمة، قد تنتج بعض الكلمات التي تحتاج بأن تدخل على قواعد الإعلال والإبدل مثل: اسم الفاعل من (أكل) يصاغ على وزن فاعل (أي: زيادة ألف بعد الحرف الأول) فتصبح(أاكل) هذا واحد من مخرجات هذا البحث، ولكن في الحقيقة ليست في اللغة العربية كلمة كهذه(أاكل) هذه تحتاج الي عمليات الإعلال والإبدال لتصبح(آكل) الكلمة العربية الصحيحة. سوف تجد هذا في نتائج تطبيق النظام، كما سوف تجد عمليات الإعلال و الإبدال تركت كتوصيات.

الفصل الثاني

الدراسات السابقة

2 - 3 الدراسات السابقة

أ. الدراسة الأولى[9]

قامت الدراسة بتوضيح إسترجاع جموع التكسير وقد ركزت دراسته على ثلاث أوزان وهي، فياعيل،تفاعيل وقد تمكن النظام بتعيين هذه الأنماط ومن ثم التعرف على الكلمات التي على هذه الأوزان.وكانت من عيوب هذا النظام :

عدم إسترجاعه للمفرد .

وقام أيضاً [1]بإستخدام نفس النموذج لإسترجاع هذه الأوزان في القرآن الكريم ولكن بالرغم من ذلك لم يستطع النظام إلا إسترجاع وزن واحد من سورتي الفاتحة والبقرة وهو وزن تصاريف.

عدم إستخدام stop word كثيرة مما زاد من الأخطاء في البرنامج.

ب. الدراسة الثانية[10]

و هو نظام للرد الآلي على الأسئلة المكتوبة بلغة طبيعية واهتم النظام بالمعالجة الآلية للغة العربية، حيث استخدم القران الكريم كقاعدة معرفة استخلصت منه بعض القوانين لعالجة الإستعلام أو السؤال التي تساعد على إستخلاص الإجابة الصحيحة المطابقة للسؤال والتي هي الأية أو عدد قليل من الآيات المرتبة والمختارة من مجموعة الآيات المسترجعة ، بحيث يمكن تعميم هذه القواعد على اللغة العربية . تم استخدام برنامج بلغة جافا للحصول على الإجابات من قاعدة بيانات مايكروسوفت أكسس والتي تحتوي على قاعدة معرفة نظام الرد على السؤال . والذي جاءت به هذه الدراسة أنها وجدت أن النظام يعطي نتائج أفضل من حيث زيادة الإجابات المتطابقة بعد تحسين النظام بواسطة إزالة الكلمات المتكررة المستبعدة (stop words) و التشكيل (diacritics)من السؤال.

ويمكن تلخيص عدد من الملاحظات عن هذه الدراسة على النحو التالي:

أن هذا النظام (الرد الآلي على الأسئلة) تم تطبيقه لفصلين فقط من القرآن الكريم :سورة الفاتحة و سورة البقرة ، ولكن يمكن القول أن هذه القواعد يمكن تعميمها على كل القرآن الكريم و أخذها أساس لنظم الرد الآلي على الأسئلة.

يتم إدخال السؤال باللغة الطبيعية واللغة العربية هي المستخدمة. ثم يتم تحليل السؤال عن طريق تقطيع السؤال الى كلمات يتم تصفيته لإيجاد الكلمات المفتاحية للبحث وإزالة الكلمات غير الضرورية و التشكيل .ثم مقارنة الكلمات المتقطعة مع قاعدة البيانات التي تحتوي على القرآن الكريم ، كل كلمة مع كل الكلمات الموجوة في كل سجل (record) الذي يحتوي على الآية .

الفوائد:

أثبتت هذه الدراسة أن إزالة الكلمات المتكررة المستبعدة (stop words) و التشكيل (diacritics)من السؤال يحسن من عملية البحث و يعطي نتائج أفضل ، وبذلك تكون هذه مسلمية لنا تم اثباتها .

أوجه القصور:

كقواعد مستخدمة:

هنالك قاعدة في هذه الدراسة تضمن استرجاع اجابات صحيحة في حالة واحدة .

المعالجة تتم على النص الخام والذي هو القرآن الكريم بينما القواعد ليست شاملة كل الإحتمالات التي تعالج النص القرآني مثال:-

كنظام :

ليست هنالك واجهات تفاعلية للمستخدم تمكنه التوصل الى الإجابة التي يريدها من خلالها.

جدول .5-Error! No text of specified style in document: مقارنة الدراسات السابقة

الدراسات السابقة	المقارنة
الدراسة[أ]	1-ركزت فقط على أوزان فياعيل وتفاعيل وتم تطبيق هذه الأوزان في النظام كما أن الدراسة لم تستطيع استرجاع المفردة كإجابة. 2- تم إستخدام الدراسة في القران الكريم ولم تتمكن الدراسة إلا من استرجاع وزن واحد وهو تصاريف. 3- لم تستخدم الدراسة معالجة الكلمات التوقيفية (words Stop).
الدراسة[ب]	1-تم إستخدام الدراسة في القران الكريم في سورتي الفاتحة والبقرة. 2- إستخدمت الدراسة مفهوم معالجة الكلمات التوقيفية(Stop words) . 3-إستخدمت الدراسة إزالة التشكيل من السؤال. 4- الإجابة هي عبارة عن أية تتم مقارنة السؤال مع قاعدة البيانات واسترجاع الإجابة.

2 - 4 الخلاصة

عبر هذا الباب تم تناول معلومات عن نظم إسترجاع المعلومات كمفاهيم أساسية و كلمات مفتاحية و، كما تم تناول دراسة اللغة العربية من الناحية اللغوية والدراسات السابقة في معالجة اللغة العربية آلياً . و في الباب التالي يتم تناول المنهجية المتبعة في هذا البحث لإسترجاع المعلومات ، و كذلك سوف يتم تناول اللغة العربية من الناحية التقنية حيث يتم دراسة قواعد الصرف من ناحية التصميم و البرمجة.

الباب الثالث

منهجية البحث

3. 1. مقدمة

يعتبر هذا الباب متن البحث حيث تتم فيه مناقشة المنهجية المقترحة كما فيه جانب من التفصيل لدراسة تصميم و برمجة قواعد الصرف للغة العربية، و ما هي الأدوات و التقنيات المستخدمة والمتبعة ، اذ انه احتوى على ثلاثة مقاطع: الأدوات و التقنيات المستخدمة، إطار البحث، تحليل الأنماط.

3. 2. الأدوات والتقنيات

3. 2. 1. لغة جافا (Java JDK_8)

الجافا هي لغة برمجة عالية المستوى، قام بتطويرها جيمس غوسلينغ من شركة (Sun Microsystems) أثناء محاولته تطوير لغة البرمجة C++ عام 1991 و كانت تسمى (OAK)، و لكن في عام 1995 قامت شركة (Sun Microsystems) بتعديلها و أطلقت عليها اسم جافا (Java)، و هي لغة تعتمد على الكائنات (Object). وتم استخدام Java JDK_8 لتوافقه مع الأداة NetBeanse Version-8 .

هنالك عدة إصدارات للجافا منها:

- (Java Standard Edition): الإصدار القياسيّ يستخدم لبرمجة برامج سطح المكتب.
- (Java Enterprise Edition): يستخدم لبرمجة الخوادم و الويب.
- (Java Micro Edition): يختص ببرمجة الهواتف المتنقلة والأجهزة المحدودة وخاصّة الألعاب.

خصائص و مميزات الجافا:

- تعّد لغة الجافا لغة سهلة لإعداد التطبيقات المختلفة، بسبب احتوائها على العديد من الميّزات التي تسهل على المبرمج عمل، فهي توفر اتصالاً مباشراً مع قواعد البيانات، والوراثة المتعددة، وغيرها.
- يمكن برمجة تطبيقات قوية وخالية من الأخطاء باستخدام الجافا، ووجود ميزة تعديل الأخطاء أثناء البرمجة.
- آمنة، فهي تعمل خلال بيئتها الافتراضية، وتعتمد إدارة المخاطر، وتعتمد التشفير.
- تعمل الجافا على عدة أنظمة تشغيل، كالويندوز والماكنتوش، ويونكس بجميع إصداراتها.
- يمكن من خلالها كتابة برامج تنفذ أكثر من عمل (task) في آن واحد بفضل ميزة (Multi-threaded).

مصطلحات هامة في لغة الجافا:

- (Object): كائن له حالة (states) ،وسلوك (behaviors) ويتمّ إنشاؤه من فئة(class) .
- (Class): يمكن تعريفها بأنّها قالب تصف السلوكيات والحالة للكائن الذي ينتمي له.
- (Methods): هي بالأساس سلوكيات، ويحتوي (class) على مجموعة (method) يتمّ من خلالها معالجة البيانات .
- (Instance Variables): المتغيرات.
- (Inheritance): التوارث، من خلال التوارث في الجافا يمكنك بناء فئة (class) من (class) آخر للإستفادة من خصائصه و مكوناته.
- (Interfaces): الواجهات، وتعرف في لغة الجافا على أنّها عقد يبيّن كيفيّة التواصل بين الكائنات بعضها ببعض، وهي هامّة جداً في ما يتعلق بالتوارث.

- (Constructor): من أهم المصطلحات التي تتعلق بـ (class) فعند انشائه يتمّ استدعاء(constructor) خاصاً به، ويمكن أن يكون أكثر من واحد، ويجب أن يكون اسم (constructor) مطابقا لاسم (class) التابع له.

بيئات عمل لغة الجافا:

- (Java Virtual Machine): بيئة جافا الإفتراضية هي وسيط بين تطبيقات جافا و نظام التشغيل و هي واحدة من مميزات جافا أنها تشتغل على جميع نظم التشغيل.
- (Java Runtime Environment): البيئة التشغيلية لتطبيقات الجافا و هو عبارة عن بيئة جافا الإفتراضية(JVM) زائداً مكتبات الجافا (Library Class). وهي لمستخدمي تطبيقات الجافا.
- (Java Development Kit): عدة جافا التطويرية من خلالها يتم عمل تطبيقات بلغة جافا تحتوي على أدوات تطوير و تصحيح الأخطاء بالإضافة الي البيئة التشغيلية لجافا(JRE). وهي لمطوري تطبيقات الجافا.

رغم أن هنالك مجموعة من اللغات التي تدعم المعالجة الآلية للنصوص - ومن بينها لغة بايثون التي تحتوي على مجموعة من المكتبات (NLP) التي تساعد في المعالجة ولكن كل المكتبات هي للغة الإنجليزية و في الحقيقة نحن لا نحتاج في بحثنا هذا الي اي من هذه المكتبات الجاهزة وبالتالي تم اسبعاد لغة بايثون، أما لغة جافا فهي فعالة جدا في معالجة الكلمات وتم إستخدام String And StringBuilder، Array of String لمعالجة النصوص.

3. 2. 2. محررالنصوص (NetBeanse_8 IDE)

واحد من بيئات التطوير المتكاملة للغات البرمجة هو من انتاج شركة (Sun Micro Systems) ، كانت بداية البرنامج في عام 1997 ، توجد اصدارات كثيره لمحرر النصوص هذا ولكن تم استخدم (NetBeans 8.1) في هذا االمشروع ، (Netbeans) متاح في 23 لغة برمجه ،ويوفر هذا المحرر وظائف مهمة تساعد المبرمجين على تطوير المشاريع الكبيرة. و هو واحد من الأدوات التي تدعم اللغة العربية.

3. 2. 3. المكتبة (ucanaccess)

هذه المكتبة تدعم القراءة من قاعدة البيانات أكسس و المكتوبة بلغات (UTF_8) ومن بينها اللغة العربية. وبالتالي تم إستخدام (jdbc:ucanaccess) Driver بدلا من استخدام (Jdbc:Odbc) الذي لا يدعم لغات (UTF_8) ومن بينها اللغة العربية. و هذه المكتبة لا تعمل إلا في النسخة (Java JDK_8).

3. 2. 4. مصمم الأشكال (Diagram designer)

يوفر للمستخدمين فرصة لتصميم الرسوم البيانية مثل التصميم بإستخدام (flow chart)،كما أن diagram designer يحتوي على واجهة سهلة الاستخدام تمكن المستخدم من التعامل معه بسهولة و دقة. و به تم رسم إطار البحث المقترح و توضيح آلية تحليل الأنماط.

3. 3 . إطار البحث

الشكل رقم3-1 يوضح إطار البحث المقترح و فيه العمليات الأساسية للنظام المقترح و جميعها متتالية و هي:

3. 3. 1. المعالجة الأولية (pre-process)

ذكر سابقاً في الباب الثاني الفصل الأول من هذا البحث عمليات المعالجة الأولية و كيف أنها تقود إلى فعالية عملية البحث (searching)، تمت المعالجة الأولية في هذا البحث على إستعلام المستخدم والآيات القرآنية من سورتي الفاتحة والبقرة التي سوف يتم البحث فيها حيث تم حذف التشكيل و تم إستبدال الحروف(أ،إ،آ) بالحرف(ا) و كذلك تم حذف الكلمات التوقيفية.

3. 3. 2. توسعة الإستعلام (query expansion)

بعد المعالجة الأولية يتم توسعة الإستعلام وهي إضافة كلمات بحث لإستعلام المستخدم ، هذه تزيد من الملفات المسترجعة وذلك عن طريق إضافة بعض الأنماط للكلمة الواحدة .

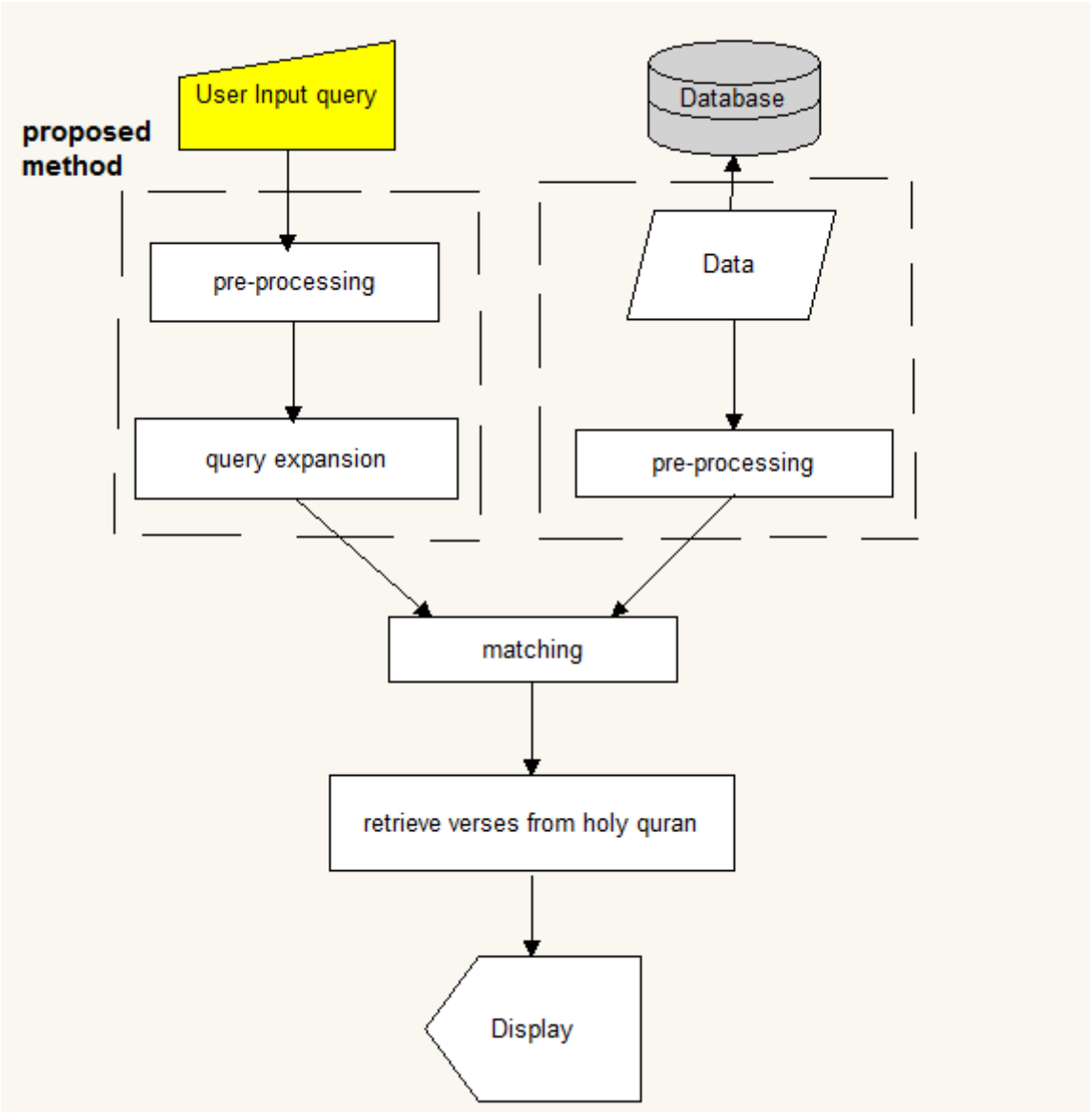
إذا النظام لم يعمل على توسعة الإستعلام هذا سوف يؤدي إلى فقد بعض الملفات أو البيانات التى تحمل الكلمات المفتاحية للبحث.

3. 3. 3. المطابقة (matching)

المطابقة - عملية البحث - تتم عن طريق مقارنة كل كلمات الاستعلام قبل و بعد التوسعة مع المعلومات المخزنة التي يجب استرجاعها (الآيات القرآنية).

3. 3. 4. إسترجاع المعلومات (retrieve verses from holy Quran)

بعد المطابقة يتم استرجاع الآية التي على الأقل تطابقها كلمة واحدة من الاستعلام و يظهر ذلك في جدول النتائج لإختبار النظام عند الفصل الرابع من هذا البحث.



الشكل 3-1: إطار البحث المقترح

3. 4 . تحليل الأنماط

سوف يتم شرح بعض المعالجات التي قمنا بها في المعالجة الطبيعية للغة العربية والمخططات التوضيحية لها .

3. 4. 1. إسم الآلة

هو كل اسم مشتق من الفعل للدلالة علي الآلة التي يقع بها الفعل، مثل مِفْتاح، مِبْرَد، مِثقَلَة، مِشْرَط، مِقَصّ، ميزان، مِرْوَحة، سَماعة، ثلاجة، شَوْكة، سَيْف.

- **إسم الآلة من الفعل الثلاثي**

هناك ثلاثة أوزان قياسية أجازها مجمع اللغة العربية وهي مفعل مفعال ومفعلة من الفعل الثلاثي:

وزن مفعل

فيه تتم الصياغة وذلك بإضافة الحرف ("م") في الموقع رقم (0).

وزن مفعال

وفيه تتم الصياغة بإضافة الحرف ("م") في الموقع رقم (0) والحرف (" ا ") في الموقع رقم (3).

وزن مفعلة

وفيه تتم الصياغة بإضافة الحرف ("م") في الموقع رقم (0) والحرف (" ة ") في الموقع رقم (4) . إذا صيغ إسم الآلة من الفعل المعتل الآخر أيضاً يصاغ على وزن مفعلة .

وأيضاً قد أقر مجمع اللغة العربية أوزان غير الأوزان الثلاثة القياسية المشهورة من الفعل الثلاثي وهي فعالة ، فعال ، فاعلة ،فاعول:

وزن فعالة

فيه تتم الصياغة وذلك بإضافة الحرف (" ا ") في الموقع رقم (2) والحرف (" ة ") في الموقع رقم (4).

وزن فعال

فيه تتم الصياغة وذلك بإضافة الحرف (" ا ") في الموقع رقم (2).

وزن فاعلة

فيه تتم الصياغة وذلك بإضافة الحرف (" ا ") في الموقع رقم (1) والحرف (" ة ") في الموقع رقم (4).

وزن فاعول

فيه تتم الصياغة وذلك بإضافة الحرف (" ا ") في الموقع رقم (1) والحرف (" و ") في الموقع رقم (3).

- **إسم** الآلة **من المضعف الثلاثي**

إذا صيغ إسم الآلة من الفعل المضعف الثلاثي الذي عينه ولامه من جنس واحد بقي الحرفان مدغمين كما كانا في فعلهما، وله وزن واحد هو مفعل ، فاسم الآلة من قص هو مقص.

وزن مفعل

فيه تتم الصياغة وذلك بإضافة الحرف (" م ") في الموقع رقم (0) . الشكل رقم 3-2 يوضح معالجة اسم الآلة آليا.

3. 4. 2. إسم الفاعل

هو إسم مشتق من الفعل للدلالة علي الحدث وفاعله، مثل محمد كاتب رسالة، (فكاتب) دلت على الكتابة ومن قام بها، نحو قوله تعالى : (" إني جاعل في الأرض خليفة ").

- **إسم الفاعل من الفعل الثلاثي**

الصحيح

تتم صياغة إسم الفاعل من الفعل الثلاثي الصحيح وذلك بإضافة الحرف (" ا ") في الموقع رقم (1) ليصبح على وزن فاعل .

المضعف الثلاثي

تتم صياغة إسم الفاعل من الفعل الثلاثي المضعف على وزن فاعل وذلك بإضافة الحرف (" ا ") في الموقع رقم (1) مع إبقاء عين الكلمة ولامها مشددة نحو: (مدً، مادً)(وعدً، وعاد).

الثلاثي الأجوف

يصاغ إسم الفاعل من الفعل الثلاثي الأجوف (الذي وسطة حرف علة) على وزن فاعل بعد قلب عين إسم الفاعل همزة سواء أكانت واوية ام يائية وذلك بإضافة الحرف (" ئ ") في الموقع رقم (2).

الثلاثي الناقص

عند صياغة إسم الفاعل من الفعل الثلاثي الناقص (الذي آخره حرف علة) يحذف آخر إسم الفاعل وهو لامه الميزان نحو : (تلا، تال) وقوله تعالى: (" كل من عليها فان ").

الثلاثي الناقص ومعرف بأل

قد يحذف حرف العلة من إسم الفاعل الناقص (وهو مقترن بأل) وهو في حالة الجر أو الرفع أو النصب إكتفاءً بالحركة قبلة

تتم صياغة إسم الفاعل من الفعل الثلاثي الناقص المعرف بأل على وزن فاعل بإضافة الحرف (" ا ") في الموقع (0) والحرف (" ل ") في الموقع (1) والحرف (" ا ") في الموقع (3) والحرف (" ي ") في الموقع (5).

كما تتم صياغة إسم الفاعل من الفعل الثلاثي الناقص وهو نكره في محل نصب بإضافة الحرف (" ا ") في الموقع (1) والحرف (" ا ") في الموقع (4).

كما تتم صياغة إسم الفاعل من الفعل الثلاثي الناقص وهو نكره في محل نصب بإضافة الحرف (" ا ") في الموقع (1) وحذف الحرف في الموقع(3).

- **إسم الفاعل من الفعل الغير ثلاثي**

فيه تتم الصياغة بإضافة الحرف (" م ") في الموقع (0). الشكل رقم 3-3 يوضح طريقة معالجة اسم الفاعل آليا.

3. 4. 3. إسما الزمان والمكان

اسم الزمان : اسم مأخوذ من حروف الفعل ليدل على زمان حدوث الفعل .

اسم المكان : اسم مأخوذ من حروف الفعل ليدل على مكان حدوث الفعل .

إسم الزمان والمكان من الفعل الثلاثي

يُصاغ اسم الزمان والمكان على وزن مَفْعَل بفتح العين وذالك بإضافة ميما مضمومه في بداية الفعل الثلاثي (من الفعل الثلاثي مضموم العين في المضارع نحو : مَدْخَل، مَخْرَج ، مَقْعَد، مكْتب، مَرْصَد، منظَر، مأخَذ).

إسم الزمان والمكان من الفعل الغير الثلاثي

يصاغ اسم الزمان والمكان من الفعل الغير ثلاثي علي وزن مفعل وذاللك بحذف الحرف الاول واضافة ميما في بدايه الكلمة. الشكل رقم 3-4 يوضح معالجة اسما الزمان والمكان آليا.

3. 4. 4. صيغة المبالغه

تصاغ صيغة المبالغة من الفعل الثلاثي من مجموعة من الأوزان :

مفعال

تتم الصياغة من الفعل الثلاثي وذالك بإضافة ("م")في الموقع رقم (0) وإضافة ("ا") في الموقع رقم (3).

فعال

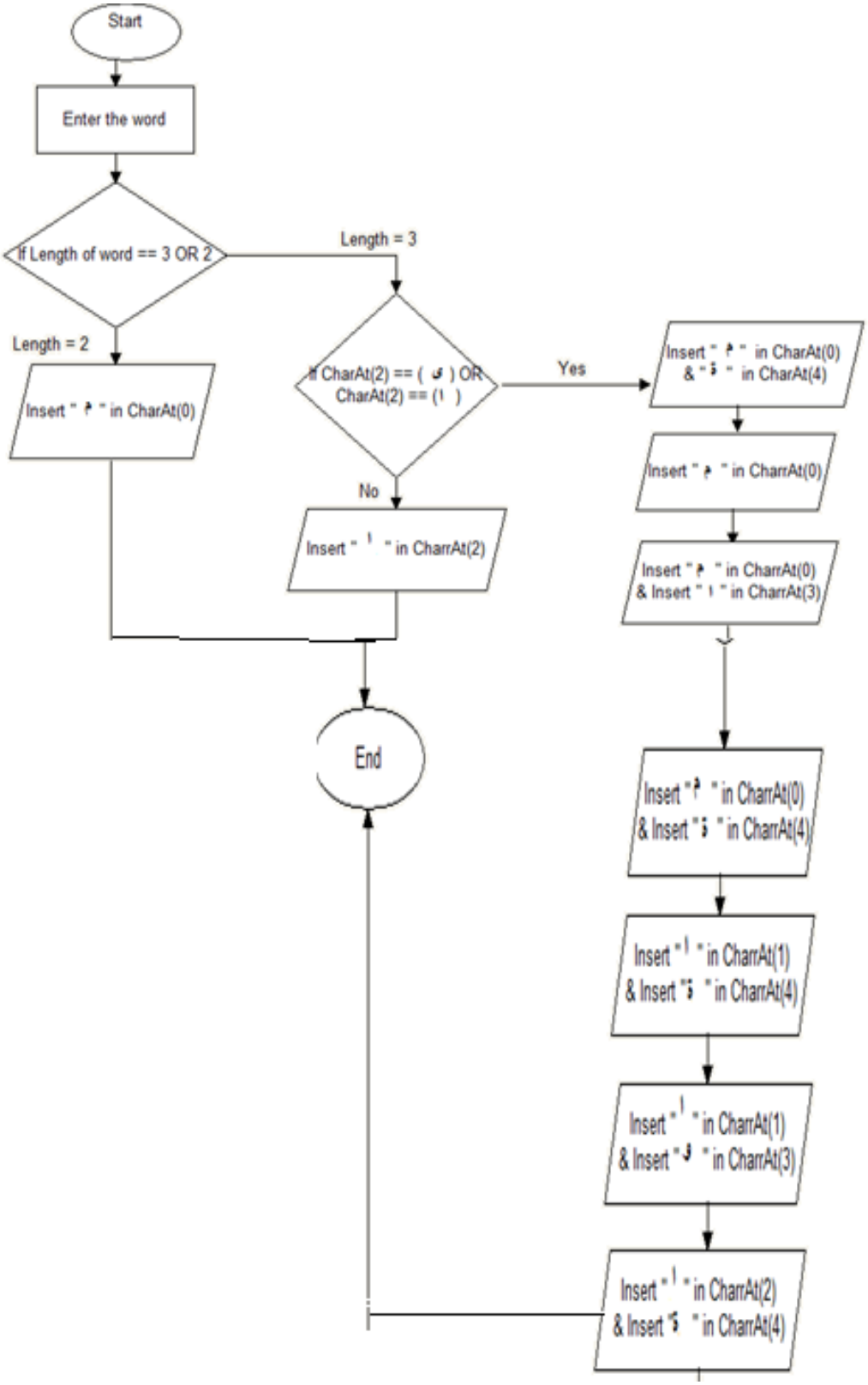
تتم الصياغة من الفعل الثلاثي وذالك بإضافة ("ا") في الموقع رقم(2).

فعيل

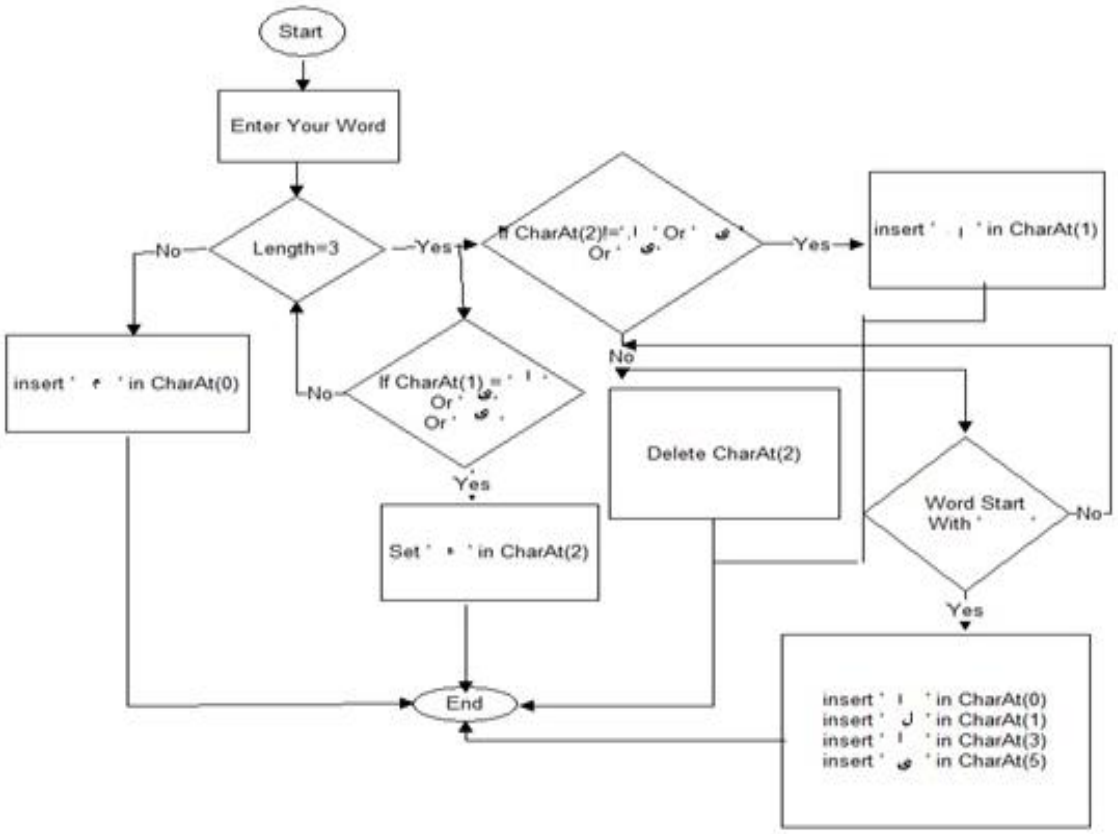
تتم الصياغة من الفعل الثلاثي وذالك بإضافة ("ي") في الموقع رقم(2).

فعول

تتم الصياغة من الفعل الثلاثي وذاللك بإضافة ("و") في الموقع رقم(2). الشكل رقم 3-5 يوضح معالجة صيغ المبالغة آليا.



الشكل3-2: طريقة معالجة اسم الآلة



الشكل3-3: طريقة معالجة اسم الفاعل آليا

3. 4. 5. الصفة المشبة

هي كل اسم صيغ من الفعل اللازم للدلالة على صفة ثابتة ملازمة لصاحبها نحو(صويل ،قصير، جميل ،جبان).

للصفة المشبهة مجموعه من الاوزان(فعلة – افعل- فعلاء- فعلى- فعلان- فعيل- فعيلة – فعال- فعالة- فاعل - فاعلة - فيعل -فعول -فعلون):

فعلة

تتم الصياغة وذالك بإضافة (" ة") في أخر الكلمة.

افعل

تتم الصياغة وذالك بإضافة (" ا") في الموقع رقم (1).

فعلاء

تتم الصياغة وذالك بإضافة (" ا") في الموقع رقم (3) وإضافة ("ء ") في الموقع رقم (4).

فعلان

تتم الصياغة وذالك بإضافة (" ا") في الموقع رقم (3) وإضافة ("ن ") في الموقع رقم (4).

فعيل

تتم الصياغة وذاللك باضافة ("ي ")في الموقع رقم (2).

فعيلة

تتم الصياغة وذالك باضافة (" ي") في الموقع رقم (2) وإضافة (" ة ") في الموقع رقم (4).

فعال

تتم الصياغة وذالك باضافة (" ا") في الموقع رقم (2).

فعالة

تتم الصياغه وذالك بإضافة (" ا") في الموقع(2) وإضافة ("ة ") في المموقع رقم (4).

فاعل

تتم الصياغة وذالك بإضافة (" ا") في الموقع رقم (1).

فاعلة

تتم الصياغة وذالك بإضافة (" ا") في الموقع رقم (1) وإضافة (" ة") في الموقع رقم (4).

فعيل

تتم الصياغة وذالك بإضافة ("ي") في الموقع رقم (2).

فعول

تتم الصياغة وذالك باضافة (" و") في الموقع رقم (2).

فعلون

تتم الصياغة وذالك بإضافة (" و") في الموقع رقم(3) وإضافة ("ن ")في الموقع رقم (4). الشكل رقم 3-6 يوضح معالجة الصفة المشبهة آليا.

3. 4. 6. الكلمات التوقيفية(Stop Words)

الكلمات التوقيفية هي مجموعه من الكلمات الشائعة التي تؤثر علي عملية البحث واسترجاع المعلومات ، وهذه الكلمات كثيرة وجميع اللغات لها مثل هذه الكلمات، يتم حذف هذه الكلمات لتحسين عملية البحث واسترجاع المعلومات والحصول على إجابات أكثر دقة، الملحق(أ) فيه جانب من الكلمات الشائعة في اللغة العربية. الشكل رقم 3-7 يوضح طريقة حذف الكلمات التوقيفية.

3. 4. 7. حذف التشكيل

في اللغة العربية عامة و القرآن الكريم خاصة توجد علامات التشكيل التي تؤثر على عملية البحث وإسترجاع المعلومات أو الأجابات وهذه العلامات هي(َ، ً، ُ، ٌ ، ِ ، ٍ). فإذا كانت الكلمة المدخلة تحتوي على إحدى هذه العلامات يتم حذفها .

تمت المعالجة بمقارنة السؤال المدخل مع الملف الذي يحتوي على علامات تشكيل "diacritics.txt" فإذا احتوى الإستعلام على واحدة أو أكثر من هذه العلامات يتم حذفها. الشكل رقم 3-8 يوضح معالة حذف التشكيل.

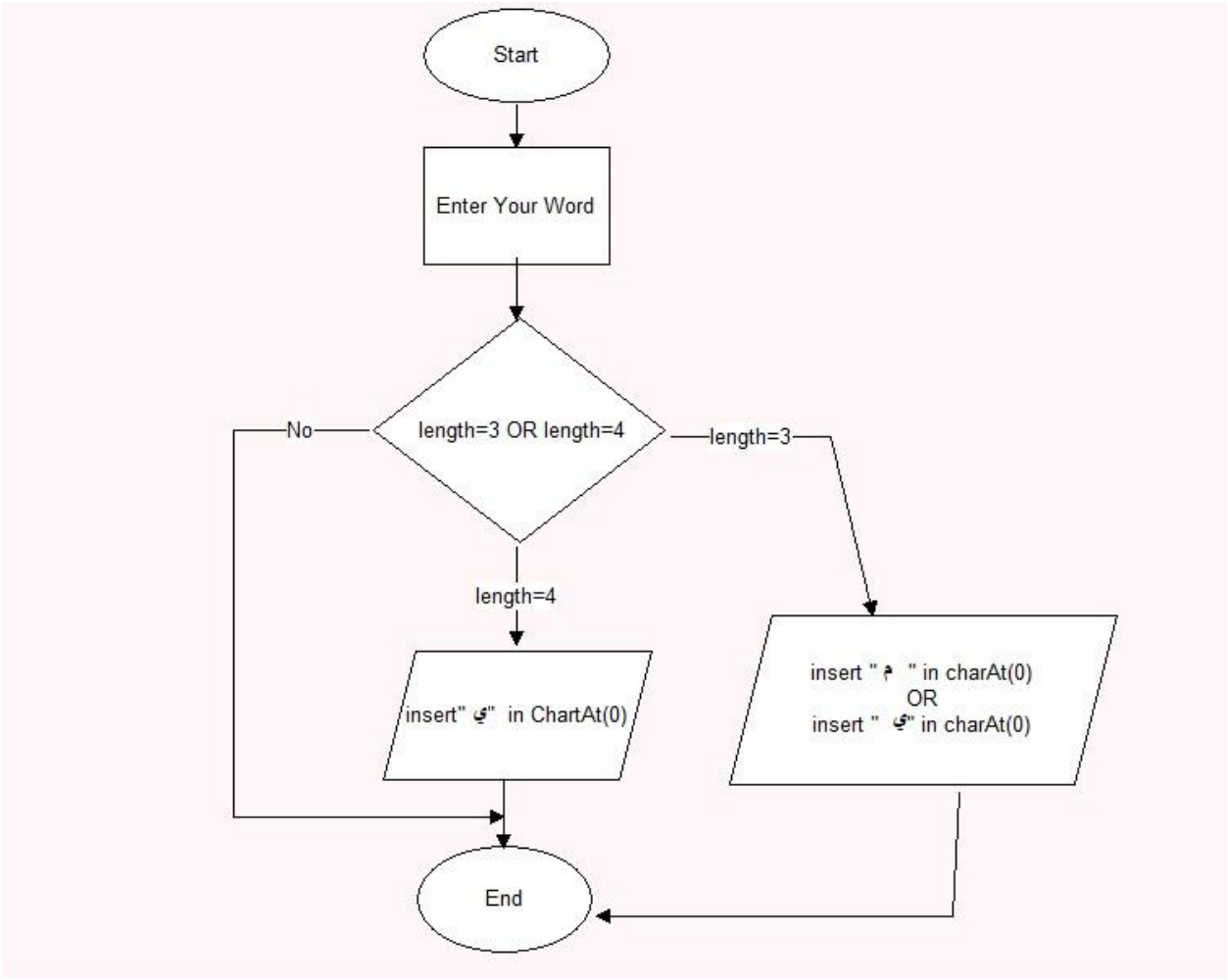
3. 4. 8. معالجة الحروف الخاصة (Special Characters)

توجد في القرآن الكريم مجموعة من الحروف تسمي الخاصة التي تؤثر على عملية البحث وإسترجاع المعلومات أو الأجابات وهذه الحروف هي(أ، إ، آ، ٰ).

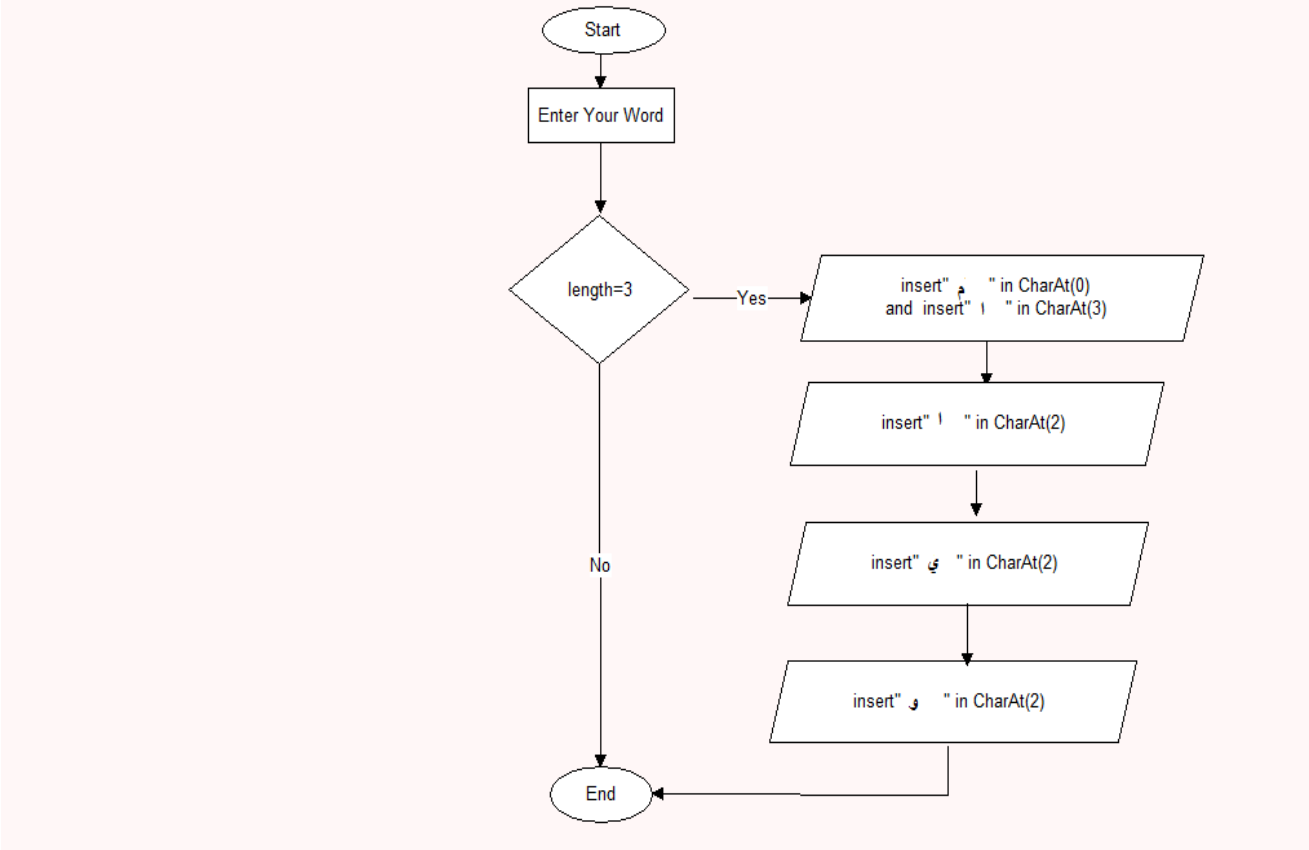
إذا كانت الكلمة المدخلة تحتوي على(أ، إ، آ، ٰ) تمت المعالجة الآلية بإستبدال الحرف (" إ ") والحرف (" آ ") والحرف (" أ ") والحرف (" ٰ") -اي جميعها- بالحرف (" ا ") .الشكل رقم 3-9 يوضح معالجة الحروف الخاصة.

3. 4. 9. التقطيع(Split)

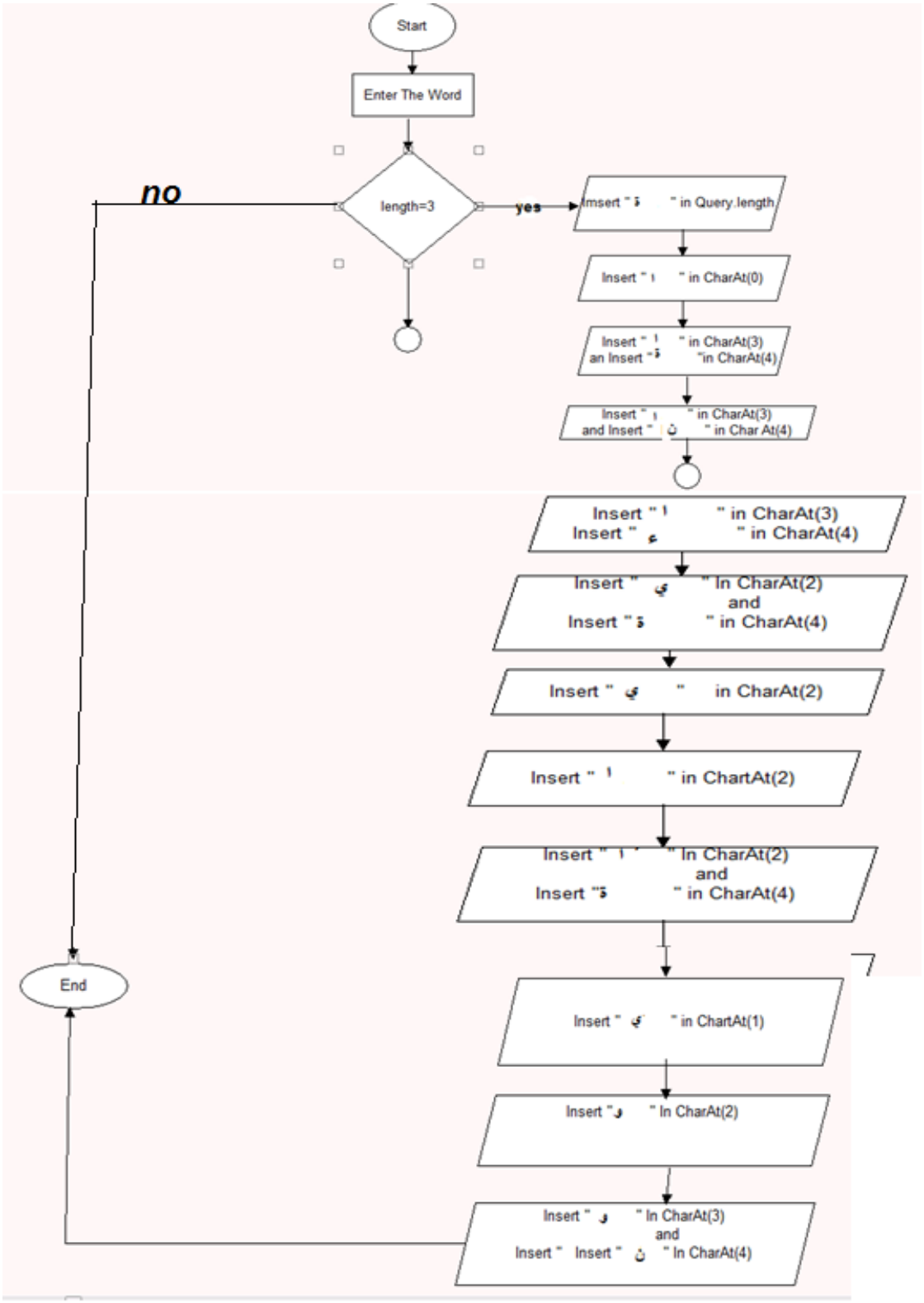
وفيه يتم الفصل بين كل كلمة والأخرى عن طريق المسافات (space) التى بين المكلمات.الشكل رقم 3-10 يوضح عملية التقطيع.



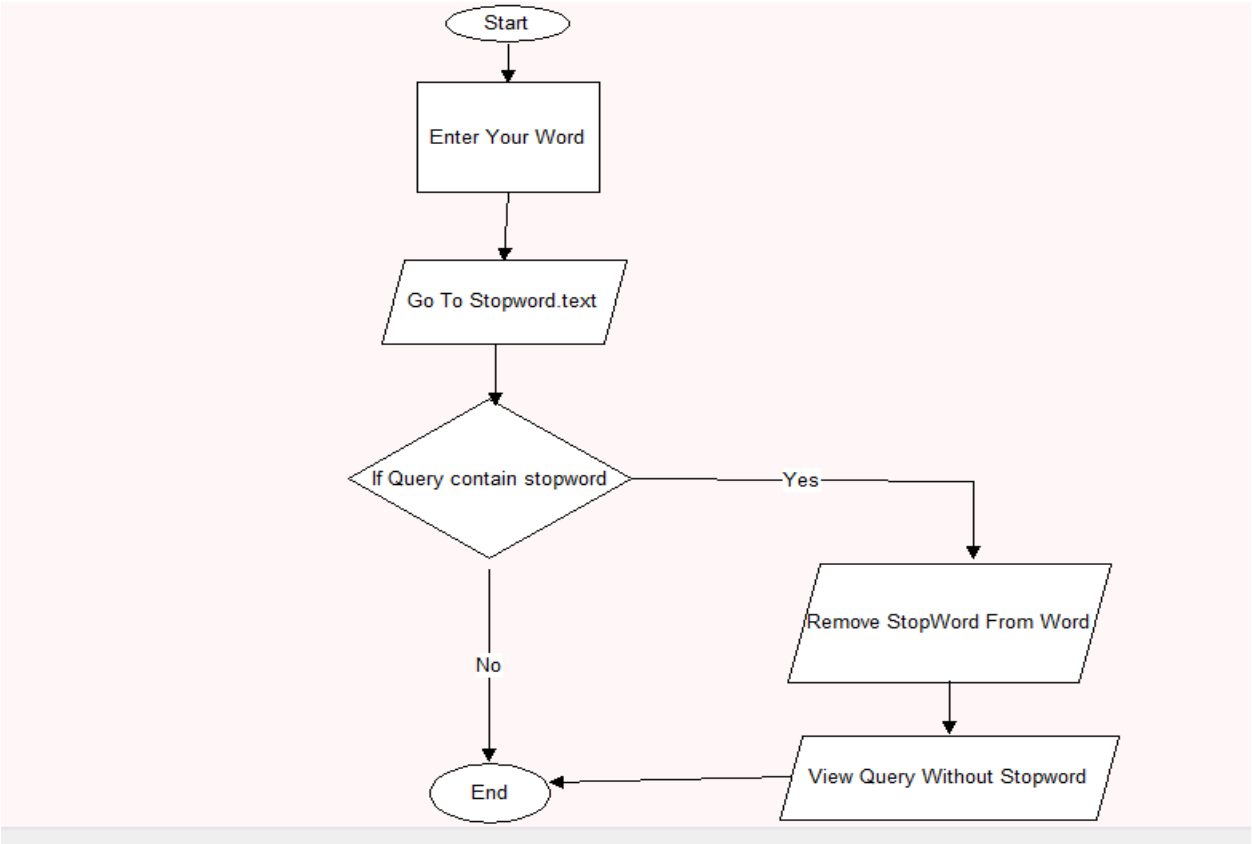
الشكل3-4: طريقة معالجة اسما الزمان والمكان



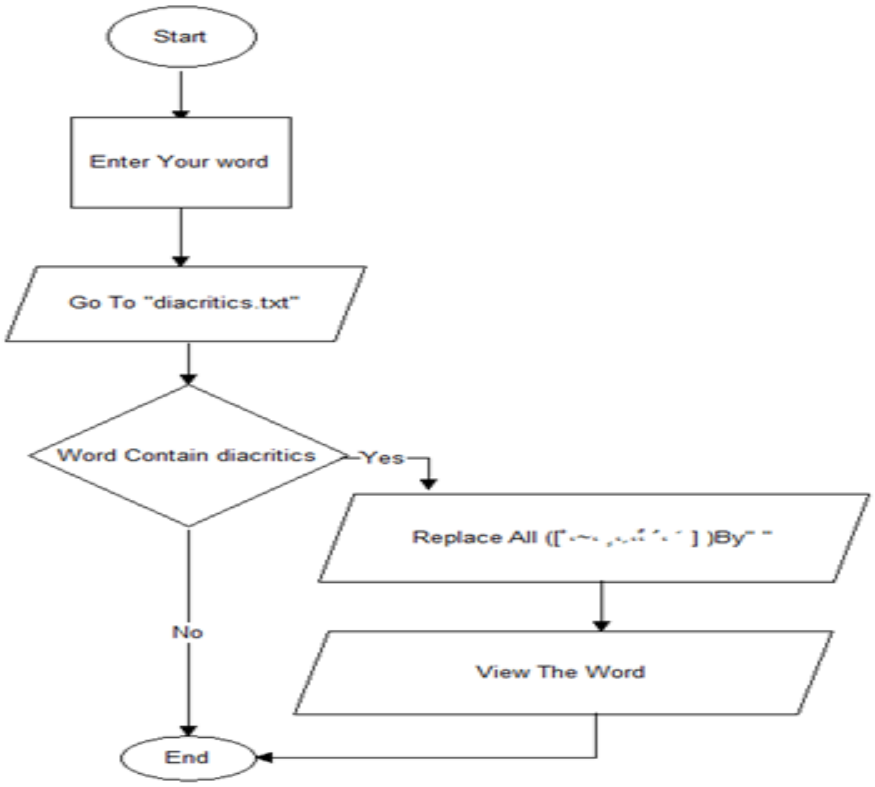
الشكل 3-5: طريقة معالجة صيغ المبالغة آليا



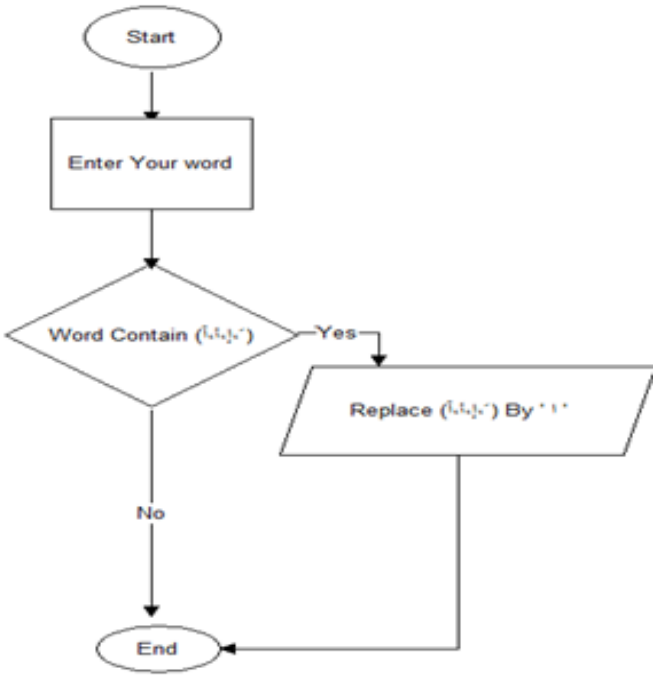
الشكل 3-6: طريقة معالجة صيغ الصفة المشبهة آليا



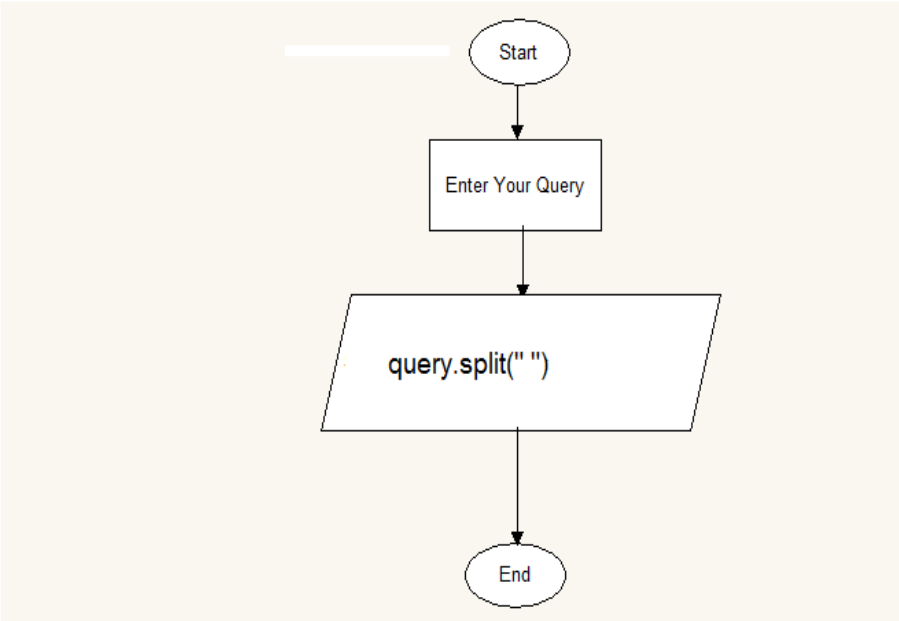
الشكل3-7: طريقة معالجة حذف الكلمات التوقيفية



الشكل 3-8: طريقة معالجة حذف التشكيل



الشكل 3-9: طريقة معالجة استبدال بعض الحروف الخاصة



الشكل 3 -10: طريقة معالجة عملية التقطيع للإستعلام

3. 5 . الخلاصة

في هذا الباب تم دراسة المنهجية المقترحة أو الإطار المقترح بشئ من التفصيل في تطبيق الإطار المقترح وذلك في تحليل الأنماط ودراسة القواعد التي صممت الي ذلك، وما هي الأدوات والتقنيات التي ساعدة على تطبيق و تشغيل النظام. في الباب التالي يتم توضيح تطبيق النظام وذلك بالصور و الواجهات الرئيسية.

الباب الرابع
تطبيق النظام

4 . 1 . المقدمة

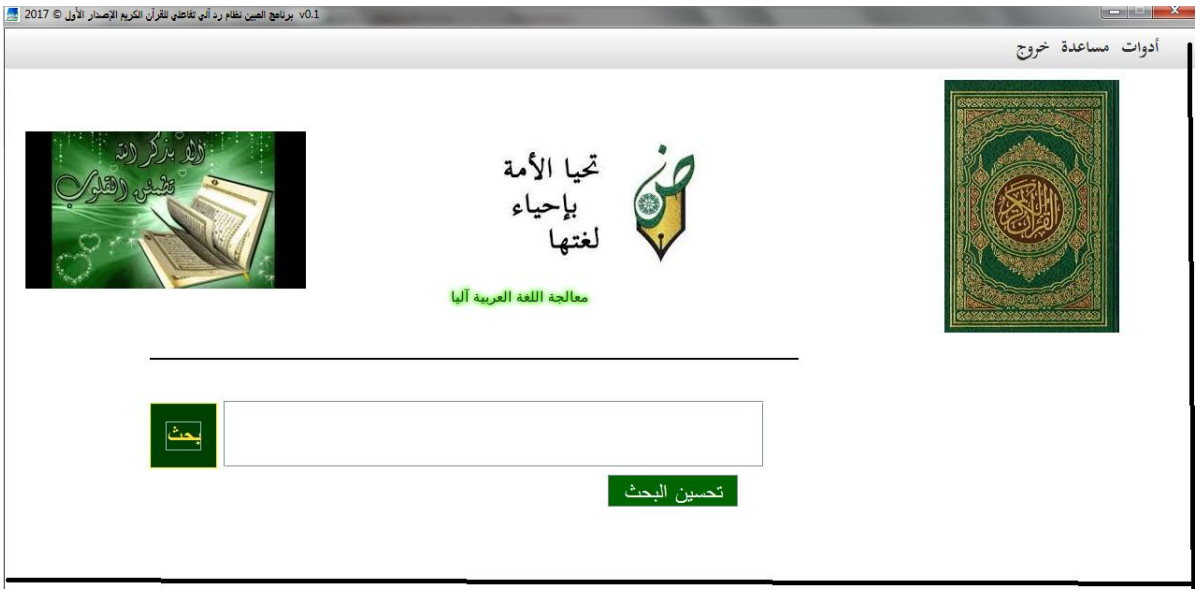
هذا الباب يحتوي على الجانب العملي للبحث، واجهات المستخدم وتفاعل المستخدم مع تلك الواجهات، مع توضيح الواجهات الأساسية للنظام. اسم البرنامج برنامج المبين و هذه هي النسخة الأولى له.

شكل 4-1 :أيقونة البرنامج

هنالك ثلاث واجهة أساسية للنظام وهي: الشاشة الرئيسية، لوحة التحكم، شاشة النتائج.

4 . 2 . الشاشة الرئيسية

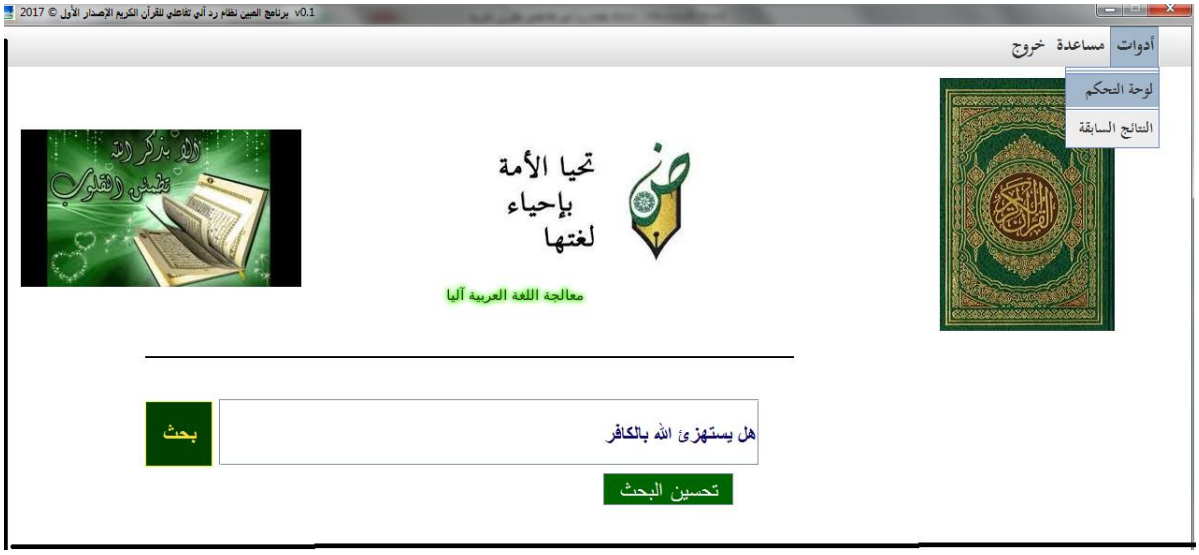
منها ينطلق البرنامج، حيث يتم منها إدخال إستعلام المستخدم وهي الواجهة الأولى للمستخدم من خلالها يمكنه إدخال الإستعلام، والتحرك الى بقية الواجهات، كما تحتوي هذا الواجهة على زر "بحث" للحصول على النتائج. الشكل رقم4-2 يوضح الشاشة الرئيسية



شكل4-2:الشاشة الرئيسية

4 . 3 . لوحة التحكم

هذه تحتوي على تطبيق منهجية البحث المقترحة التطبيق الفعلي للإطار المقترح ، فيها كل القواعد والأنماط التي درسة وصممت لهذا البحث. و من خلالها يتم تطبيق هذه القواعد على إستعلام المستخدم لتوسعته لضمان إسترجاع كل الملفات ذات الصلة بعملية البحث. الشكل رقم4-4 يوضح شاشة لوحة التحكم.



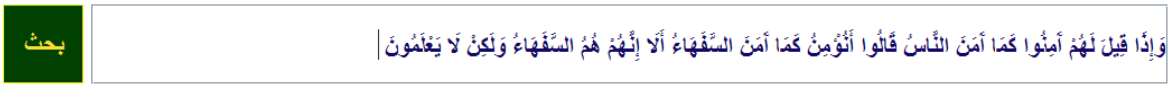
شكل4-3: الذهاب الي لوحة التحكم

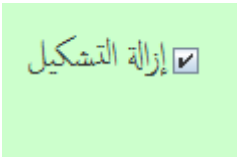
شكل 4-4: لوحة التحكم

4.4. شاشة النتائج

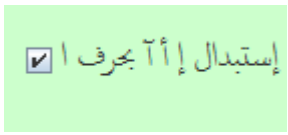
بعد كتابة الأستعلام، يمكن للمستخدم الذهاب الى "لوحة التحكم" لإختيارالمعالجة التي سوف يتم تطبيقها على الإستعلام مع ضمان بقاء إستعلام المستخدم، ثم بعد الضغط على الزر بحث يعمل البرنامج على أخذ إستعلام المستخدم و المعالجة التي أختيرت و الذهاب الي عملية البحث و إعطاء النتائج على شاشة النتائج. الشكل رقم 5:4 يوضح شاشة النتائج

4 . 5 . تنفيذ شاشة لوحة التحكم جزء بجزء

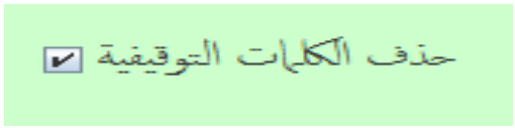




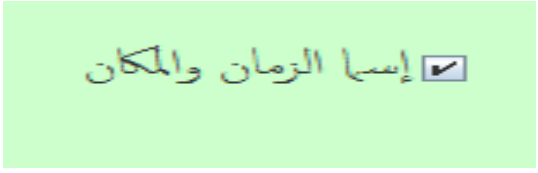

```
run:
وإذا قيل لهم آمنوا كما آمن الناس قالوا أنؤمن كما آمن السفهاء ألا إنهم هم السفهاء ولكن لا يعلمون
BUILD SUCCESSFUL (total time: 0 seconds)
```



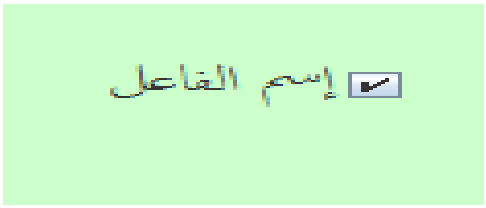

```
run:
وَإِذَا قِيلَ لَهُمْ آمِنُوا كَمَا آمَنَ النَّاسُ قَالُوا أَنُؤْمِنُ كَمَا آمَنَ السُّفَهَاءُ أَلَا إِنَّهُمْ هُمُ السُّفَهَاءُ وَلَكِنْ لَا يَعْلَمُونَ
BUILD SUCCESSFUL (total time: 1 second)
```



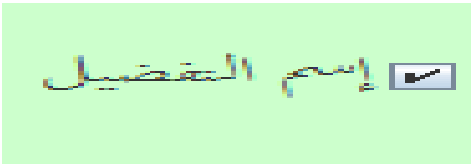

```
run:
فان
BUILD SUCCESSFUL (total time: 1 second)
```



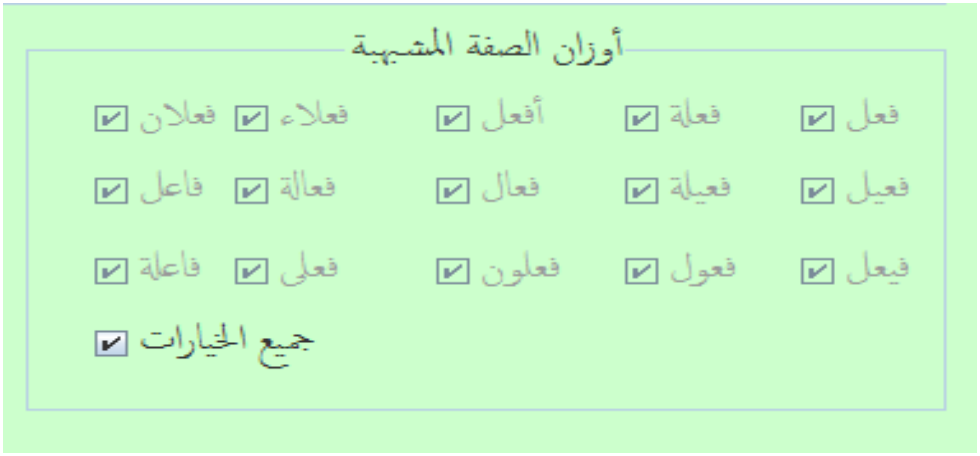
```
run:
مغرب
BUILD SUCCESSFUL (total time: 0 seconds)
```



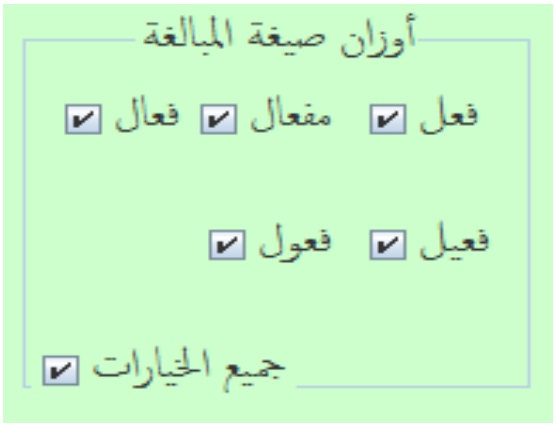

```
run:
راحم راحم
BUILD SUCCESSFUL (total time: 2 seconds)
```



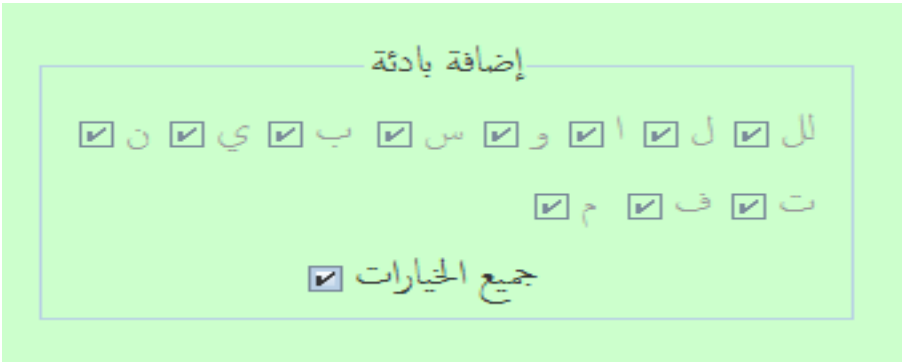

```
run:
أرحم
BUILD SUCCESSFUL (total time: 2 seconds)
```



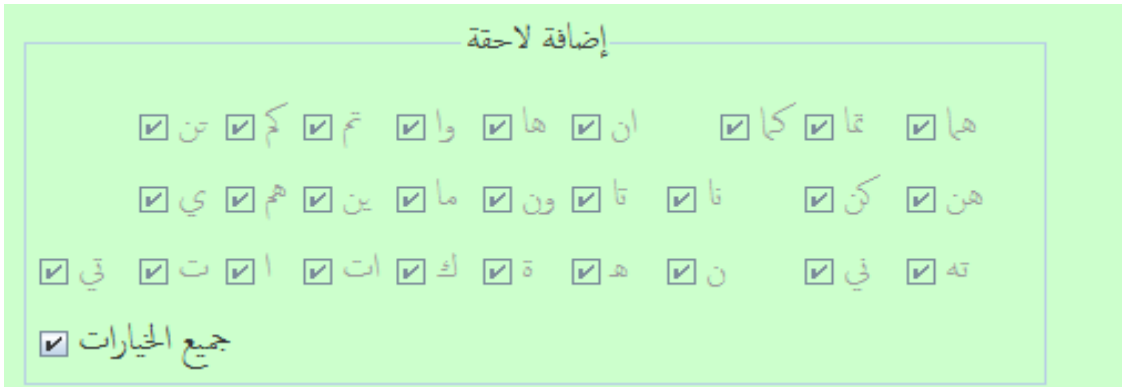

```
run:
 رحم رحمة أرحم رحماء رحمى رحمان رحيم رحيمة رحام رحامة راحم راحمة ريحم رحوم رحمون
BUILD SUCCESSFUL (total time: 2 seconds)
```



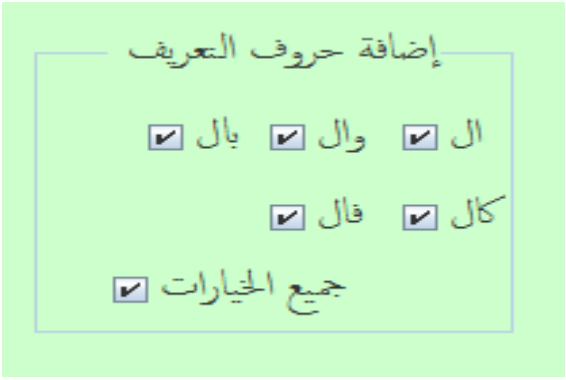

```
run:
   رحم  مرحام   رحام   رحيم   رحوم
BUILD SUCCESSFUL (total time: 1 second)
```



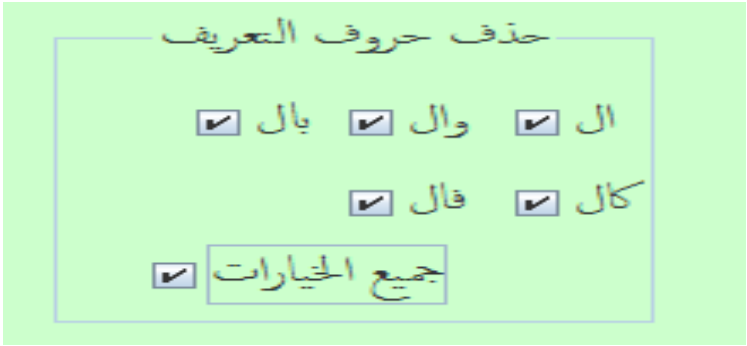

```
run:
  لـرحم  ارحم  ورحم  سرحم  بـرحم  يـرحم  نـرحم  مـرحم  تـرحم  فـرحم
BUILD SUCCESSFUL (total time: 3 seconds)
```




```
run:
رحمنا رحمكما رحمان رحمها رحموا رحمتم رحمكم رحمتن رحمكن رحمنا رحمنا رحمنا رحمون رحمين رحمهن رحمهم رحمنه رحمني رحمني رحمن رحمك رحمه رحمة رحمت رحما رحمي
BUILD SUCCESSFUL (total time: 0 seconds)
```




```
run:
 الرحم والرحم بالرحم كالرحم فالرحم
BUILD SUCCESSFUL (total time: 2 seconds)
```




```
run:
 مسلم مسلم مسلم مسلم
BUILD SUCCESSFUL (total time: 0 seconds)
```

4 . 6 . نتائج إختبار النظام

للتأكد من الوصول الـي الهدف و التحقق من عمل و دقة النظام تم إختبار النظام بعدد من الأسئلة المعروف إجابتها مسبقاً و عدد أسئلة الإختبار عشرة سؤال، بحيث كانت نتائج الإختبار كما في الجدول رقم.

جدول 4-1: أسئلة الإختبار والإجابات الصحيحة لها.

الإختبار	صيغة السؤال	رقم الآية\ الآيات الصحيحة	اسم السورة
.1	من هم الذين إعتدوا في السبت	65	البقرة
.2	من أشد الناس ظلم	114، 140	البقرة
.3	من الذي حاج ابراهيم في ربه	258	البقرة
.4	مـا هـو أجـر مـن ينفـق بالليـل و النهار سراً و علانيةً	274	البقرة
.5	هـل يكتـب الـدين اي كاتـب ام لـه	282	البقرة

	صفات		
.6	من هم الذين اذنوا بحرب من الله و رسوله	279	البقرة
.7	من اين يخرج الطاغوت الكافر و الي اين	257	البقرة
.8	لون البقرة لبنو اسرائيل	69	البقرة
.9	هل يستهزئ الله بالكافر	15	البقرة
.10	من هو ولي الذين امنوا		البقرة

جدول 4-2: نتائج الإختبار بدون معالجة.

رقم الإختبار	عدد الإجابات	عدد الإجابات الصحيحة	الكلمات التي طابقة في الصحيحة
1	261	1	الذين
2	151	1	من
3	205	6	من، الذي، حاج، ابراهيم، في، ربه
4	203	1	بالليل
5	31	1	يكتب
6	405	4	من، بحرب، من ، الله
7	148	2	من، الطاغوت
8	6	0	ــــــــــــــــــــــــــــــ
9	133	1	يستهزئ
10	218	4	من ، ولي، الذين، امنوا

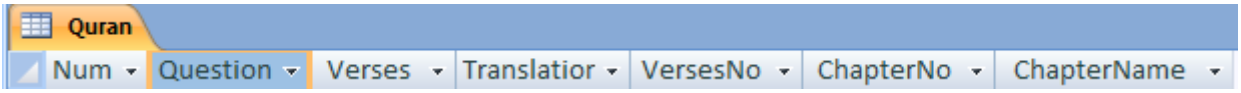
جدول 4-3: نتائج الإختبار بإختيار جميع المعالجات

رقم الإختبار	عدد الإجابات	عدد الإجابات الصحيحة	الكلمات التي طابقة في الصحيحة
1	2	1	اعتدوا
2	36	2	اظلم
3	12	1	حاج
4	13	1	بالليل
5	6	1	يكتب
6	1	1	بحرب
7	10	1	الطاغوت
8	10	1	لونها

9	2	1	يستهزئ
10	33	1	ولي

4 . 7 . وصف مدونة القرآن الكريم

مدونة القرآن الكريم التي إعتمد عليها التطبيق هي قاعدة البيانات أكسس وهي من إنشاء مشرف هذا البحث.الشكل رقم4-5 يوضح الأعمدة التي في قاعدة البيانات.



شكل4-5:أعمدة مدونة القرآن الكريم

شكل رقم 4-5 يوضح أعمدة مدونة القرآن الكريم حيث:

Num: رقم تسلسلي.

Question: تفيدنا في الإختبار فقط.

Verses: الآية.

Translation: الآية باللغة الإنجليزية.

VeresesNo : رقم الآية في السورة.

ChapterNO: رقم السورة حسب ترتيب المصحف الشريف.

ChapterName: إسم السورة.

فقط العمود Vereses هو الذي يدخل في عمليات المعالجة لإسترجاع الإجابة و التي تحتوي على بقية الأعمدة .

فعلى الرغم من أن القرآن الكريم نص لغوي يتصف بما تتصف به اللغة العربية من خصائص صوتية وصرفية وتركيبية ودلالية، فإنه يزيد عليها ما يجعله معجزا خارجًا عن قدرات البشر في الإبداع اللغوي، وإن كان بإمكانهم فهمه وتبينه، وهذا يعني أن المعالجة الآلية للنص القرآني تختلف في مفهومها وأهدافها عن المعالجة الآلية لعموم نصوص اللغة العربية، حيث يتوجه اهتمامها الأول إلى تحري الدقة ما استطاعت، أما غاية تعميم النتائج على عموم النص القرآني فيمكن بلوغها عن طريق بناء قواعد بيانات شاملة تستوعب جميع مفردات النص القرآني، وهذا ما لا يمكن إدراكه في معالجة عموم نصوص اللغة العربية.

4 . 8 . الخلاصة

تم توضيح واجهات النظام،وفي الباب التالي أهم النتائج التي توصل إليها البحث بعد عمليات التنفيذ و الإختبار و المراقبة، و توصيات الباحث، وخاتمة البحث .

الباب الخامس

النتائج والتوصيات

5 . 1 . مقدمة

في هذا الفصل سنتحدث عن المشكلة و خلفيتها و كيف قمنا بحلها و ما هي الأهداف التي برزت لحل المشكلة، و توصيات الباحث.

بدأت المشكلة بكيفية معالجة اللغة العربية آليا وكيفية تفاعل المستخدم مع النظام لإدراك المعالجة الآلية، و نظم إسترجاع المعلومات التقليدية لم تعد كافية لإرضاء حاجة المستخدم من المعلومات مع الكم الهائل من المعلومات والبيانات المخزنة سواء على الأنترنت أو الوسائط الألكترونية، درسنا في هذا البحث الجانب الصرفي للغة العربية و كيفية المعالجته كحل من حلول مشكلة البحث.

5 . 2 . النتائج

قامة الدراسة على بناء نظام رد آلي تفاعلي مع المستخدم مبني على كل القواعد التي تم إستخلاصها و تحليلها في هذا البحث، هذا النظام قادر على استقبال إستعلام مستخدم النظام وعمل ما تقوم به نظم إسترجاع المعلومات من المعالجات الأولية للإستعلام ثم عملية المطابقة أو البحث في الملفات الموجودة وإسترجاع المعلومات. بالإضافة إلى أنه يتم في هذا النظام توسعة الإستعلام بالقواعد التي صممت ليضمن للمستخدم كل الملفات ذات الصلة إضافة إلى كون المستخدم يتفاعل مع هذه القواعد من خلال واجهة المستخدم التفاعلية.

وفيما يلي أهم النتائج التي توصل إليها:

- تمت برمجة الأنماط و القواعد للجانب الصرفي للغة العربية، وهي: منها المشتقات: اسم الفاعل، اسم التفضيل، الصفة المشبهة، اسما الزمان والمكان، اسم الآلة، اسم المرة اسم الهيئة و صيغ المبالغة. ومنها إضافة حروف التعريف، إضافة السابقة، إضافة اللاحقة و حذف حروف التعريف. ذلك لتحسين نظم إسترجاع المعلومات، هذا ما لم يتم للغة العربية قبل هذا.
- بناء نظام رد آلي تفاعلي للقرآن الكريم فعال و كفء يعمل على إسترجاع الإجابة وهي الآية أو الآيات من القرآن الكريم بناءً على سؤال (استعلام) المستخدم ، و تفاعل المستخدم مع النظام من خلال واجهة البرنامج الذكية.

5 . 3 . التوصيات

- في إزالة الجذع، تم إزالة حروف التعريف(ال، بال، كال، فال، وال) لتوسعة الإستعلام، إكتفينا بهذا ولم نتطرق إلى إزالة اللاحقة و السابقة نسبة لصعوبة إزالة اللاحقة و السابقة من الكلمة نظرا لأنها قد تكون من أصل الكلمة نحو:"سنة الله" فالاحقة التاء المربوطة هي من أصل كلمة سنة ولكن النظام لا يتعرف على هذا فتحذف و تصير (سن) وهذه ليست من النتائج المرجوة، وكذلك في كلمة (حيينا) اللاحقة (نا) هي من أصل الكلمة.
- و أيضاً ما لم نطرقه في البحث هو إيجاد أصل الكلمة (root) هذا أمر صعبٌ جدا، يحتاج إلى جهود متضافرة مع اللغويين والحاسوبيين وبذل كل الجهد لإتمام المهة.
- عمليات الإعلال و الإبدال.
- في اللغة العربية عادة ما يتم دمج بعض حروف العطف مع الكلمة نحو: والذي ، والجامعات . وكذلك في القرآن الكريم بعض الكلمات مثل:"ويقيمون" "واسماعيل" "وابراهيم" "واياك نستعين" "والذين" هذا كله من الكتابة الألكترونية أو نمط الكتابة الحديثة على الحاسوب وليس من اصل القرآن . إذا أدخل المستخدم كلمة(اسماعيل)

-
- وأراد البحث عنها لا تظهر له نتائج لأنها مكتوبة (واسماعيل) في الملف أو قاعدة البيانات والعكس صحيح. إذن لابد من المعالجة الآلية.
- بما أنه يتم حذف التشكيل فهذا يفقدنا خاصية التعرف على الثلاثى المضعف ، والقواعد بنيت على الكلمة المكونة من ثلاثة أحرف فما فوق، أما الكلمة المضعفة من حرفين و تسقط من هذه القاعدة. في البحث تم التغلب على هذا الإستثناء بأن جعلنا القاعدة تستقبل الكلمة المكونة من ثلاثة أحرف أو حرفين . ولكن قد يكون هناك حل أفضل.
- استخدام القواعد المقترحة بصورة متداخلة اي تطبيق أكثر من قاعدة على الكلمة الواحدة، نحو: كلمة(مسلم) لتصبح (المسلمين) يجب تنفيذ قاعدتين عليها هما إضافة حرف التعريف (ال) ثم إضافة اللاحق (ين) فتصبح(المسلمين). أو إضافة اللاحقة (ين) فتصبح(مسلمين) ثم إضافة حرف التعريف فتصبح(المسلمين)، وهكذا لجميع القواعد.
- مواصلة لمعالجة اللغة العربية آليا يجب الإهتمام بالجوانب الاخرى لعلم الصرف التي لم نتطرق لها، و كذلك لبقية التراكيب الأخرى للغة العربية و الجانب الدلالي.
- مشكلة لغة البرمجة ظهور الحرف الذي يمثل في الآسكي كود و هو (65279) على حرف التعريف (ال) حلت المشكلة بقراءة الآسكي كود للحرف و في كل مرة تتم القراءة من الملف يتم إختبار هذا الحرف إن وجد يتم حذفه.
- إن أمكن إضافة عمود آخر في المدونة فيه تفسير الآية، عند مطابقة الآية الكريمة من القرآن الكريم مع سؤال المستخدم يتم إسترجاع الآية مع تفسيرها لتكون المعلومات المسترجعة أكثر فائدة للمستخدم.
- و أخيرا نوصي الجامعات العربية و كليات علوم الحاسوب خاصة الإهتمام كل الإهتمام بمعالجة اللغة العربية آليا و ذلك بإدخال منهج حوسبة اللغة العربية سيؤدي ذلك لزيادة البحوث ودعم اللغة العربية آلياً.
- كما إن أُتيحت لنا الفرصة بأن نطلب من علماء اللغة العربية الإنضمام مع علماء الحاسوب للإستنهاض باللغة العربية آليا.

5 . 4 . الخاتمة

نحمد الله تعالى حمدا كثيرا، كما ينبغي لجلال وجهه و عظيم سلطانه لما وفقنا إليه من إتمام هذا البحث المتواضع (وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ ۚ)(هود،88)، إن لكل بداية نهاية و إن كل نهاية تعني أن هنالك بداية جديدة، و خير العمل ما أحسن آخره، وخير الكلام ما قل و دل، نقدم لكم هذا العمل المتواضع على أملٌ أن نكون قد قدمنا عملاً صالحاً لنا و لكل المسلمين، وما كان من تقصير فمن أنفسنا و الشيطان، ونرجوا من الله المغفرة، والقبول و السداد، ونسأله تعالى أن يجعل هذا العمل خالصا لوجهه الكريم، و أن تكون قد وصلت رسالتنا العلمية إلى درجة القبول لديكم.

المراجع

[1] ".منتديات ستار تايمز، "المعالجة الآلية للغة العربية،جهود وتحديات [Online]. Available: http://www.startimes.com/?t=21168801. [Accessed: 21-Oct-2017].

[2] RT news، "ترتيب لغات العالم من حيث الانتشار - RT Arabic." [Online]. Available: https://arabic.rt.com/news/786982-ترتيب-لغات-العالم-الانتشار/. [Accessed: 21-Oct-2017].

[3] غسان مراد، "المعالجة الآلية للغة العربية : الامكانات والفوائد والآفاق،"*رقم العدد: 15578*، 2005 [Online]. Available: http://daharchives.alhayat.com/issue_archive/Hayat INT/2005/26/11/المعالجة-الآلية-للغة-العربية-الام%D9. [Accessed: 16-Oct-2017].

[4] A. L. I. Farghaly، "Arabic Natural Language Processing : Challenges and Solutions،" vol. 8، no. 4، pp. 1–22، 2009.

[5] م. ا. ت. ا. محمد، "بحث لنيل درجة الماجستير في علوم الحاسوب، بعنوان 'تصميم طريقة لمعالجة جموع التكسير لتحسين أداء نظم إسترجاع المعلومات العربية' اسم الدارس:محمد المؤيد تاج السر محمد اشراف : د.البراء ابوعبيدة محمد علي، sudan university of science and technology."

[6] ".د. عبد الله علي علي الثوري رئيس، "خصائص تراكيب اللغة العربية ~ مدونة الثوري [Online]. Available: http://thory1.blogspot.com/2016/02/blog-post.html. [Accessed: 21-Oct-2017].

[7] ".د. ج. ع. العزيز، "قواعد الصرف

[8] ش. خوجة، "Khoja Stemmer."

[9] MohamedAlmoayed and TajAlsir، "Designing broken plurals Processing Method for Enhancing the Performance of Arabic Information Retrieval Systems (Thesis)،" 2015.

[10] E. Mohamed Adeny *et al.*، "Islamic Applications of Question-Answering System:Compartive Study Vol.7 NO.1 2017،29-41:،" *J. Adv. Comput. SICENSE Technol. Res.* .، 2017.

الملاحق

الملحق أ

الكلمات التوقيفية

{ان بعد ضد يلي الى في من حتى وهو يكون به وليس أحد على وكان تلك كذلك التي وبين فيها عليها إن وعلى لكن عن مساء ليس منذ الذي أما حين ومن لا ليسب وكانت أي ما عنه حول دون مع لكنه ولكن له هذا والتي فقط ثم هذه أنه تكون قد بين جدا لن نحو كان لهم لأن اليوم لم هؤلاء فإن فيه ذلك لو عند اللذين كل بد لدى وثي أن ومع فقد بل هو عنها منه بها وفي فهو تحت لها أو إذ علي عليه كما كيف هنا وقد كانت لذلك أمام هناك قبل معه يوم منها إلى إذا هل حيث هي اذا او و ما لا الي إلي مازال لازال لايزال مايزال اصبح أصبح أمسى امسى أضحى اضحى ظل مابرح مافتئ ماانفك بات صار ليس إن كأن ليت لعل لاسيما ولايزال الحالي ضمن اول وله ذات اي بدلا اليها انه الذين فانه وان والذي وهذا لهذا الا فكان ستكون مما أبو بإن الذي اليه يمكن بهذا لدي وأن وهي وأبو آل الذي هن الذى}.

السابقة في اللغة العربية

{لل ل أ و س ب ي ن م ت ف}.

اللاحقة في اللغة العربية

{ هما تما كما ان ها وا تم كم تن كن نا تا ما ون ين هن هم ته تي ني ن ك ه ة ت ا ي ات }