منظومة قراءة آلية لمخطوطات التراث العربى

د.محمد يسرى النحاس
قسم هندسة النظم والحاسبات
كلية الهندسة، جامعة الأزهر

أ.د. محمد يونس الحملاوى
قسم هندسة النظم والحاسبات
كلية الهندسة، جامعة الأزهر

نبذة

يقدم هذا البحث منظومة آلية للتعرف على مخطوطات التراث العربى. تعتمد هذه المنظومة منهج التعلم بدون إشراف للتعرف على نمط الحروف المكتوبة. وقد صممت إجرائية تعرف وتعلم تعمل "باستراتيجية تعلم كلما مضيت" ، حيث تتم عمليتا التعرف والتعلم بصورة مستمرة وديناميكية، أى أنه بدءاً من أساسيات مقترحة لأشكال الحروف يقوم الحاسوب بتحليل الصفحة ثم البحث فى نافذة معزولة من الصفحة عن الحروف وفى هذه الأثناء يقوم بتطوير معرفته لهذه الأساسيات بما يستنبطه من الأساسيات الموجودة فى صورة الحروف . وباستمرار عمليات التحليل والتعرف يتم تطوير منظومة الحروف ويقل معدل أخطاء التعرف، وعند الوصول إلى نسبة عالية من الإستقرار فى التعلم تعيد المنظومة تحقيق نتائجها فى بدايات التعرف. طبقت هذه المنظومة على مجموعة من الحروف العربية فى مخطوطات التراث ، وقد أظهرت نتائج التجارب الأولية بالإضافة إلى إرتفاع معدل التعرف ، ثبوت هذا المعدل بتغير أسلوب الكتابة داخل المخطوطة ذاتها.

١ مقدمة

يعتبر موضوع القراءة الآلية لمخطوطات التراث من الموضوعات التى تشهد نشاطا بحثيا واسعا فى الآونة الأخيرة[٢]،[٣]،[٤]،[٥]،[٦] . ويرجع ذلك إلى الأهمية التى تمثلها هذه المخطوطات والحاجة إليها فى بناء قواعد المعرفة المتكاملة ،ومن ناحية أخرى فإن التطور السريع الذى حدث فى تقنيات المعلومات أتاح للباحثين فى مجال التعرف على الأنماط من تنفيذ الكثير من التطبيقات بنجاح مثل التعرف على الحروف المطبوعة والتعرف على الأصوات،إلخ...

إن معالجة الوثائق والتعرف عليها يحتوى على الكثير من المشاكل التى مازالت قيد البحث والتجريب مثل ترميم صور الوثائق ، والفصل بين الأنواع المختلفة من المعلومات المطبوعة سواء أكانت حروفا أو رسومات أو جداول ،إلخ ...، وهذا التعدد فى الأنماط والانتشار الكبير للكائنات الممثلة لها يمثل تحديا حقيقيا من الناحية النظرية مما يتطلب إيجاد نظرية متكاملة للتحليل والتعرف على هذه النوعية من الصوربمعدلات عالية وثابتة. كما أن المنظومات التقنية المطلوب استخدامها لابد من أن توفر درجة عالية من المعالجات المتوازية حتى يمكن التعرف على الوثائق فى الوقت الفعلى بالنسبة للمستخدم .

وتتميز مخطوطات التراث المكتوبة بالعربية عن الكتابات المطبعية بتنوع أسلوب الخط

أنه خط يد في المقام الأول ، مع افتقاد النمطية في أسلوب عرض النصوص في الصفحات ، هذا بالإضافة إلى وجود بعض التذييلات والحواشي في متن الصفحة. كل هذه الخصائص تمثل صعوبات نظرية وتقنية في عمليتي التعلم والتعرف على هذا النوع من الوثائق بالنسبة لمثيلاتها من الوثائق المطبوعة.

يهدف هذا البحث إلى توصيف المشاكل المرتبطة بالقراءة الآلية لمخطوطات التراث العربية وتصميم منظومة تحقق القراءة الآلية في الوقت الفعلي للمستخدم. والبدائل المطروحة في منظومات التعرف على الأنماط كثيرة مثل مصنفات "بايز"[١] أو الشبكات العصبية[١] أو سلاسل ماركوف المخبأة [٥] والاختيار بين أي من هذه المنظومات يفرضه نوع التطبيق والمشاكل النظرية والعملية الخاصة به. كما أن أسلوب عمل المنظومة قد يختلف من النصف آلية إلى الآلية الكاملة أيضا حسب القيود التي يحددها التطبيق.

وأمام كثرة الأنماط الموجودة في صفحات مخطوطات التراث والانتشار الواسع لكل نمط منها وتغير خواصه الإحصائية من صفحة إلى أخرى فقد وجدنا أنه من الأنسب وضع منظومة تستخدم آلية عشوائية محدودة في التعرف على الأنماط ، وتعتمد على منهج التعلم بدون إشراف الذي يتيح للمنظومة تعلم نمط الكتابة في المخطوطة وأساسيات الحروف بالمواءمة.

٢ تعريف المشكلة

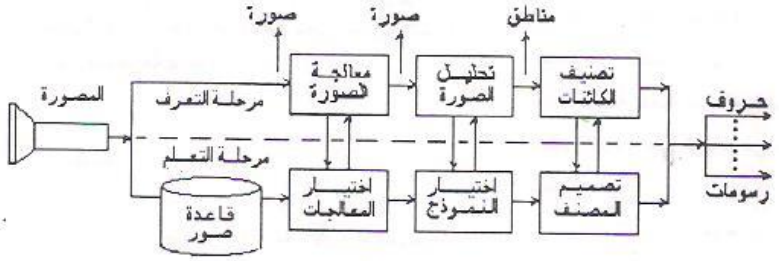
إن منظومة التعلم والتعرف الآلي على الوثائق يمكن تمثيلها بالشكل ١. تعمل هذه المنظومة عادة على مرحلتين:مرحلة التعلم، ثم مرحلة التعرف.وتتكون مرحلة التعرف من العناصر والعمليات الآتية:

١-المصورة ، ٢-معالجة الصورة ، ٣-تحليل الصورة ، ٤-تصنيف كائنات الصورة.

وتتكون مرحلة التعلم من العناصر والعمليات الآتية:

١-قاعدة صور، ٢-اختيار معالجات الصورة، ٣- اختيار نموذج الصورة، ٤-تصميم المصنف.

إن مشكلة تصميم منظومة القراءة الآلية هي وضع نماذج العناصر، وتوصيف العمليات، وترتيب عملها في هذه المنظومة. وحل مشكلة التصميم هذه يتحدد بالمدخل المستخدم في التعرف على الأنماط. ونحن في هذا البحث نعتمد المدخل



شكل ١ منظومة القراءة الآلية للمخطوطات

الإحصائي-النحوي في عمليتي التعلم والتعرف كما سنوضح في تصميم عناصر هذه المنظومة .

هناك ثلاث مشكلات جزئية على التحديد يجب حلها عند التصميم : ١-معالجة عيوب الصورة ، ٢-فصل مناطق الكتابة ، ٣-التعرف على الحروف.ولتحديد نوعية هذه المشاكل الجزئية المطروحة سنستعين بصورة الصفحة الممثلة في الشكل ٢ وهو يمثل إحدى صفحات كتاب الكاشي[١١].

إن النظرة الفاحصة لهذه الصورة تبين أنها تحتاج إلى نوعين أساسيين من المعالجات :

-١ترشيح الشوشرة ، ٢-وترميم الأجزاء الناقصة في الحروف والكلمات. ومشكلة ترشيح الشوشرة في هذه الحالة يمكن حلها بأحد المرشحات التقليدية المعروفة[٩]. أما مشكلة ترميم الأجزاء الناقصة فهي أكثر صعوبة ولن نتعرض لها في هذا البحث لأنها تتطلب المزج بين التعرف على أشكال الحروف والتعرف على الكلمات المكتوبة ، الذي يتطلب استخدام معارف الصرف والنحو والمضمون[١]. والحل الأمثل في هذه الحالة هو ارجاء هذه المرحلة لما بعد التعرف على الحروف الواضحة في المخطوطة.

أما المشكلتان الأساسيتان في معالجة هذا النوع من الصور فهما كيفية فصل مناطق الكتابة في الصفحة، ثم تليها مشكلة التعرف على حروف الكتابة، ونلاحظ في هذه الصورة وجود نوعين من المعلومات المرئية تتمثل في الكتابة والجداول. كما أنه توجد ثلاثة أنماط من الكتابة تتمثل في :الكتابة الأفقية، والكتابة العمودية، والكتابة المائلة. أما النمط المستخدم فهو خط يد حروفه متصلة ومتداخلة ، وإن كان يتميز بالنمطية.

وهذا البحث يقدم بالتفصيل كيفية فصل مناطق الكتابة من الصفحة ثم تجزيئها إلى أسطر ثم إلى كلمات وحروف. سواء كانت الكتابة تمثل سطرا أفقيا أو سطرا مائلا ، وسواء كانت الكتابة شرحا أو تذييلا للصفحة. أما الجداول، أوالرسومات بصفة عامة، والكتابات الموضحة عليها فتمثل بصورتها كما هي. ومشكلة التعرف على الحروف هي تجزيئ كل سطر وتصنيف هذه الأجزاء إلى إحدى طبقات الحروف التسع والعشرين الممثلة لحروف الهجاء العربية.ومن المعروف أن عملية التصنيف هذه

شكل ٢. صورة رقمية لصفحة من مخطوطة الكاشي[١١]

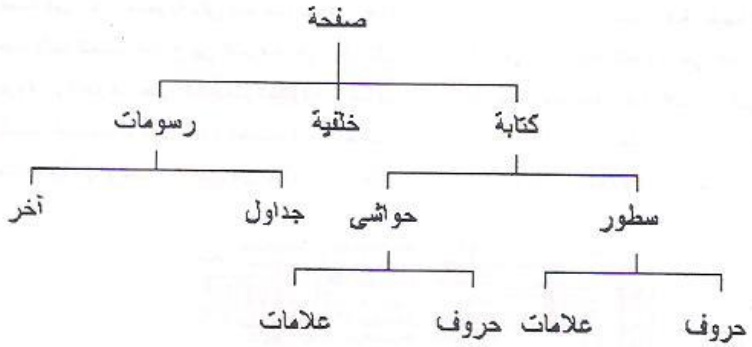
تعتمد على دقة عملية التحليل السابقة وعلى مدى تكامل المعارف الخاصة باللغة في منظومة التعرف سواء بالنسبة لشكل الحرف ، أو بالنسبة لقواعد الصرف والنحو إلى حد كبير.

ولحل هاتين المشكلتين فقد رأينا من الأنسب وضع آلية تعمل بأسلوب المواءمة [٨] أثناء عمل المنظومة مما يسمح بثبات أداء المنظومة مع تغير نوع المخطوطات. كما أن هذه الآلية يجب أن تستخدم بصورة متكررة في حل المشكلتين.

٢ نموذج الصفحة

يهدف تحليل الصورة إلى تقسيمها إلى طبقات متباينة من الكائنات وكشف وقياس سمات كل منها. وهذا يتطلب وضع نموذجا لوصف مكونات الصورة، والنموذج الذى نستخدمه هنا لوصف مكونات الصورة هو النموذج الهرمى. أى أننا نفترض أن الصفحة مكونة من مجموعات تتقسم بصورة مكررة إلى مجموعات أصغر فأصغر من مستوى لآخر كما هو ممثل في الشكل ٣.

وبتطبيق عملية التحليل والتقسيم حتى المستوى الثانى للمخطوطة المصورة فى الشكل ٢ ينشأ عنها التقسيم الموضح فى الشكل ٤. فى هذا الشكل أحيطت كائنات كل منطقة بنافذة أفقية مستطيلة لسهولة العرض وإن كان شكل هذه النافذة وميلها يعتمد على الشكل الهندسى للمنطقة وميل محاوره الأساسية بالنسبة لمحاور الصفحة.



شكل ٣. النموذج الهرمى لصفحة من المخطوطة.

شكل ٤. تقسيم صورة الصفحة إلى مناطق من الكتابة والجداول

٤ تقسيم صورة الصفحة

يمكن تعريف عملية تقسيم الصورة بإنها عملية تصنيف نقاط الصورة ، حيث تقسم الصورة إلى مناطق بتوزيع النقاط على الطبقات التى تنتمى إليها. ولتحقيق عملية التقسيم نفترض أن الصورة مكونة من مناطق تختلف فى خواصها مثل المكانية، أو الهندسية ، أو القياسية ،أو شدة إضاءتها أو نسيجها أو فى كل هذه الخواص معا[٩] . وعند توافر المعلومات الإحصائية عن هذه المناطق مسبقا نشرع فى قياس كل من هذه الخواص عند كل نقطة من نقاط الصورة ونضعها فى هيئة متجه يعرف بمتجه السمات، والسمات التى يمكن استخدامها فى تقسيم الصفحة هى :مستوى الرمادية للنقطة، القيمة المتوسطة لمستوى الرمادية فى نافذة محلية،إلخ... . ثم يصنف متجه السمات باستخدام منظومة تصنيف مصممة على أسس قريبة من مصنف بايز المثالى. وتوفير المعلومات الإحصائية عن مناطق الصورة يتم من خلال المسح الشامل لصورة أو أكثر تحتوى على المناطق المفترضة فيما يعرف بالتعلم تحت الإشراف. عند غياب مثل هذه المعلومات لكل منطقة على حدة يصبح منهج التعلم بدون إشراف هو البديل الأفضل فى هذه الحالة. ومفهوم التعلم هنا يقصد به إيجاد التقسيم الأمثل لمجموعة من الكائنات ممثل كل منها بمتجه السمات بحيث تكون درجة التشابه بين كائنات المجموعة الواحدة أكبر مايمكن ودرجة التباين بين المجموعات أكبر مايمكن[١].

ويمكننا تعريف مشكلة تقسيم صورة الصفحة وضعيا على النحو التالى:

عند كل نقطة من نقاط الصورة يعرف متجه السمات، **سق** ،

$$\text{سق} = (س_1, س_2, \ldots, س_ب) \qquad (1)$$

وتمثل نقاط الصورة فئة كائنات التعلم المطلوب. توزيعها على الطبقات المختلفة ، ويرمز لفئة كائنات التعلم بـ س،

$$س = \{ \text{سق}_1, \text{سق}_2, \ldots, \text{سق}_ن \} \qquad (2)$$

وتعرف مشكلة التقسيم على أنها كيفية تقسيم فراغ السمات ، غ ب ، إلى عدد م من المناطق، س ل، تحقق الشرطين الآتيين:

$$س_1 \cap س_2 \cap \ldots \cap س_م = غ_ب$$

$$س_ل \neq \emptyset \quad ، ل = 1، 2، \ldots، م \qquad (3)$$

س١ ∩ س٢ ∩ ... ∩ س ل ... ∩ س م = ϕ

ل= ١، ٢، ... ، م (٤)

والقاعدة التى تربط منطقة الكائنات، س ل ، بالطبقة الخاصة بها، ق ل، هى:

س ∈ ق ل إذا س ∈ س ل (٥)

ولايجاد مناطق التقسيم، س ل، نبحث عن النسق الأمثل، ق، الذى يحقق المعيار الآتى:

ع = ع(ق ، س) (٦)

حيث ق هى متجه النسق،

ق = (ق١، ق ٢، ... ، ق م) (٧)

و يربط هذا النسق كل كائن من الكائنات، س، بطبقة من الطبقات، ق ل. والنسق الأمثل ، ق *، هو الذى يحقق العلاقة:

ع(ق *، س) = صغرى ع(ق ، س) (٨)

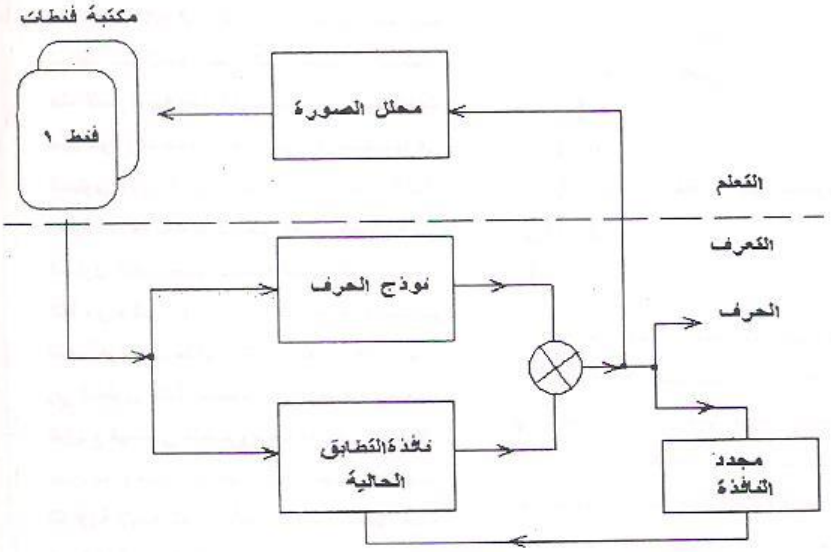
والمعيار، ع، هنا يمثل انفصالية المجموعات وهو علاقة بين الانتشار الداخلى فى كل منطقة والانتشار البينى للمناطق ، والبحث عن هذا النسق الأمثل، ق*، يتم باستخدام بإحدى الطرق التكرارية المعروفة كالبرمجة الديناميكية [١].

٥ تعلم أساسيات الحروف

تهدف عملية تعلم أساسيات الحروف إلى معرفة شكل الحرف وقياس سماته الأساسية وتحديد نوع الفنط الذى ينتمى إليه. وحيث أننا لانعرف مسبقا نوعية الفنط المستخدم فى صفحات المخطوطة فلابد من اتباع منهج التعلم بدون إشراف لمعرفة أشكال الحروف والفنط الذى تنتمى إليه.وفى هذا الإطار تصبح المشكلة هى كيفية تحليل منطقة الكتابة وتقسيمها إلى مناطق جزئية يمثل كل منها حرفا من الحروف، أى من الناحية النظرية تصنيف كل منطقة جزئية إلى طبقة من الطبقات. هذا المنهج يسمح لنا استخدام نموذج التقسيم المعروض سابقا كنموذجا للتعلم والتعرف على أشكال الحروف. ولن كانت السمات المستخدمة تعتمد على خواص أخرى، والخاصية المعتمدة فى الفصل بين الحروف هى التوزيع الهندسى لأساسيات الحرف داخل نافذة .حيث تمثل نقاط هذه النافذة متجه السمات فى هذه الحالة.

والنموذج الذى نستخدمه هنا لتمثيل الحروف هى نافذة على شكل متوازى مستطيلات تحتوى على نموذج لأساسيات الحرف[٧]. يتم تحديد حجم وشكل واتجاه هذه النافذة لكل مجموعة من السطور بعد عملية المسح الأولى للصورة وكنتيجة لعملية التقسيم المصاحبة لها. أما النموذج المخزن فى كل نافذة فيتم اكتشافه من خلال عملية تقسيم مماثلة لتلك المستخدمة فى تقسيم الصفحة إلى سطور من الكتابة ورسومات، أو يعرف بمعلومات مسبقة عند بداية مرحلة التعلم. وكما أشرنا من قبل فإن مرحلتى تقسيم الصفحة والتعرف على حروف الكتابة مرتبطتان فى عملهما ونتائجهما،يوضح ذلك منظومة التعرف والتعلم الممثلة فى الشكل ٥.

فى بداية عملية التعلم يوجد نسقان للطبقات، ق١، ق٢، يمثل كل منهما فنطا للحروف. والفنط الأول هو فنط قياسى يمثل فيه الحرف بخط نحيف داخل نافذة حجمها متوسط ويصلح للاستخدام مع غالبية أنواع الفنطات المعروفة. والفنط الثانى يمثل حروف الكتابة المستخدمة فى المخطوطة. وفى البداية تكون فئة النسق الثانى خاوية، وتستخدم حروف فئة النسق الأول كما هى. وبنهاية عملية



شكل ٥. منظومة تعلم أساسيات الحروف.

التعلم تمثل فئة النسق الثانى وينعدم استخدام فئة النسق الأول.

عند كل نقطة من نقاط مناطق الكتابة تقارن محتويات نافذة التطابق الحالية مع أشكال الحروف فى الفنط القياسى ويصغر أويكبر إلى حجم نافذة التطابق الحالية. ثم تصنف محتويات نافذة التطابق الحالية بقياس مسافة ماهلانوبيس بين نموذج الحرف والشكل الموجود بداخل النافذة، ونتيجة عملية التصنيف هى إحدى الخطوات الآتية:

١- يصنف الحرف إلى الطبقة الأقرب فى النسق الثانى ،وتعدل السمات الممثلة لشكل الحرف المكتشفة لاستخدامها فى عملية التعرف التالية،وكذلك مكان وحجم نافذة التطابق الحالية.

٢-يصنف الحرف إلى الطبقة الأقرب فى النسق الأول ، وتنشأ طبقة جديدة فى النسق الثانى نواتها الحرف المكتشف ، ويلغى استخدام هذا الحرف فى النسق الأول.

٣-لا يمكن تصنيف محتويات النافذة فيؤجل اتخاذ القرار حتى يتوافر عدد مناسب من الكائنات الممثلة لكل طبقة فى النسق الثانى.أو ينشأ نسق جديد يمثل فنطا جديدا.

واختيار مسافة ماهالانوبيس هنا للتصنيف يضمن ثبات معدل التصنيف مع تغير موضع أوحجم أوميل الحرف فى الصفحة، وإن كان يتطلب عددا أكبر من الكائنات الممثلة لشكل الحرف قبل استقرار نتائج التصنيف.

٦ إجرائية التعلم والتعرف

إن الإجرائية المقترحة للتعلم والتعرف تستخدم استراتيجية " تعلم كلما مضيت" وتستخدم هذه الاستراتيجية عند كل مستوى من مستويات تمثيل نموذج الصفحة ، وهذا يعنى أننا نستخدمها فى المستوى الأول لتقسيم الصفحة إلى مناطق كتابة، وخلفية ، ورسومات ثم تستخدم نفس الإجرائية فى المستوى الثانى لتقسيم منطقة الكتابة إلى سطور كتابة ، وحواشى ، ولكننا فى هذه المرحلة نستخدم سمات أخرى مثل مكانية الكائنات للفصل فيما بينها. وفى المستوى الثالث نستخدم نفس الإجرائية وسمات التوزيع الهندسى للتعلم والتعرف على أشكال الحروف، وهذه الإجرائية هى تنفيذ للطريقة التكرارية لإيجاد الحل الأمثل[١٠]. وسنمثل هذه الإجرائية بالبرمجية الآتية:

إجرائية البحث_عن_الطبقات (س ، ق)
مدخلات: فئة الكائنات المجهولة، س ،
مخرجات: نسق التصنيف، ق ،
{
طالما(حصل_كائن(س ، سم) == لاشيئ)
{
صنف = خطأ ؛
طالما ((صنف == حقيقى) واو (ل < عدد الطبقات))
{
ف = مسافة_ماهالانوبيس(سم ، س ل) ؛
إذا (ف <= ف١) { حدث_الطبقة (سم ، س ل) ؛ صنف = حقيقى ؛ }
وإلا إذا (ف > ف١)
}
إذا(الرصة == مملوءة)
صنف_المخزون (رصة ، ق) ؛
دفع_ الكائن(سم ، رصة) ؛
صنف = حقيقى؛
{
}؛
إذا (صنف = خطأ) انشأ_طبقة(سم ، س ل+١ ، ق) ؛
{
{

هذه الإجرائية تعمل كما يلى: بداية تعتبر فئة التعلم خاوية وكذلك يعتبر نسق الطبقات خاويا، ق، . عند دخول كائن جديد إلى فئة التعلم تقوم الدالة " حصل_ كائن" باكتسابه ويعتبر مجهول الطبقة أى غير مصنف ثم تقاس مسافة ماهالانوبيس بين متجه السمات ،سم ، ومتجه سمات الوسط الممثل لمنطقة الطبقة ، م ، فإذا كانت هذه المسافة أقل من الحد الأصغر ، ف، ، صنف الكائن مع كائنات هذه الطبقة ، ثم تجدد معلومات هذه الطبقة ممثلة فى متجه الوسط ، ومصفوفة الانتشار ،ش ،. ماعدا ذلك إذا كانت هذه المسافة أقل من الحد الأكبر ،ف١، يعتبر الكائن الجديد غير محدد الطبقة مؤقتا ويدفع به إلى رصة ؛ بالدالة " دفع_الكائن" ؛عندما تمتلئ هذه الرصة تقوم الدالة "صنف_المخزون" بتصنيف جميع كائناتها لأقرب جار من الطبقات الموجودة حين ذلك ، وتفريغ الرصة فى نفس الوقت؛ أما إذا لم يمكن تصنيف هذا الكائن أى أنه يوجد خارج منطقة الحد الأكبر لجميع الطبقات تنشأ طبقة جديدة نواتها هذا الكائن.

ونلاحظ هنا أن اجرائيتى التعرف والتعلم اندمجتا فى إجرائية واحدة تحقيقا للاستراتيجية

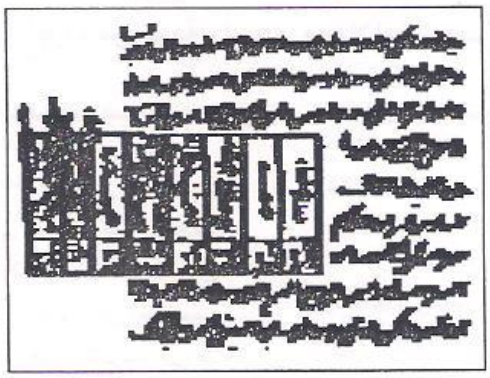
المقترحة فى البداية وهى "تعلم كلما مضيت".وهذه الاجرائية هى نواة البحث عن النسق الأمثل للأنماط الموجودة فى الصفحة ؛ أى أنها تعمل داخل حلقة مستمرة للبحث تبدأ بنسق افتراضى وتنتهى عند الوصول إلى النسق الأمثل والذى يحقق أصغر معدل للخطأ فى التعرف على الكائنات الجديدة.

٨ النتائج التجريبية

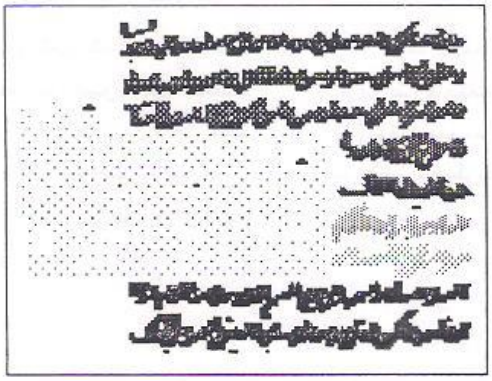
نفذت منظومة التعلم والتعرف المقترحة على حاسوب شخصى .استخدمت لغة الـ "سى++" فى تنفيذ عناصر هذه المنظومة، كما اعتمدت فى التصميم منهجية التصميم بالكائنات مما يسمح ببناء النواة الأولى بسرعة واعادة استخدام عناصرها فى مراحل التطوير اللاحقة.

وقد استخدمت المنظومة المقترحة فى التعرف على صفحات من مخطوطة كتاب الكاشى [١١]. وفى هذه المرحلة من التصميم والاختبار تعمل المنظومة بطريقة نصف آلية حيث يتم تشغيل مراحل تقسيم الصفحة إلى مناطق وسطور والتعرف على الحروف بعد تدخل المستخدم. وينحصر تدخل المستخدم هنا فى اختيار واحدا من عدة اختيارات معروضة عليه تمثل حجم نافذة التطابق والحدود الصغرى والكبرى لتصنيف كائنات كل طبقة.

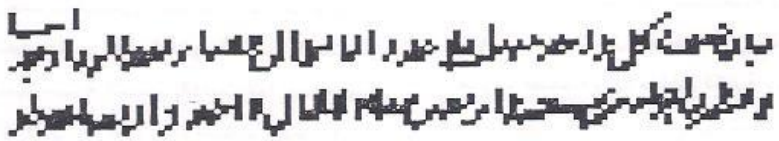
يمثل الشكل ٦ مراحل معالجة صورة إحدى عينات الصفحات المستخدمة فى اختبار المنظومة .عند تشغيل المرحلة الأولى للتقسيم تخرج صورة الصفحة وبها مناطق الكتابة ، والخلفية ، والرسومات ،كما هو موضح فى الشكل ٦-أ. وفى المرحلة الثانية تقسم مجموعة الكتابة إلى سطور بناء على خاصية المكانية لينشأ عنها الصورة الممثلة فى الشكل ٦-ب.



شكل ٦-أ صورة الصفحة بعد مرحلة التقسيم الأولى



شكل ٦-ب صورة الصفحة بعد مرحلة التقسيم الثانية



شكل ٧-أ صورة مكبرة للسطرين الأولين من الصفحة فى شكل -٢.

فإن قسمنا كل منهما على عدد الأموال فصار بعد الرد أربعة
وعشرون أجزاء من تسعة وأربعين معادلا لمال واحد وأربعة أعداد

شكل ٧-ب الحروف المتعرف عليها فى السطرين الأولين من الصفحة فى شكل -٢

يمثل الشكل ٧-أ السطرين الأولين من الصفحة قبل التعرف على الحروف المكتوبة فيهما. والشكل ٧-ب بين الحروف التى تم التعرف عليها وتلك التى لم يمكن التعرف عليها والتى تحتها خط. ويرجع السبب الرئيسى فى فشل التعرف على هذه الحروف هو تداخل كلماتها مع حروف كلمات أخرى. وتشير النتائج الأولية أن معدل التعرف يصل إلى ٧٧٪.

٩ الخلاصة

إن الهدف من هذا البحث هو تصميم نموذج تجريبى لمعالجة مخطوطات التراث. لهذا الغرض وضعت إجرائية تعلم وتعرف تعمل بالمواءمة وبدون إشراف . وقد تم تنفيذ هذا النموذج التجريبى على حاسوب شخصى باستخدام منهجية التصميم بالكائنات مما يسمح بتطويرها فيما بعد. وقد جربت هذه المنظومة على بعض المخطوطات [١١]

ولقد حققت المنظومة معدلا مقبولا من التعرف علىالحروف يصل إلى ٧٧٪ . ويجرى تطوير المنظومة حاليابتوصيف نماذج أساسيات الحروف بصورة أدق تسمح برفع معدل التعرف على الحروف . ويهدف البحث مستقبلا لمعالجة أنماط أخرى أكثر تعقيدا فى المخطوطات كالجداول مثلا.

إمتنان

يتقدم الباحثان بجزيل الشكر للأستاذ الدكتور/ محمود خشان لتوفيره صورة المخطوطة محل التطبيق.

المراجع

[1] R.O.Duda , P.E. Hart," Pattern Classification and Scene Analysis", Wiley,New York,1973,

[2] Fujihara, H.; Babiker, E.; Simmons, D.B.,"Fuzzy Approach to Document Recognition" Second IEEE International Conference on Fuzzy Systems ,p.980-5 vol.2,

[3]Jianying Hu; Pavlidis, T., "Hierarchical curve retrieval with application to aerial image alignment and cursive handwriting recognition",Proceedings. Second Annual Symposium on Document Analysis and Information Retrieval,p. 411-27,

[4] Mohamed Kamel and Aiguo Zhao,"Extravtion of Binary/Graphics Images from Grayscale Document Images",graphical models and image processing,vol 55,no.3,may,pp 203-217,1993,

[5] Gary E Kopec,Philip A Chou,"Document Image Decoding Using Markov Source Models",IEEE transactions on Pattern Analysisi and Machine Intelligence,vol 16, no 6,june 1994,

[6] Su Liang and M.Ahmadi.,"A morphological Approach to Text String Extraction from Regular Periodic Overlapping Text/Background Images", graphical models and image processing,vol 56;no.5,september,pp 402-418,1994,

[7] M. Y. Mahmoud El Nahas, Mohamed A. El Hamalaway & Ali Ali Fahmy, "A Statistical Approach for Arabic Character Recognition", 12th International Congress for Statistics, Computer Science, Social and Demographic Research; Cairo; pp. 243-250; March 28th-April 2nd, 1987,

[8] M.Y.Mahmoud (El Nahas) : " Object Decomposition by a Nondeterministic finite automaton ",Fifth ORMA Conference, MTC Cairo, November 23-25, 1993,

[9] Azriel Rosenfeild , Avinach C. Kak , "Digital Picture Processin",Academic Press,London,1982,

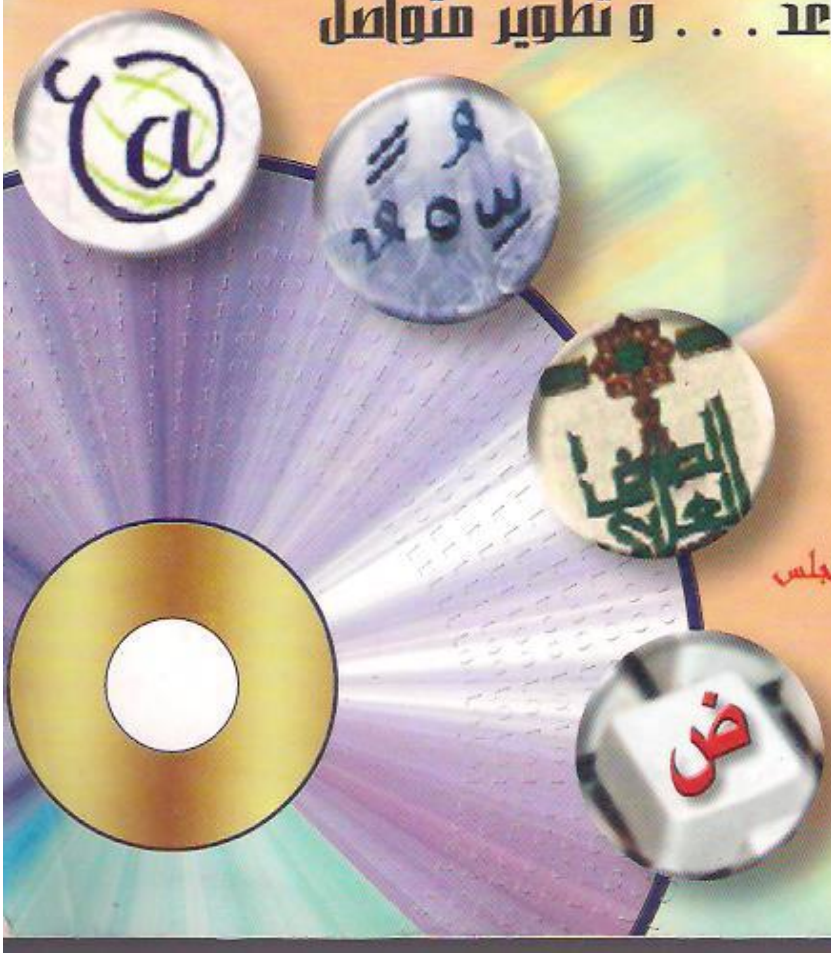
[10] Stephen P. Banks,"Signal Processing, Image Processing and Pattern Recognition", Prentice Hall, London,1990,

[١١] جمشيد غياث الدين الكاشى ، "مفتاح الحساب والرسالة المحيطية ٣ ، دار الطبع والنشر للأدب الفنى والعلمى للدولة ، موسكو ١٩٥٦م.

المجلس الأعلى للغة العربية

اللغة العربية في تكنولوجيا المعلومات

تطور واعد . . . و تطوير متواصل



منشورات المجلس
2005

دراسة مقارنة بين أشكال الحروف العربية والحروف الإنجليزية

أ. د. محمد يونس الحملاوى
أستاذ هندسة الحاسبات، كلية الهندسة، جامعة الأزهر، القاهرة
5 شارع سليمان محمد أباظة، الحى الثانى، مصر الجديدة، القاهرة
هاتف 6321465 ناسوخ (فاكس) 6377446
بريد إلكترونى mhamalwy@hotmail.com

الكلمات المفتاحية: الحروف العربية، الحروف اللاتينية، الحروف الإنجليزية، التعرف الآلى على الحروف المطبوعة OCR، المعالجة الآلية للغة العربية.

البحث يدخل ضمن المحور التالى من محاور المؤتمر كما فى النشرة باللغة الإنجليزية:

Natural Processing of the Arabic Language: e-tools, OCR,

C:\Documents and Settings\محمد الحملاوى\My Documents\مستندات\الجزائر\المجلس الأعلى للغة العربية\الحروف العربية والإنجليزية.doc

نبـــــذة:

يدرس هذا البحث أساساً أشكال الحروف المكتوبة للغة العربية وللغة الإنجليزية. وقد تم استنباط سمات أساسية للأشكال المختلفة للحرف العربى الواحد بحيث يمكن الاستفادة منها فى عملية التعرف على الحروف. كما تمت دراسة الشكلين الأساسيين للحرف الإنجليزى الواحد، واتضح أنه لا يوجد للشكلين الأساسيين للحرف الإنجليزى الواحد فى المجمل سمة أساسية تربط الشكلين، بل أن بعض هذه الحروف لا توجد علاقة البتة بين شكلى الحرف المختلفين. ومن المأمول أن تساعد هذه الدراسة فى وضع برمجيات للتعرف الآلى على الحروف العربية المكتوبة بطريقة سهلة تؤدى إلى الحصول على نتائج جيدة. كما أظهر البحث منطقية استنباط الأشكال المختلفة للحرف العربى تبعاً لموقعه من الكلمة على عكس الحروف الإنجليزية التى لا يوجد فى الغالب رابط منطقى بين الشكل الكبير (المستعمل فى بداية أسماء الأعلام) والشكل الصغير (المستعمل فى وسط الكلمة) لها.

1. مقدمة:

يقوم الكثيرون بعمل محاولات للتعرف على الحروف المكتوبة في اللغات المختلفة وقد تم وضع برمجيات كثيرة في لغات مختلفة تقوم بهذه الوظيفة بدرجات نجاح متفاوتة، بل ولقد تم وضع برمجيات للتعرف على الخطوط المكتوبة باليد أيضا لهذه اللغات؛ وان لم يزل الكثير منها في طور التطوير[1]. على أنه بالنسبة للغة العربية فمازالت نتائج تلك البرمجيات غير مرضية وتحتاج لعمل دؤوب للوصول إلى نتائج أفضل[2-3-4-5].

وفي هذا الإطار فإن أول خطوة للتعرف على الحروف هي دراسة أشكالها وأنماط كتابتها المختلفة. بالنسبة للغة العربية توجد خطوط عديدة للكتابة وللطباعة منها النسخ، الرقعة، الكوفي، الثلث....وهي تشترك جميعها في سمات يسهل معها تمييز الحرف الواحد مهما كان نوع الخط المكتوب به. والدراسة الحالية لأشكال الحروف تيسر عملية التعرف الآلي عليها من خلال وضع قواعد منطقية تسير عليها عملية التعرف.

ويركز البحث على حروف اللغة العربية بالإضافة إلى حروف اللغة الإنجليزية والتي هي مجموعة جزئية من حروف اللغة اللاتينية[6-7-8]. وبالرغم من أن الحروف العربية والتي تأخذ أشكالا متعددة باختلاف موقعها من الكلمة؛ مما كان يشكل صعوبة في عملية التعرف؛ إلا أنه قد أمكننا التوصل إلى أنه ذلك توجد سمات أساسية لكل حرف تجرى عليه إضافات بشكل ما للحصول على الأشكال الأخرى لنفس الحرف على حسب موقعه من الكلمة. كما تمّت دراسة علاقة شكلي الحرف الإنجليزي ببعضهما البعض بهدف الوصول إلى النسق الذي يربط بين هذين الشكلين.

---

[1] Ching Y. Suen; Advances in Optical Character Recognition; Canadian Computer Conference; Edmonton, Canada; May 23-25th, 1978.

[2] H. Y. Abdelazim and M. Y. Hashish; Arabic Reading Machine; 10th National Computer Conference; Riyad, Saudi Arabia; 1988; pp. 733-744.

[3] M. A. Sharkawy et all; Forier Descriptors for Printed Arabic Character Recognition; 13th International Conference on Statistics and Computer Science; Cairo, Egypt; March 26-31st, 1988.

[4] M. F. Tolba et all; A Recognition Algorithm for Arabic Printed Characters; 11th International Congress for Statistics and Computer Science; Cairo, Egypt; 1986

[5] M. Y. Mahmoud et all: A Statistical Approach for Arabic Character Recognition; Twelfth International Congress for Statistics and Computer Science; Cairo, Egypt; March 28th - April 2nd, 1987.

[6] ISO/IEC 10646 International Standard; ISO/IEC; Geneve, Switzerland; 1993.

[7] Mohamed A. El Hamalaway; AFGUST: A Standard for Coding Arabic Character Sets; 8th International Congress for Statistics and Computer Science; Cairo, Egypt; March 26-31st, 1983; also in Egyptian Computer Science Journal; Vol. 6, No. I; Jan. 1983.

[8] ISO/IEC 8859 International Standard; ISO/IEC; Geneve, Switzerland; 1999.

2. السمات العامة لحروف اللغة العربية:

من المعروف أن حروف اللغة العربية تتخذ أشكالا تختلف باختلاف مواقعها من الكلمة أى بحسب كونها فى حالة إفراد أو فى أول الكلمة أو وسطها أو نهايتها[9]. ويوضح جدول (1) هذه الأشكال لجميع الحروف العربية. فى عملية مثل عملية التعرف الآلى على الحروف لابد من الأخذ فى الاعتبار هذا التغير فى الشكل[10]. على انه بفحص هذه الأشكال يمكن التوصل إلى أن لكل حرف من الحروف العربية سمة أساسية يجرى التغيير فيها بالإضافة أو الحذف فى اولها او آخرها لإيجاد شكل مختلف فإذا أخذنا الحرف "ا" مثلا تجد أن السمة الأساسية له جدول رقم 1.

الحرف مفردا	الحرف في بداية الكلمة	الحرف في وسط الكلمة	الحرف في نهاية الكلمة	السمة الأساسية للحرف	الإضافة
ا	ا	ـا	ـا	ا	ـ
ب	بـ	ـبـ	ب	بـ	ـ
ت	تـ	ـتـ	ت	تـ	ـ
ث	ثـ	ـثـ	ث	ثـ	ـ
ج	جـ	ـجـ	ج	جـ	ـ/C
ح	حـ	ـحـ	ح	حـ	ـ/C
خ	خـ	ـخـ	خ	خـ	ـ/C
د	د	د	د	د	
ذ	ذ	ذ	ذ	ذ	
ر	ر	ر	ر	ر	
ز	ز	ز	ز	ز	

[9] M. C. Vanwormhoudt and Mohamed A. El Hamalaway; Remarks about printing and Displaying some Non- Latin Characters; International Federation of Automatic Control (IFAC) Conference; Cairo, Egypt; Nov. 26-29th, 1977.

[10] Mohamed A. El Hamalaway and Salwa H. El Ramly; A Language Dependent Arabic Character Recognition Approach; 14th International Congress for Statistics and Computer Science; Cairo, Egypt; March 25-30th, 1989.

س	سـ	سـ	س	سـ	ـس/U
ش	شـ	شـ	ش	شـ	ـش/U
ص	صـ	صـ	ص	صـ	ـص/U
ض	ضـ	ضـ	ض	ضـ	ـض/U
ط	ط	ط	ط	ط	
ظ	ظ	ظ	ظ	ظ	
ع	عـ	عـ	ع	عـ	ـع/C
غ	غـ	غـ	غ	غـ	ـغ/C
ف	فـ	فـ	ف	فـ	ـ
ق	قـ	قـ	ق	قـ	ـق/U
ك	كـ	كـ	ك	كـ	
ل	لـ	لـ	ل	لـ	ـل/U
م	مـ	مـ	م	مـ	
ن	نـ	نـ	ن	نـ	ـن/U
هـ	هـ	هـ	ـه	هـ	ـهـ/U
و	و	و	و	و	
ي	يـ	يـ	ي	يـ	

هي "ا" ويتم إضافة جزء أفقي "ـ" إذا ما وقع الحرف في وسط الكلمة. وبالنسبة للحرف "ب" تجد أن له سمة أساسية هي "بـ" ّدـ يضاف إليها جزء على شكل " ـا " في آخر الحرف وذلك في حالة إفراد الحرف أو وجوده في آخر الكلمة. وتطبيق هذه الملاحظة على الحروف "تـ"، "ثـ"، "ف". أما الحرف "جـ" فان له سمة أساسية هي "جـ" يضاف إليها الجزء " ـ " في حالة إفراد الحرف أو وقوعه في نهاية الكلمة. وتنطبق نفس الملاحظة على الحروف "ح"، "خ"، "ع"، "غ". ونلاحظ أنها نفس القاعدة المطبقة في المجموعة "ب"، "ت"، "ث"، "ف" مع تغيير شكل الإضافة.

بالإضافة إلى ذلك فانه يمكن الملاحظة أن الحروف "ش"، "س"، "ص"،

"ض"، "ق"، "ل"، "ن" لها سمات هي "سـ "، "شـ"، "صـ "، "ضـ"، "ةـ"، "لـ" ويجري إضافة إما جزء على شكل"↔"إذا كتب الحرف بصورة مفردة أو إذا كان في نهاية الكلمة، أو إضافة جزء أفقي "ـــ" في حالة كتابة الحرف في اول الكلمة أو في آخرها.

كما يمكن ملاحظة أن الحرف "د"، "ذ"، "ر"، "ز"، "ط"، "ظ"، "مـــ"، "و" تحفظ بشكل واحد أيا كان موقعها في الكلمة وهو نفس شكل السمة الأساسية لها.

يبقى ثلاثة حروف عربية ينفرد كل منها بخصائص ولا يشترك مع أي مجموعة أخرى وهي الحرف "ك" حيث له سمة أساسية "لـــ" يتم إضافة جزء علوى "/" في حالة وجود الحرف في اول او وسط الكلمة، أما إذا وقع الحرف في نهاية الكلمة أو جاء منفردا فتتم عليه إضافتين هما "ء"، " ـ ". أما الحرف "ه" وهي نفس شكل الحرف في حالة الإفراد، أما إذا كان الحرف في نهاية الكلمة فيتم إضافة "ـ" قبله، أماإذا جاء الحرف في أول أو في وسط الكلمة فيتم شكل "ه" بالإضافة إلى المد "ـ" فيصبح الشكل "هـــ". يبقى من الحروف العربية الثمانية والعشرين حرف واحد هو الحرف "ى" الذي لم يكن إستنباط سمة أساسية له لكن يبقى الحرف على شكله "ى"في حالة وجوده منفردا أو وقوعه في آخر الكلمة، بينما يصير شكله "يـــ" في حالة وروده في أول الكلمة أو في وسطها.

- **السمات العامة لحروف اللغة الإنجليزية:**

بتفحص الحروف الإنجليزية المعروضة في الجدول "2" نجد شكلين لكل حرف شكل صغير وشكل كبير. ونجد أن مجموعة من الحروف تحتفظ بنفس شكلها في المجموعتين مع تغيير جدول رقم "2".

أشكال حروف اللغة الإنجليزية

الحرف الكبير Capital	الحرف الصغير Small	السمة الأساسية	الإضافة
A	a		
B	b	b	
C	c	c	
D	d		
E	e	E	
F	f	F	-
G	g		
H	h	h	⌋
I	i	I	
J	j	J	
K	k	K	
L	l	l	-
M	m		
N	n		
O	o	O	
P	p	P	
Q	q		
R	r		
S	s	S	
T	t	T	⌊
U	u	U	
V	v	V	
W	w	W	
X	x	X	
Y	y	Y	
Z	z	Z	

حجمها من صغير إلى كبير وهي المجموعة "W"، "V"، "U"، "S"، "O"، "C"، "Z"، "X" ويمكن اتخاذ هذا الشكل للحرف على أنه السمة الأساسية له.

وعلى خلاف ذلك فتوجد مجموعة من الحروف لا يمكن أستنباط سمة أساسية لها حيث لا توجد علاقة بين الشكلين مثال ذلك مجموعة الحروف "R"،"Q" ، "N"، "M"، "G"، "D"، "A"كما يمكن ملاحظة أن مجموعة من الحروف هى "K"، "F"، جرى تحريف فى شكلها الأصلي لتصبح "k"، "f"، وأخرى مثل الحرفين "T"، "E"، جرى لها تحريف فى الشكل الأصلى مع إضافة. بالإضافة إلى ذلك يمكن ملاحظة أن الحرفين التاليين "Y"، "P" قد جرى تحريف فى شكلهما الأصلى وإنزالهما للمستوى الأسفل. أما الحرف "J" فقد جرى له تحريف فى الشكل الأصلى مع تحريك للمستوى لأسفل مع إضافة نقطة ليصبح "j"، وكذلك الحرف"I" فتم تصغيره مع إضافة نقطة ليصبح "i". كما توجد ثلاثة حروف يمكن استنباط سمة أساسية لها وهى الحرف "B" (سمة أساسية b مع إضافة جزء علوى لتصبح "B") والحرف H (سمة أساسية h مع إضافة جزء علوى لتصبح "H") والحرف "L" (سمة أساسية l مع إضافة "ـ" لتصبح"L"). وبهذا فانه توجد سمات أساسية لأحد عشر حرفا فقط من ستة وعشرين حرفاً هى مجموعة حروف اللغة الإنجليزية على عكس ما تبين بالنسبة لحروف اللغة العربية التى كان لجميع حروفها ما عدا حرفاً واحداً سمات أساسية.

ويمكن من الجدول "2" ملاحظة أن الحروف الإنجليزية لا يوجد فى الغالب رابط منطقى بين شكلها الكبير (Capital) وشكلها الصغير (Small). كما يمكننا ملاحظة وجود لَبْس منطقى بين الشكل الكبير لحرف (آى) "I" والشكل الصغير لحرف آخر هو حرف (إل) "l".

3- ملاحظات على صوتيات الحروف فى اللغة العربية مقارنة باللغة الإنجليزية

باعتبار الحروف المرسومة تعبر عن أصوات بعينها لذلك من المفيد إثبات بعض الملاحظات على أصوات حروف اللغة العربية وحروف اللغة الإنجليزية كى تتضح صورة ما نحن بصدد رسمه. ونلاحظ بالتالى تميز الحروف العربية بأن لكل منها صوتاً واحداً لا يتغير بموقع الحرف فى الكلمة وبتتابع الحروف قبله وبعده. كما توجد حركات محددة يمكن عن طريقها تنغيم الحرف وهى الفتحة والكسرة والضمة والشدة والسكون، بالإضافة إلى أن التنوين يلعب دوراً فى تنغيم الكلمة. وعلى خلاف ذلك فإننا نجد أن اللغة الإنجليزية لا يوجد لكل حرف صوت واحد، بل أكثر من صوت مثل الحرف "S" والذى ينطق بطرق متعددة على حسب ما قبله

وما بعده من حروف. وينطبق نفس الشيء على حرف "C" الذى ينطق س، ش، ك على حسب الكلمة بدون قواعد محددة لهذا.

أضف إلى ذلك أنه توجد حروف لا تنطق ولا تتبع قاعدة محددة لذلك كما في حرف "G". كما يلاحظ أن حرفاً مثل حرف "U" له ثلاثة عشر صوتاً ليس لها قاعدة منطقية فى توليدها.

ولم يتعرض البحث لرسم الأصوات المختلفة فى كلتي اللغتين انطلاقاً من تلك الملاحظات اللغوية الصوتية رغم أنها تستحق البحث إلا أننا اكتفينا برسم الحروف المكتوبة الموثقة فى كلتي اللغتين مع تسجيل أن هذه النقطة تستحق بحثا مستقبليا.

4- الخلاصة:

يقدم هذا البحث دراسة على أشكال الحروف المكتوبة للغة العربية وللغة الإنجليزية. وقد تم استنباط سمات أساسية للأشكال المختلفة للحرف العربى الواحد بحيث يمكن الاستفادة منها فى عملية التعرف على الحروف. ومن المأمول أن تساعد هذه الدراسة فى وضع برمجيات للتعرف الآلى على الحروف العربية المكتوبة بطريقة سهلة تؤدى إلى الحصول على نتائج جيدة.

تشترك مختلف خطوط اللغة العربية للكتابة فى سمات مشتركة يسهل معها تمييز الحرف الواحد مهما كان نوع الخط المكتوب به. والدراسة الحالية لأشكال الحروف تضع قواعد منطقية تسير عليها عملية التعرف. ولقد تم توليد السمة الأساسية لكل حرف عربى التى يجرى عليها إضافات بشكل ما للحصول على الأشكال الأخرى لنفس الحرف على حسب موقعه من الكلمة. وبالتالى فإن عملية التعرف تفضى إلى التعرف على تلك السمات الأساسية للحرف وليس على شكله، آخذين فى الاعتبار وجود قواعد لتركيب الكلمة العربية لا يمكن الحيود عنها.

كما أظهر البحث منطقية استنباط الأشكال المختلفة للحرف العربى تبعاً لموقعه من الكلمة على عكس الحروف الإنجليزية التى لا يوجد فى الغالب رابط منطقى بين الشكل الكبير (المستعمل فى بداية أسماء الأعلام) والشكل الصغير (المستعمل فى وسط الكلمة) لها.

كما اتضح من دراسة الشكلين الأساسيين للحرف الإنجليزى الواحد، أنه لا يوجد للشكلين الأساسيين للحرف الإنجليزى الواحد فى المجمل سمة أساسية تربط الشكلين، بل أن بعض هذه الحروف لا توجد علاقة البتة بين شكلي الحرف المختلفين. كما يتضح وجود لبْس منطقي بين الشكل الكبير لحرف (آي) "I" والشكل الصغير لحرف آخر هو حرف (إل) "l".

5- عرفان:

يشكر الباحث الأستاذة الدكتورة سلوى الرملى على مساعدتها القيمة فى إنجاز هذا البحث.

الجمعية المصرية لهندسة اللغة

المؤتمر الخامس لهندسة اللغة

مجلد الأوراق البحثية

١٤-١٥ سبتمبر ٢٠٠٥م
القاهرة - جمهورية مصر العربية

التعرف على حروف اللغة العربية باستخدام خوارزمات التنحيف والتقطيع

أ.د/محمد يونس الحملاوى
كلية الهندسة-جامعة الأزهر

أ.د/محمد يسرى النحاس
كلية الهندسة-جامعة الأزهر

أ.د/إسماعيل عبد الغفار إسماعيل
الكلية الفنية العسكرية

م/محمد لبيب يوسف
إدارة نظم المعلومات

ملخص الورقة :

نقدم فى هذه الورقة منظومة تجريبية للتعرف على حروف اللغة العربية. الإطار المستخدم فى تصميم هذه المنظومة هو المنهج التركيبى للتعرف على الحروف. هذا المنهج يتعامل مع الحروف على أساس أنها أشكال هندسية مكونة من مقاطع أساسية و إضافية لها سمات قياسية وعلاقات مكانية مميزة لكل مقطع وبنية تركيبية لكل حرف. البرنامج المقترح للتعرف على الحروف العربية يقوم بمعالجة أولية للنص المأخوذ من الماسح الضوئى ثم يقوم بعملية تنحيف لهذه الحروف و بالتالى تحويلها إلى سلسلة متصلة من النقط ثم تصنيفها إلى مجموعات من المقاطع، و بتحليل العلاقات و المسافات بين نقاط المقاطع تتم عملية تمييزها ثم تركيبها للتعرف على الحروف. فى هذا الإطار تم تطوير خوارزم التنحيف "المحور الأوسط" وخوارزم تقطيع الكتابة باستخدام الوصف التركيبى لحروف اللغة العربية. تشير النتائج التجريبية الأولى إلى تحقيق دقة عالية فى تنحيف خطوط الكتابة وتقطيعها و معدل جيد من التعرف على حروف اللغة العربية.

الكلمات الدليلية :

التعرف على الحروف –المنهج التركيبى–خوارزم التنحيف – المحور الأوسط—تقطيع الكتابة– القارئ الآلى– البرمجة الشيئية.

١ –مقدمة:

يعتبر تاريخ البحث فى مجال التعرف على الحروف العربية قصير نسبياً مقارنة بنظيره فى التعرف على الحروف اللاتينية. إذ سجلت المحاولات الأولى للتعرف على الحروف العربية فى بداية الثمانيات [١،٢]. قدم الباحثون(أمين وآخرون) طرق تفاعلية للتعرف على الحروف العربية[٢]. واستخدم الباحثون(المعلم وآخرون) المدخل التركيبى لتقطيع الحروف إلى مقاطع والتعرف عليها باستخدام السمات الهندسية والمكانية للمقاطع[٢]. وقدم الباحثون (التكيرتى وآخرون) منطق الغموض فى التعرف بعض الحروف العربية المكتوبة باليد[٢]. واستخدم (الشيخ) تحويلات فورييه فى توصيف حدود الحروف العربية واستخدم (ماجت) عزوم زرنيك فى توصيف الحروف العربية[٣].

وبالرغم من هذه الجهود فما زال التعرف على حروف اللغة العربية يمثل صعوبة نسبية والمنظومات التى نجحت مازالت محدودة التطبيق لنوع معين من الكتابات العربية. وأهم الصعوبات التى تواجه البحث فى التعرف على الحروف العربية هى:

- أن حروف الكتابة العربية متصلة سواءً المطبوع منها أو المكتوب باليد
- أن كثيراً من هذه الحروف يحتوى على مكونات مركبة من النقط والهمزات وعلامات التشكيل.
- أن أشكال الحروف العربية تختلف باختلاف موقعها من السطر.
- كثرة وتنوع فنطات وأساليب الكتابة العربية. فقد قدرت عدد الفنطات المستخدمة فى الكتب بـــ ٤٥٠ ، وعدد الفنطات المستخدمة فى الصحف بـــ ١٥٠ [٢].

ويمكن تصنيف المنهجيات المتبعة فى التعرف على حروف اللغة العربية الطبوعة أو المكتوبة باليد إلى:

- المنهج التركيبى: وهى تتميز بالدقة والمرونة فى التعرف على الفنطات المختلفة للحروف ولكنها تعتمد إلى حد كبير على التحليل الدقيق للغة والتوظيف الجيد لطرق التطبيق
- المنهج الإحصائى: وهى تتميز بالدقة والمرونة فى التعرف على الحروف فى المضامين المختلفة وأشهرها الطرق المبنية على سلاسل ماركوف الخفية التى أثبتت نجاحات أخرى فى مجال التعرف على الكلام.
- الشبكات العصبية: و هى تتميز بسهولة بالتنفيذ وخاصة بالمعالجات المتوازية وإن كانت تحتاج إلى إمكانيات حسابية فائقة ووقتاً طويلاً للتعلم
- مناهج توفيقية بين أكثر من منهج مما سبق ذكرها

ونحن فى هذه الورقة نقدم منظومة تجريبية للتعرف على الحروف المطبوعة للغة العربية يمكن تطبيقها فى مجالات متعددة من تقنيات المعرفة. صممت هذه المنظومة فى إطار المنهج التركيبى للتعرف على الحروف بحيث لا يعتمد

معدل التعرف على الحروف على نوع أو حجم الفنط المستخدم. والورقمة البحثية منسقة كالآتى. فى المقطع الثانى نعرض الإطار المنهجى للتعرف على حروف اللغة العربية. فى المقطع الثالث نقدم الإجرائيات التفصيلية لبرنامج التعرف على الحروف العربية. فى المقطع الرابع نعرض النتائج التجريبية الأولية وتقييم معدلات التعرف فى الحالات المختلفة. والمقطع الخامس يلخص نتائج البحث واتجاهات البحث المستقبلية.

٢-الإطار المنهجى للتعرف على الحروف:

الإطار المستخدم فى تصميم منظومة التعرف على الحروف العربية هو المنهج التركيبى للتعرف على الحروف وهو ينقسم إلى ثلاث مراحل هى:

المرحلة الأولى: مرحلة ما قبل المعالجة:

فى هذه المرحلة تتم العمليات الآتية :

١. قراءة ملف النص المأخوذ من الماسح الضوئى
٢. تنقية بيانات النص بمعنى إزالة النقط السوداء الزائدة التى تنتج أثناء عملية المسح الضوئى وذلك باستخدام مرشح "ملح وفلفل" على سبيل المثال.
٣. تحويل الصورة إلى التمثيل الثنائى :الغرض من هذه الخطوة تخصيص أحد اللونين فقط لكل نقطة من النقط المضيئة التى يتكون منها الملف ، ويتم ذلك بتحديد القيمة المتوسطة لشدة إضاءة جميع النقط ويخصص اللون الأسود لجميع النقط التى يزيد وزنها عن هذه القيمة المتوسطة ، ويخصص اللون الأبيض لباقى النقط.
٤. إزالة الحزات التى تظهر أحيانا على أطراف الحروف على هيئة دوائر صغيرة جدا يمكن التعرف عليها بأنها نقطة واحدة بيضاء محاطة من جميع الجهات بنقط سوداء وسيتخدم لذلك مرشح "الحزات".

المرحلة الثانية: مرحلة المعالجة و التحليل:

فى هذه المرحلة تتم العمليات الآتية :

١. تنحيف النص: تطبيق خوارزم التنحيف على تجمعات النقط السوداء التى يتكون منها النص فينتج عن ذلك تقليص سمك هذه التجمعات ليصبح كل منها عبارة عن سلسلة متلاصقة من النقط السوداء لايزيد سمكها عن نقطة واحدة
٢. تحديد مواصفات السطر-المحور الرئيسى: تحديد مواصفات السطر وأهمها السمك ويقصد به الفرق بين أكبر وأصغر إحداثى أفقى لنقطتين لهم نفس الإحداثى الرأسى ، وبفرض أن جميع سطور النص مكتوبة بحجم ثابت فإنه يمكن إدراك حدود السطور التالية حتى وإن كانت لاتضم الحد الأدنى من سلاسل النقط المتلاصقة
٣. تقطيع النص: تمييز كل سلسلة من النقط المتلاصقة واعتبارها شىء منفصل بذاته ، وكل واحدة من هذه السلاسل قد تعبر عن حرف منفصل أو مقطع من كلمة يضم حرفين أو أكثرأو نقطة منعزلة (وهى التى تظهر أعلى أو أسفل الحروف ب ، ج ،خ ، ن ، ض ، ظ ، غ ، ف ، ذ ، ز) أو نقطتين متلاصقتين (وهى التى تظهرأعلى أو أسفل الحروف ت ، يـ) أو ثلاثة نقط متلاصقة (وهى التى تظهر على الحروف ث ،ش).

المرحلة الثالثة: مرحلةالتعرف:

فى هذه المرحلة تتم العمليات الآتية :

١. تقدير سمات المقاطع: بالنسبة لكل مقطع يتم تحديد المحور الرئيسى، والمستطيل المحدد لكل نقاط سلسلة المقطع، و مركزالمقطع أى النقطة المتوسطة فى هذه السلسلة و الجزء الواقع منه على المحور الرئيسى المذكور أعلاه ، ويجب ألايقل عرض هذا الجزء عن حد معين ، وبمعلومية هذه الأجزاء لجميع مقاطع السطرالواحد يتم تحديد المحور الرئيسى لهُ ، وفى حالة تعذر ذلك يستفاد من معرفة سمك السطور السابقة.
٢. التعرف على الحروف: والفكرة الأساسية لأسلوب تمييز الحروف هى أن الأشكال الهندسية للحروف المتصلة فى اللغة العربية عبارة عن تراكيب يجمعها محور أفقى ، يلتصق بعضها بهذا المحور من أعلاه وبعضها من أسفله ، ويتميز كل حرف بالطول النسبى لهذه التراكيب واتجاه ميلها على هذا المحور ، وكذا موقع الحرف فى المقطع سواء فى بدايته أومنتصفه أو آخره ، أو أن يأتى الحرف منعزلا بمفرده

يتم التعرف على الحروف المحتوية لكل مقطع من خلال مواصفاته الهندسية والمكانية التى يمكن إيجازها فى المحددات الآتية:

١. علاقة النقطة الطرفية للحرف بنقطة التلاقى مع المحور الرئيسى (مقارنة الإحداثيات)
٢. تحديد النبرات (حروف ب ، ت ، س ، نـ ، .....) وعصا حرف الطاء والظاء بمقارنة طولها بحرف الألف.
٣. تحديد الدوائر المغلقة (حروف ص ، ض ، ط ، ظ ، ف ، ق ، م ، هـ ، و) .
٤. بناء على هذه المحددات يمكن تحديد المجموعة التى ينتمى إليها هذا الحرف (ب ، ت ، ث) أو (ج ، ح ، خ) وهكذا ، وتميز هذه المجموعات بأكواد خاصة.
٥. يستكمل التصنيف طبقا للقواعد الآتية:

- ربط النقط المنعزلة بالحروف التى تقع فى الجوار فتتحول النبرة إلى ب أو ن ، والدائرة المغلقة إلى ف أو ض ، وحرف ح إلى ج أو خ.
- ترقية بعض الحروف مثل ن لتصبح ت أو ب لتصبح بـ أو س لتصبح ش.
- ربط الهمزات بالحروف التى تقع فى الجوارلتتحول ا إلى أ والنبرة إلى نـ والواو إلى ؤ وهكذا.

والقواعد التى نطبقها فى توصيف حروف اللغة العربية محددة فى الجدول (١) والجدول (٢).

جدول (١) توصيف حروف اللغة العربية :الحروف البسيطة

الحرف		موقعه فى المقطع	مكوناته	علاقة ق.م بالمحور	علاقة ق.م بالق.م الإحداثى السينى
ا	ا	منفصل		أعلى	=
	ـا	آخره			
ب ت ث	ب	منفصل	نبره	أعلى	=
	بـ	أوله			
	ـبـ	منتصفه			
	ـب	آخره			
ج ح خ	جـ	أوله		أعلى	>
	ـجـ	منتصفه			
د ذ	د	منفصل	نبره	أعلى	=
	ـد	آخره			
ر ز	ر	منفصل		أسفل	>
	ـر	آخره			
س ش	سـ	أوله	ثلاث نبرات	أعلى	=
	ـسـ	منتصفه			
ص ض	صـ	أوله	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	ـصـ	منتصفه	نبره	أعلى	=
ط ظ	ط	منفصل	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	طـ	أوله			
	ـطـ	منتصفه	ا	أعلى	=
	ـط	آخره			
ع غ	عـ	أوله	نبره	أعلى	<
	ـعـ	منتصفه		أعلى	=
ف ق	ف	منفصل	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	ـف	آخره	نبره	أعلى	=
	فـ	أوله	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	ـفـ	منتصفه			
ك	ك	منفصل	ا	أعلى	=
	ـك	آخره	همزة		
	كـ	أوله		أعلى	<
	ـكـ	منتصفه			
ل	ل	منفصل	ا	أعلى	=
	ـل	آخره	نبره		
	لـ	أوله	ا	أعلى	=
	ـلـ	منتصفه			
م	مـ	أوله	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	ـمـ	منتصفه			
ن	ن	منفصل	نبره	أعلى	=
	نـ	أوله			
	ـنـ	منتصفه			
هـ	ه	منفصل	دائرة مغلقة	ليس لها نقطة نهاية	لها نقطتى تقاطع
	ـه	آخره			
ى	يـ	أوله	نبره	أعلى	=
	ـيـ	منتصفه			

جدول (٢) توصيف حروف اللغة العربية :الحروف المركبة

الحرف		موقعه فى المقطع	الجزء العلوى		الجزء السفلى	
			علاقة ق٢ بالمحور	علاقة ق١ـق٢ الإحداثى السينى	علاقة ق٢ بالمحور	علاقة ق١ـق٢ الإحداثى السينى
ج ح خ	ح	منفصل	أعلى	>	أسفل	<
	ـحـ	أخره				
س ش	س	منفصل	أعلى	=	أسفل	>
	ـسـ	أخره				
ص ض	ص	منفصل	ليس لها نقطة نهاية	لها نقطتى تقاطع	أسفل	>
	ـصـ	أخره				
ع غ	ع	منفصل	أعلى	<	أسفل	<
	ـعـ	أخره	أعلى	=		
ق	ق	منفصل	ليس لها نقطة نهاية	لها نقطتى تقاطع	أسفل	>
	ـقـ	أخره				
ل	ل	منفصل	ا		أعلى	
	ـل	أخره	نبره			
	لـ	أوله	ا		أعلى	
	ـلـ	منتصفه				
م	م	منفصل	دائرة مغلقة		أسفل	=
	ـمـ	أخره				
ن	ـن	أخره	أعلى	=	أسفل	>
هـ	هـ	أوله	دائرتان مغلقتان			
	ـهـ	منتصفه	دائرة مغلقة		دائرة مغلقة	
و	و	منفصل	دائرة مغلقة		أسفل	>
	و	أوله				
	ـو	أخره				
ى	ى	منفصل	نبره		أسفل	<
	ـيـ	أخره			أسفل	>

وسنقدم فيما يلى من البحث شرحا للطرق المستخدمة فى المرحلتين الثانية والثالثة

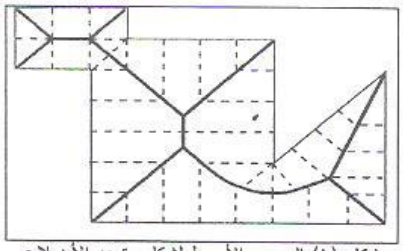
٢-١ تنحيف النص:

عملية تنحيف النص تعنى تحويل خطوط الكتابة السميكة إلى خطوط ومنحنيات سمك كل منها نقطة واحدة. وتعتبر عملية التنحيف من العمليات الأساسية فى طرق التعرف على الحروف باستخدام سماتها الشكلية الثابتة مع تغير فنط الكتابة أو درجة ميل سطور الكتابة. والمهتم بمعرفة هذه الطرق سيجد فى البحث المنشور[٥] مسح مرجعى وافى لطرق تنحيف الحروف. وتنحصر طرق التنحيف فى: طرق تتابعية[٤،٧،٦،٥]، و طرق متوازية[٩]. و تعتمد الطرق التتابعية على استخدام القوالب والنوافذ فى تنحيف و تقليم الأشكال. وتعمل الطرق المتوازية باستخدام نفس أساسيات التنحيف السابقة ولكنها تطبق على مراحل متوالية وبالتالى فيمكنها استخدام المعالجات المتوازية مما يزيد من سرعة المعالجة.وكذلك توجد طرق تعتمد على الشبكات العصبية[١٠].ومازالت كفاءة أداء هذه الطرق و دقتها وسرعتها محل بحث فى مجالات تحليل صور الوثائق. والكثير منها يتأثر بدوران الأشكال بالنسبة لمحاور الصفحة. واقترحت مؤخراً طرق لاتعتمد على دوران أشكال الحروف[١١].

فى هذا البحث نقترح استخدام خوارزم "المحور الأوسط" لتنحيف أشكال حروف النصوص العربية. فقد وجد فى دراسة سابقة عن طرق توليد المسارات لحركة الروبوت أن طريقة المحور الأوسط لها سمة أساسية وهى البحث عن المسار الوسيط وتحقق دقة عالية فى إيجاد هذا المسار أيا كانت درجة دوران الأشكال فى المستوى أو مدى ضيق المسافة بين العوائق مما يجعلها مناسبة لتنحيف الحروف العربيّة التى تتميز بتعقيد أشكالها.

خوارزم المحور الأوسط:

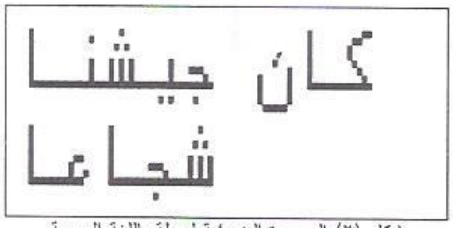
يعرف المحور الأوسط لشكل متعدد الأضلاع بأنه المحل الهندسى لجميع النقط التى يتساوى بعدها عن جميع أضلاع هذا الشكل[١٢] ، ويوضح الشكل (١) مثال لشكل متعدد الأضلاع و المحور الأوسط له ، ويلاحظ أن هذا المحور الأوسط يمثل "تنحيف" لهذا الشكل ولذا نستطيع أن نطلق عليه خوارزم التنحيف.



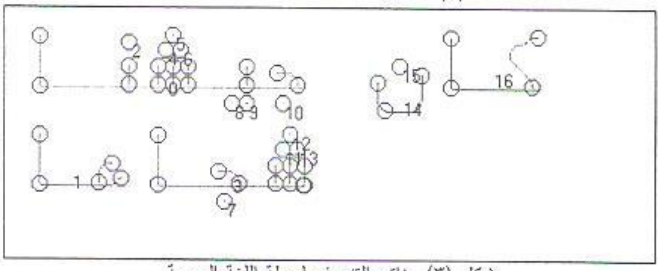
شكل (١) المحور الأوسط لشكل متعدد الأضلاع

يمكن استخدام "خوارزم المسافات الإقليدية" أو "خوارزم الهيكلة المورفولوجية" فى إيجاد المحور الأوسط للأشكال الهندسية. وهى عمليات متكافئة فى النتيجة. و فى هذا البحث تم تطوير "خوارزم تنحيف مورفولوجى" يعمل تكراريا على حذف نقاط حدود الشكل مستخدما نافذة مورفولوجية من ٣×٣ عنصرا حتى يتوقف التغيير فى هيكل الشكل.

ويوضح الشكل (٢) أن حروف اللغة العربية بعد أن نحصل على الصورة الضوئية لها تكون عبارة عن تجمعات غير منتظمة من الأشكال ، ولذا فإنه يمكننا أن نتعامل معها بخوارزم التنحيف بحيث تتحول إلى سلسلة متصلة من النقط كما يبين الشكل (٣)



شكل (٢) الصورة الضوئية لجملة باللغة العربية



شكل (٣) ناتج التنحيف لجملة اللغة العربية

٢-٢تقطيع النص:

يعتبر تقطيع الحروف من العمليات الهامة والأكثر صعوبة فى التعرف على حروف اللغة العربية. والغرض من عملية التقطيع تقسيم كلمات الكتابة إلى أشكال مميزة يمكن استخدامها فى التعرف على الحروف والكلمات. وعملية التقطيع تتلخص فى الخطوات العامة الآتية:

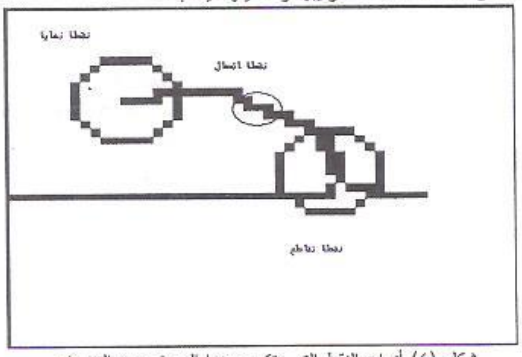
١. بدءاً من نقطة فى صورة الصفحة
٢. أوجد مكان الشكل التالى
٣. استخرج السمات المميزة لصورة الشكل الحالى
٤. أوجد الرمز الأكثر تماثلاً من مجموعة الرموز مع سمات الشكل الحالى
٥. استمر حتى نهاية صور الأشكال فى الصفحة

و يوجد ثلاث استراتجيات متبعة فى تقطيع النصوص[١٣] وهى:

١. الطريقة التقليدية والتى تقطع الصورة إلى مكونات مميزة بناءً على خواص الحروف
٢. التقطيع بالتعرف على أشكال الحروف
٣. الطريقة الكلية والتى تبحث عن الكلمات ككل

ونحن فى هذا البحث ننهج الطريقة التقليدية مع استخدم الخواص المميزة لمقاطع الحروف العربية فى عمليات تقطيع صورة النص العربى. وتتلخص هذه الطريقة فى الخطوات التالية:

١. تصنف النقط المكونة لكل مقطع إلى أربعةً أصناف كما هو موضح بالشكل (٤)
   - نقطة منعزلة : وهى تلك التى ليس لها أى اتصال بنقط أخرى
   - نقطة نهاية : وهى تلك التى لها اتصال بنقط أخرى فى اتجاه واحد فقط ق هـ
   - نقطة اتصال : وهى تلك التى لها اتصال بنقط أخرى فى اتجاهين.
   - نقطة تقاطع : وهى تلك التى لها اتصال بنقط أخرى فى أكثر من اتجاهين
٢. بمقارنة الإحداثيات الرأسية لنقط الإتصال فى كل مجموعات النقط ، يتم تحديد المحور الرئيسى ثم فصله
٣. نتيجة عملية الفصل تتبقى أجزاء من مجموعات النقط المتلاصقة السابق تحديدها كأشياء منفصلة كما يوضح الشكل (٥).
٤. بدءاً من كل نقطة من نقط النهاية يتم تتبع النقط الملاصقة لها إلى أن نصل إلى آخر نقطة كانت تربط هذه السلسلة بالمحور الرئيسى ق ر.
٥. ترقم كل سلسلة من النقط المتلاصقة لترتيبها وتصنيفها فيما بعد.



شكل (٤) أنواع النقط التى يتكون منها الحرف بعد التنحيف

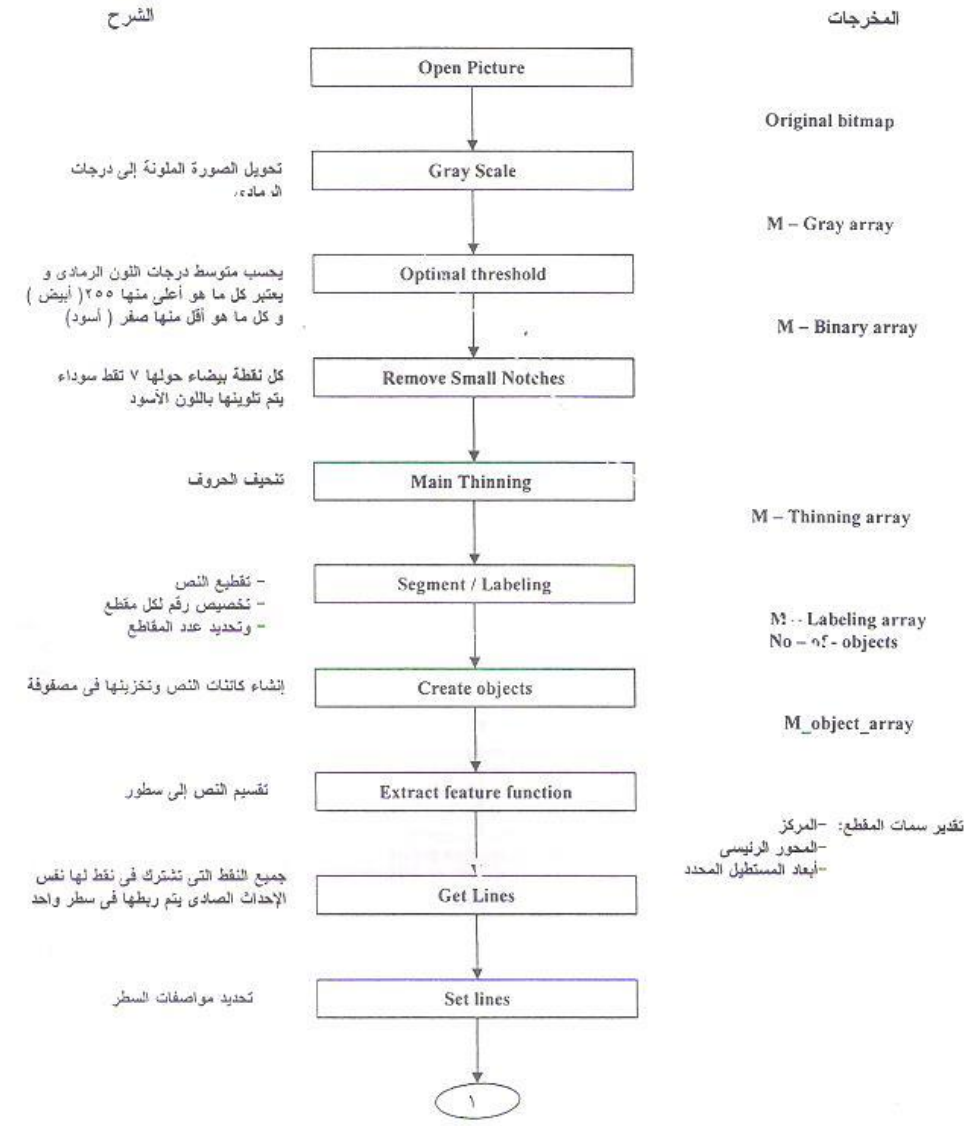
شكل (٥) شكل الجملة بعد التنحيف وفصل المحور الرئيسى

٣-برنامج التعرف على حروف اللغة العربية :

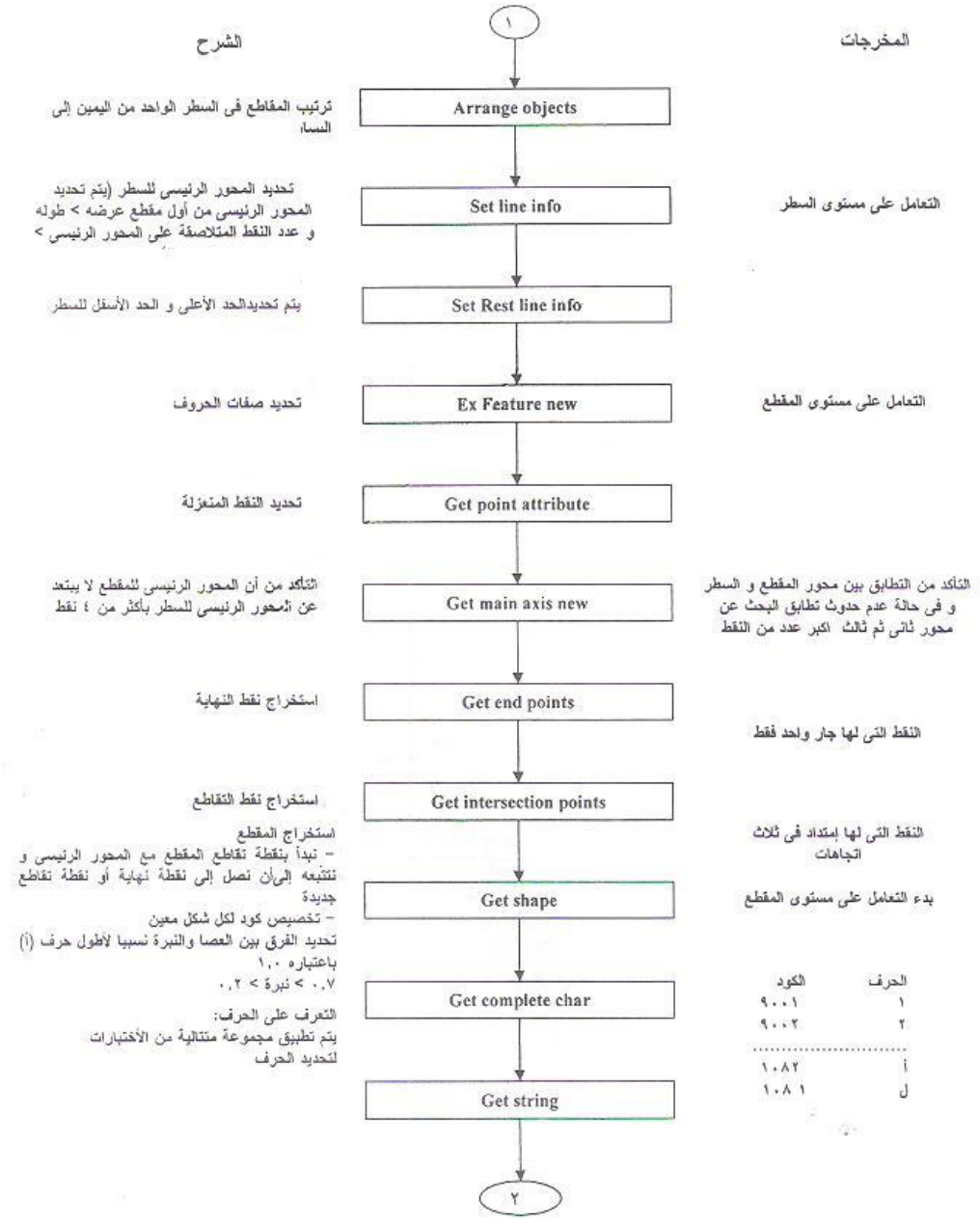
بعد قراءة صورة الصفحة ومعالجتها يعمل برنامج التعرف طبقا للتسلسل التالى :

١. تنحيف النص باستخدام خوارزم **التنحيف** ليتحول النص إلى سلاسل من النقط المتلاصقة
٢. تمييز كل سلسلة واعتبارها "كائن" منفصل قائم بذاته .
٣. تقطيع كائنات السلاسل الناتجة من عملية التنحيف إلى مقاطع فرعية باستخدام خوارزم **التقطيع**
٤. استخراج سمات كل مقطع وهى:
   - المحور الرئيسى
   - المستطيل المحدد لكل نقاط سلسلة المقطع
   - مركز المقطع أى النقطة المتوسطة فى هذه السلسلة قم
٥. ترتيب المقاطع فى السطر الواحد من اليمين إلى اليسار
٦. يصنف كل مقطع إلى حرف من الحروف طبقاً للقواعد الآتية:
   - المقطع ينتمى إلى المجوعة الأولى من الحروف السيطة و يوضحها جدول (١) إذا كان يقع بكملها أعلى أو أسفل المحور الرئيسى
   - المقطع ينتمى إلى المجوعة الثانية من الحروف المركبة و يوضحها جدول (٢) إذا كان يقع جزء منه أعلى المحور الرئيسى و جزء منه أسفل المحور الرئيسى
   - بناء على اتجاه النقطة قم بالنسبة للمحور الرئيسى و بمقارنة الإحداثى السينى للنقطتين قمـ ، قم و كذا علاقة كل مقطع بما قبله و ما بعده يتم التعرف على الحرف طبقاً للتوصيف الموضح بالجدول (١) و الجدول (٢).

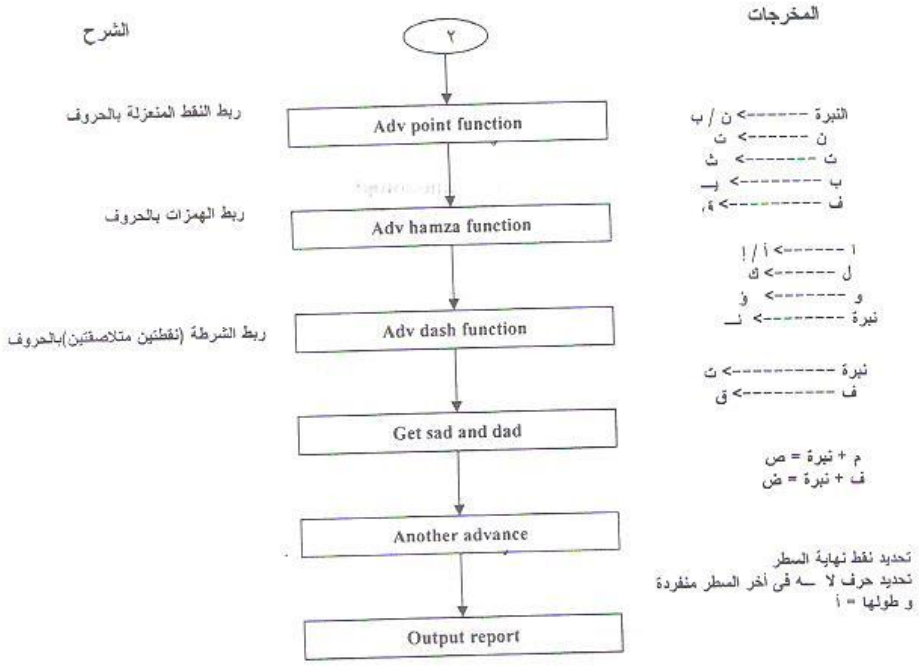
وشرح الإجرائيات التفصيلية لهذا البرنامج ومخرجات كل إجرائية موضح فى مخطط التدفق بالشكل (٦)



شكل (٦) مخطط تدفق البيانات لبرنامج التعرف على حروف اللغة العربية



تابع شكل (٦) مخطط تدفق البيانات لبرنامج التعرف على حروف اللغة العربية



تابع شكل (٦) مخطط تدفق البيانات لبرنامج التعرف على حروف اللغة العربية

٤-النتائج العملية:

للتحقق من صحة المنهج المقترح فى التعرف على الحروف العربية واختبار أداء البرنامج تم اختيار عينات مختلفة ممثلةللنصوص العربية المطبوعة وتحتوى جميع التراكيب الممكنة للحروف العربية داخل الكلمات وكتابتها بثلاث فنطات شائعةالاستخدام (Simplified Arabic-Arabic Transparent-Times New Roman) ثم إدخالها إلى الحاسب باستخدام الماسح الضوئى. وقد استخدمت هذه العينات فى مرحلة التعلم وكانت نتائجها كما فى الجدول (٣).

جدول (٣) نسبة تمييز الحروف فى عينات التعلم

الفقرة	عدد الحروف	الحروف المقروءة	نسبة التمييز
١	٣٢١	١٨٦	%٥٨
٢	٢٧٧	٢١٣	%٧٦
٣	٢٩٧	١٠٧	%٦٩
٤	٢٢٩	١٢٥	%٥٤
٥	١٦٥	١٢٠	%٧٢
٦	١٣٠	١٠٤	%٨٠
٧	١٦١	١٣٠	%٨٠
الإجمالي	١٥٨٠	١٠٨٥	%٦٨

وفى مرحلة تقييم أداء البرنامج اختيرت عينات للنصوص المستقطعة من الصحف اليومية والكتب المدرسية. والشكل(٧) به صورة نص مستقطع من إحدى الصحف اليومية المصرية. و الشكل (٨) به نتيجة تنحيف وتقطيع هذا النص. ونتيجة التعرف موضحة فى الشكل (٩).

سجلت أسعار الجملة للسلع الزراعية انخفاضا ملحوظا خلال شهر مايو الماضي لهذا العام مقارنة بالأسعار في نفس الفترة من العام الماضي بلغت نسبته١،٢% أشار الي ذلك تقرير الأرقام القياسية لأسعار الجملة الذي يصدره الجهاز المركزي للتعبئة العامة والاحصاء بصفة شهرية.

وأشار التقرير الي تراجع أسعار مجموعة من المنتجات من أهمها الخضراوات والفواكه والبيض وغيرها حيث انخفضت اسعار الخضراوات بنسبة٨١,١ % والفواكه بنسبة٦,٨ % والبيض بنسبة١١,٣ % .

وصرح اللواء أبو بكر الجندي رئيس الجهاز المركزي للتعبئة العامة والاحصاء بأن أسعار بعض السلع الاخري في اسواق الجملة سجلت ارتفاعا قدره٤,٢ % خلال شهر مايو الماضي مقارنة بالاسعار في نفس الفترة من العام الماضي

مثل المواد الغذائية والوقود ومنتجاته فقد ارتفعت اسعار الوقود بنسبة١٣,٦ % والمواد الخام بنسبة١,٨ % والسلع الوسيطة بنسبة٦,٨ % والسلع الاستهلاكية غير المعمرة بنسبة٣,١ % والسلع الاستهلاكية المعمرة بنسبة٣ % والسلع الاستثمارية

شكل (٧) صورة النص المستقطع من إحدى الصحف اليومية

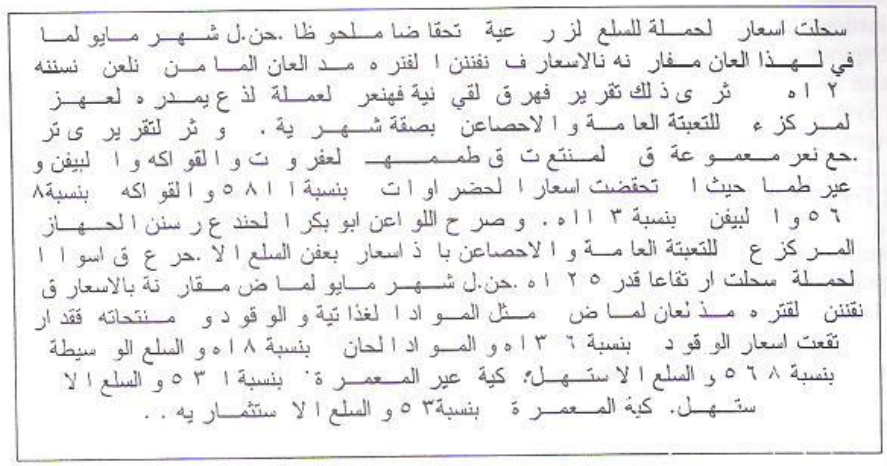
سجلت أسعار الجملة للسلع الزراعية انخفاضا ملحوظا خلال شهر مايو الماضي لهذا العام مقارنة بالأسعار في نفس الفترة من العام الماضي بلغت نسبته١،٢% أشار الي ذلك تقرير الأرقام القياسية لأسعار الجملة الذي يصدره الجهاز المركزي للتعبئة العامة والاحصاء بصفة شهرية.

وأشار التقرير الي تراجع أسعار مجموعة من المنتجات من أهمها الخضراوات والفواكه والبيض وغيرها حيث انخفضت اسعار الخضراوات بنسبة٨١,١ % والفواكه بنسبة٦,٨ % والبيض بنسبة١١,٣ % .

وصرح اللواء أبو بكر الجندي رئيس الجهاز المركزي للتعبئة العامة والاحصاء بأن أسعار بعض السلع الاخري في اسواق الجملة سجلت ارتفاعا قدره٤,٢ % خلال شهر مايو الماضي مقارنة بالاسعار في نفس الفترة من العام الماضي

مثل المواد الغذائية والوقود ومنتجاته فقد ارتفعت اسعار الوقود بنسبة١٣,٦ % والمواد الخام بنسبة١,٨ % والسلع الوسيطة بنسبة٦,٨ % والسلع الاستهلاكية غير المعمرة بنسبة٣,١ % والسلع الاستهلاكية المعمرة بنسبة٣ % والسلع الاستثمارية

شكل (٨) نتيجة تحليف وتقطيع صورة النص المستقطع من إحدى الصحف اليومية


سجلت اسعار لحمـــلة للسلع لز ر عية تحقا ضا مـــلحو ظا .حن.ل شـــهـــر مـــايو لمـــا في لـــهـــذا العان مـــفار نه نالاسعار ف نفننن ا لفتر ه مـــد العان المـــا مـــن نلعن نسننه ٢ ا ه ثر ى ذ لك تقر ير فهر ق لقي نية فهنعر لعمـــلة لذ ع يمـــدر ه لعـــهـــز لمـــر كز ء للتعبئة العا مـــة و ا لاحصاعن بصفة شـــهـــر ية . و ثر لتقر ير ى تر .حع نعر مـــعمـــو عة ق لمـــنتع ت ق طمـــمـــهـــ لعفر و ت و ا لقو اكه و ا لبيفن و عير طمـــا حيث ا تحفضت اسعار ا لحضر او ا ت بنسبة ا ا ٨ ٥ ه و ا لقو اكه بنسبة٨ ٥ ٦ و ا لبيفن بنسبة ٣ ا ا ه . و صر ح اللو اعن ابو بكر ا لحند ع ر سنن ا لحـــهـــاز المـــر كز ع للتعبئة العا مـــة و ا لاحصاعن با ذ اسعار بعفن السلع ا لا .حر ع ق اسو ا ا لحمـــلة سجلت ار تفاعا قدر ٢ ٥ ا ه .حن.ل شـــهـــر مـــايو لمـــا ض مـــقار نة بالاسعار ق نفننن لفتر ه مـــذ نعان لمـــا ض مـــثل المـــو اد ا لغذا ئية و الو قو د و مـــنتحاته فقد ار تفعت اسعار الو قو د بنسبة ٦ ٣ ا ه و المـــو اد ا لحان بنسبة ٨ ا ه و السلع الو سيطة بنسبة ٨ ٦ ٥ و السلع ا لا ستـــهـــل، كية عير المـــعمـــر ة بنسبة ا ٣ ٥ و السلع ا لا ستـــهـــل. كية المـــعمـــر ة بنسبة٣ ٥ و السلع ا لا ستثمـــار يه . .

شكل (٩) نتيجة التعرف على الحروف فى صورة النص المستقطع من إحدى الصحف اليومي

وقد جاءت نتائج تقييم معدلات التعرف على الحروف العربية بناء على هذه العينات كما هو مبين فى الجدول (٤).

جدول (٤) نسبة تمييز الحروف على عينات الاختبار

عدد الحروف	الحروف المقروءة	نسبة التمييز
٧٢٣	٥٧٢	٧٨,٧٨٧٨٨

هذه النتيجة ٧٨,٧٨٧٨٨% أفضل من الأداء المتوسط المحسوب ٦٨% لأنها تمثل عينة واحدة من عينات الاختبار مقارنة بالمتوسط المحسوب بالنسبة لـــ ٢٧ عينة من النصوص.

ومعظم الأخطاء نجمت من الخلط بين الحروف المنقوطة ومن الخلط بين حروف الـــ س و الـــ ش مع النبرات المتكررة لحروف بـــ ، نـــ ، تـــ والخلط بين الهمزات وحرف ع . وكذلك تعريف بعض الرموز الرياضية مثل %.هذه الأخطاء تتطلب الأخذ فى الاعتبار السمات المورفولوجية للحروف بجانب السمات التوبولوجية. وهو اتجاه التطوير المقترح.

٥-الخلاصة :

هذه الورقة تلخص نتائج تصميم منظومة تجريبية للتعرف على حروف اللغة العربيةتعتمد على المنهج التركيبى للتعرف على الحروف. تم تنفيذ البرنامج بلغة (C#). للتعرف على الحروف العربية يقوم بمعالجة أولية للنص المأخوذ من الماسح الضوئى ثم يقوم بعملية تنحيف لهذه الحروف و بالتالى تحويلها إلى سلسلة متصلة من النقط ثم تصنيفها إلى مجموعات من المقاطع، و بتحليل العلاقات و المسافات بين نقاط المقاطع تتم عملية تمييزها ثم تركيبها للتعرف على الحروف. فى هذا الإطار تم تطوير خوارزم التنحيف "المحور الأوسط" وخوارزم "تقطيع الكتابة" باستخدام الوصف التركيبى لحروف اللغة العربية. تشير النتائج التجريبية الأولى إلى تحقيق دقة عالية فى تنحيف خطوط الكتابة وتقطيعها و معدل جيد من التعرف على حروف اللغة العربية يصل إلى ٧٨% عمليا.

هذه النتائج أقل من النتائج المنشورة [١٤] عن أداء برنامج صخر ٩٠,٣٣% وبرنامج "أومنيبيج" ٨٦,٨٩%.

هذا البرنامج متاح على شبكة الإنترنت وهو تحت التطوير ومحاور التطوير المقترحة هى تدقيق قواعد تقطيع النص وتطوير قواعد توصيف الحروف العربية لتحتوى السمات المورفولوجية بالإضافة إلى السمات التوبولوجية.

المراجع :

[1] Gheith A. Abandah , Mohammed Zeki Khedher, "Printed and Handwritten Arabic Optical Character Recognition – Initial Study", University of Jordan, Faculty of Engineering and Technology, Amman, August 2004,

[2]H. Al-Yousefi and S. S. Udpa, "Recognition of Arabic Characters", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14,no. 8, Aug.. 1992,

[3] Maget Mahomed Mahmoud Fahmy*, Hytham El-Messiry**, "Automatic Recognition of Typewritten Arabic Characters Using Zernike Moments as a Feature Extractor", *Computer Sience Department, College of Science, University of Bahrayn, Isa Town,

**Informatics Research Institute, Mubarak City for Scientific Research and Technological Applications, Mansheyat Al-Olama, Alexandria, Egypt,
web:http//www.ici.ro/ici/revista/sic2001_3/art4.html,
[4]Issam Bazzi, Richard Schwartz, and John Makhoul , "An Omnifont Open-Vocabulary OCR System for English and Arabic", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 6, Jun. 1999,
[5] L. Lam, S. Lee, and C. Suen, "Thinning Methodologies—A Comprehensive Survey," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, no. 9,pp. 869-885, Sept. 1992,
[6] Paul C.K.Kwok, "A Thinning Algorithm by Contour Generation", Department of Computer Science, University of Calgary.2500 University Drive NW, Calgary, Canada T2N 1N4,
[7] Tal Steinherz, Nathan Intrator and Ehud Rivlin, "A Special Skelonization Algorithm for Cursive Words,
[8] Jung-Me Park, Hui-Chuan Chen, Shu T. Huang., "A New Gray Level Edge Thinning Method", University of Alabama, Tuscaloosa, AL 35487,
[9] L. Lam and C. Suen, "An Evaluation of Parallel Thinning Algorithms for Character Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence,vol. 17, no. 9, pp. 914-919, Sept. 1995,
[10] M. Altuwaijri and M. Bayoumi, "A New Thinning Algorithm for Arabic Characters Using a Self-Organizing Neural Network," Proc. IEEE Int'l Symp. Circuits and Systems, vol. 3, pp. 1824-1827, 1995,
[11]Maher Ahmed and Rabab Ward, "A Rotation Invariant Rule-Based Thinning Algorithm for Character Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24,no. 12, Dec. 2002,
[12]Steven S. Skiena, "The algorithm design manual",Springer Telos, NY. 1998.
[13]Richard G. Casey and Eric Lecolinet, "Survey of Methods and Strategies in Character Segmentation", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 7, Jul. 1996,
[14] Tapas Kanungo, Gregory A. Marton, Osama Bulbul, " Performance Evaluation of Two Arabic OCR Products", Center for Automation Research University of Maryland, College Park, MD 20742, web:http//www.cfar.umd.edu/~kanungo

المؤتمر الثانى لهندسة اللغة

تنظمه

الجمعية المصرية لهندسة اللغة

تحت رعاية

الأستاذ الدكتور / حسن أحمد غلاب

رئيس جامعة عين شمس

١٨ أبريل ١٩٩٩

القاهرة – جمهورية مصر العربية

الأبحاث المقبولة

الجمعية المصرية لهندسة اللغة
THE EGYPTIAN SOCIETY
OF LANGUAGE ENGINEERING

The Second Conference
on Language Engineering
18 April 1999

عنوان البحث

تجانس شكلى منظومة الأرقام العربية مع أشكال الحروف العربية وأشكال حروف لغات أخرى

أ. د / محمد يونس الصلاوى

د / محمد يسرى النحاس

تجانس شكلىّ منظومة الأرقام العربية مع أشكال الحروف العربية وأشكال حروف لغات أخرى

أ. د. محمد يونس الحملاوى
قسم هندسة النظم والحاسبات
كلية الهندسة جامعة الأزهر
القاهرة

د. محمد يسرى النحاس
قسم هندسة النظم والحاسبات
كلية الهندسة جامعة الأزهر
القاهرة

نبذة

يهدف هذا البحث إلى دراسة مدى التجانس بين فئتى منظومة الأرقام العربية (العربية المشرقية والغبارية المغربية) وأشكال الحروف فى اللغة العربية وبعض اللغات الأخرى. ولقد أجرينا دراسة على كل من أشكال حروف اللغات الآتية: اللغة العربية، واللغة اللاتينية، واللغة اليونانية، واللغة السنسكريتية (الهندية) ومقارنة ذلك مع أشكال منظومة الأرقام العربية بفئتيها وصنفت الأشكال من خلال شجرة وراثية. وهذا الاختيار يتيح لنا تمحيص وجه الحقيقة فى الجدل الدائر حول الارتباط التاريخى بين شكل الأرقام العربية والحضارة الهندية القديمة والحضارة الإسلامية العربية.

بالنسبة للعينة الأولى والمكونة من أشكال الحروف العربية، أظهرت النتائج تجانساً أكبر لفئة الأرقام العربية المشرقية مع أشكال الحروف، بينما كان التجانس بين فئة الأرقام الغبارية المغربية وأشكال الحروف العربية أقل. ولقد تم عمل شجرة وراثية تضم أشكال فئتى الأرقام وحروف اللغة العربية مع حوف اللغة اللاتينية وصنفت الأشكال من خلال تلك الشجرة الوراثية وأدى ذلك إلى ظهور تجانس للأرقام العربية المشرقية مع الحروف العربية عند المستوى الأدنى من الشجرة الوراثية مما يبين أن أصولها التاريخية أقرب إلى العربية. وعلى النقيض من هذا ظهر تجانساً بين الأرقام الغبارية المغربية وأشكال الحروف العربية عند المستوى الأعلى من الشجرة مما يبين أن أصولها التاريخية أبعد عن العربية.

أما بالنسبة للعينة الثانية والمكونة من أشكال الحروف السنسكريتية (الهندية)، فقد كانت النتائج على العكس من التجربة السابقة حيث أظهرت النتائج تجانساً أكبر لفئة الأرقام الغبارية المغربية مع أشكال الحروف السنسكريتية الهندية، بينما كان التجانس بين فئة الأرقام العربية المشرقية وأشكال الحروف السنسكريتية الهندية أقل.

وبالنسبة للعينة الثالثة والمكونة من أشكال الحروف اللاتينية، فقد أظهرت النتائج تجانساً أكبر

لفئة الأرقام الغبارية المغربية مع أشكال الحروف اللاتينية، بينما كان التجانس مع فئة الأرقام العربية المشرقية أقل.

وبالنسبة للعينة الرابعة والمكونة من أشكال الحروف اليونانية، فقد أظهرت النتائج تجانساً أكبر لفئة الأرقام الغبارية المغربية مع أشكال الحروف اليونانية، بينما كان التجانس مع فئة الأرقام العربية المشرقية أقل.

١ مقدمة:

يهدف هذا البحث إلى عمل دراسة مقارنة بين درجة التوافق الشكلى لكل من فئتى منظومة الأرقام العربية المستخدمتين في منظومة الأرقام العشرية مع منظومة الكتابة العربية. هاتان الفئتان تسميان خطأً فئة الأرقام الهندية وفئة الأرقام العربية[١]. بينما الأصح تسميتهما بالأرقام العربية المشرقية والأرقام المغربية الغبارية. والبحث يستهدف وضع الأسس العلمية لتوصيف منظومة الكتابة العربية آخذين فى الاعتبار كل الخواص والمعايير.

والخاصية التى نبحثها هنا هى درجة توافق بنية أشكال الأرقام فى كل فئة مع منظومة الكتابة العربية وكتابات عض اللغات الأخرى. إن علم التعرف على الأنماط يتيح لنا استخدام القياس الكمى فى دراسة الخواص البنيوية الشكلية للأنماط من خلال عدة معايير موضوعية. والحكم الموضوعى على درجة التوافق يخضع لعدة معايير منها:

١-درجة تنافر نمط أشكال الأرقام فى كل فئة مع نمط أشكال الحروف العربية.

٢-درجة توافق نمط أشكال الأرقام فى كل فئة مع نمط أشكال الحروف العربية.

....

في هذا البحث نستخدم المدخل الإحصائى لعلم التعرف على الأنماط كوسيلة للمقارنة بين منظومتى كتابة الأرقام: منظومة الكتابة العربية المشرقية، ومنظومة الكتابة الغاربة المغربية. هذا المدخل الإحصائى يعتمد على الاختيار الأفضل لثلاث أركان للمنهاج الموضوعى فى المقارنة وهى : التمثيل، المعيار، والإحصائيات. ولإيضاح هذا المنهاج نبدأ بتعريف نمط الأشكال والكائنات المنتمية له وكيفية تمثيله من خلال متجه السمات، ونحدد المعيار المستخدم في عملية المقارنة بحيث تكون نتائج عملية المقارنة موضوعية غير متحيزة ومعتمدة عملياً. ثم نعرض الإحصائيات المختلفة الممكن استخدامها فى المقارنة المطروحة. وبعد أن نحدد أركان منهاج المقارنة نعرض كيفية تطبيق هذا المنهاج لإجراء المقارنة الكمية المطلوبة

---

[١] محمود فهمى حجازى ومحمد يونس الحملاوى ومحمد يسرى النحاس؛ أرقامنا العربية: الأرقام المشرقية والأرقام المغربية؛ المؤتمر السنوى الثانى لتعريب العلوم؛ القاهرة؛ ٢٠-٢١ مارس ١٩٩٦م

ونعرض نتائج التطبيق ممثلة فى الإحصائيات الناتجة ونوضح كيفية تفسير هذه النتائج.
ولتدعيم نتائج البحث فلقد تم تصنيف أشكال الأرقام والحروف فى شكل شجرة وراثية معيار التفرع فيها هو مدى التجانس بين أشكال الحروف والأرقام. للبحث عن هذه الشجرة رتبنا بطريقة عشوائية خليطاً من جميع أشكال فئتى منظومة الأرقام العربية (المشرقية والمغربية) مع أشكال الحروف العربية (خط النسخ) والحروف اللاتينية ثم اتبعنا منهج التعرف على الأنماط لإيجاد التجمعات المتجانسة من الأشكال فى هذا الخليط العشوائى، وترتيب هذه التجمعات فى مستويات هرمية عند كل درجة تجانس.
ولقد أجرينا هذه الدراسة على أربع عينات تمثل أشكال حروف الخط العربي (خط النسخ) و الخط الهندى (السنسكريتى) والخط اللاتينى والخط اليونانى.

٢ منهاج الدراسة:

٢-١ تعريف المشكلة:

يمكن صياغة مشكلة إيجاد التجمعات المتجانسة على النحو التالى: بدءاً من فئة كائنات مجهولة الهوية؛ ش ؛

$$ش = \{ ش_{١} ، ش_{٢} ، \dots ، ش_{ن} \}$$

مطلوب البحث عن بنية التجمعات المتجانسة من هذه الكائنات؛ ج؛

$$ج = \{ ج_{١} ، ج_{٢} ، \dots ، ج_{ص} \}$$

لحل هذه المشكلة يجب تعريف معيارا لقياس درجة التجانس داخل كل تجمع ثم البحث عن فئة التجمعات التى تحقق القيمة المثلى لهذا المعيار. بالرغم من أن هذه المسألة قد تجد حلا مثاليا ومباشرا فى بعض المشاكل العملية، إلا أنه من الصعب إيجاد حلاً لها فى الكثير من المشاكل العملية الأخرى وذلك للمشاكل الناجمة عن كبر حجم فئة الكائنات المطلوب تجميعها وكبر حجم متجه السمات الممثل لكل كائن منها. ويكفى لمعرفة مدى صعوبة البحث عن هذه الحلول حساب عدد التقاسيم الممكنة؛ ت $_{ج}$ ؛ من العلاقة التوفيقية التالية[٢]:

$$ت_{ج} = \frac{١}{ن!} \sum_{ر=١}^{ص} ق_{ر} (-١)^{ن-ر} ر^{ص}$$

هذه العوامل يجب مراعاتها بدقة عند البحث عن أفضل تقسيم للتجمعات المتجانسة والذى

[٢] R.O. Duda & P. Hart, "Pattern Classification and Scene Analysis", A. Wiley Interscience Publication , New York , 1974

يحقق درجة عالية من المصداقية فى الحكم على الكائنات الجديدة المطلوب دمجها فى هذه التجمعات. ويقدم هذا البحث طريقة غير متحيزة للبحث عن العلاقات بين أشكال الحروف والأرقام كما يلى شرحه.

٢-٢ تمثيل الأشكال:

حيث أن أشكال الأرقام والحروف هى التى تهمنا فى هذا البحث فإننا سنمثل منظومة الأرقام والحروف بفئة الأشكال النمطية المستخدمة؛ ز؛ والمكونة من تسع وثلاثون شكلاً كالآتى :

ز = { ش$_{١}$ ، ش$_{٢}$ ، ... ، ش$_{ا}$ ، ... ، ش$_{٣٩}$ }

وحيث إن الاختلافات المطروحة واقعياً فى النمط أو الخط المستخدم كثيرة ومتنوعة فقد قمنا باختيار الخط النسخ نظراً لشيوعه فى الكتابة والطباعة. و قد اختير حجم الخلية الممثلة للحرف أو الرقم أكبر من الحد الأدنى الضرورى لتمييز شكله مما يسمح بدراسة الفروق الدقيقة فى بنية الأشكال.

ويمثل كل نمط من الأشكال بمتجه السمات؛ ش ا ؛ المكون من عدد؛ ن ؛ من السمات الأساسية؛ س ل ؛ كالآتى:

ش ا = (س١ ، س٢ ، ... ، س ل ، ... ، س ن)^{م}

تعتبر مجموعة نقاط الصورة الثنائية هى السمات الذاتية لذلك الشكل. هذه السمات لا تصلح للمقارنة كما هى حيث أنها تتغير بالنقل أو الدوران أو الانكماش فى فراغ الصورة. لذلك فإن شكل الحرف يجب معالجته أولاً باستخدام استحالة خطية؛ لقياس بعد نقاط الشكل عن مركز ثقله ومعايرتها بالنسبة لمحاور الشكل الأساسية وحجمه[٣]. الشكل فى الصورة الثنائية الناتجة يتميز بأنه لا يتغير مع أية تحويلات هندسية فى فراغ الصورة ويمكن تمثيل كل نمط بمتجه السمات، شَ، المكون من نفس العدد ن من السمات.

٢-٣ قياس التشابه:

اخترنا في هذا البحث دالة الارتباط؛ ر ؛ لقياس درجة التشابه. وتعرف دالة الارتباط؛ ر؛ بين شكلين ش$_{١}$ ، ش$_{٢}$ بالعلاقة الآتية:

$$ر(م ، ل) = \frac{\text{مج مج}(س، ص)\ د\ ق_{ش_{١}}(س - م ، ص - ل)\ ش_{٢}(س ، ص)}{[\text{مج مج}(س، ص) د\ ش_{١}^{٢}(س، ص)]\ [\text{مج مج}(س، ص) د\ ش_{٢}^{٢}(س، ص)]}$$

[٣] K. Fukunaga , "Statistical Pattern Recognition", Academic Press, inc., San Diego, 1990

حيث يتم حساب المجموع فى التعريف السابق لكل النقاط (س ، ص) داخل المنطقة المشتركة للشكلين؛ ق. هذه الدالة تمثل النسبة بين الطاقة الكامنة فى النقاط المشتركة بين الشكلين والطاقة الكامنة فى نقاط كل منهما. وكما هو واضح من التعريف فهذا القياس لا يعتمد على عدد النقاط الممثلة لكل شكل. ولكنه يعتمد على الوضع النسبى للشكلين معاً فى فراغ الصورة. وهذه الدالة توافقية عادة ولها قيمة عظمى عامة عند أقصى درجة ارتباط وعدة قيم عظمى جانبية تنتشر حول القيمة العظمى العامة. ولزيادة مصداقية الاعتماد على نتائج قياس هذه الدالة يمكن قياس معامل الارتباط الأعظم؛ ر(م ع ، ل ع). أما فى حالة معايرة الأشكال بالنسبة لنظام المحاور المستخدم فى كتابة المنظومة العربية فإنه يمكن الاكتفاء بقياس معامل الارتباط عند مركز ثقل الشكل أى النقطة (م ث ، ل ث).

٢-٤ تصميم شجرة التجمعات المتجانسة:

هناك عدة طرق لإيجاد التجمعات المتجانسة من الأنماط لخليط من الكائنات المجهولة. والطريقة التى نستخدمها فى هذا البحث تندرج تحت طرق التجميع التراكمى الهرمى. فى هذه الطريقة تبدأ تتابعية التجميع بتتابعية بها عدد من التجمعات مساو لعدد الكائنات ثم عند كل مستوى من مستويات التجميع تدمج كائنات التجمعات مع بعضها البعض، وحيث أن الكائنات المدمجة عند مستوى ما تبقى سويا فى المستوى الأعلى فهذه التتابعية تسمى بالتجميع التراكمى الهرمى. وصحة شجرة التجمعات المتجانسة تعتمد بصورة عامة على الترتيب الابتدائى للتجمعات، وعلى ترتيب الكائنات داخل كل تجمع، وعلى اختيار معيار التشابه بين التجمعات. لذا فإن أى تصميم لهذه الإجرائية لابد أن يأخذ فى الاعتبار كل هذه العوامل. لذلك فقد قمنا بتطوير هذه الطريقة لتلافى تأثير هذه العوامل بإضافة التالى:

١-عملية خلط عشوائى تام للكائنات أو التجمعات قبل مقارنتها مع بعضها البعض.

٢-استخدام معيار الزوج الأكثر تشابها لاستقطاب التجمعات ودمجها وهو فى بحثنا هذا يمثل الوضع الأسوأ فى الحكم على صحة النتائج.

سنستخدم هنا اللغة المعروفة ب" شبه شفرة " لتمثيل إجرائية تصميم الشجرة. فى ما نعرضه هنا تمثل الكلمات ذات المدلول الحاسوبى بالخط الكوفى العريض والكلمات ذات المدلول الرياضى بخط النسخ المائل العريض.

إجرائية التجميع التراكمى الهرمى:

مدخلات: *فئة أشكال الأرقام والحروف؛ ش*

مخرجات: *شجرة التجمعات المتجانسة؛ ت*

الخطوات:

خلق عدد ن من التجمعات يحتوى كل منها على شكل واحد فقط

ابدأ بعدد التجمعات الحالية ص = ن

كرر ما يلى **طالما** عدد التجمعات الحالية ص > ١

بداية

اخلط التجمعات الحالية عشوائيا

ابحث عن أقرب زوج من التجمعات؛ (ج م ، ج ل)

أدمج زوج التجمعات (ج م ، ج ل)

احذف التجمع الأدنى ج ل

ارفع مستوى التجمع الأعلى ج م

أنقص عدد التجمعات الحالية واحدا

نهاية

٣ التطبيق والنتائج:

لإجراء التجارب المطلوبة فى هذا البحث تم تصميم وتنفيذ تطبيق حاسوبى يعمل فى بيئة النوافذ يتكون من الوحدات التالية:

١- واجهة التطبيق، ٢- مولد الأشكال، ٣- معالج الصور، ٤-مصنف الكائنات، ٥- عارض البيانات.

عند إجراء تجربة ما على أشكال الأرقام وحروف الخط العربى تتبع الخطوات التالية:

١- تولد أشكال منظومة كتابة الأرقام والحروف؛ ز؛ بواسطة برنامج مولد الأشكال.

٢- تعالج الأشكال الناشئة بحصرها داخل نافذة ثم تنحف هذه الأشكال وتحدد بنافذة مكونة من ١٧×٢٠ نقطة.

٣- يحسب لكل شكل من الأشكال؛ ش ١ ؛ متجه مركز الثقل؛ م ١ ؛ ومصفوفة الانتشار؛ ك ١ ؛ ودالة الاستحالة؛ د ١ .

٤- يعاد توقيع النقاط السوداء فى كل شكل من الأشكال؛ ش ١ ؛ باستخدام دالة الاستحالة الخطية؛ د ١ ؛ لينشأ عن ذلك الشكل المعاير؛ شَ ١ .

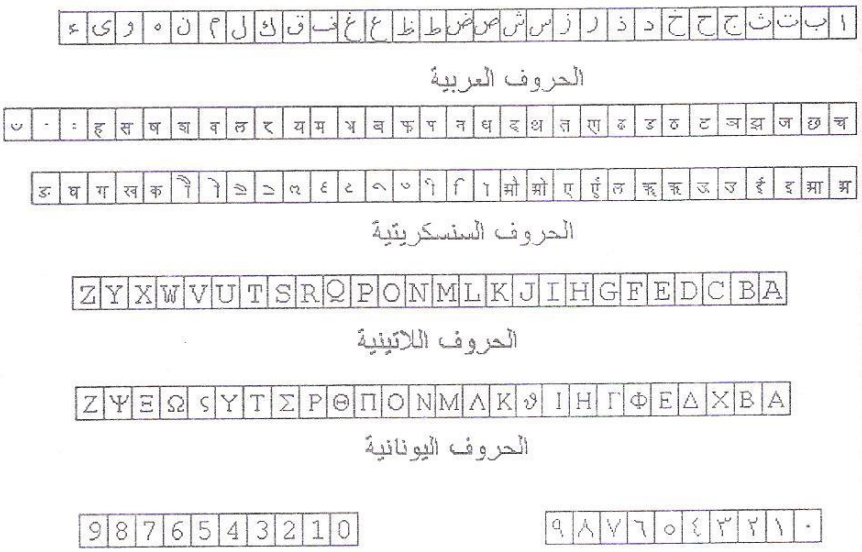
٥- تحسب دالة الارتباط بين جميع الثنائيات الممكنة للأشكال؛ (شَ ١ ، شَ ٢)؛ فى كل منظومة

على حدة.

٦- يسجل أقصى معامل ارتباط لكل ثنائية من الثنائيات الممكنة للأشكال؛ ر ع(ش ١ ، ش ٢)؛ في مصفوفة الارتباط المتماثلة.

٧- تطبق إجرائية التجميع الهرمي السابق شرحها لإيجاد التجمعات المتجانسة والعلاقات الهرمية بينها.

ولقد طبقنا هذا المنهج على أربع عينات يمثل كلاً منها شكلاً من أشكال الحروف العربية والهندية واللاتينية واليونانية بالإضافة إلى شكلى الأرقام العربية المشرقية والمغربية الغبارية كما هو موضح فى الشكل رقم (١).



MAPS1

شكل ١ أشكال منظومات الأرقام والحروف العربية والغربية.

ولقياس درجة الانتماء وعمل المقارنة نحتاج هنا إلى حساب ورسم ستة منحنيات هى:

١- منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية.

٢- منحنى كثافة الاحتمال لفئة الأرقام المغربية الغبارية.

٣- منحنى كثافة الاحتمال لفئة الحروف العربية (خط النسخ).

٤- منحنى كثافة الاحتمال لفئة الحروف الهندية (السنسكريتية).

٥- منحنى كثافة الاحتمال لفئة الحروف اللاتينية.

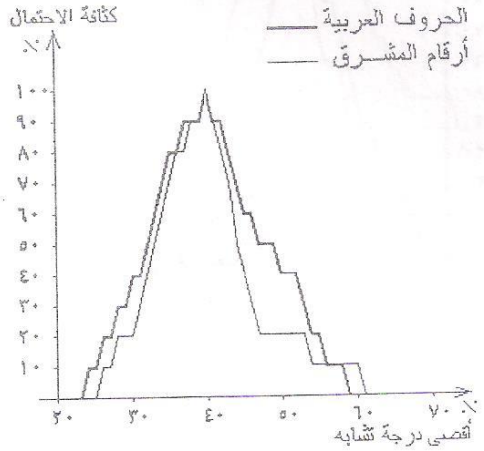
٦- منحنى كثافة الاحتمال لفئة الحروف اليونانية.

وبرسم منحنَيَ كثافة الاحتمال لكل منظومة من منظومات الحروف مع احدى منظومتى الأرقام يمكننا معرفة إذا ما كانت فئة الأرقام المستخدمة تمثل نمطاً مستقلاً عن نمط بقية عناصر المنظومة أم أنها تمثل مجرد فئة جزئية من عناصر المنظومة. ويستدل على ذلك بظهور وسطين حسابيين فى منحنى كثافة الاحتمال لدرجات التشابه، وكلما زاد الفرق بين هذين الوسطين كلما قلت درجة انتماء كل من النمطين إلى الآخر، أى نمط الحروف ونمط الأرقام.

ولقد تم حساب منحنيات كثافة الاحتمال السابقة ومثلت النتائج فى الأشكال من رقم (٢) إلى رقم (١٠).

٣-١ نتائج القياسات على الأرقام العربية المشرقية والحروف العربية:

يمثل الشكل رقم (٢) منحنى كثافة الاحتمال لفئة الحروف العربية مع منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية.



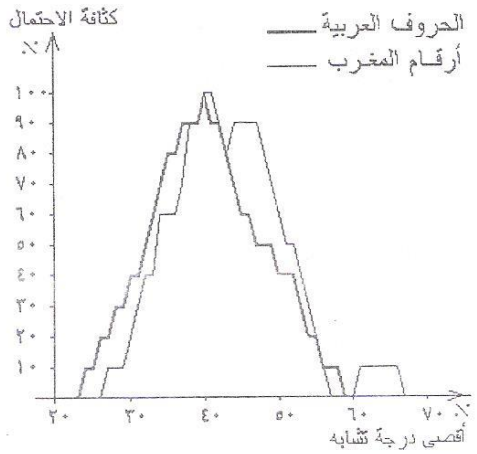
شكل ٢ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف العربية والأرقام العربية المشرقية

ويلاحظ فى هذا الشكل أن الفئتين لهما نفس متوسط أقصى درجة تشابه، كذلك يُلاحظ أن منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية يقع بالكامل تحت منحنى كثافة الاحتمال للحروف العربية. وهذا يعنى فى نظرية علم التعرف على الأنماط أن الفئتين يصعب التمييز بينهما. أى أن أشكال الأرقام العربية المشرقية تمثل فئة جزئية من فئة أشكال الحروف العربية مما يعنى

أن السمات الأساسية للأرقام العربية المشرقية والحروف العربية واحدة مما قد يدل على أن الثقافة التى أنتجت كلاً من الفئتين واحدة. والجدير بالذكر أيضاً أن منحنى كثافة الاحتمال للأرقام العربية المشرقية باحتوائه على قمة واحدة يشير إلى أن هذه الأرقام متناسقة فيما بينها مما يعنى أنها وليدة ثقافة واحدة.

٣-٢ نتائج القياسات على الأرقام الغبارية المغربية والحروف العربية:

لقد تم إجراء نفس القياسات بالنسبة للأرقام الغبارية المغربية حيث يمثل الشكل رقم (٣) منحنى كثافة الاحتمال لفئة الحروف العربية مع منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية.



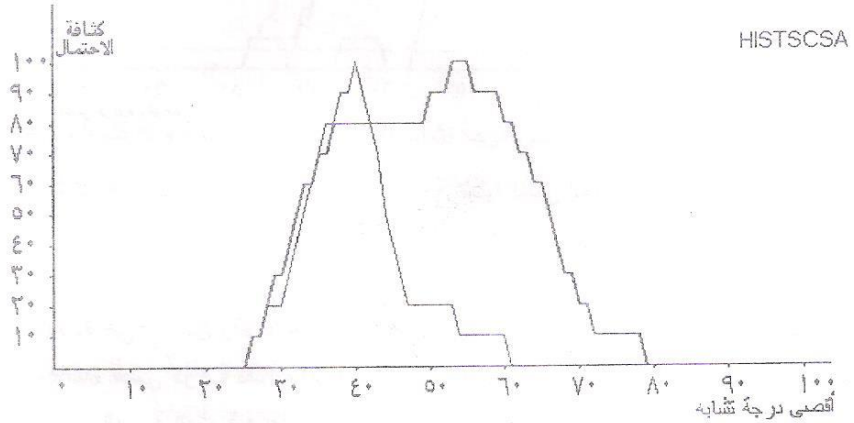
شكل ٣ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف العربية والأرقام الغبارية المغربية

ويُلاحظ فى هذا الشكل أن متوسط أقصى درجة تشابه لفئة الأرقام الغبارية المغربية أكبر من متوسط أقصى درجة تشابه لفئة الحروف العربية. وكذلك يُلاحظ أن منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية ممتد يمينا خارج منحنى كثافة الاحتمال للحروف وبه أكثر من قمة مما يعنى احتواء مجموعة أرقام المغرب على فئات جزئية غير ظاهرة. وهذا يعنى فى نظرية علم التعرف على الأنماط أن الفئتين يسهل التمييز بينهما. أى أن أشكال الأرقام الغبارية المغربية تمثل فئة مستقلة عن فئة أشكال الحروف العربية، مما يعنى اختلاف السمات الأساسية بين الأرقام الغبارية المغربية والحروف العربية. والجدير بالذكر أيضاً أن منحنى كثافة الاحتمال

للأرقام الغبارية المغربية باحتوائه على أكثر من قمة يشير إلى أن هذه الأرقام غير متناسقة فيما بينها مما يعنى أنها وليدة أكثر من ثقافة.

٣-٣ نتائج القياسات على الأرقام العربية المشرقية والحروف الهندية السنسكريتية:

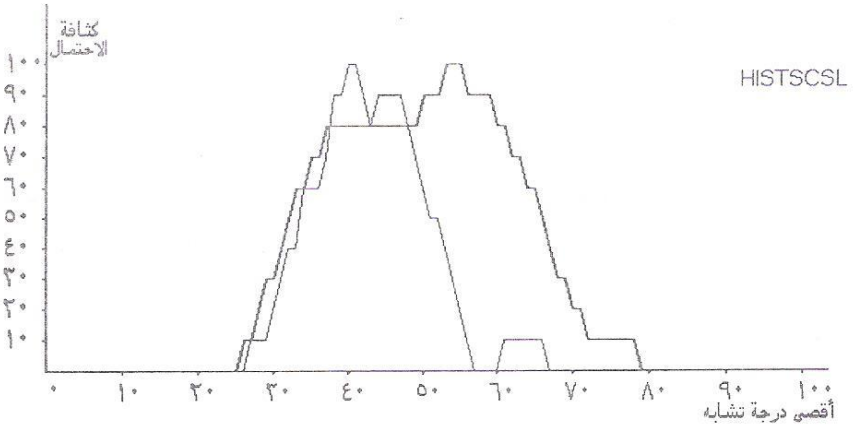
يمثل الشكل رقم ٤ منحنى كثافة الاحتمال لفئة الحروف الهندية السنسكريتية مع منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية.



شكل ٤ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف الهندية والأرقام العربية المشرقية

٣-٤ نتائج القياسات على الأرقام الغبارية المغربية والحروف العربية:

لقد تم إجراء نفس القياسات بالنسبة للأرقام الغبارية المغربية حيث يمثل الشكل رقم ٥ منحنى كثافة الاحتمال لفئة الحروف العربية مع منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية.

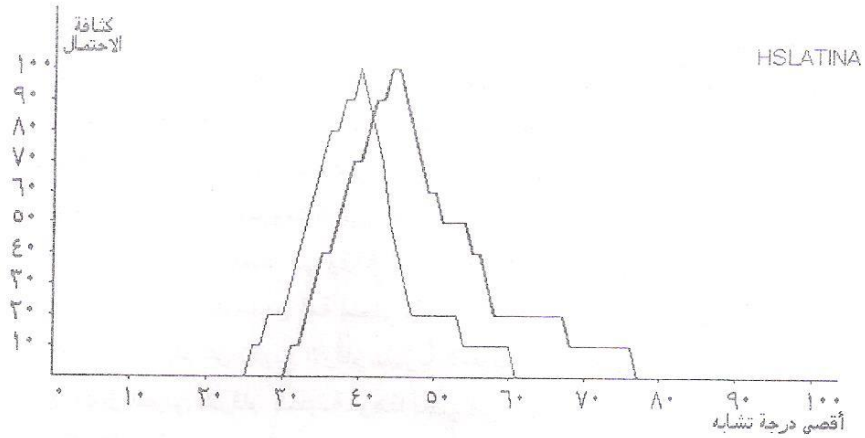


شكل ٥ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف الهندية والأرقام الغبارية المغربية

ويلاحظ فى الشكلين رقم ٥،٤ أن متوسط أقصى درجة تشابه لكل فئة من فئتى الأرقام أقل من متوسط أقصى درجة تشابه لفئة الحروف الهندية، إلا أن متوسط أقصى درجة تشابه لفئة الأرقام العربية المشرقية أقل من متوسط أقصى درجة تشابه لفئة الأرقام الغبارية المغربية وبالتالى يبعد أكثر عن متوسط أقصى درجة تشابه لفئة الحروف الهندية. كما أن المساحة المشتركة بين منحنى الحروف الهندية والأرقام العربية المشرقية أقل منها فى حالة الأرقام الغبارية المغربية. وهذا يعنى فى نظرية علم التعرف على الأنماط أن فئة الأرقام الغبارية المغربية أقرب إلى فئة الحروف الهندية من فئة الأرقام العربية المشرقية من ناحية السمات. كما يعنى اختلاف السمات الأساسية بين الأرقام العربية المشرقية والحروف الهندية بصورة أكبر عنها فى حالة الأرقام الغبارية المغربية والحروف الهندية.

٣-٥ نتائج القياسات على الأرقام العربية المشرقية والحروف اللاتينية:

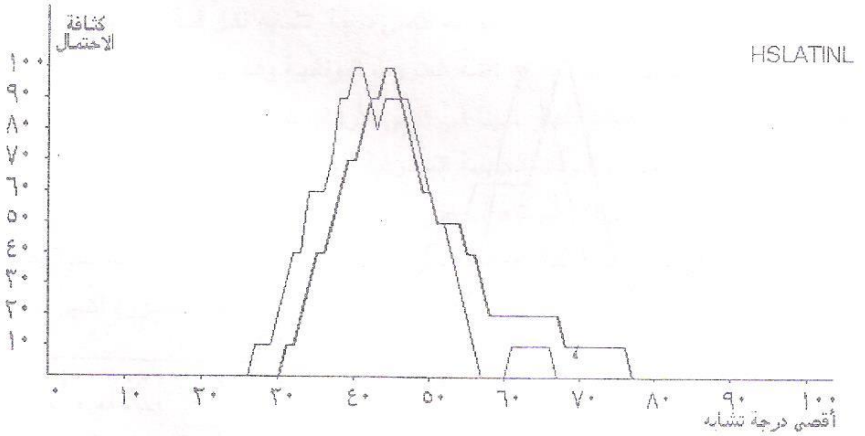
يمثل الشكل رقم (٦) منحنى كثافة الاحتمال لفئة الحروف اللاتينية مع منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية.



شكل ٦ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف اللاتينية والأرقام العربية المشرقية

٣-٦ نتائج القياسات على الأرقام الغبارية المغربية والحروف اللاتينية:

لقد تم إجراء نفس القياسات بالنسبة للأرقام الغبارية المغربية حيث يمثل الشكل رقم (٧) منحنى كثافة الاحتمال لفئة الحروف اللاتينية مع منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية.

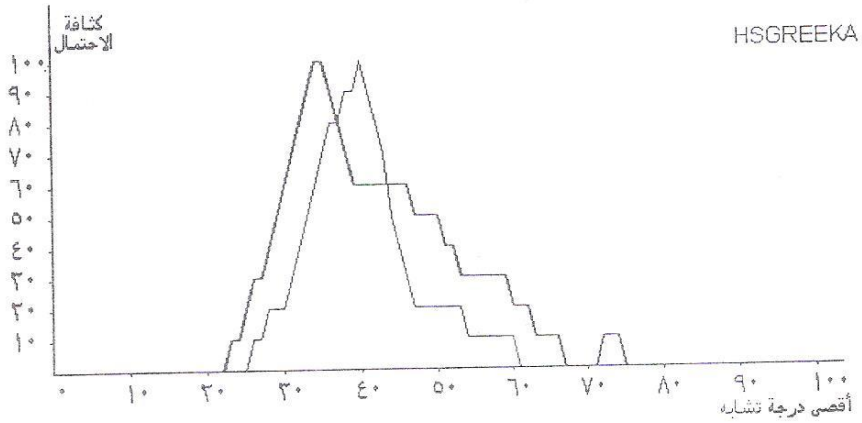


شكل ٧ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف اللاتينية والأرقام الغبارية المغربية

ويلاحظ فى الشكلين رقم ٧،٦ أن متوسط أقصى درجة تشابه لكل فئة من فئتى الأرقام أقل من متوسط أقصى درجة تشابه لفئة الحروف اللاتينية، إلا أن متوسط أقصى درجة تشابه لفئة الأرقام العربية المشرقية أقل من متوسط أقصى درجة تشابه لفئة الأرقام الغبارية المغربية وبالتالى يبعد أكثر عن متوسط أقصى درجة تشابه لفئة الحروف اللاتينية. كما أن المساحة المشتركة بين منحنى الحروف اللاتينية والأرقام العربية المشرقية أقل منها فى حالة الأرقام الغبارية المغربية، بالإضافة إلى وقوع القمة الأصغر فى منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية بالقرب من قمة منحنى كثافة الاحتمال لفئة الحروف اللاتينية مما يثير بعض علامات الاستفهام عن تطويع الأرقام الغبارية المغربية لتصل لصورتها الحالية وإن لم تتخلص من الأصل العربى للأرقام الحديثة. وهذا يعنى فى نظرية علم التعرف على الأنماط أن فئة الأرقام الغبارية المغربية أقرب إلى فئة الحروف اللاتينية من فئة الأرقام العربية المشرقية من ناحية السمات. كما يعنى اختلاف السمات الأساسية بين الأرقام العربية المشرقية والحروف اللاتينية بصورة أكبر عنها فى حالة الأرقام الغبارية المغربية والحروف اللاتينية.

٣-٧ نتائج القياسات على الأرقام العربية المشرقية والحروف اليونانية:

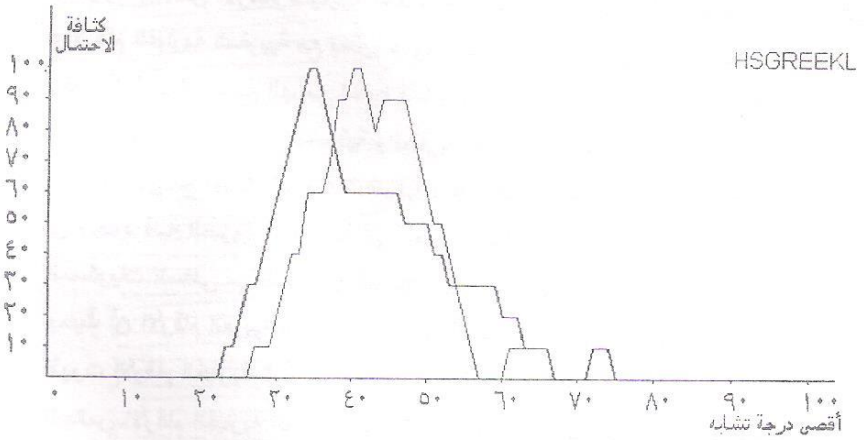
يمثل الشكل رقم (٨) منحنى كثافة الاحتمال لفئة الحروف اليونانية مع منحنى كثافة الاحتمال لفئة الأرقام العربية المشرقية.



شكل ٨ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف اليونانية والأرقام العربية المشرقية

٣-٨ نتائج القياسات على الأرقام الغبارية المغربية والحروف اليونانية:

لقد تم إجراء نفس القياسات بالنسبة للأرقام الغبارية المغربية حيث يمثل الشكل رقم (٩) منحنى كثافة الاحتمال لفئة الحروف اليونانية مع منحنى كثافة الاحتمال لفئة الأرقام الغبارية المغربية.



شكل ٩ منحنى كثافة الاحتمال لأقصى درجة تشابه لفئتى الحروف اليونانية والأرقام الغبارية المغربية

ويلاحظ كذلك فى الشكلين رقم ٨،٩ أن متوسط أقصى درجة تشابه لكل فئة من فئتى الأرقام أكبر من متوسط أقصى درجة تشابه لفئة الحروف اليونانية وقد يشير هذا إلى تشابهات واضحة بين الحروف اليونانية أكثر منها فى فئتى الأرقام. كما أن المساحة المشتركة بين منحنى الحروف اليونانية والأرقام العربية المشرقية أقل منها فى حالة الأرقام الغبارية المغربية. وهذا يعنى فى نظرية علم التعرف على الأنماط أن فئة الأرقام الغبارية المغربية أقرب إلى فئة الحروف اليونانية من فئة الأرقام العربية المشرقية من ناحية السمات. كما يعنى اختلاف السمات الأساسية بين الأرقام العربية المشرقية والحروف اليونانية بصورة أكبر عنها فى حالة الأرقام الغبارية المغربية والحروف اليونانية.

٣-٩ نتائج القياسات على خليط أشكال حروف اللغة العربية وحروف اللغة اللاتينية والأرقام العربية المشرقية والغبارية المغربية:

لحسم النتائج السابقة فقد كونا خليطاً من ١٠١ عنصراً تمثل الأرقام العربية المشرقية والأرقام

الغبارية المغربية مع حروف خط النسخ العربى وحروف اللغة اللاتينية الكبيرة والصغيرة. ولقد أجريت التجربة بنفس الشروط السابقة وكانت النتيجة هى ظهور تجمعين من الأشكال: التجمع الأول مكون من معظم حروف الخط العربى والأرقام العربية المشرقية مع بعض الحروف اللاتينية وبعض الأرقام الغبارية المغربية، والتجمع الثانى شمل معظم الحروف اللاتينية والأرقام الغبارية المغربية مع بعض حروف الخط العربى والأرقام المغربية. ويوضح الشكل رقم ١٠ شجرة التجمع الهرمى لخليط أشكال حروف اللغة العربية (خط النسخ) وحروف اللغة اللاتينية والأرقام العربية المشرقية والغبارية المغربية. والشكل يؤكد التجانس الواضح لحروف الخط العربى مع الأرقام العربية المشرقية وعدم تجانسها مع الأرقام الغبارية المغربية. ويلاحظ أن وجود فئة الحروف اللاتينية فى الخليط استقطب الأرقام الغبارية المغربية سريعاً عند المستويات السفلى من التجمعات الهرمية مؤكداً تجانسهما معاً.

وحيث أن الأرقام العربية المشرقية أقدم فى النشوء حيث ظهرت لأول مرة عام ٢٠٤ هـ[٤] بينما ظهرت الأرقام الغبارية المغربية لأول مرة عام ٤٨٠ هـ[٥] بفارق حوالى ثلاثة قرون، وحيث تتجانس الأرقام الغبارية المغربية بصورة أكبر مع أشكال الحروف الهندية واللاتينية واليونانية لذا فمن المنطقى الحكم على تلك الأرقام الغبارية المغربية بأنها تطويع لمنظومة الأرقام العربية كى تلائم تلك اللغات الموجودة فى فترة نشوء تلك الأرقام. ومما يدعم ذلك تاريخياً أن تلك الأرقام الغبارية قد تطورت على ثلاثة مراحل منذ نشأتها حتى وصلت إلى صورتها الحالية[٦]. والجدير بالذكر أن كلمة غبارية التى اشتهرت بها تلك المنظومة فى الكتابات العربية أتت من طريقة كتابة الهنود حيث كانوا ينثرون الغبار على الألواح ثم يكتبون عليها[٧].

ونستنتج من المقارنات الإحصائية السابقة أن فئة الأرقام العربية المشرقية يمكن دمجها مع فئة الحروف العربية لتنتج فئة متجانسة فى خواصها الإحصائية. أما فئة الأرقام الغبارية المغربية عند دمجها مع فئة الحروف العربية تنتج فئة غير متجانسة، حيث يتميز منحنى كثافة الاحتمال الخاص بها بوجود قمتين تشيران إلى عدم تجانس تلك الفئة فى ذاتها وبالتالى مع فئة الحروف العربية. وبالتالى نستنتج أن فئة الأرقام العربية المشرقية أكثر تجانساً مع فئة الحروف العربية من تجانس فئة الأرقام الغبارية المغربية مع فئة الحروف العربية.

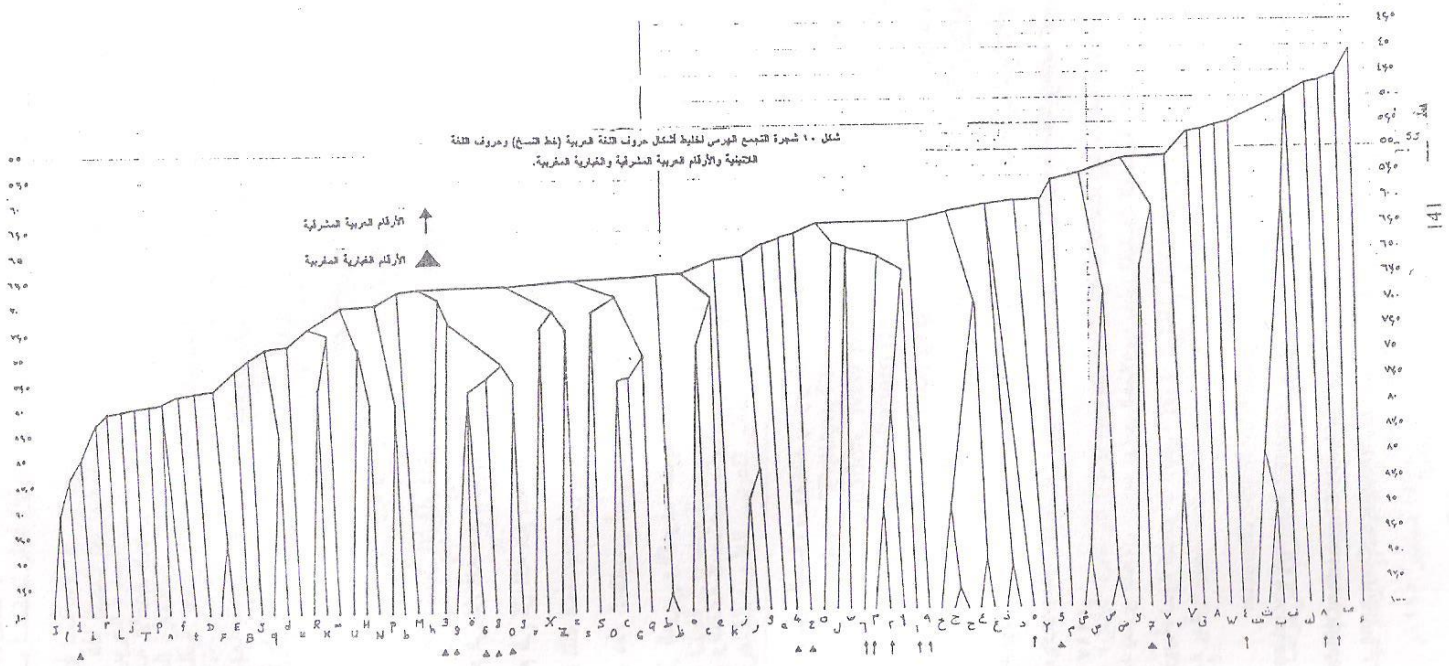
---

[٤] محمد بن موسى الخوارزمى؛ الجبر والمقابلة؛ دار الكاتب العربى للطباعة والنشر؛ القاهرة؛ ١٩٦٨م؛ صفحة ٢٤

[٥] لجنة الرياضة؛ مجمع اللغة العربية؛ الدورة الحادية والأربعون؛ القاهرة؛ ١٩٧٤-١٩٧٥م

[٦] موريس شربل؛ الرياضيات فى الحضارة الإسلامية؛ جروس برس؛ طرابلس، لبنان؛ ١٩٨٨م صفحة ٩١

[٧] قاسم على سعد؛ الأرقام العربية: تاريخها وأصالتها وما استعمله المحدثون وغيرهم منها، القسم الأول؛ مجلة الأحمدية؛ العدد ٢؛ ١٩٩٨م؛ دبى



شكل ١٠ شجرة التجمع الهرمي لخليط أشكال حروف اللغة العربية (خط النسخ) وحروف اللغة اللاتينية والأرقام العربية المشرقية والغبارية المغربية.

٤ الخلاصة:

تم فى هذا البحث إجراء دراسة مقارنة بين فئتى منظومة الأرقام العربية المستخدمتين في منظومة الأعداد العشرية ومدى تناسقهما مع منظومة الكتابات العربية والهندية واللاتينية واليونانية. هاتان الفئتان تسميان خطأً فئة الأرقام الهندية وفئة الأرقام العربية. بينما الأصح تسميتهما بالأرقام العربية المشرقية والأرقام الغبارية المغربية.

ولقد أتاح المبحث الهندسى لنا استخدام القياس الكمى لسمات كل من الفئتين وإجراء المقارنات الموضوعية بينهما من خلال بعض المعايير الغير متحيزة، حيث تم حساب درجة انتماء نمط أشكال الأرقام فى كل فئة مع نمط أشكال الحروف العربية. كما تم تغطية نتائج قياس درجة تشابه نمط أشكال الأرقام داخل كل فئة من فئتى الأرقام[٨] ونتائج قياس درجة توافق نمط أشكال الأرقام فى كل فئة مع نمط أشكال الحروف العربية[٩] وكذلك نتائج قياس درجة توافق نمط أشكال الأرقام فى كل فئة مع نمط أشكال حروف بعض اللغات غير العربية[١٠]؛ حيث سبق للباحثَين إجراء تلك الحسابات فى بحوث أخرى[١١].

وتمثل هذه المقاييس أهمية خاصة لأنها تؤثر بصورة مباشرة فى حل مشاكل توصيف الرقم والرمز العربى والتعرف عليهما بالتقنيات الحديثة كالحاسوب[١٢] بالإضافة إلى تصميم منظومات الرسم والكتابة المستعينة بالحاسوب[١٣].

---

[٨] محمد يونس الحملاوى ومحمد يسرى النحاس؛ الأرقام الهندعربية، دراسة مقارنة: التاريخ، السمات، الاستخدام، التقنية؛ المؤتمر الدولى السادس عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ٣- ٥ سبتمبر ١٩٩٦م

[٩] محمد يسرى النحاس ومحمد يونس الحملاوى؛ تجانس الأرقام الهندعربية مع أشكال حروف الخط العربى؛ المؤتمر الدولى السابع عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ٢- ٤ سبتمبر ١٩٩٧م

[١٠] محمد يونس الحملاوى ومحمد يسرى النحاس؛ تجانس الأرقام الهندعربية مع أشكال الحروف العربية وأشكال حروف لغات أخرى؛ ندوة الأرقام ومكانتها فى قضية التعريب؛ مجمع اللغة العربية؛ القاهرة؛ ٢٠ فبراير ١٩٩٧م

[١١] محمد يسرى النحاس ومحمد يونس الحملاوى؛ بعض القياسات الهندسية والعلمية على مجموعتى الأرقام الهندعربية؛ ندوة الخطوات العملية لإقرار استخدام الأرقام العربية؛ جامعة الأزهر؛ القاهرة؛ ١ مارس ١٩٩٨م

[١٢] M.Y. Mahmoud (El Nahas) et al : "A statistical approach for Arabic character recognition", 12th international congress for statistics, computer science, social and demographic research, Cairo, 1987

[١٣] M.A. El Hamalaway & S.H. El Ramly , "A Novel Arabic Numerals Shapes for Optical Character Recognition", Thirteen International Congress for Statistics, Computer Science, Social and Demographic Research, 26-31 March 1988

ولقد تبين أن الأرقام المشرقية (٠، ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩) أكثر تجانساً مع أشكال الحروف العربية من الأرقام الغبارية المغربية (9, 8, 7, 6, 5, 4, 3, 2, 1, 0). كما أن الأرقام العربية المشرقية تنتمى بدرجة أكبر إلى الحضارة العربية منها إلى الحضارة الغربية؛ بينما الأرقام الغبارية المغربية تنتمى بدرجة أكبر إلى الحضارة الغربية منها إلى الحضارة العربية.

وبناء على النتائج التى ظهرت من البحث فإن فئة الأرقام العربية المشرقية التى تنتمى إلى حضارة واحدة تبين أنها ليست الهندية وليست اللاتينية وليست اليونانية، بل هى الحضارة العربية وبالتالى لا يوجد سبب قوى أو ضعيف يبرر نبذ سلسلة الأرقام المستعملة فى المشرق العربى واستعمال الأرقام المستعملة فى المغرب العربى محلها، وذلك إضافة إلى عدة عوامل أهمها أن الأرقام المشرقية هى الأقدم فى الاستعمال والشيوع فى مختلف أرجاء الأمة العربية والإسلامية واستعمالها فى أكثر التراث العربى[١٤]، بالإضافة إلى كفاءتها عن المجموعة المغربية[١٥]، زيادة على توافقها بصورة أكبر مع أشكال حروف اللغة العربية.

٥ المراجع:

١.محمود فهمى حجازى ومحمد يونس الحملاوى ومحمد يسرى النحاس؛ أرقامنا العربية: الأرقام المشرقية والأرقام المغربية؛ المؤتمر السنوى الثانى لتعريب العلوم؛ القاهرة؛ ٢٠-٢١ مارس ١٩٩٦م

2. Duda & P. Hart, "Pattern Classification and Scene Analysis", A. Wiley Interscience Publication , New York , 1974

3. Fukunaga , "Statistical Pattern Recognition", Academic Press, inc., San Diego, 1990

٤.محمد بن موسى الخوارزمى؛ الجبر والمقابلة؛ دار الكاتب العربى للطباعة والنشر؛ القاهرة؛ ١٩٦٨م؛ صفحة ٢٤

٥.لجنة الرياضة؛ مجمع اللغة العربية؛ الدورة الحادية والأربعون؛ القاهرة؛ ١٩٧٤-١٩٧٥م

٦.موريس شربل؛ الرياضيات فى الحضارة الإسلامية؛ جروس برس؛ طرابلس، لبنان؛ ١٩٨٨م صفحة ٩١

٧.قاسم على سعد؛ الأرقام العربية: تاريخها وأصالتها وما استعمله المحدثون وغيرهم منها،

---

[١٤] محمد يونس الحملاوى ومحمد يسرى النحاس؛ استعمال الأرقام العربية المشرقية فى تراثنا العلمى؛ المؤتمر السنوى الرابع لجمعية لسان العرب؛ جامعة الدول العربية؛ القاهرة؛ ١٥- ١٦ نوفمبر ١٩٩٧م

[١٥] محمد يسرى النحاس ومحمد يونس الحملاوى؛ قياس درجة التشابه فى مجموعتى الأرقام الهندعربية؛ المؤتمر الخامس عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ١٢-١٤ سبتمبر ١٩٩٥م

القسم الأول؛ مجلة الأحمدية؛ العدد ٢؛ ١٩٩٨م؛ دبى
٨.محمد يونس الحملاوى ومحمد يسرى النحاس؛ الأرقام الهندعربية، دراسة مقارنة: التاريخ، السمات، الاستخدام، التقنية؛ المؤتمر الدولى السادس عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ٣- ٥ سبتمبر ١٩٩٦م
٩.محمد يسرى النحاس ومحمد يونس الحملاوى؛ تجانس الأرقام الهندعربية مع أشكال حروف الخط العربى؛ المؤتمر الدولى السابع عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ٢- ٤ سبتمبر ١٩٩٧م
١٠.محمد يونس الحملاوى ومحمد يسرى النحاس؛ تجانس الأرقام الهندعربية مع أشكال الحروف العربية وأشكال حروف لغات أخرى؛ ندوة الأرقام ومكانتها فى قضية التعريب؛ مجمع اللغة العربية؛ القاهرة؛ ٢٠ فبراير ١٩٩٧م
١١.محمد يسرى النحاس ومحمد يونس الحملاوى؛ بعض القياسات الهندسية والعلمية على مجموعتى الأرقام الهندعربية؛ ندوة الخطوات العملية لإقرار استخدام الأرقام العربية؛ جامعة الأزهر؛ القاهرة؛ ١ مارس ١٩٩٨م
12.Mahmoud (El Nahas) et al : "A statistical approach for Arabic character recognition", 12th international congress for statistics, computer science, social and demographic research, Cairo, 1987
13.El Hamalaway & S.H. El Ramly , "A Novel Arabic Numerals Shapes for Optical Character Recognition", Thirteen International Congress for Statistics, Computer Science, Social and Demographic Research, 26-31 March 1988
١٤.محمد يونس الحملاوى ومحمد يسرى النحاس؛ استعمال الأرقام العربية المشرقية فى تراثنا العلمى؛ المؤتمر السنوى الرابع لجمعية لسان العرب؛ جامعة الدول العربية؛ القاهرة؛ ١٥- ١٦ نوفمبر ١٩٩٧م
١٥.محمد يسرى النحاس ومحمد يونس الحملاوى؛ قياس درجة التشابه فى مجموعتى الأرقام الهندعربية؛ المؤتمر الخامس عن الحاسب الآلى بين النظرية والتطبيق؛ الإسكندرية؛ ١٢- ١٤ سبتمبر ١٩٩٥م

امتنان:

يشكر الباحثان الجمعية المصرية لتعريب العلوم؛ ص. ب ٥٣٠١ غرب مصر الجديدة، القاهرة ١١٧٧١، مصر على دعمها لجهود التعريب مما كان له أبلغ الأثر فى إنجاز هذا البحث.

LNGENG22.DOC