الجمهوريـــــة الجـــزائريـــة الديمقراطيـــة الشعبيـــــة
وزارة التعليم العالي والبحث العلمي

جامعـــة حمـــة لخضـــر الـــوادي كلية العلوم الدقيقة قسم الإعلام الآلي
مذكرة نهاية التخرج
تدخل ضمن متطلبات الحصول على شهادة

ليســـانس أكاديمـي

الميدان: رياضيات وإعلام آلي
الشعبة: الإعلام الآلي
التخصص: أنظمة معلوماتية

تصميم و تطوير نظام استخراج الخريطة المكانية و الزمانية للأحداث من النصوص و المراجع النصية

من انجاز الطلبة:

***زغيدي هناء**
***سخري بدرة**
***غيادي رزيقة**

تحت إشراف الأستاذ المحترم :

***محمد مفتاح شرف الدين**

السنة الجامعية: 2020 /2021

شكـــــر وتقديـــر

قبل كل شيء، أشكر الله القوي على الصحة الجيدة والشجاعة

الذي أتاحها لنا لإعداد هذا العمل المتواضع أما بعد :

نشكر الأستاذ المشرف السيد: **مفتاح محمد شرف الدين** الذي

قبلنا لقيادتنا بفضل انتقاداته و ملاحظاته و توجيهاته .

نشكر جميع الذين ساهموا بشكل مباشر أو غير مباشر في نجاح هذا

البحث و خاصة الوالدين الذين ساعدونا طوال حياتنا حتى اليوم

دون نسيان بالطبع جميع الإشراف الفني و الإداري لجامعة الشهيد

حمة لخضر في الوادي .

الإهـــداء

إلى من أُفضِّلها على نفسي، ولِمَ لا؛ فلقد ضحَّت من أجلي

ولم تدَّخر جُهدًا في سبيل إسعادي على الدَّوام

(أُمِّي الحبيبة).

نسير في دروب الحياة، ويبقى من يُسيطر على أذهاننا في كل مسلك نسلكه

صاحب الوجه الطيب، والأفعال الحسنة.

فلم يبخل عليَّ طيلة حياته

الطلبة :

* زغيدي هناء

* سخري بدرة

* غيادي رزيقة

ملخص:

إن الباحثين في مجال السير (جمع سيرة) والأحداث السياسية و التاريخية يعتمدون في أبحاثهم على كم كبير من المراجع المختلفة في دراساتهم التاريخية مما يجعل من الصعب على الباحث الربط بين التواريخ (الأزمنة) و الأماكن و الأحداث التي وقعت فيها و التزامن بينهم .

النظام المقترح في هذا الموضوع سيكون عبارة عن وسيلة مساعدة لإستخراج خريطة أو خرائط الأحداث و إرتباطاتها المكانية و الزمانية و أيضا الشخصيات ... بحيث تكون مدخلات هذا النظام مجموعه مراجع (كتب "نصوص") و مخرجاتها خريطة أو خرائط يمكن تصنيفها عبر الأماكن أو الأزمنة أو الشخصيات ..

وقدا عتمدنا في تطويرهذا البرنامج على العديدمن التقنيات مثل:

*start UML لوصفأجزاءومكوناتالبرنامج

*قواعد البيانات SQL Server للتخزين

*ولغة java لتطوير البرنامج

الفهرس

شكـــــر وتقديــر .....ب

ملخص: .....ث

قائمة الأشكال : ..... خ

قائمة الصور : .....د

قائمة الجداول : .....ذ

مقدمة عامة : ..... ر

الفصل الأول :معالجة النصوص الطبيعية .....

مقدمة :.....**1**

1-تعريف (معالجة اللغات الطبيعية): **.... .....1**

2-الطريقة التي تمكن الحاسوب من فهم اللغة : **.....** خطأ! الإشارة المرجعية غير معرّفة.

3-بناء تسلسل معالجة اللّغات الطّبيعية خطوة بخطوة : **. 2**

3-1تجزئة الجمل Sentence Segmentation : 2

3-2الحصول على الوحدات اللغوية Tokenization : 3

3-3 التنبؤ بأقسام الكلام Part of Speech لكل وحدة لغوية : 4

3-4أصل كلمات النص Lematization : 5

3-5تحديد كلمات التوقف : 7

3-7 التعرف على الكيانات المسماة Named-Entity Recognition NER : 11

خلاصة : **..... .....13**

الفصل الثاني :التصميم **.....14**

مـقدمـة: **.....16**

1-لغة النمذجة uml (modeling language) **.....16**

2-مخطط الحالة (diagramme de cas d' utilisation): **.....17**

3-مخطط النشاط (diagrammed Activity): **.....18**

4-مخطط التسلسل (Diagramme séquence) **.....19**

5-مخطط الفئات(digramme de classe) **.....21**

خلاصة : **.....21**

المقدمة: **.....23**

1-نتبينز (Net Beans) : **.....23**

2-لغة الجافا (JAVA) : **.....23**

3-قواعد البيانات: **23**
1-1 جدول المستخدمين : 24
2-1جدول الاحداث : 24
3-1جدول المكان : 25
4-1 جدول الشخصيات : 25
5-1 جدول الربط الاحداث _ المكان : 25
6-1 جدول الربط الحدث _ الشخصيات : 26
4-قاعدة البيانات المنجزة : **26**
5-هيكلة الملفات : **27**
6-الإنجاز: **29**
1-4 الواجهة الرئيسية للتطبيق : 29
2-4 واجهة إضافة ملف : 30
7-قاعدة البيانات : **32**
8-خاتمة : **32**
خاتمة عامة : 33
المصادر: 34

قائمة الأشكال :

الشكل-1-مخطط إستخراج الأماكن و الأشخاص و الأحداث 17

الشكل-2- مخطط الحالة.....18

الشكل -3- مخطط النشاط.....19

الشكل -4- مخطط التسلسل 20

الشكل -5- مخطط الفئات.....21

الشكل -6-قاعدة البيانات (الربط).....26

الشكل -7- صورة موجزة للملفات الاساسية.....28

الشكل-8- الواجهة الرئيسية للتطبيق.....29

الشكل-9- واجهة إضافة ملف.....30

الشكل -10- واجهة إدخال نص 31

الشكل -11- واجهة تنفيد التطبيق 31

الشكل-12- قاعدة البيانات.....32

قائمة الصور :

الصورة -1- التنبؤ بأقسام الكلام باستخدام نموذج مدرب 5

الصورة -2-" الهيئات التركيبية للجملة "صديقي وُلِد في دمشق.....8

الصورة -3- شجرة التحليل.....9

الصورة -4- شجرة التحليل للتنبؤ بالعلاقات.....10

الصورة -5- العبارات الاسمية.....11

قائمة الجداول :

الجدول:01 :جدول الكلمات المفتاحية.....24

الجدول:02 المستخدمين.....24

الجدول:03 الأحداث.....24

الجدول:04 جدول المكان.....25

الجدول:05 جدول الشخصيات.....25

الجدول:06جدول ربط الاحداث_المكان.....25

الجدول:07جدول ربط الحدث_الشخصيات.....26

مقدمة عامة :

تبدو الحواسيب رائعة في معالجة البيانات المهيكلة مثل الجداول الإلكترونية و قواعد البيانات ، لكننا نحن كبشر نتواصل مع بعضنا من خلال الجمل لا الجداول .

وهذا غير مألوف للحاسوب الذي لا يفقه إلا أرقاما و بيانات منظمة فالكثير من المعلومات في الواقع هي غير مهيكلة ، مثلا النصوص المكتوبة باللغة الإنجليزية أو العربية أو غيرها من اللغات الطبيعية .

يعتمد الباحثين كثيرا على المراجع و الكتب مما يجعل هذا الأمر صعب و عسير لهذا لجأنا للبحث على طريقة تسهل لهم أبحاثهم بحيث نلجأ لمعالجة اللغات الطبيعية و ننشأ نظام يساعدهم على استخراج الأماكن و الشخصيات و الأحداث

الفصل الأول :معالجة النصوص الطبيعية

الفصل الأول : معالجة النصوص الطبيعية

مقدمة :

تعتبر معالجة اللغات الطبيعية من تطبيقات اللغات الذكاء الإصطناعي التي تركز على تمكين الحاسوب من فهم و معالجة اللغات البشرية .

1- تعريف (معالجة اللغات الطبيعية):

معالجة اللغات الطبيعية (NLP) , اختصارا بإسم NLP، هي تقنية تستخدم أجهزة الكمبيوتر لمعالجة أنواع مختلفة من معلومات اللغة الطبيعية الفريدة للبشر في كل من الأشكال المكتوبة و الشفوية .

- تبني المهام أنظمة تؤدي مهاما معينة من خلال معالجة اللغة و فهمها .
- هي حقل فرعي من الذكاء الإصطناعي وعلم اللغات و علوم الحاسوب ، يهتم بدراسة تعلم الآلة للغات البشرية وتفاعلها معها ، و تتناول معالجة اللغات الطبيعية فهم اللغة و توليدها و التعرف على النصوص و الكلام

[1]

2- الطريقة التي تمكن الحاسوب من فهم اللغة :

تستمر محاولات المبرمجين في كتابة برمجيات تفهم اللغات البشرية طالما يستمر استخدام الحواسيب في حياتنا . و السبب وراء ذلك واضح ، فالبشر سيكون من المفيد حقا إذا أصبحت أجهزة الحواسيب تقرأ و تفهم تلك البانات المكتوبة .

ما زالت الحواسيب لا تستطيع فهم اللغة البشرية بشكل صحيح تماما مثل الطريقة التي يفهم بها الإنسان . ستشعر أن معالجة اللغات الطبيعية في بعض المجالات تشبه أعمال السحر . فأنصحك أن تطبق

تقنيات معالجة اللغات الطبيعية على مشاريعك الخاصة و أن تستخدم مكتبات بايثون مفتوحة المصدر الخاصة بمعالجة اللغات الطبيعية لتتابع أحداث التطورات في معالجة اللغات الطبيعية مثل **textacy** ، **SpaCy** و **neuralcoref** فبأسطر قليلة ستحقق عملا رائعا .

[1]

3- بناء تسلسل معالجة اللّغات الطّبيعية خطوة بخطوة :

لمعالجة نصوص يجب علينا إتباع خطوات كثيرة نتطرق لها في هذا الجزء من معالجة اللغات الطبيعية ، دعونا نأخذ هذا الجزء من النص :

"مدينة دمشق هي عاصمةُ الجمهورية العربية السورية و أكبر المدن السورية من حيث الكثافة السكانية؛ تعتبر مدينة دمشق من أقدم العواصم المأهولة في العالم، وقد رجّح المؤرّخون أن تاريخ المدينة يعود إلى ما قبل الألف السابع قبل الميلاد، حيث تم العثور على بعض الحفريات في منطقة تل الرماد، ودلّت هذه الحفريات أن تاريخ المدينة يعود إلى تسع آلاف سنة قبل الميلاد. يعود اسم (دمشق) إلى أصولٍ آشوريّةٍ قديمةٍ و يعني الأرض العامرة والزّاهِرة، دَلالةً على جمال طبيعتها وتضاريسها الخلّابة، ويُقال بأنها سُميت (شام) نسبةً إلى سام بن نوح عليه السلام. تقع مدينة دمشق في الجزء الجنوبي الغربي من الجمهورية السوريّة، يحدّ المدينة سهول حوران وجبال القلمون والبادية السورية و تحيط بالمدينة بساتين الغوطة، وربوة دمشق، وجبل قاسيون، كما تطل المدينة على ضفاف نهر بردى".

تتألف هذه الفقرة النصية من عدة حقائق مفيدة. وسيكون شيئاً عظيماً إذا استطاع الحاسوب أن يفهم من قراءته للنص أن دمشق هي عاصمةُ الجمهورية العربية السورية و هي من أكبر المدن السورية من حيث الكثافة السكانية، وكذلك تعدّ من أقدم العواصم المأهولة في العالم، وذات أصولٍ آشوريّةٍ قديمةٍ. لنحقق ذلك، علينا أن نعلّم الحاسوب أولاً المفاهيم الأساسية للغة المكتوبة، ثم ننتقل إلى المراحل الأخرى للوصول إلى فهم النص.[1]

3-1 تجزئة الجمل Sentence Segmentation :

إن المرحلة الأولى في مراحل معالجة اللغات الطبيعية هي تقطيع النص إلى الجمل المكون منها، فمن النص السابق المدروس تكون الجمل:

مدينة دمشق هي عاصمةُ الجمهورية العربية السورية و أكبر المدن السورية من حيث الكثافة السكانية

تعتبر مدينة دمشق من أقدم العواصم المأهولة في العالم

وقد رجِّح المؤرِّخون أن تاريخ المدينة يعود إلى ما قبل الألف السابع قبل الميلاد....

إنّ كل جملة من جمل النص تعبر عن فكرة معينة كما هو الأمر بالنسبة لأي نص في اللغة الطّبيعية، لذلك يكون من الأسهل أن نجعل الحاسوب يفهم فكرة جملة إلى أن يصل إلى فهم كامل النص المؤلّف من هذه الجمل. ومن الشّائع في تقطيع النصوص الاعتماد على علامات التنقيط، حيث تشكّل علامات التنقيط الفاصل بين جمل النص، وفي حال لم يكن النص منسّقاً باستخدام علامات التنقيط فيمكن اعتماد طرق أكثر تعقيداً لتقطيع النص والحصول على جمله. [1]

2-3 الحصول على الوحدات اللغوية Tokenization :

في المرحلة الثانية نقوم بتقطيع الجمل التي حصلنا عليها إلى الوحدات اللغوية Tokens ، في حالتنا تتمثل token بالكلمة word ، ويمكن أن تتم هذه المرحلة من المعالجة على التسلسل أي معالجة جملة تلو الجملة الأخرى. لنأخذ الجملة الأولى من النص المدروس على سبيل المثال :

‘‘مدينة دمشق هي عاصمةُ الجمهورية العربية السورية و أكبر المدن السورية من حيث الكثافة السكانية’’

تتمثل هذه المرحلة التي تدعى Tokenization (الحصول على الوحدات اللغوية) بتقطيع الجمل إلى كلماتها، فإذا طبقنا عملية الحصول على الوحدات اللغوية على الجملة المدروسة سنحصل على الكلمات التالية:

"مدينة"، "دمشق"، "هي"، "عاصمةُ"، "الجمهورية"، " العربية"، "السورية"، "و"، "أكبر"، "المدن"، "السورية"، "من"، "حيث"، "الكثافة"، "السكانية"، ".".

إن عملية الحصول على الكلمات في اللغة الإنجليزية بسيط لأنه يمكن أن يكون الفراغ هو الفاصل بين كلمات الجمل أما بالنسبة إلى اللغة العربية فالأمر أكثر تعقيداً، لوجود ضمائر متصلة بالكلمة وكل ضمير يحمل معنىً مختلفاً. تُفصل علامات التنقيط كوحدة لغوية Token مستقلة، وذلك لأننا نعرف أن كل علامة تنقيط لها معنى مختلف عن الأخرى. [1]

3-3 التنبؤ بأقسام الكلام Part of Speech لكل وحدة لغوية :

نقوم في هذه المرحلة بإسناد كل وحدة لغوية إلى قسم الكلام المناسب لها سواء كان اسماً، أو فعلاً أو صفة. إن معرفة الدور الذي تؤديه كل كلمة في الجملة سوف يساعد الحاسوب على فهم ما تعنيه هذه الجملة. سنحصل على أقسام الكلام من خلال تمرير الكلمة والسياق المجاور لها إلى نموذج تصنيف أقسام الكلام المدرب مسبقاً.

الصورة -1- التنبؤ بأقسام الكلام باستخدام نموذج مدرب

يتم تدريب نماذج أقسام الكلام على مجموعة كبيرة من الجمل بحيث تكون كل جملة موسومة tagged بأقسام الكلام المناسب لها . لنأخذ باعتبارنا أن النموذج يعتمد بشكل كلي على الإحصاءات ومهما كانت نتائجهم صحيحة لكنه بالتأكيد لن يفهم ماذا تعني الكلمات في الحقيقة كما نفهمها نحن كبشر. إن النموذج يعرف فقط كيف يتنبأ بأقسام الكلام بناء على الجمل والكلمات التي مررت له سابقاً.

بعد معالجة الجملة التالية ‘‘تعتبر مدينة دمشق من أقدم العواصم المأهولة في العالم’’ يمكن أن نحصل على النتيجة التالية:

أقدم	من	دمشق	تعتبر
’’اسم‘‘	’’حرف جر‘‘	’’اسم علم‘‘	’’فعل‘‘
العالم	في	المأهولة	العواصم
’’اسم‘‘	’’حرف جر‘‘	’’اسم‘‘	’’اسم‘‘

عند حصولنا على هذه المعلومات ، سنستطيع البدء بجمع المعاني الأساسية، على سبيل المثال بعد معرفتنا أن الأسماء تضم "دمشق" و "العواصم" فإن الجملة ربما تتحدث عن دمشق. [1]

4-3 أصل كلمات النص Lematization :

تظهر الكلمة في اللغات الطبيعية بأشكال مختلفة ، مثلاً في اللغة الإنجليزية يضاف حرف s إلى آخر بعض الكلمات لتعبر عن الجمع، أما في اللغة العربية فتضاف بعض الحروف وتحذف بعضها لتعبر عن حالات مختلفة في علم الصرف للغة العربية، وأصل الكلمة هو الشكل الأساسي لها على سبيل المثال فإن الشكل الأساسي لكلمة ‘‘يحتاجون’’ في اللغة العربية هي ‘‘احتاج’’ وكذلك الشكل الأساسي لكلمة ‘‘studied’’ في اللغة الإنكليزية ‘‘study’’.

ومن المفيد أن يعرف الحاسوب أصل الكلمات أي شكلها الأساسي كي يعرف أن الجمل التي تحتوي هذه الكلمات تدل على مفهوم واحد. فكما تعلم عزيزي القارئ أن الحاسوب يفهم الكلمات كلها كمحارف فإذا اختلفت هذه المحارف ولو اختلافاً بسيطاً اعتبرها الحاسوب كلمة مختلفة تماماً.

سنسمي هذه المرحلة في معالجة اللغات الطبيعية إيجاد "أصل الكلمة" أي اكتشاف الشكل الأساسي "Lemma" لكل كلمة في الجملة. تطبق عملية إيجاد أصل الكلمة على الأفعال والأسماء. فنحن نعلم أن الأفعال تتغير أشكالها بتغير الزمن الذي تدل عليه أو حسب الضمائر المتصلة بها.

عادة يتم الاعتماد على جدول مفهرس لأشكال الكلمات وأصلها بناء على قسمها الكلامي من أجل عملية إيجاد أصل الكلمة، ويمكن أن يكون هناك قواعد مخصصة للكلمات التي لا ترد كثيراً، وتعتبر خوارزميات إيجاد أصل الكلمة مجالاً مفتوحاً للبحث العلمي. مثال يوضح عملية الحصول على أصل الكلمة:

أقدم	من	دمشق	تعتبر
"أقدم"	"من"	"دمشق"	"اعتبر"
العالم	في	المأهولة	العواصم
"عالم"	"في"	"مأهول"	"عاصمة"

5-3 تحديد كلمات التوقف :

نريد الآن أن نفكر بأهمية كل كلمة في الجملة، ففي أغلب اللغات تتكررحروف العطف والتعريف كثيراً خلال الجمل النصية. وعند بناء نماذج إحصائية للغة ستعطي هذه الكلمات تشويشاً للنتيجة حيث أنها تظهر أكثر من باقي الكلمات في الجمل. يمكن تسمية هذه الكلمات بكلمات التوقف Stop Words وهي الكلمات التي يجب حذفها من النص قبل البدء بأي عملية إحصائية. تصبح الجملة المدروسة بالشكل التالي بعد إزالة حروف التوقف "تعتبر مدينة دمشق أقدم العواصم المأهولة العالم".

يمكن تحديد حروف التوقف بإضافة سلسلة من كلمات التوقف المعروفة، لكن لا يوجد سلسلة قياسية لحروف التوقف مناسبة لكل التطبيقات وذلك لأن كلمات التوقف تختلف من تطبيق لآخر.

على سبيل المثال في اللغة الإنجليزية ، إذا كنت تقوم بإنشاء محرك بحث لموسيقى الروك ، فأنت تريد التأكد من أنك لا تتجاهل كلمة "The". وذلك لأنه لا تظهر كلمة "The" في أسماء مجالات الموسيقى فقط ، فهناك أيضًا فرقة روك شهيرة من ثمانينيات القرن العشرين تسمى The The. [1]

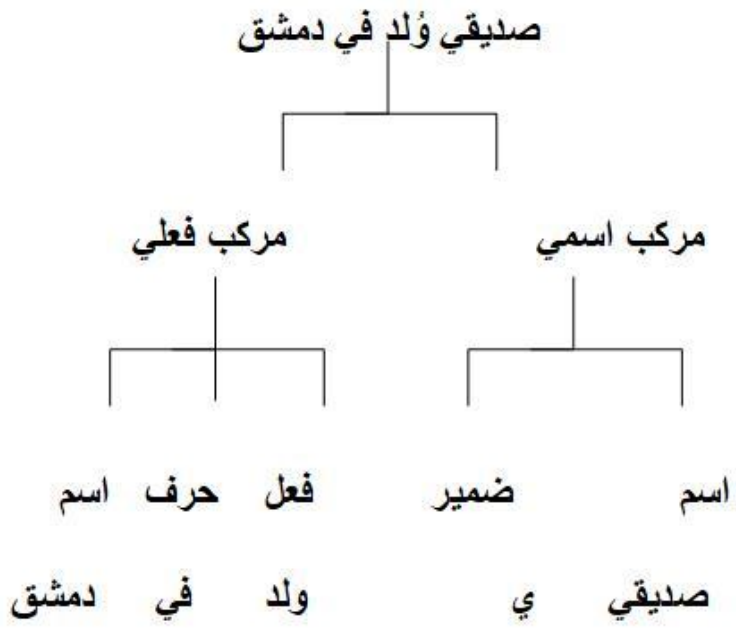
6-3 المرحلة السادسة :

أ- الإعراب الاعتمادي Dependency Parsing :

إنها لغة طبيعية ومعالجتها ليست بالأمر السهل فنحن نحاول أن نجعل الحاسوب - الآلة الجامدة - تقرأ لغتنا وتفهم معانيها.

في هذه المرحلة سنكتشف علاقات كلمات الجملة مع بعضها البعض وهذا ما يسمّى الإعراب الاعتمادي.

تتميز اللغة العربية بنظام إعرابي خاص بها، كونها لغة غنية بالمكونات النحوية، فمثلاً لتكن لدينا الجملة التالية: "صديقي وُلِد في دمشق" يمكن أن تكون الهيئات التركيبية للجملة – أي ترتيب المكونات النحوية التي تتألف منها الجملة- كما يلي:

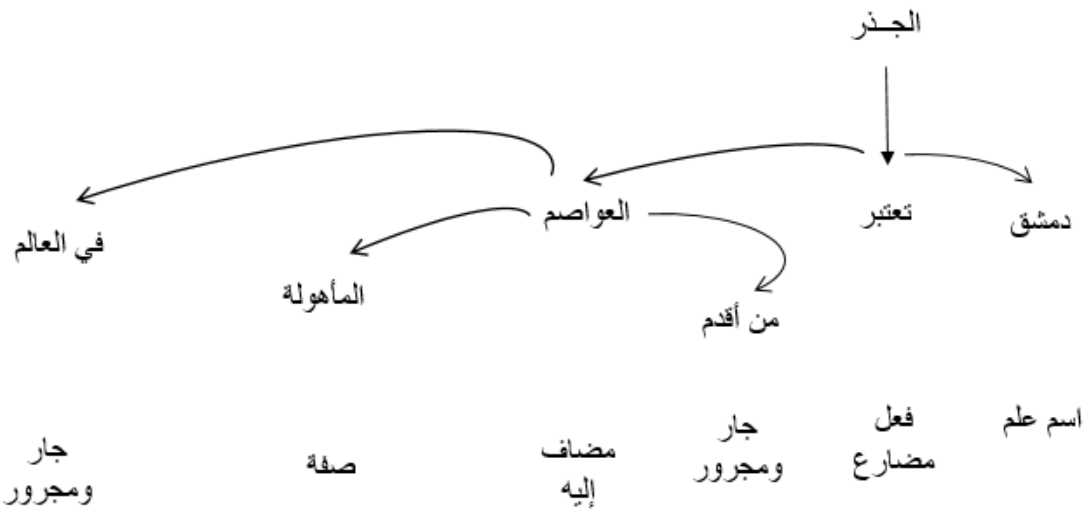


الصورة -2-" الهيئات التركيبية للجملة "صديقي وُلِد في دمشق

سنأخذ الجملة التالية"*تُعتبر دمشق من أقدم العواصم المأهولة في العالم، وهي عاصمة سورية*"

الهدف من مرحلة الإعراب الاعتمادي هو بناء شجرة تعيّن كلمة أصل واحدة لكل كلمة في الجملة. وسيكون جذر الشجرة هو الفعل الرئيسي في الجملة.

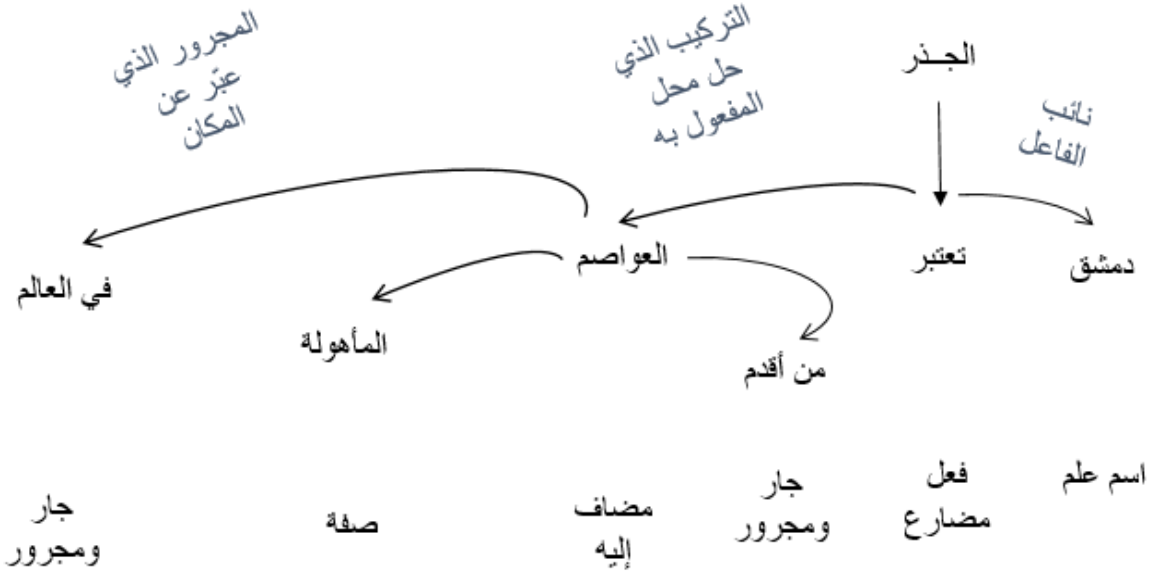
ستبدو شجرة التحليل كما يلي:



الصورة -3- شجرة التحليل

بالإضافة إلى تحديد الكلمة الأب لكل كلمة، نستطيع أيضاً التنبؤ بنمط العلاقة الموجودة بين هذه الكلمات.

[1]



الصورة -4- شجرة التحليل للتنبؤ بالعلاقات

توضح لنا هذه الشجرة التحليلية أن نائب الفاعل في الجملة هو الاسم "دمشق" وله علاقة "تُعتبر" مع "العواصم"، ونحن نعلم سابقاً من مرحلة إيجاد أصل الكلمة أن أصل العواصم "عاصمة"، وبالتالي علمنا شيئاً مفيداً وهو أن دمشق عاصمة وإذا تابعنا شجرة التحليل الكاملة لهذه الجملة، فسنكتشف أن دمشق هي عاصمة سورية.

يعمل الإعراب الاعتمادي أيضًا على تمرير الكلمات إلى نموذج التعلم الآلي وإخراج النتيجة، تمامًا مثل الطريقة التي تنبأنا بها عن أقسام الكلام في المرحلة السابقة باستخدام نموذج التعلم الآلي. ولكن الإعراب الاعتمادي مهمة معقدة بشكل خاص وتتطلب مقالات أخرى لشرحها بالتفصيل.

يمكن اعتبار هذا الأسلوب في الإعراب الاعتمادي هو طريقة قياسية لكن قد يجد البعض أنه أصبح قديماً، حيث أصدرت Google في عام 2016 محللًا جديدًا للاعراب يُدعى Parsey McParseface والذي تفوق على المعايير السابقة باستخدام منهج جديد للتعلم العميق انتشر سريعًا في جميع الأنحاء. بعد ذلك بعام ، أطلقوا نموذجًا أحدث أطلق عليه اسم ParseySaurus مما زاد من تحسين الأمور. وبذلك يمكننا القول، أن تقنيات الإعراب لا تزال مجالًا للبحث العلمي.

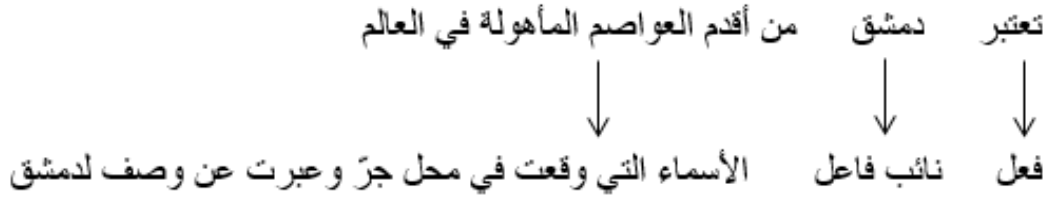
من المهم أن نتذكر أن اللغة العربية غنية بالمكونات النحوية وتحتاج إلى دراسة تحليلية تخصصية لمعرفة إعرابها الصحيح، وكذلك هناك عدة جمل إنجليزية غامضة ويصعب تحليلها، وفي هذه الحالات ، سيقدم النموذج تخمينًا اعتماداً على ما يبدو إعراب الجملة ، لكنه ليس مثالياً وأحيانًا سيكون النموذج خاطئًا بشكل كبير. لذلك مع مرور الوقت، ستستمر نماذج معالجة اللغات الطبيعية في تحسين إعراب النص بطريقة مقبولة، على مستوى جميع اللغات، ولا سيّما اللغة العربية التي هي قيد البحث والتطوير.

ب- إيجاد العبارات الاسمية :

حتى الآن، تعاملنا مع كل كلمة في الجملة ككيان منفصل. لكن في بعض الأحيان يكون من المنطقي تجميع الكلمات التي تمثل فكرة أو شيء واحد معًا. يمكننا استخدام المعلومات من شجرة الإعراب الاعتمادي لتجميع الكلمات التي تتحدث عن نفس الشيء تلقائيًا.

على سبيل المثل بدلاً من هذا الشكل:

يمكننا تجميع الكلمات التي وقعت في محل جرّ وعبرت عن فكرة واحدة، لتوليد ذلك الشكل:



الصورة -5- العبارات الاسمية

إن هذه المرحلة اختيارية وذلك بناء على الهدف المطلوب. وغالبًا تكون هذه المرحلة طريقة سريعة وسهلة لتبسيط الجملة إذا لم نكن بحاجة إلى المزيد من التفاصيل. [1]

7-3 التعرف على الكيانات المسماةNamed-Entity Recognition NER :

بعد أن أنهينا القسم الصعب، الذي كان يهتم بالقواعد، ننتقل إلى استخراج الأفكار من الجمل.

في جملتنا المدروسة لدينا الأسماء التالية:

"تعتبر دمشق من أقدم العواصم المأهولة في العالم، وهي عاصمة سورية"

المأهولة	العواصم	أقدم	دمشق
سورية	عاصمة		العالم

تشير بعض هذه الأسماء على أشياء من الواقع مثلا دمشق ، سورية لها مواقع جغرافية على الخريطة يمكنك تحديدها. وبذلك سنقوم بالاستخراج الآلي للائحة من الأماكن الحقيقية المذكورة في المستند النصي باستخدام معالجة اللغات الطبيعية.

الهدف من التعرف على الكيانات المسماة هو تحديد وتسمية label هذه الأسماء Nouns بمفاهيم حقيقية تمثّلها.

تصبح جملتنا بعد تمريرها على نموذج تحديد الكيانات المسماة بالشكل التالي:

المأهولة	العواصم	أقدم	دمشق
			موقع جغرافي
سورية	عاصمة		العالم
			موقع جغرافي

لكن أنظمة التعرف على الكيانات المسماة لا تقتصر مهمتها على إنجاز قاموس بسيط فحسب، إنما تستخدم سياق الكلام لتعرف كيفية ظهور الكلمة و النموذج الإحصائي للتنبؤ بنوع الاسم الذي تمثله الكلمة.

فنظام التعرف على الكيانات المسماة الجيد يستطيع تمييز ‘‘شام’’ (اسم لفتاة) أنّها شخص و ‘‘شام’’ مكان ذو موقع جغرافي وذلك من السياق.

سنورد لكم بعض الأشياء التي تستطيع أنظمة التعرف على الكيانات المسماة تسميتها.

- أسماء الأشخاص
- أسماء الشركات.
- المواقع الجغرافية(الفيزيائية والسياسية)
- أسماء المنتجات
- التاريخ والوقت
- كميات النقود
- أسماء الأحداث.

إن التعرف على الكيانات المسماة له الكثير من الاستخدامات لأنه يسّهل كثيراً عملية الحصول على البيانات المنظمة من النص، والوصول إلى فهمه [1].

خلاصة :

تعتبر اللغات الطبيعية من المواضيع المثيرة للإهتمام في عصر أصبحت به التكنولوجيا شريكة البشر في كل المجالات . تتم معالجة اللغات الطبيعية من خلال مراحل متكاملة مع بعضها للوصول إلى االهدف المطلوب.

الفصل الثاني :التصميم

مـقدمـة:

تلعب لغة النمذجة الموحدة (uml) دور كبير في تطوير البرامج , و انظمة غير برمجية في العديد من الصناعات ,كوسيلة لاظهار سلوك النظام او عملية او هيكلها بشكل مرئي ,تساعد uml على عرض الاخطاء المحتملة في هياكل التطبيقات و سلوك النظام و عمليات اخرى.

سوف نتطرق في هذا الفصل الى معرفة متطلبات برمجة هذا التطبيق كما سوف يقدم لنا تحليلا يتوافق مع مخططات uml .

1- لغة النمذجة uml (modeling language)

المعروف أيضا بلغة النمذجة الموحدة

لغة النمذجة او لغة التعديل الموحدة و التي يمكن ترجمتها إلى لغة نمذجة موحدة و هي علامة يمكن استخدامها لنمذجة مشكلة بطريقة قياسية .

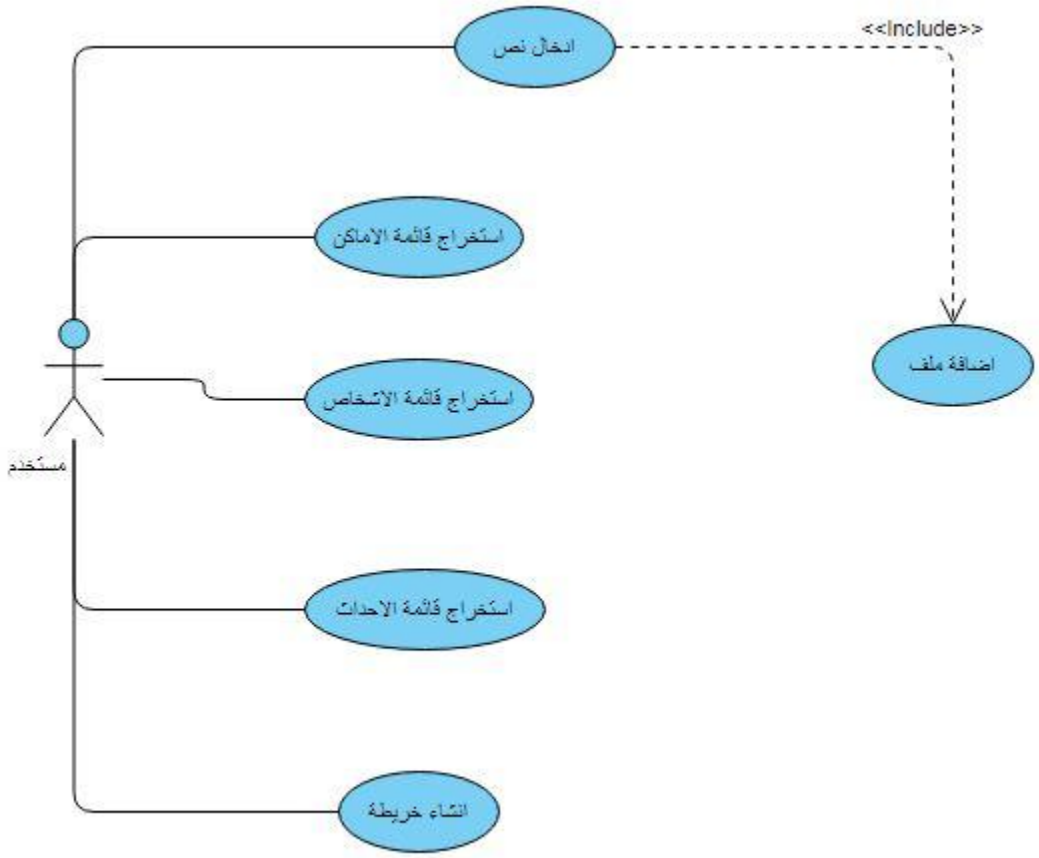
ولدت هذه اللغة من مزيج من العديد من الأساليب الموجودة سابقا و أصبحت المرجع من حيث نمذجة الكائنات. لتحديد وظائف هذا التطبيق اخترنا uml الذي يحتوي قواعد و أسس ثابتة و موحدة .

(Unified) Uml

2- مخطط الحالة (diagramme de cas d’ utilisation):

مخطط حالة الاستخدام هو ابسط عرض لتفاعلات المستخدم على شكل دوائر و اشكال بيضاوية , وهو يعرض العلاقات بين المستخدم و حالات استخدام المختلفة التي تشارك فيها , ويمكن لمخطط الاستخدام ايضا ان يعرض انواع المستخدمين و حالات الاستخدام المختلفة .

حيث يوضح المخطط الصلاحيات الممكنة للمستخدم النظام حيث يقوم بإدخال نص مما يصبح لديه صلاحية بالقيام بجميع العمليات الممكنة داخل النظام من إضافة وبحث وغيرها .



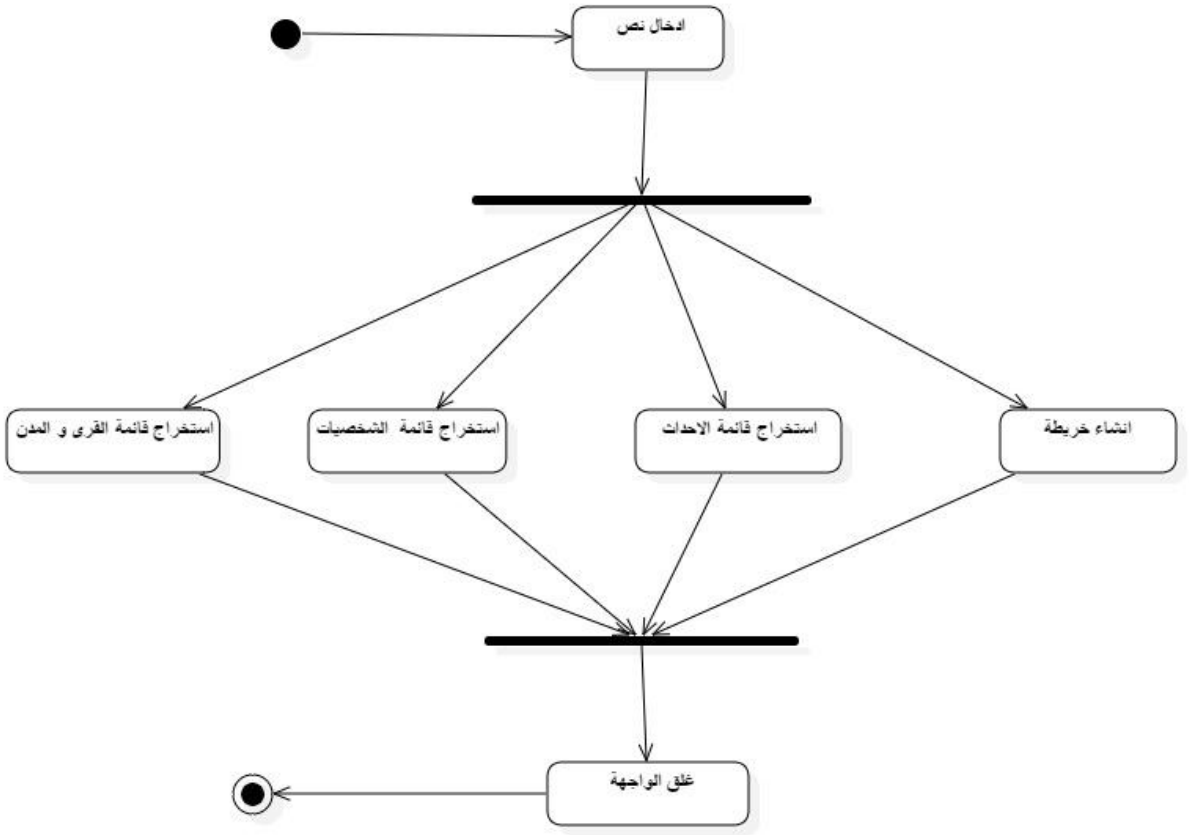
الشكل-1- مخطط الحالة

3- مخطط النشاط (diagrammed Activity):

هي عبارة عن مخططات تصف سلوك سير العمل للنظام وهذا من خلال إظهار تسلسل الأنشطة التي تم القيام بها ،مخطط النشاط " activitydiagramd" يمكن أن تظهر الأنشطة المشروطةconditional" " أو المتوازيةparalle "."

مخططات النشاط تصف سلوك عمل النظام من خلال اظهار تسلسل انشطة التي تم القيام بها.

في البداية يحتاج المستخدم الى ادخال النص لنظام حيث يتمكن من القيام بجميع العمليات المذكورة .

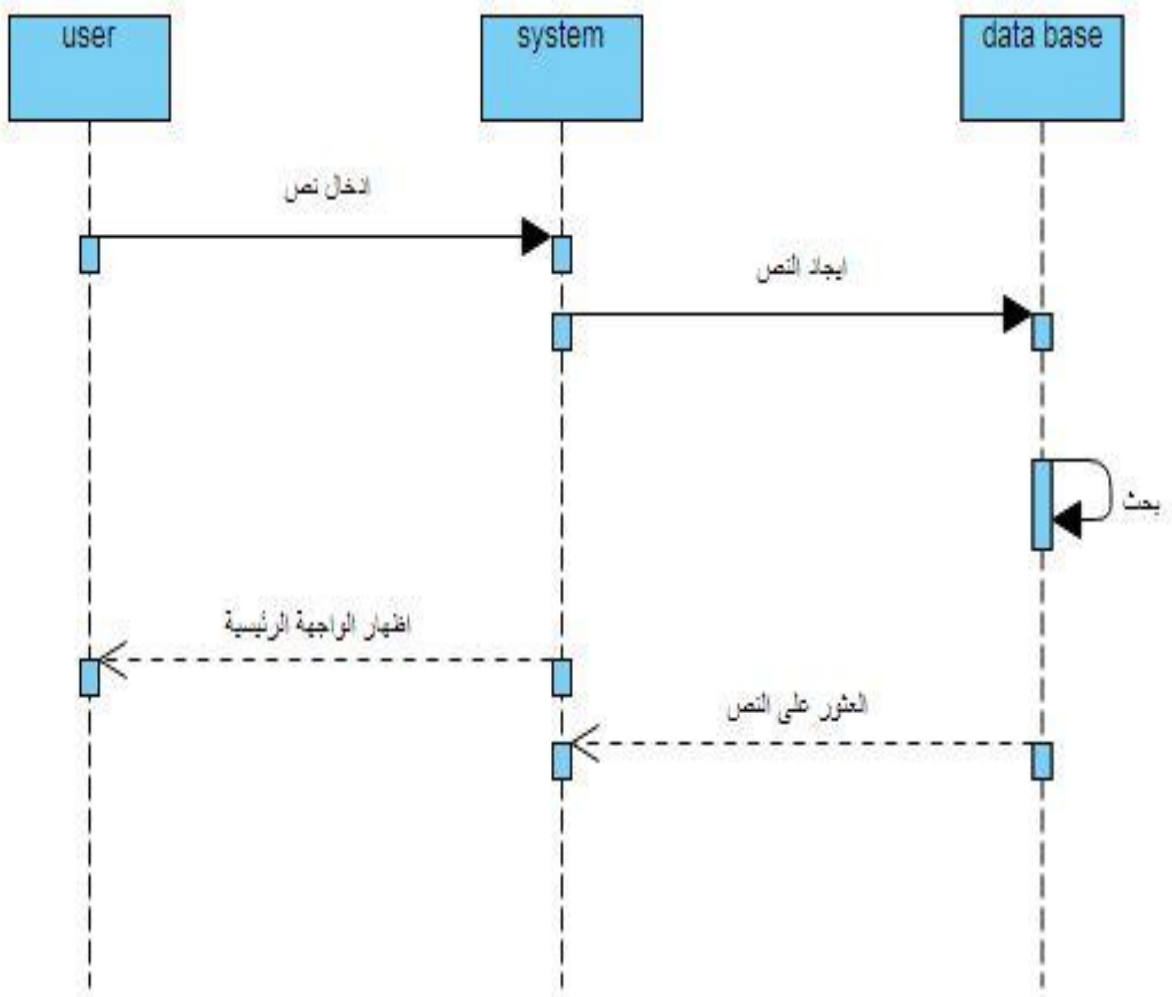


الشكل -2- مخطط النشاط

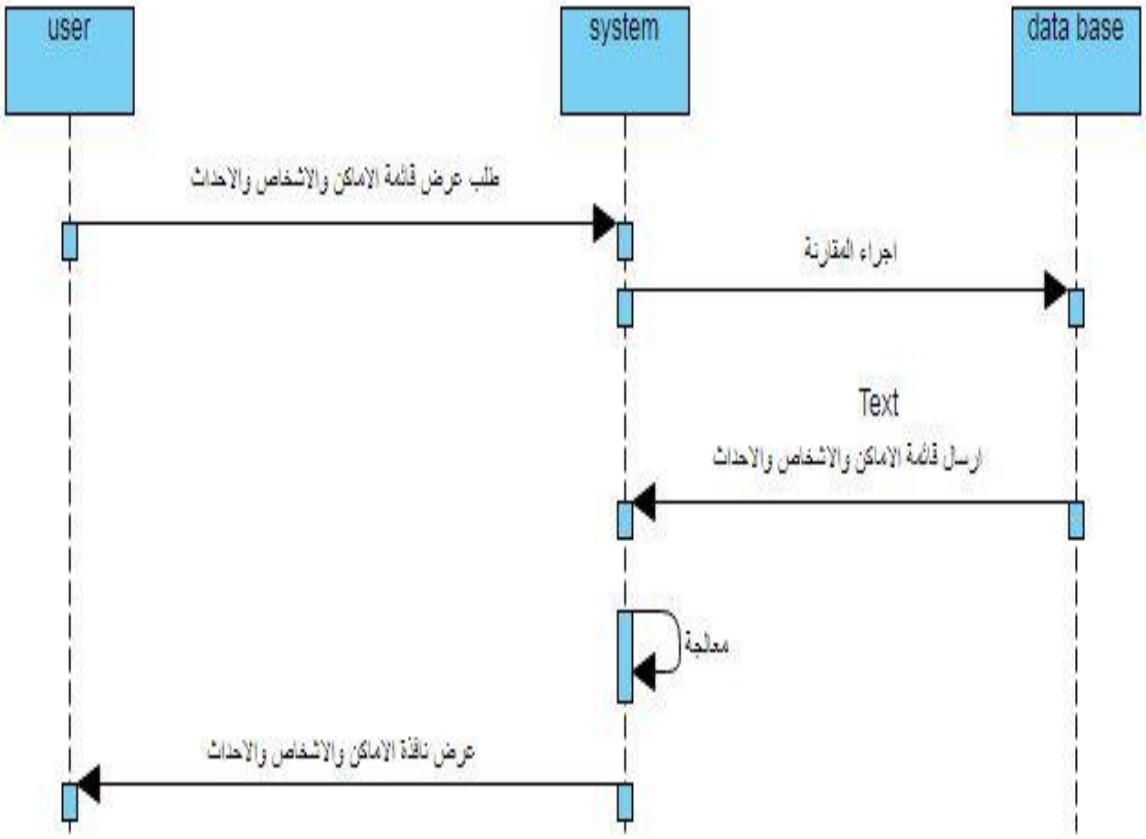
4- مخطط التسلسل (Diagramme séquence)

هو التمثيل الرسومي للتفاعلات بين الجهات الفاعلة و النظام حسب الترتيب الزمني في صيغة لغة النمذجة الموحدة

يمثل المخطط مصادقة المستخدم فيما يتعلق بالنظام بحيث يقوم المستخدم بإدخال نص فيتم الدخول إلى الواجهة الرئيسية .



الشكل -3- مخطط التسلسل



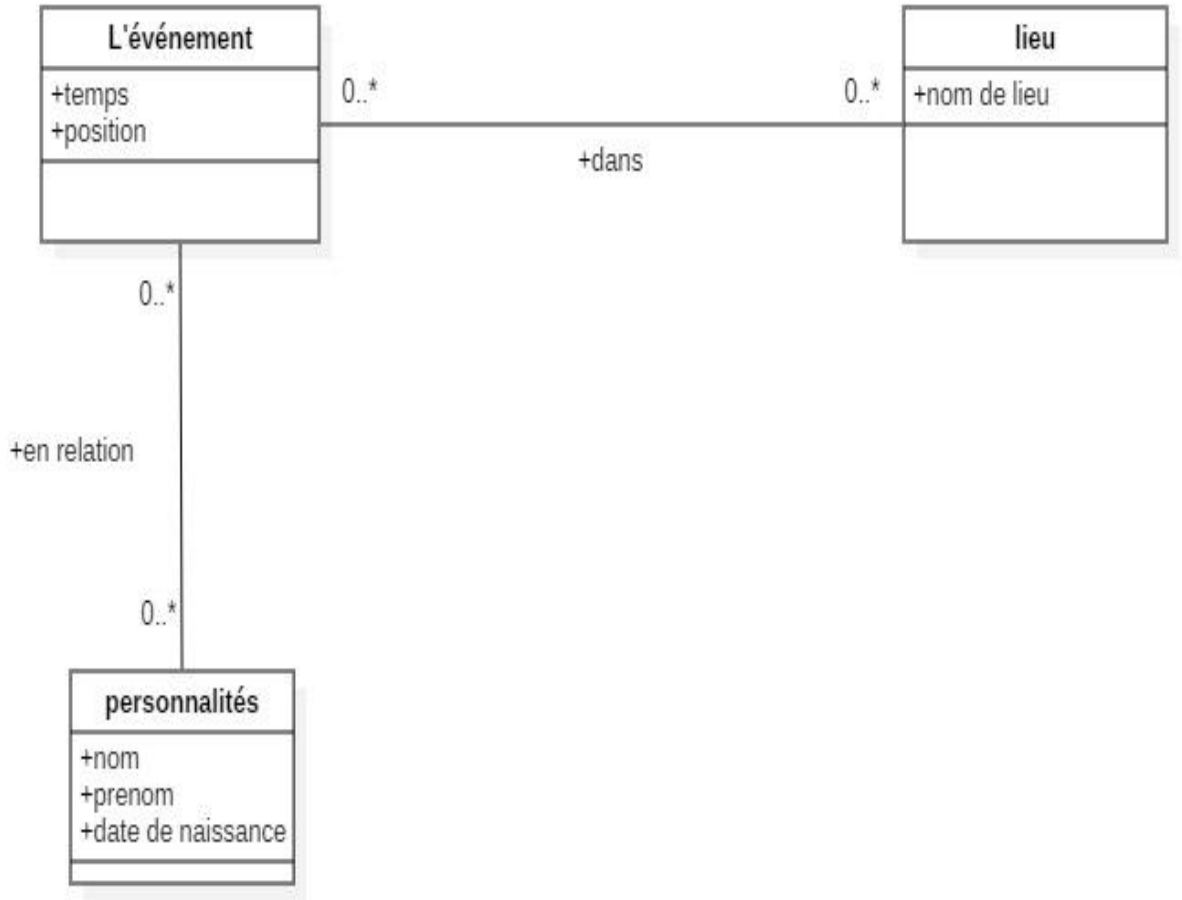
الشكل -4- مخطط إستخراج الأماكن و الأشخاص و الأحداث

في مخطط التالي يقوم المستخدم بطلب عرض قائمة الأماكن و الأشخاص و الأحداث فيتم إجراء مقارنة بين النص المدخل و قائمة مراجع الموجودة في قاعدة البيانات تتم المعالجة و عرض نافذة المكان و الشخص و الحدث.

5- مخطط الفئات(digramme de classe)

هو جزء مهم جدا من لغة النمذجة الموحدة و هو مخطط هيكلي ل uml مهمته عرض الفئات بنظام معين مع جميع العلاقات التى تربط بينهما .

في هذا المخطط قدمنا صورة كاملة عن جميع الفئات المؤثرة في النظام بحيث وضعنا العلاقات بين مختلف هذه الفئات.



الشكل -5- مخطط الفئات

خلاصة :

لقد تطرقنا في هذا الفصل إلى تقديم مختصر و كافي لشرح هذا النظام ، وهذا بعد إنجاز المخططات اللازمة ومن خلال هذا الأخير لقد توصلنا إلى تصور كافي عن هذا النظام الجاري إنجازه ، وبعدما توصلنا إلى ما نحن عليه الآن من التقدم بقي لنا القيام بالدراسة التقنية والتي من خلالها نصل إلى هدفنا .

و في الختام يمكن القول بان uml تقدم لنا نموذجا يمكننا من خلاله التخطيط لكيف سيتم تجهيز لبرنامجنا ..

الفصل الثالث:الإنجاز

الفصل الثالث : الإنجاز

المقدمة:

لقد أصبحت لغات البرمجة عالية المستوى قريبة جدا من لغة الإنسان، بحيث وصل تطورها مع مرور الأيام إلى جعل أغلبها سريعة وقوية وبسيطة خاصة المتبنية من قبل شركات كبيرة كقوقل وفيس بوك وغيرها في نفس السياق، حيث يسعى جميع مطوري اللغات إلى التحسين وتطوير التقنيات من كل هذا يحتار مطور البرمجيات في اختيار لغة مناسبة له ولمشاريعه ومن هنا حاولنا في هاذا المشروع البحث عن أبرزهم وأهم التقنيات الحديثة الامنة التي تشهد تطور ملحوظ والتي وظفناها في هذا المشروع و وضعنا لها شرح مبسط في هذا الفصل، كما سوف نتطرق إلى عرض الواجهات الرسومية لهذا النظام

1- نتبينز (Net Beans) :

هي بيئة مجانية ومفتوحة المصدر، يقوم باستخدامها الكثير من المطورين ومع مجتمع متنامي والكثير من الشركاء العالميين ، وعادة ما تستخدم لتطوير البرامج من قبل المطورين ،وهي أداة للمبرمجين لكتابة وبناء واختيار ومتابعة الأخطاء وتجهيز البرامج للعمل ومن الأمثلة على اللغات التي تدعمها البيئة :

(JAVA;HTML;Fortan;C++;C)

2- لغة الجافا (JAVA) :

الجافا هي لغة برمجة عالية المستوى ، قام بتطويرها جيمس جوسلينغ من شركة (Sun Micro Systems)

اثناء محاولته تطور لغة البرمجة C++ عام 1995 , وكانت تسمى (OAK), لكن تم تعديلها واطلق عليها اسم جافا(JAVA) , وهي لغة تعتمد على الكائنات

3- قواعد البيانات:

لقد إعتمدنا في نظامنا هذا على قاعدة بيانات مفتوحة المصدر SQL Server وهو نظام إدارة قواعد البيانات متسلسل.

الكلمات المفتاحية:

Code	Translation
PK	PRIMARY KEY

الجدول : 01 :جدول الكلمات المفتاحية

3-1 جدول المستخدمين :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id		int	PK
Villes	المدينة	Varchar 50	
Villages	القرية	Varchar 50	
Personnalites	الشخصيات	Varchar 50	

الجدول : 02 المستخدمين

3-2 جدول الاحداث :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id_evenement		Int	PK
Name_eve	اسم الحدث	Varchar 50	
Date	التاريخ	Date	

الجدول : 03 الأحداث

3-3 جدول المكان :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id_lieu		int	PK
Ville	المدينة	varchar 50	
Village	القرية	varchar 50	

الجدول : 04 جدول المكان

4-3 جدول الشخصيات :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id_personnalites		int	PK
Prenom 1	الاسم الأول	Varchar 50	
Prenom 2	الإسم الثاني	Varchar 50	

الجدول : 05 جدول الشخصيات

5-3 جدول الربط الاحداث _ المكان :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id_lieu		int	PK
Id_evenement		int	PK

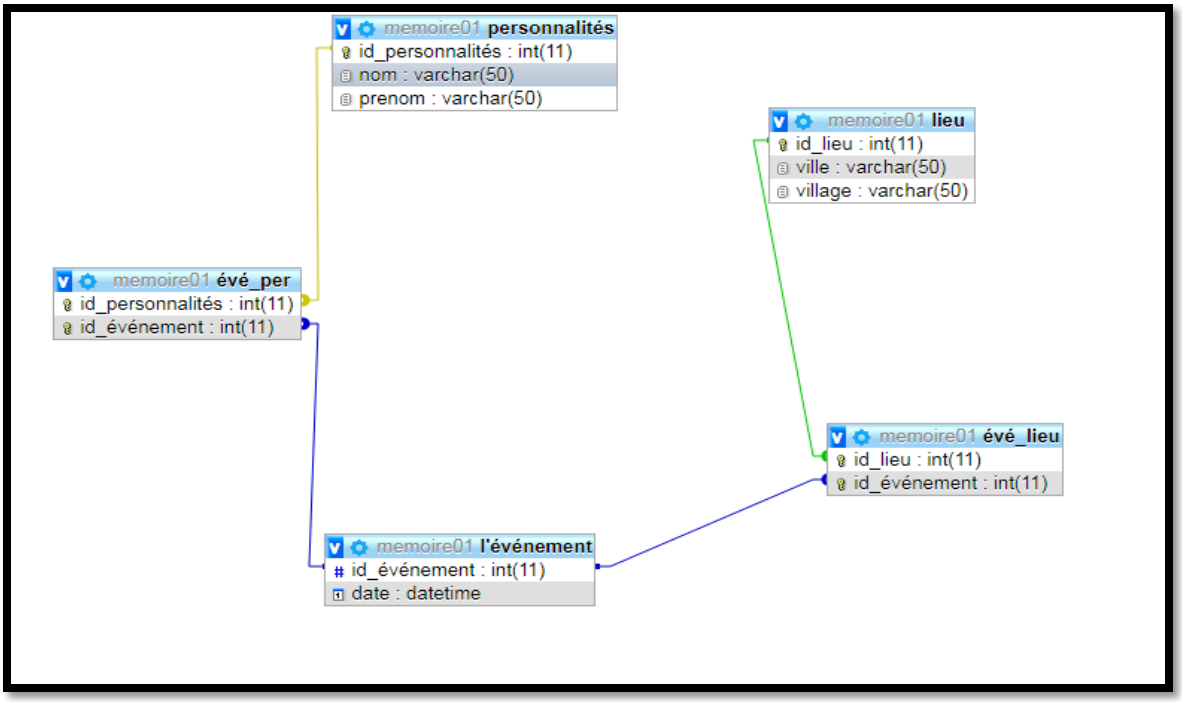
الجدول : 06 جدول ربط الاحداث_المكان

6-3 جدول الربط الحدث _ الشخصيات :

اسم الحقل	وصف الحقل	نوع البيانات	حجم الحقل
Id_personnalites		int	PK
Id_evenement		int	PK

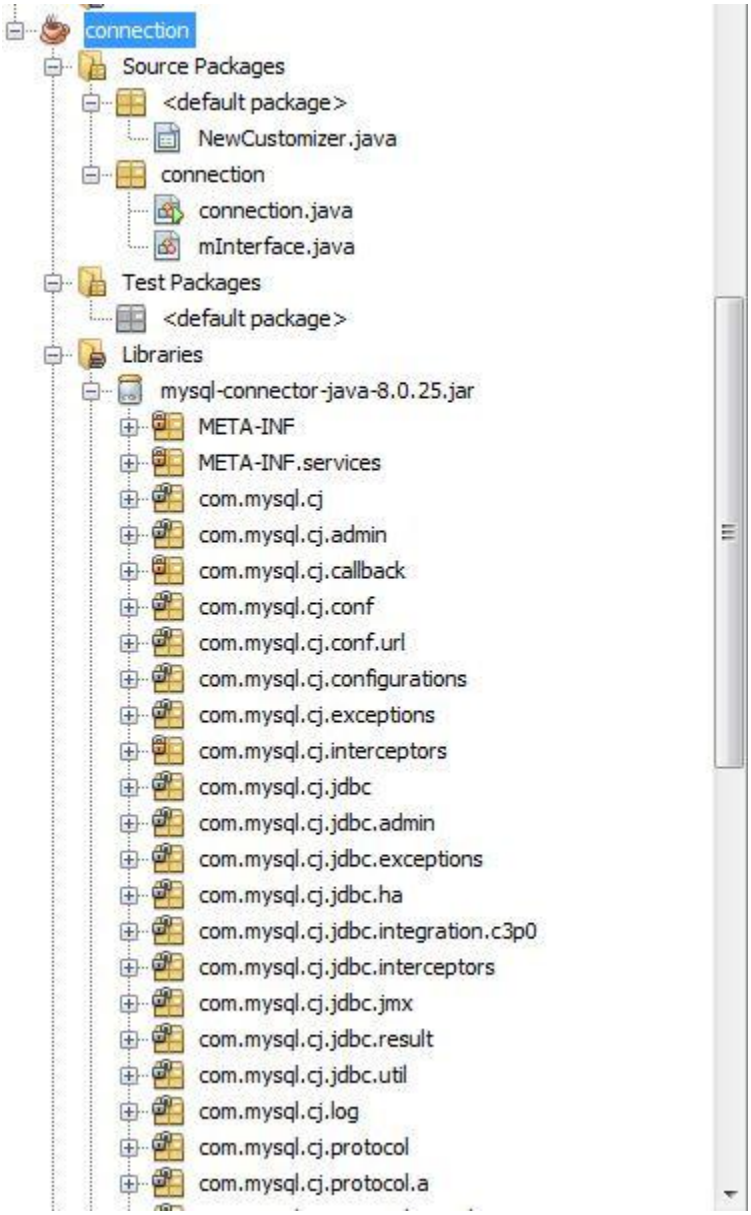
الجدول : 07 جدول ربط

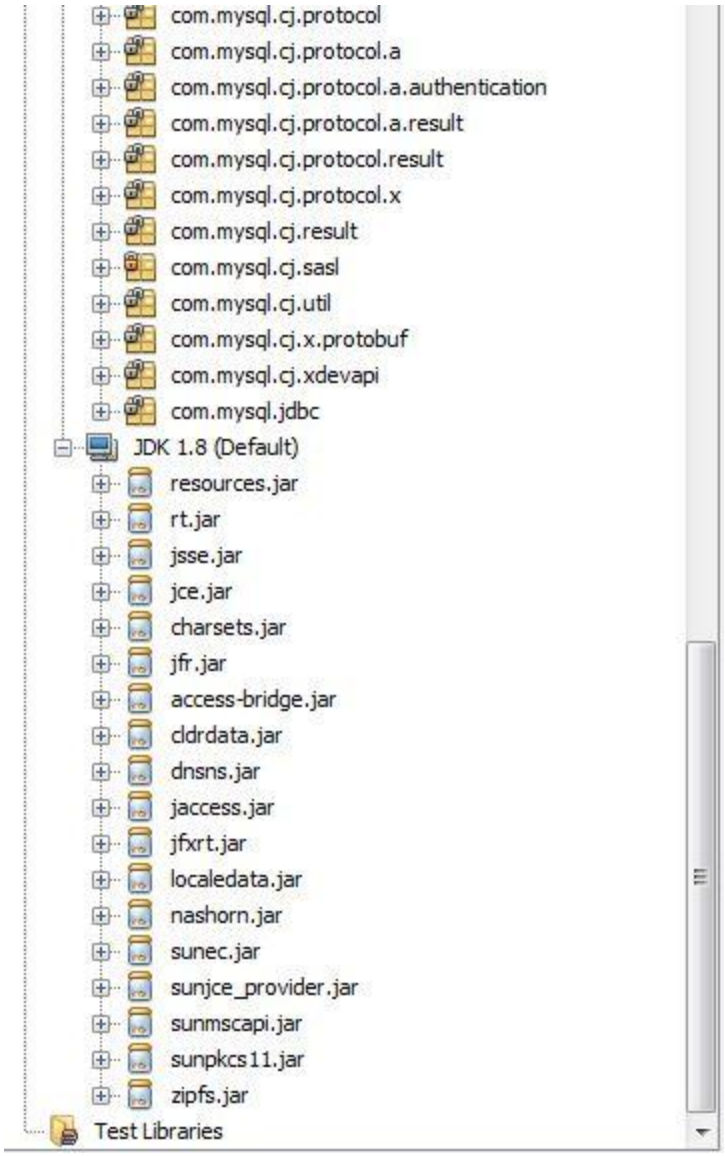
4- قاعدة البيانات المنجزة :



الشكل -6- قاعدة البيانات (الربط)

5- هيكلة الملفات :



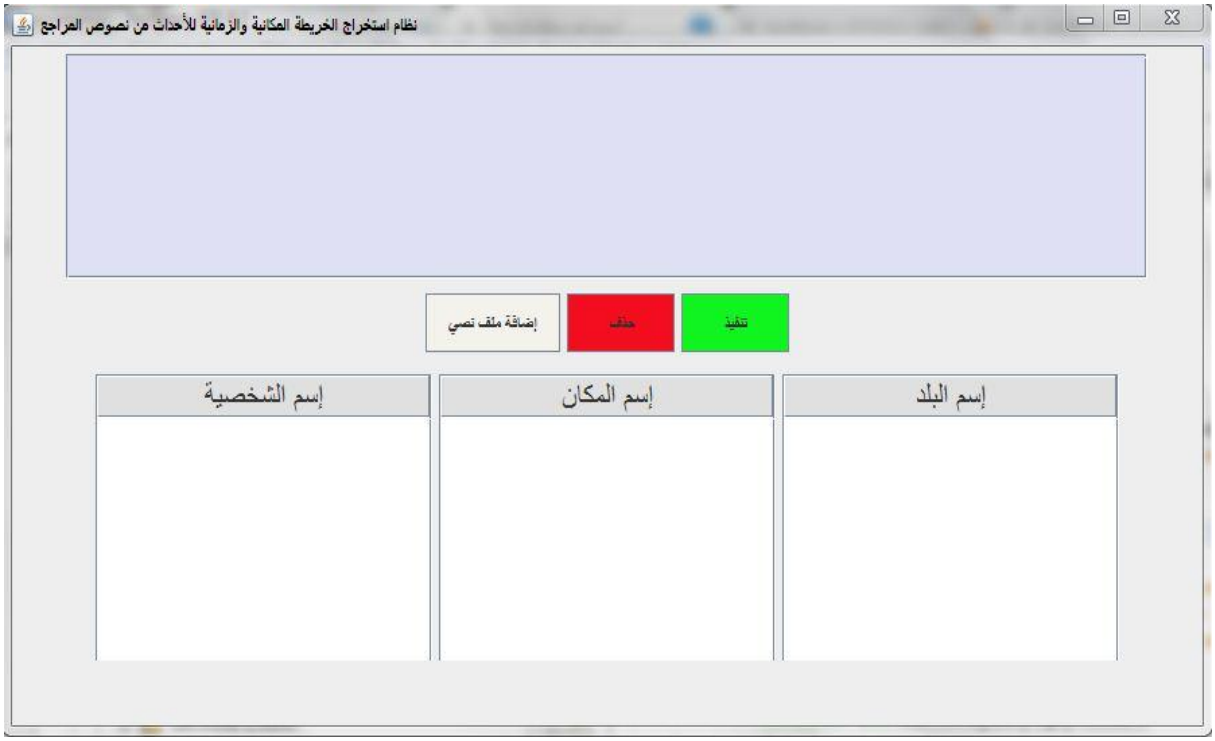


الشكل -7- صورة موجزة للملفات الاساسية

6- الإنجاز:

هذا الجزء مخصص لوصف مرحلة تنفيذ هذا المشروع وسوف نقدم بعض الواجهات لتوضيح الاستخدامات المختلفة للتطبيق بشكل أوضح

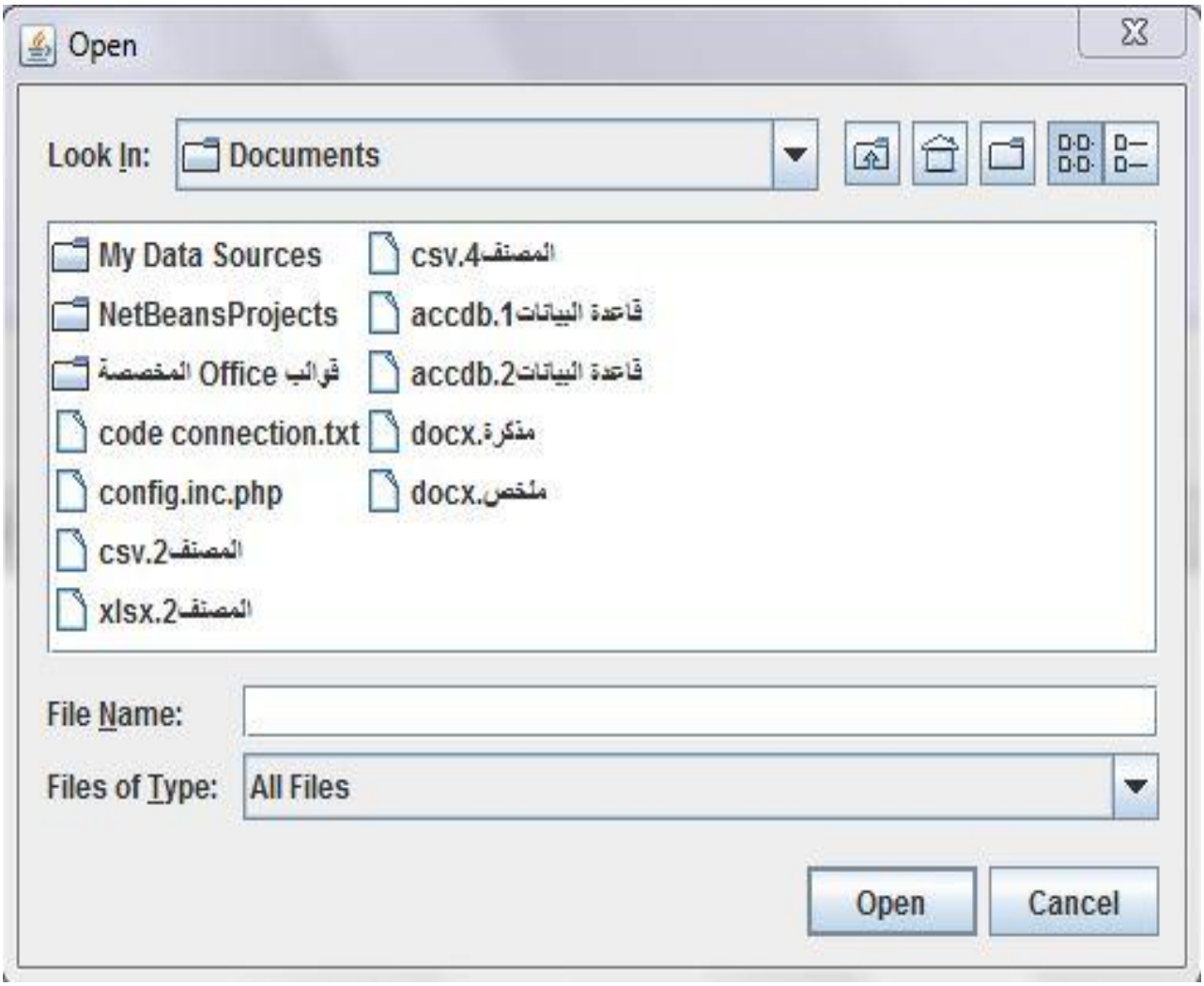
6-1 الواجهة الرئيسية للتطبيق :



الشكل-8- الواجهة الرئيسية للتطبيق

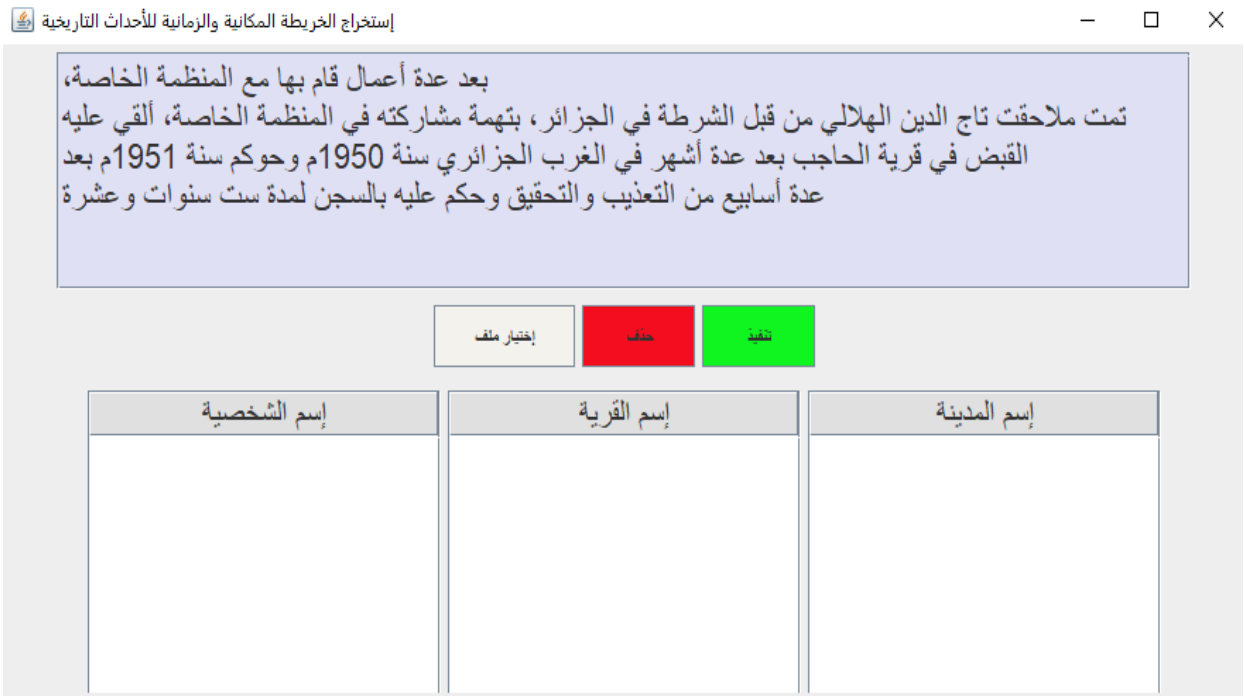
يقوم المستخدم بالضغط على إضافة ملف شخصي ثم يضغط على النص الذي نريد تطبيقه ثم الضغط على زر التنفيذ

3-6 واجهة إضافة ملف :

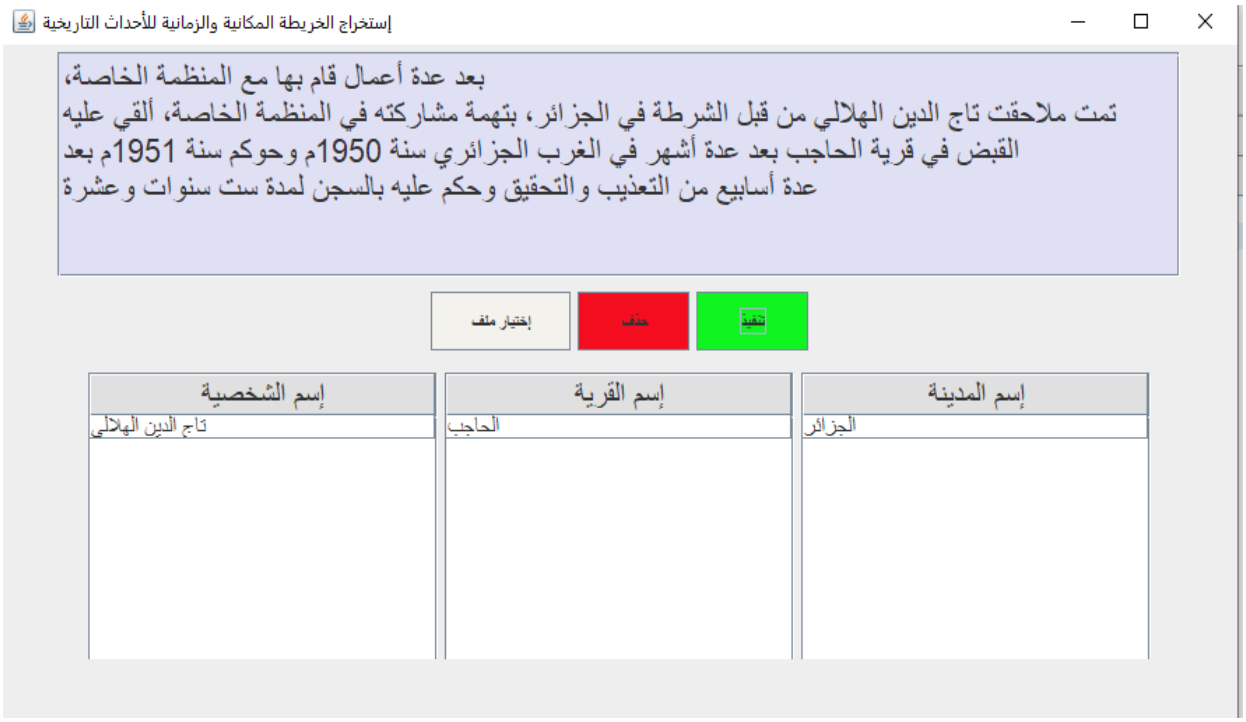


الشكل-9- واجهة إضافة ملف

2-6 واجهة توضح عمل التطبيق :



الشكل -10- إدخال نص للتطبيق



الشكل -11- نتيجة تنفيذ التطبيق

7- قاعدة البيانات :

personnalites	villages	villes
تاج الدين الهلالي	عبان رمضان	الجزائر
فهمي ناجي الامام	المستور	الجزائر
محمد التوحيدي	الحاجب	الجزائر
اسكندر بروتشاي	برج السنوسي	الجزائر
بوجار سباهيو	بلبل	الجزائر
رشاد خليفة	قصبة كاحرار	الجزائر
المنصور الحسين بن القاسم	جوقار	تونس
المحسن بن احمد	سمنجة	تونس

الشكل-12- قاعدة البيانات

8- خاتمة :

لقد توصلنا في نهاية عملنا هذا إلى الشكل النهائي للمشروع ،وهذا بعدما أتممنا إنجاز الواجهات اللازمة والتي تعبر عن الشكل النهائي للمطلوب ،وبعدما توصلنا إلى ما نحن عليه الان فإننا بلغنا الهدف.

خاتمة عامة :

بعد إكمالنا للفترة المقررة قدمنا هذا العمل المتواضع الذي كان خلاصة لمجهوداتنا المكثفة والمتواصلة فتحت أمامنا افاقا كبيرة استطعنا من خلالها إلى فكرة نظام الي مصمم .الهدف الأساسي منه أن يعمل على نظام استخراج الخريطة المكانية والزمانية للأحداث من نصوص المراجع ويتكون هاذا العمل من مرحلتين :
المرحلة النظرية وتشمل الفصلين الأول والثاني والمرحلة العملية والتي تناولناها في الفصل الأخير(الثالث).
وفي الاخير نتمنى من الله عز وجل أن يوفقنا ويسدد خطانا لهذا العمل وان نكون قد أضفنا شيئا ولو بسيطا إلى هذا العالم في حياتنا بالنظر الى سرعة التطور التي يشهدها .كما نام لان يحظى هذا العمل بالإعجاب الذي كنا نطمح إليه.

المصادر:

[1]- دانيا صغير ، (2019/12/01) ،لمحة عامة عن معالجة اللغات الطبيعية ، (2021/03/14) ، رابط الموقع : https://aiinarabic.com/natural-language-processing-is-fun/

[2]-menobot ، (2021/05/15) آخر يوم للتعديل ، قادة دينيون مسلمون ، (2021/02/09) ، رابط الموقع :
https://ar.m.wikipedia.org/wiki/%D8%AA%D8%B5%D9%86%D9%8A%D9%81:%D9%82%D8%B1%D9%89_%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9?fbclid=IwAR1MaZ29-Qyq2SDiQT9m0vq7NfVVmv9Vd0emt459iDHQK9U1fe0SiY_i2Ec

[3] -إسلام ، (2018/06/14) ، قادة جيوش إسلامية ، (2021/02/09) ، رابط الموقع :
https://ar.m.wikipedia.org/wiki/%D8%AA%D8%B5%D9%86%D9%8A%D9%81:%D9%82%D8%A7%D8%AF%D8%A9_%D8%AC%D9%8A%D9%88%D8%B4_%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9?fbclid=IwAR3YAZhR925VILm-lvw80y79rBBrGYgiqDzcXLf0QIJEk0s4OeecYsrKm0U