أستاذة المادة: نادية زيد الخير

المقياس: اللغة العربية والحاسوب.

المستوى: السنة الثانية ماستر

التخصص: لسانيات تطبيقية (ف1+ف2).

## عنوان المحور: الخوارزميات ورسم الحرف العربي

### أولا: مفاهيم عامة حول اللسانيات الحاسوبية:

من المعلوم أن الحاسوب لا يمكنه فهم اللغات الطبيعية (لغات البشر)، كما أن الإنسان لا يستطيع فهم لغة الحاسوب (اللغة الرقمية) نظرا لما يقتضيه من حفظ ودقة في كتابة سلسلة طويلة من الأرقام الثنائية (0.1) بترتيب معين، مما يودي به إلى ارتكاب أخطاء في البرنامج، وبالتالي فقد تطلّب الأمر وجود **لغات** تعمل كوسيط بين الإنسان والحاسوب، تسمى **لغات البرمجة (progamming languages).**

**أ- ما معنى البرمجة، وما هي لغة البرمجة؟**

البرمجة هي عملية كتابة تعليمات (instuctions) وأوامر لجهاز الحاسوب باستخدام لغة برمجة لتوجيهه وإعلامه بكيفية التعامل مع البيانات، وطريقة تنفيذ سلسلة من الأعمال المطلوبة.

هذه العملية تقتضي وجود مجموعة من البرمجيات التي تعمل كوسيط بين المستخدم والحاسوب، حيث يستخدمها المبرمجون لكتابة تعليمات تترجم باستعمال برمجيات خاصة إلى لغة يفهمها الحاسوب.

**ب- أنواع لغات البرمجة:**

تصنّف لغات البرمجة إلى أنواع حسب سهولة التعامل معها من قبل الإنسان وهي:

1–**لغات منخفضة المستوى (Low-level programming languages):** وهي اللغة الوحيدة التي يفهمها الحاسوب، وتتكون من سلسلة من الأرقام الثنائية (0.1).

**2–لغات متوسطة المستوى (Medium- level programming languages):** وهي لغات قريبة من لغة الآلة، إلا أنها أكثر فهما وأسهل استخداما بالنسبة للإنسان. وتتمثل في لغة التجميع (Assembly language).

**3–اللغات عالية المستوى (High-level programming languages):** وهي لغات قريبة نسبيا إلى لغة الإنسان، منها: لغة الفيجوال، البيسك، الباسكال، الفورتران، السي++...إلخ.

## ثانيا: مفهوم اللسانيات الحاسوبية (Computational Linguistics):

هي إطار تقني تنصهر فيه تجليات اللغة الطبيعية وتمظهراتها في تفاعل يسعى إلى ربط اللغة بالحاسوب وأنظمته.

وتعرّف بأنها: "الفرع الذي يبحث في مختلف طرق وتقنيات الاستفادة من قدرات الحاسوب في تحليل اللغة ومعالجتها وتعليمها وتعلّمها".

إذا، فاللسانيات الحاسوبية علم بينيّ يتعلّق الشق الأول منه باللغة (وهي النواة)، أما الثاني فيتصل بما هو حاسوبي (وهو المحدِّد)، وكل ما يتعلق بالتقانات الصناعية والذكاء الاصطناعي...

## ثالثا: جوانب الدراسة اللسانية الحاسوبية:

ينبني أساس الدراسة العلمية لإطار اللسانيات الحاسوبية على جانبين هما: النظري والتطبيقي.

**الجانب الأول** منها يبحث في حيثيات دراسة هذا الإطار العلمي، فيتيح لنا إمكانية افتراض كيفية عمل الدماغ الإلكتروني لحل بعض المشكلات اللغوية، كما أنه يسعى إلى تحصيل الحاسوب بكفاية لغوية تشبه ما يكون للإنسان حين يستقبل اللغة ويدركها ثم يعيد إنتاجها وفق المطلوب.

**أما الجانب الثاني** فيعنى بكل العمليات الخوارزمية، سواء كانت رياضية أو رمزية (مجموعة من القواعد التي تنطلق من البسيط إلى المعقد إلى الأعقد، أي أنها مسألة تقنية).

## رابعا: بعض أهداف اللسانيات الحاسوبية:

هدفها محاولة تلقين الحاسوب لغة بشرية تمكنه من التحاور مع مستخدميه شفهيا وكتابيا، حتى يصبح قادرا على فهم الكلام وإنتاجه وتوليده (مجاراة التفكير والإنجاز البشريين).

الدرس الثاني

المحور : الخوارزميات ورسم الحرف العربي:

### أولا: التحدّيات التي تواجه معالجة اللغة العربية والتعرُّف عليها آليا:

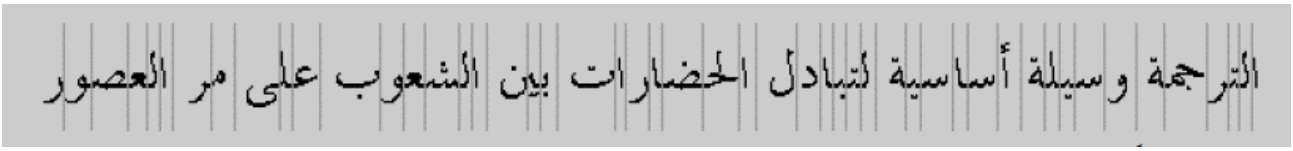
تمتلك اللغة العربية الكثير من السمات والخصائص التي تميزها عن باقي اللغات، ومع كون هذه السمات تمثّل إحدى عناصر قوّة لها، إلا أنها تمثل في الوقت ذاته نقاط تحدٍّ إذا ما رمنا حوسبتها.

### 1. الكتابة/الخِطاطة (paléographie) العربية:

أ- **اتّصال حروفها:** الكتابة العربية لا تُفهم إلا إذا كانت حروفها متصلة، غير أن هناك بعض الحروف التي لا تقبل الاتصال، إذ بينها وبين متجاوراتها من الحروف في الكلمة الواحدة مسافة قصيرة نسبيا. مثل: **قرأت – مسؤولية – تقارب – كبيرة.**

**بخلاف اللغات اللاتينية التي تكتب بحروف منفصلة / أو متصلة عند الكتابة باليد.**

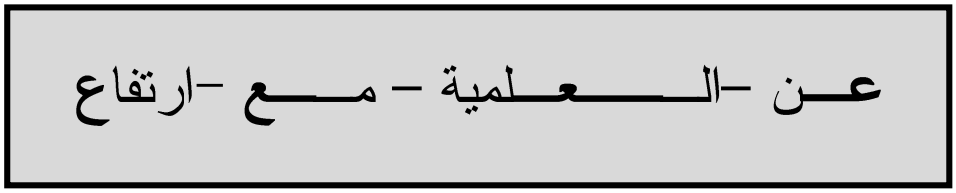
**ومن وجهة النظر الحاسوبية، فإن التعرُّف على الأنماط الرسومية/ الخطية وهي منفصلة أيسر من التعرف عليها وهي متصلة، حيث يتوجّب في هذه الحال (الاتصال) مسألة تعيين حدود كل رمز خطي (Grapheme)، باعتماد آليتين هما: التقطيع (segmentation) والتعرُّف (recognition) على الحرف الذي يرمز إليه كل جرافيم. مثل النموذج أدناه:**




ب- **التداخل بين حدود الجرافيمات:** حيث نلاحظ في بعض الأحيان التداخل/التراكب الطفيف بين حدود الصور الخطّية للحروف في بعض الخطوط العربية وهو ما يصعّب من معالجتها والتعرّف عليها آليا. مثل:

ت- **تغيُّر رسم الحرف بتغير موقعه في البنية:** وذلك نتيجة للكتابة المتّصلة الحروف، حيث يؤدي إلى زيادة كبيرة في عدد الرموز الرسومية/الصور الخطية للحرف الواحد، التي يستلزم التعامل معها للتعرف الآلي عليها مقارنة باللغات اللاتينية الأخرى.

فمن وجهة النظر الحاسوبية فمسألة التعرف على الحروف/ الأنماط الرسومية يرتبط بعددها، فكلما قلّت كان التعرف عليها أيسر وأسهل وكلما كثرت ازدادت صعوبة التعرف عليها آليا. مثل:



ث- **اللغة العربية لا تميز في كتابتها بين أسماء الأعلام والأشخاص...** ولا عند بدايات الجمل بحروف كبيرة كما الحال في اللغات اللاتينية التي تميز بينها على أساس الرسم بـ (الحروف الاستهلالية Capital letters)، مثل: زرت **كـــــندا (I have visited Canada)**. فهذه العلامة في الكتابة ترفع نسبة التعرف عليها آليا.

ج- **الجرافيمات المركبة من حرفين فأكثر:** حيث تحتوي العربية على الكثير من الصور الخطية التي تتركب من حرفين أو ثلاثة مثل:

| |
|---|
| مـــ ، لجـــ ، لا ، لمحـــ |

فهذه الجرافيمات تزيد من عدد الأنماط الشكلية التي يتوجّب على النظام التعرف عليها مما يزيد في درجة الصعوبة.

ح-**النّقط:** فكلما كانت الفروق الشكلية بين الأنماط الرسومية واسعة، كلما ارتفعت نسبة التعرف عليها آليا، وكون جرافيمات الخِطاطة العربية متشابهة شكليا إلى حدّ بعيد ولا تتمايز إلا بوجود/ غياب النقاط، فإن ذلك يرفع التحدي أمام النظام للتعرف عليها آليا. مثل:

| |
|---|
| بـــ ، نـــ ، تـــ/ ســـ ، شـــ / حـــ ، خـــ ، جـــ |

**خ-علامات الضّبط الصوتيّ / التّشكيل:**

تعدّ هذه الميزة إحدى أوجه التعقيد الإضافية للخِطاطة العربية أمام النظام للتعرف الآلي على الحرف؛ لأنها لا تقع في سياق تسلسل خطيّ/ أفقي مثل تسلسل الجرافيمات، وإنما تقع في مواضع رأسية إما فوق الحرف أو تحته. مثل:

# اللِّسَانِيَاتُ الْحَاسُوبِيَّةُ عِلْمٌ بَيْنِيٌّ

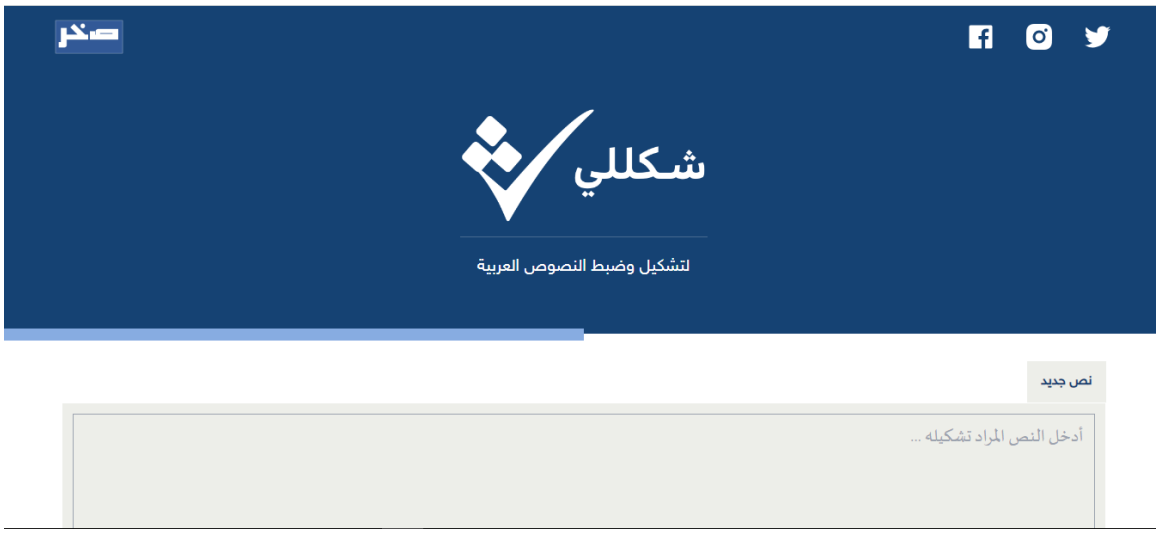
وفي بعض الأحيان قد يغيب تشكيل بنية الكلمة فتزداد نسبة اللّبس في التعرف على بنية الكلمة وبذلك تحتمل قراءات متعددة، ومنه تتعدد أوجه دلالتها، مثل كلمة: (وجد)، التي تحتمل أن تكون: [وَجَدَ / وُجِدَ/ وَ+جِدّ / وَ+جَدّ...].

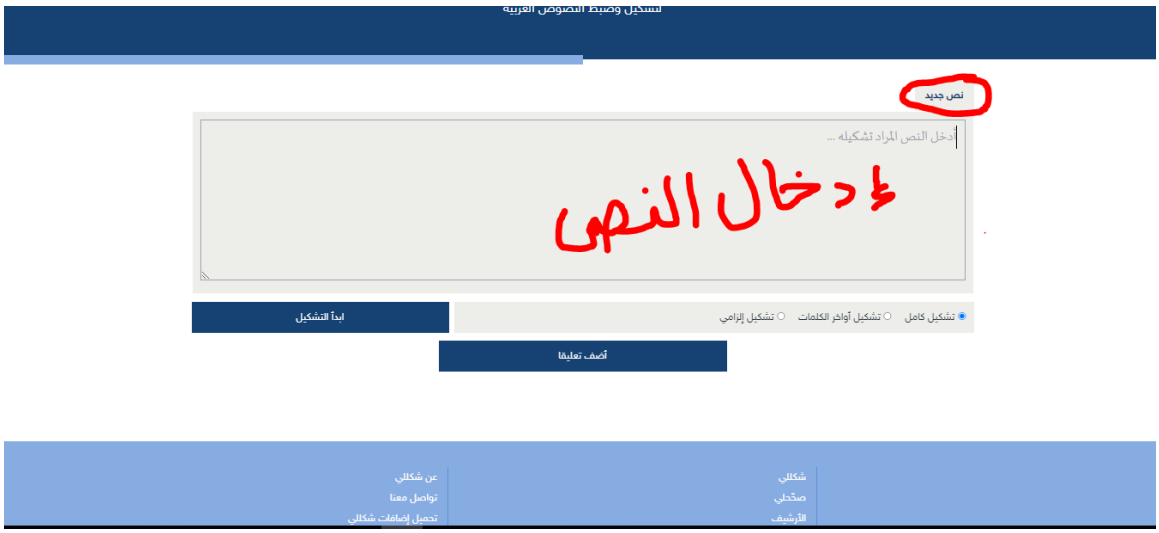
إذاً، فمشكلة غياب التشكيل تمثّل عقبة أساسية أمام حوسبة ومعالجة النصوص العربية آليا، وقد قدّمت شركة صخر برنامجا للتشكيل الآلي (automatic diacritization) للنصوص العربية.

**خامسا: بعض النماذج التطبيقية للبرنامج التشكيل والتصحيح للنص العربي (موقع صخر):**

*أولا تدخل لموقع الشركة **(alsharekh.org)** عبر محرك البحث Google، فيفتح الموقع كما في الصورة:

*نضغط على أيقونة (شكلي) تفتح معنا النافذة كما في الصورة أدناه:






*نضغط على أيقونة (**نص جديد**) ثم نتجه للمربع أسفلها (**أدخل النص المراد تشكيله**)، ثم نختار في الأسفل نوع التشكيل (تشكيل كامل/ تشكيل أواخر الكلمات/ تشكيل إلزامي)، بعدها نضغط على أيقونة (ابدأ التشكيل).

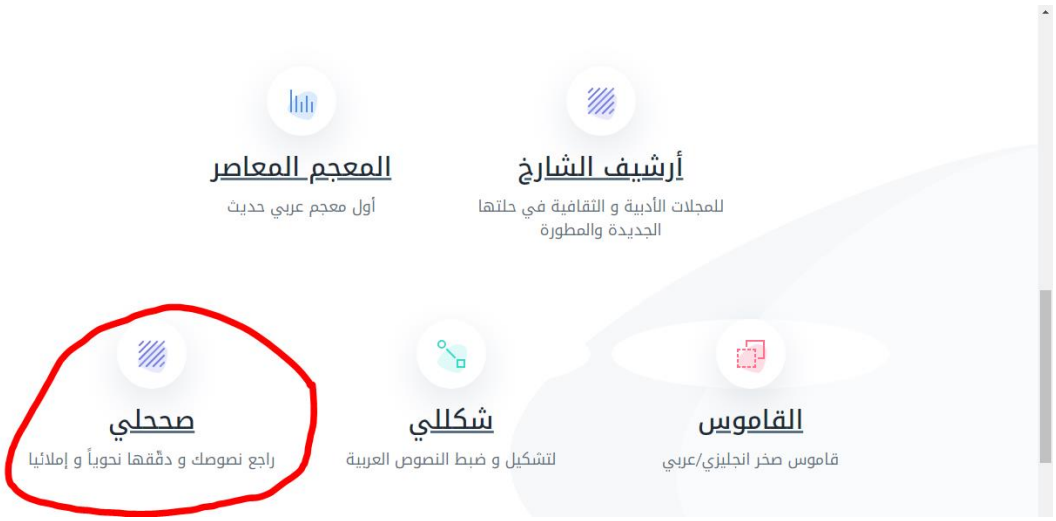
مثال توضيحي:

فبعد اختيار (تشكيل كامل) نضغط على (ابدأ التشكيل) فنحصل على النص مشكّل آليا كما يلي:

كما يتيح هذا البرنامج خاصية (النسخ) و (الطباعة) بعد التشكيل الآلي للنص المُدخل.

وقد قامت شركة صخر بتطوير برنامج للتصحيح الآلي للنصوص العربية (صحلي)، ويمكن تطبيق ذلك كما يلي:

*ندخل للموقع (**alsharekh.org**) ونضغط على أيقونة (صحلي) كما يلي:




*ثم نقوم بإدخال النص:

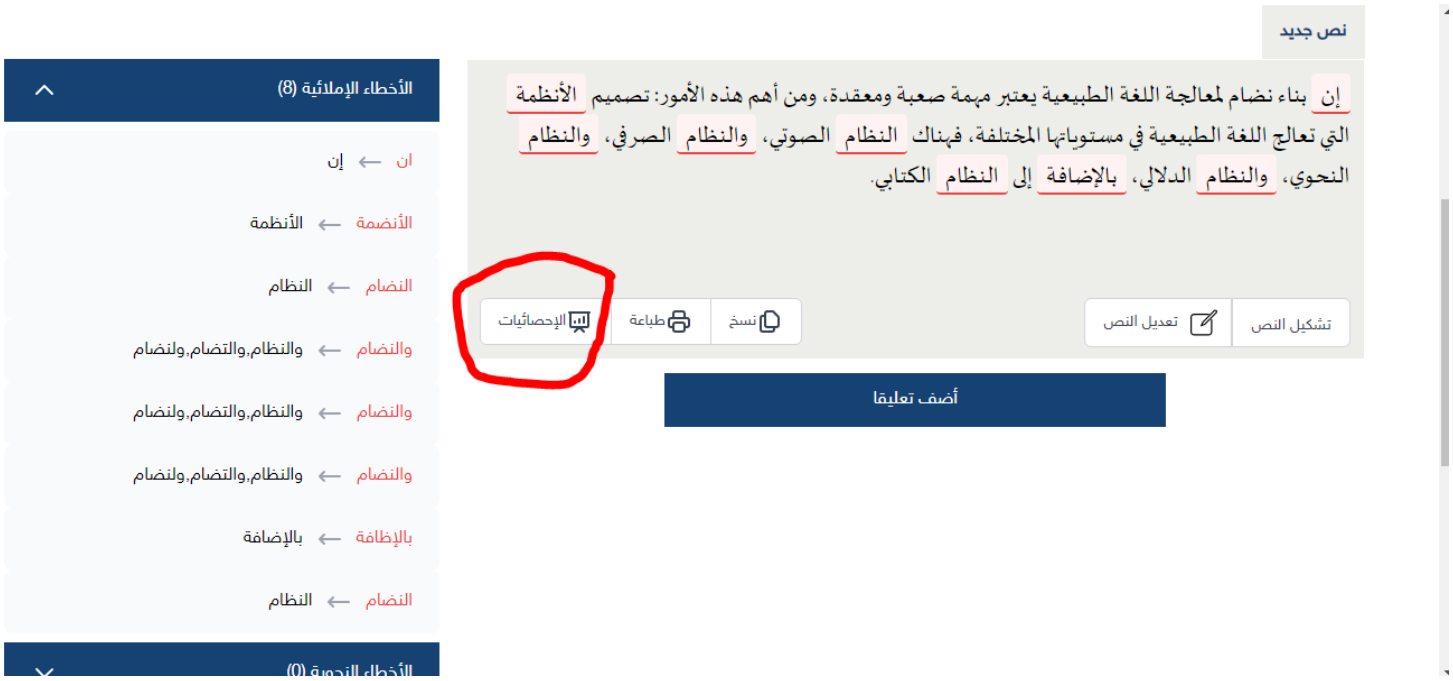
ونختار إما (تدقيق إملائي) أو (تدقيق إملائي ونحوي)، بعدها نضغط على أيقونة (ابدأ التصحيح)، فنجد في المخرجات إحصاء الإملائية، والنحوية، والشائعة منها مع التنبية على الكلمة الخاطئة في النص. كما يظهر في الصورة:

النتيجة:

عندما نضغط على أيقونة (الأخطاء الإملائية) نحصل على:

حيث نلاحظ أنه يقابل بين الكلمة الخاطئة وتصحيحها.

ويمكننا الاطلاع كذلك على الإحصائيات من خلال الأيقونة (الإحصائيات) كما يلي:




لنحصل على المعلومات المبينة في الصورة أدناه:

## قائمة المصادر والمراجع:

1–حميدي بن يوسف، مفاهيم وتطبيقات في اللسانيات الحاسوبية، مركز الكتاب الأكاديمي، عمان، الأردن، ط1، 2019م.

2–دخيسي لطيفة، من التعرف البصري على الحرف العربي إلى تصنيف النص، ضمن كتاب "اللغة العربية والتقانات الحديثة"، ج1، منشورات المجلس الأعلى للغة العربية، الجزائر، 2018م.

3– رضوان تلمساني، التعرف الآلي والآني على أحرف اللغة العربية، ضمن كتاب "اللغة العربية وبرامج الذكاء الاصطناعي الواقع والرهانات"، منشورات المجلس الأعلى للغة العربية، الجزائر، 2019.

4– نبيل علي، العرب وعصر المعلومات، سلسلة عالم المعرفة، المجلس الوطني للثقافة والفنون والآداب، الكويت، أبريل 1994م.

**أستاذة المادة: نادية زيد الخير**

**المقياس: اللغة العربية والحاسوب.**

**المستوى: السنة الثانية ماستر**

**التخصص: لسانيات تطبيقية (ف1+ف2).**

## عنوان المحور: التعرف البصري على الحروف:

## التعرف البصري على الحروف (الحرف العربي المكتوب بخط اليد أنموذجا)

يتم التعرف البصري على الحروف العربية (OCR) من خلال تقنية المسح الضوئي التي تحوّله من نص مكتوب إلى نصّ معالج آليا، والذي يتم ترميزه باستعمال نظام التمثيلية.

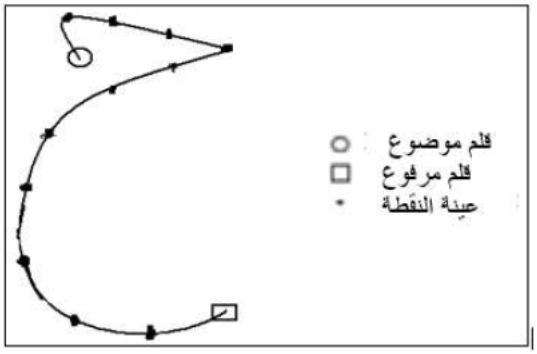
ويمر التعرف البصري باستخدام هذه النظم عبر مراحل:

1-عمليات ما قبل المعالجة: وذلك بتطبيق عزل الكلمات عن طريق تجزئة الكلمات والخطوط وإزالة الانحراف...إلخ.

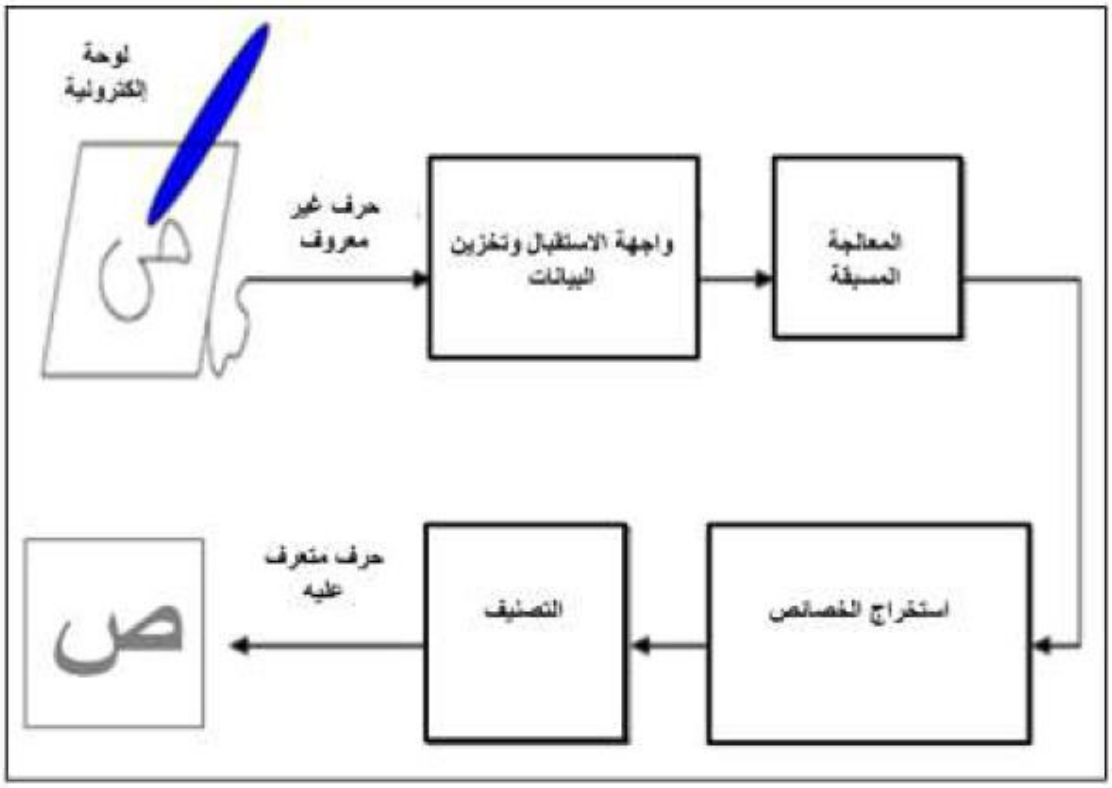
1.1. تجزئة السطور: إن الصور النصية المراد معالجتها/التعرف عليها آليا عبارة عن صفحة ذات أسطر، وبالتالي فإن تطبيق هذه التجزئة تكون من خلال عزل سطر عن سطر آخر.

2.1. تقسيم الكلمات: وذلك من خلال القيام بتجزئة كل سطر إلى كلمات.

2-مرحلة المراجعة: وذلك من خلال ربط السمات الشكلية/الهيكلية للحرف العربي المكتوب كنقطة البداية ونقاط الفروع ونقطة النهاية برقم الحيّز (كما هو مبيّن في الشكل أسفله)، الذي يتم إنشاؤه بأصغر مستطيل يحيط بالكلمة المكتوبة بخط اليد.




وفي هذه المرحلة يتم استخراج السمات والخصائص الشكلية للتعرف على الحرف آليا، كما هو موضّح في الشكل الآتي:




**قائمة المصادر والمراجع:**

1–حميدي بن يوسف، مفاهيم وتطبيقات في اللسانيات الحاسوبية، مركز الكتاب الأكاديمي، عمان، الأردن، ط1، 2019م.

2–دخيسي لطيفة، من التعرف البصري على الحرف العربي إلى تصنيف النص، ضمن كتاب "اللغة العربية والتقانات الحديثة"، ج1، منشورات المجلس الأعلى للغة العربية، الجزائر، 2018م.

3– رضوان تلمساني، التعرف الآلي والآني على أحرف اللغة العربية، ضمن كتاب "اللغة العربية وبرامج الذكاء الاصطناعي الواقع والرهانات"، منشورات المجلس الأعلى للغة العربية، الجزائر، 2019.

4– نبيل علي، العرب وعصر المعلومات، سلسلة عالم المعرفة، المجلس الوطني للثقافة والفنون والآداب، الكويت، أبريل 1994م.