اعتماد نمذجة الدماغ لتمييز الحروف العربية المطبوعة

علي فاضل مرهون و ساهرة عبيد سعد

جامعة البصرة/ كلية العلوم/قسم الحاسبات

ISSN –1817 –2695
الاستلام 2006/5/24, القبول 2006/9/10

الملخص

يقدم هذا العمل تصميما لنظام تمييز الحروف العربية قادرا" على تجاوز بعض الصعوبات والمشاكل التي تواجهها أنظمة التمييز الأولى وذلك لاعتماد العمل الحالي على أسلوب خاص في مزج السلوك الإنساني في عملية التمييز. .

يتضمن تصميم النظام المقترح خمس مراحل رئيسة ، يتم في المرحلة الأولى قراءة الوثائق العربية المطبوعة باستعمال جهاز الماسح الضوئي Hp Scanner Jet إذ يولد ملفا" من نوع TIF (Tagged Image File) ،وفي المرحلة الثانية يتم اختيار درجة الانتباه المناسبة لعمل النموذج ، وفي المرحلة الثالثة يتم معالجة تأثيرات النص المدخل على فعالية بعض مكونات الدماغ ، وفي المرحلة الرابعة يتم استخراج القمم للحروف على سطح وثيقة النص ،أما في المرحلة الخامسة فيتم تقطيع النص المدخل إلى الحروف، وقد اقترحت عدة أفكار وطرائق جديدة لغرض التقطيع التي أعطت بتطبيقها نتائج جيدة ومقنعة ويمكن بوساطتها تبسيط عملية التقطيع المستعملة في طرائق معروفة أخرى ،ثم بعد ذلك يتم تصنيف الحروف المقطوعة اعتمادا على الخبرة المتولدة نتيجة لتجارب عينية سابقة ، وتم تجهيز النظام بطور للتعلم بأسلوب المشرف عليه .

Key Words: Segmented Characters,Characters Recognition , Brain Modeling

المقدمة

كان للتقدم الهائل الذي حصل على صعيد الأجهزة والبرامجيات دورا" مهما" في إيجاد تطبيقات برامجية تحاكي طبيعة البشر وذكاءه[1،2،3،4] , يعد الذكاء الاصطناعي Artificial Intelligence أحد المجالات العلمية التطبيقية في علوم الحاسبات الذي يحاول إيجاد تطبيقات كثيرة منها تمييز الأنماط[1،2،3] ,وقد أدت أسباب كثيرة إلى تبني هذا الحقل في علوم الحاسبات وبخاصة مسألة تمييز الحروف العربية ,حيث تعد من المسائل المثالية لتمييز الأنماط للحاجة اكثر إلى آله تقرأ المطبوع أو المكتوب [4], ونلاحظ أن اغلب البحوث في تمييز الحروف العربية اتجهت إلى ما يسمى بنظام التمييز الاصطناعي الذي يتكون عادة من خمس مراحل ,وهي مرحلة التحويل ومرحلة التقطيع، ومرحلة المعالجة الأولية،ومرحلة استخلاص الصفات ، ثم مرحلة التصنيف[1،5،6]، تلاحظ أن هذا النظام يعمل على محاكاة لعمل التمييز أو الإدراك الذي يحدث عند الإنسان ,بحيث لجأت تلك التقنيات إلى تتبع النقاط pixels (عناصر النص) سواء في مرحلة استخلاص الصفات أو مرحلة التقطيع وغيرها من المراحل التي يمر بها نظام التمييز الاصطناعي ولغرض الوصول إلى تمييز الحروف، أن محاولة الرجوع إلى طبيعة التمييز التي تحصل في الدماغ واستغلال بعض المعالجات التي تحصل هو طريق آخر لتمييز الحروف , حيث يتطلب هذا الاتجاه دراسة باتجاهين الدراسة البيولوجية والدراسة السيكولوجية ,إذ تتضمن الدراسة السيكولوجية دراسة التصرفات الفطرية للإنسان عند ملاحظته أنماط الكتابة والعوامل التي تتداخل خلال عملية عرض النص المكتوب على الشخص والتي تمكنه من التمييز في البحث الحالي تم استغلال هذه الميزات الإنسانية لبناء منظومة تمييز جديدة للحروف العربية المطبوعة ذات الحجم والشكل الثابت ,أما الدراسة البيولوجية فأنها تتضمن دراسة معمقة في تركيب دماغ الإنسان وأسلوب المعالجة فيه.

خلاصة النموذج الدماغي المعتمد في التمييز [7]

لغرض فهم النموذج الدماغي المعتمد في التمييز سوف نذكر مجموعة من الحقائق البيولوجية التي تعد أساس لفهم النموذج:-

1-الكتلة الدماغية والقشرة Brain Mss and Cerebral Cortex [8]

الدماغ هو كتلة من مجموعة كبيرة من الخلايا تحتل منطقة داخل الجمجمة تحاط بالقشرة ومن ثم تحاط بالسائل المحيطي , القشرة هي المنطقة التي حددها العلماء لأغراض الذاكرة ومعالجة المعلومات ولها ارتباطات مباشرة بجميع المؤثرات والتأثيرات Afferent and Efferent ولاتزال المعرفة فيها غير كاملة لوظائف القشرة. تتكون القشرة من نوعين من الخلايا ،الخلايا الحبيبية Granular Cells والخلايا الهرمية Pyramidal Cells تخصص الخلايا الهرمية لوظائف التأثيرات في حين تخصص الخلايا الحبيبية لوظائف المؤثرات في الجهاز العصبي.

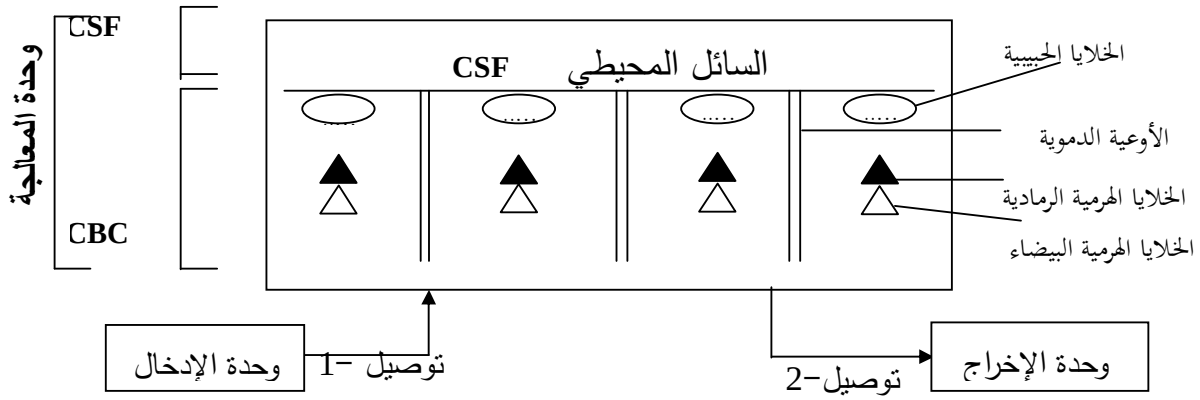
PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

2-السائل المحيطي Cerebrospinal Fluid (CSF)

هو أحد السوائل الخلوية الخارجية حيث تصنف السوائل في جسم الإنسان على نوعين ، سوائل داخلية وسوائل خارجية ،السوائل الداخلية تكون داخل أجسام الخلايا والسوائل الخارجية تكون خارج أجسام الخلايا ، يحتوي هذا السائل على مواد بتراكيز مختلفة منها الصوديوم ،والبوتاسيوم، والماء والكار بون ،والكلوريد أكثرها تركيزا هو أيونات الصوديوم التي تنفذ إلى أعماق القشرة عبر ظاهرة الانتشار.

النموذج الدماغي

تم صياغة النموذج من خلال جمع الكتلة الدماغية والسائل المحيطي معا كما يبينة الشكل(1).

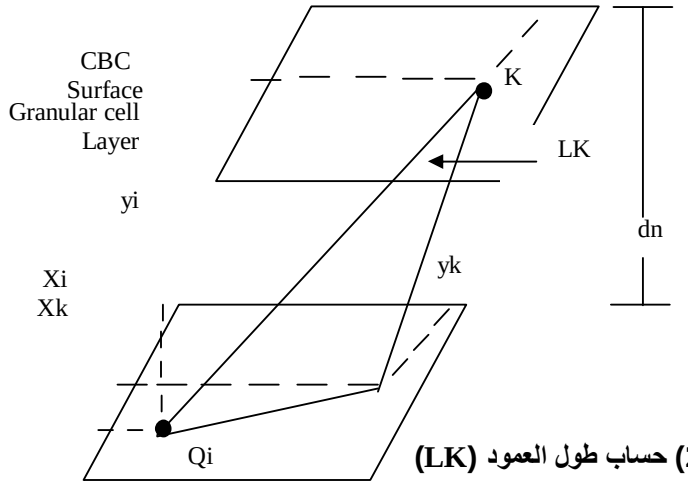


شكل (1) المخطط الكتلي للنموذج الدماغي

يتكون النموذج من ثلاث وحدات رئيسة هي وحدة الإدخال ، ووحدة المعالجة، ووحدة الإخراج ، وحدة الإدخال(التحسس) و وحدة الإدخال هي التي تستلم المنبه من البيئة الخارجية وتحوله على شكل إشارات كهربائية الى وحدة المعالجة من خلال التوصيل -1 ، ووحدة الإخراج تشبة وحدة الإدخال لكنها تستلم بياناتها من وحدة المعالجة من خلال التوصيل – 2 ، ووحدة المعالجة هي الوحدة الرئيسة والمهمة تتكون من جزئين جزء علوي صغير يدعى Cerebrospinal Fluid(CSF) وجزء سفلي كبير يدعى Cerebral Cortex(CBC) ويتكون هذا الجزء من مجموعتين من الطبقات , المجموعة العليا تمثل الخلايا الحبيبية والمجموعة السفلى تمثل بالخلايا الهرمية وتكون على شكل أزواج من الخلايا ،خلية هرمية رمادية اسفل منها خلية هرمية بيضاء كما يبينها الشكل(1).

ميكانيكية النموذج

عند استلام المنبه من البيئة الخارجية نفرض انه كان نصا" كتابيا" كما في نظام التمييز الحالي فان عناصر النص تنقل من وحدة الإدخال إلى وحدة المعالجة والبضبط إلى الخلايا الحبيبية(في وحدة المعالجة) عناصر النص تمثل على شكل شحنات وتمثل الخلايا الحبيبية مراكز هذه الشحنات ، هذه الشحنات تؤثر في أيونات الصوديوم الموجودة في السائل المحيطي بمجال فيعمل على تنظيم هذه الأيونات ويأخذ هذا التنظيم عدة أشكال اعتمادا" على عامل التمييز dn الذي يمثل المسافة البينية بين الخلايا الحبيبية والحافه العليا للخلايا المغلفة للقشرة الدماغية ، الشكل(2) يوضح ذلك.



شكل (2) حساب طول العمود (LK)

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

وفي جميع تلك التنظيمات تنفذ أعمدة من أيونات الصوديوم إلى أعماق القشرة بنسب وكميات اعتمادا على وجود الأيونات وانتشارها ، وباستعمال قانون كولوم بين الشحنات الموجودة في الخلايا وشحنات الحافة العليا وجدت العلاقة(1) [7] الآتية:-

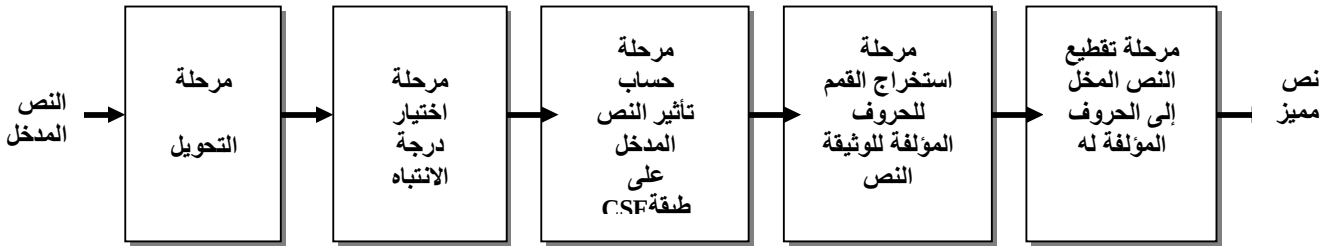
$$\mathbf{Lk}\ \alpha\ \sum_{i=1}^{n} Qi * dn/(d(i,k))^3 \quad \text{---------- (1)}$$

حيث أن n , ... i=1,2,..... تمثل كل مراكز تجمعات الخلايا الحبيبية .
K:- يمثل الموقع للنقطة الفضائية الممثلة بالإحداثيات (x,y,dn) .
dn :- يمثل عامل التمييز .
إما d(i, k) تحسب كالآتي :-

$$d(i,k) = \sqrt{(xi - xk)^2 + (yi - yk)^2 + dn^2}$$

نظام التمييز المقترح

يبين الشكل(3) المراحل الرئيسة لنظام التمييز المقترح ، وهي مرحلة التحويل ،مرحلة اختيار درجة الانتباه ،مرحلة حساب تأثير النص المدخل على طبقة CSF، مرحلة استخراج القمم للحروف المؤلفة لوثيقة النص ومرحلة تقطيع النص المدخل إلى الحروف المؤلفة له.



الشكل(3) المراحل الرئيسة للنظام التمييز المقترح

1-مرحلة التحويل

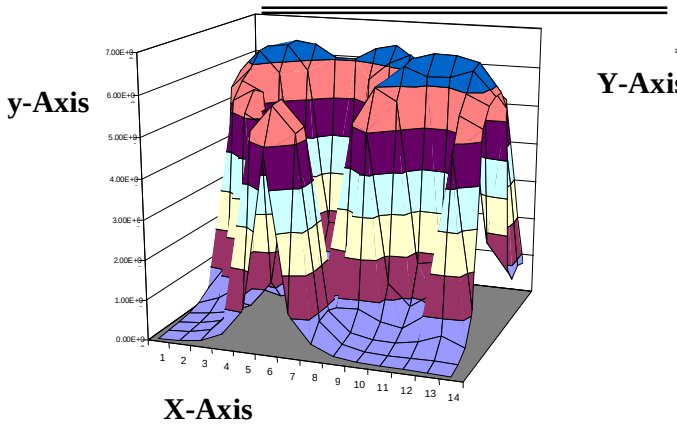
هي أول مرحلة في أي نظام تمييز أنماط ، وقد استعمل للعمل الحالي الماسح الضوئي HP Scanner Jet(اسود/ ابيض) كوسيلة لإدخال الوثائق ، ناتج هذه المرحلة هو تحويل ملف الوثيقة الذي يكون من نوع TIF من صيغة التماثلية إلى الصيغة الرقمية والتي تأخذ القيمتين(0,1) عادة.

2-مرحلة اختيار درجة الانتباه

يتم في هذه المرحلة اختيار درجة الانتباه المناسبة لعمل النموذج الدماغي التي تتم من خلال الاختيار المناسب لعامل التمييز dn، حيث أنجزت هذه العملية من خلال عمل عدة تجارب على مقطع كلمة مدخلة إلى النموذج وبدرجات مختلفة من الانتباه ، أي اخذ قيم مختلفة إلى عامل التمييز dn وتطبيقها بالعلاقة(1) ورسم نتيجة كل تجربة وبعد عدة تجارب تمكن من الحصول على درجة الانتباه المناسبة حيث كانت (0.5).

ومن الجدير الذكر أن هذه الدرجة افضل درجة انتباه لعمل النموذج تم الحصول عليها لعمل التمييز للنموذج،ولتوضيح هذه المرحلة نأخذ كمثال مقطع كلمة متكون من حرفين مثلا كلمة "بن" ، لنفرض في

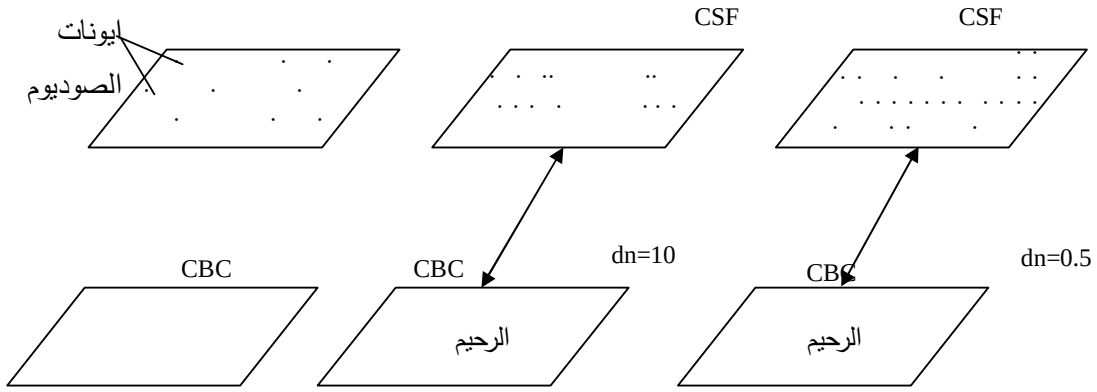
PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com



(أ)توحد القمم لحرفي(بـ،ن)عند dn= 20 (ب) تمييز قمم الحرفين (بـ،ن) عندما (dn=0.5)

شكل (4) المقطع "بن" بدرجات مختلفة للانتباه

التجربة الأولى تم اختيار قيمة إلى عامل التمييز dn هي (20) بعد الرسم نلاحظ أن القمم للحرفين (بـ ، ن) لا يمكن التمييز بينهما ، حيث نلاحظ توحد القمتين (بـ ،ن) بقمة واحدة تقريبا مما يشير إلى أن درجة الانتباه المختارة قليلة ، في التجربة الثانية عند اختيار درجة للانتباه كانت (0.5)مثلا نلاحظ من خلال الرسم تمييز القمم للحرفين (بـ،ن) مما يشير إلى أن درجة الانتباه المختارة كبيرة،الشكل(4) يوضح ذلك. يبين الشكل (5) تنظيم أيونات الصوديوم في طبقة (CSF)تبعا لاختيار الـ (dn) لكلمة " الرحيم" ,في الشكل((5(a) تكون الأيونات في طبقةCSF متناثرة قبل إدخال كلمة " الرحيم" ,أما الشكل((5(b تتشكل الأيونات على هيأة كومتين ,الكومة الأولى مقابل المقطع "الر" والكومة الأخرى مقابل المقطع "حيم" ولا نستطيع هنا تمييز المعالم الرئيسة لكلمة "الرحيم" وذلك بسبب أن درجة الانتباه المختارة قليلة((dn =10), أما في الشكل((5(c فنلاحظ أن الأيونات في طبقة CSF تشكلت على هيأة مماثلة لكلمة "الرحيم" وذلك لانه درجة الانتباه المختارة هنا كبير (dn= .0.5)



(a)أيونات الصوديوم قبل إدخال المقطع إلى طبقة (CBC)

(b) تنظيم أيونات الصوديوم بعد إدخال المقطع إلى طبقة (CBC)مع (dn=10)

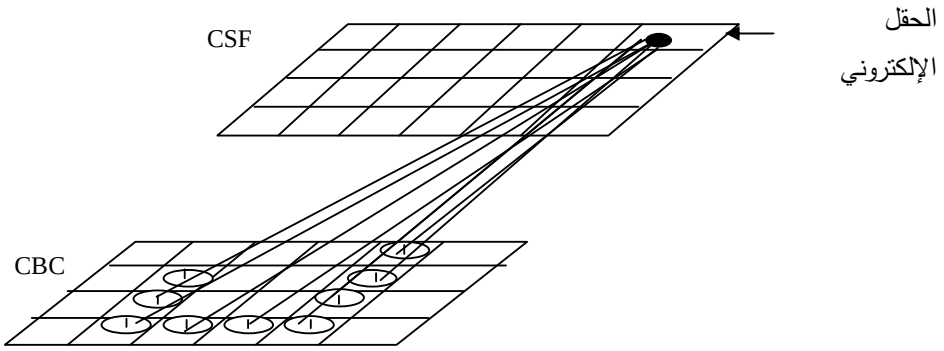
(c)تنظيم أيونات الصوديوم بعد إدخال المقطع إلى طبقة (CBC) مع (dn=0.5)

شكل (5) تأثير اختيار عامل التمييز (dn) في تنظيم أيونات الصوديوم لكلمة "الرحيم"

1- مرحلة حساب تأثير النص المدخل في طبقة CSF

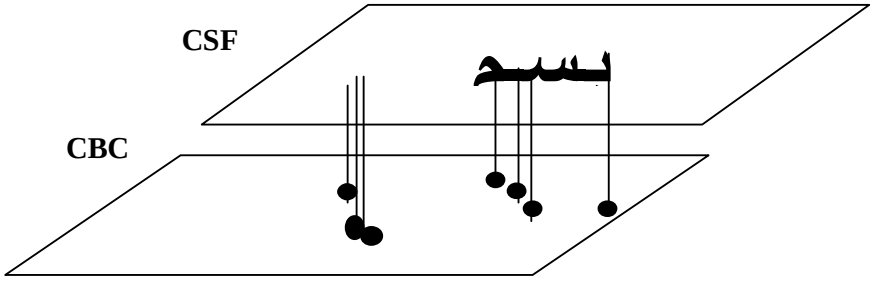
النص المدخل يحدث تأثيرات وقد تم توضيحها في ميكانيكية النموذج الدماغي ، وفي هذه المرحلة تم حساب التأثير الذي يحدثه هذا النص وباستعمال العلاقة(1)، الشكل(6) يبين عملية الحساب.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com



شكل (6) عملية حساب كل شدة حقل في طبقة CSF

حيث نلاحظ من خلال الشكل أن عملية حساب شدة كل حقل تشترك عناصر النص جميعها في حساب شدة كل حقل فضلا" عن ذلك فانه عملية حساب شدة كل حقل تمثل عملية حساب أطوال عمود الصوديوم النافذ في طبقة CBC فتتولد بذلك أعمدة نافذة مختلفة الأطوال يوضحها الشكل(7) لمقطع كلمة "بسم".



شكل (7) مناطق نفوذ أطوال أعمدة الصوديوم في طبقة (CBC) لمقطع كلمة "بسم "

4- مرحلة استخراج القمم للحروف المكونة لوثيقة النص المدخل

أن ترتيب البيانات على طبقة CSF الناتجة من المرحلة السابقة تشير إلى ضرورة إيجاد طريقة لاستخراج القيم العظمى على هذه الطبقة ،أن الفكرة المقترحة لاستخراج القيم العظمى يتم بالحركة على طبقة CSFمن اليسار إلى اليمين ومن الأعلى إلى الأسفل ، حيث يتم اختيار العنصر مع مجاوراته الأربع إذا كان اكبر من مجاوراته الأربع يتم استخراجه،بحيث نهاية هذه المرحلة نحصل على قائمة بالقيم العظمى ودلائلها ،فمثلا المقطع "بسم" في الشكل(7) نحصل على القائمة الممثلة بالجدول(1).

جدول (1) القمم المستخرجة مع دلائلها لمقطع "بسم "

Num Point	Peak Values	Peak Locations	
		x-axis	y-axis
1	6.8350	5	4
2	6.8486	5	25
3	6.6242	6	21
4	6.8132	7	40
5	5.4341	11	1

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

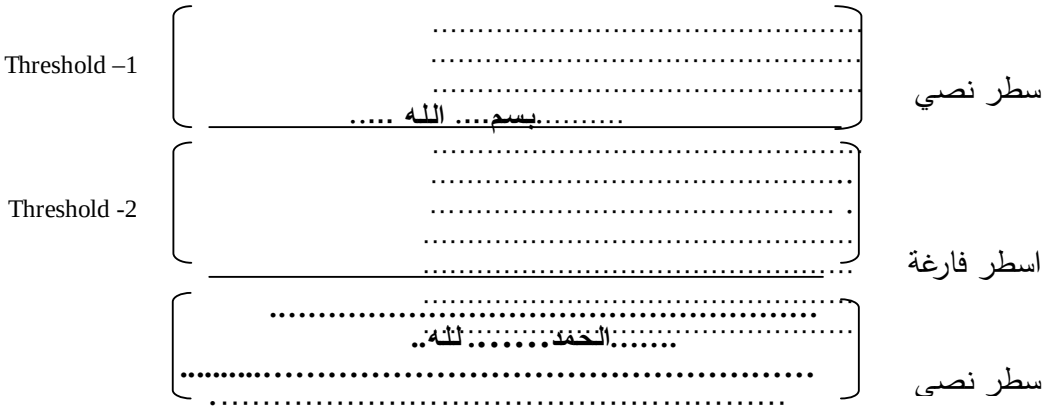
6	5.3554	12	2

5- مرحلة تقطيع النص المدخل إلى الحروف المؤلفة له

بالاعتماد على دلائل القمم الناتجة من المرحلة السابقة يتم تقطيع الوثيقة بالكامل إلى الحروف المكونة لها ، حيث تم من المرحلة السابقة اختزال الحجم الهائل لعناصر وثيقة النص بمجموعة من النقاط التي تمثل نقاط القمم للحروف ، وقد تضمنت هذه المرحلة مجموعة من العمليات الثانوية على نقاط القمم لغرض تقطيع الوثيقة إلى الحروف المؤلفة لها ومن ثم تصنيف هذه الحروف ،هذه العمليات تشمل :-

1- عملية تحديد دلائل القمم الخاصة بكل سطر نصي في الوثيقة

تساعد في هذه العملية المميزات التي توفرها طبيعة الكتابة ،حيث أن كل سطر نصي في الوثيقة يتكون من مجموعة من الأسطر النقطية تحدد بعتبة معينة Threshold -1 ،فضلاً عن ذلك تفصل السطور النصية الواحدة عن الأخر بفراغ هذا الفراغ أيضا يترجم بمجموعة من الأسطر النقطية Point Lines السوداء وتحدد عتبة معينة لها Threshold-2 كما يبنها الشكل (9)



شكل (9) تنظيم السطور النصية والسطور النقطية في وثيقة النص

وعليه يتم تحديد دلائل القمم الخاصة بكل سطر نصي بمساعدة المميزات التي وفرتها طبيعة الكتابة التي فرضت مقدار محدد لسعة الانتشار الطولي للسطر النصي إذ تؤخذ فقط دلائل النقاط الواقعة في ضمن المدى المحدد بين السطر النقطي الأول والسطر النقطي الأخير للسطر النصي المراد استخراج دلائل القمم الخاصة به ،وان العلاقات التي تحدد المدى لكل سطر تكون كما يأتي:-

بداية السطر النصي الحالي (LS) = نهاية السطر النصي السابق (LEP) +Threshold-2 ..... (2)

نهاية السطر النصي الحالي(LE) = LS + Threshold-1 ..... (3)

2- عملية ترتيب دلائل القمم للسطر النصي المحدد

من الضروري إجراء عملية ترتيب إلى دلائل القمم الخاصة بكل سطر تم تحديده من العملية السابقة ، حيث تكون دلائل القمم المستخرجة والمحددة التي تكون غير مرتبة بسبب طريقة استخراجها التي تم توضيحها في المرحلة السابقة ، مثلا القائمة الناتجة في الجدول(1) تم ترتيبها كما في القائمة المبينة في الجدول (2) ،ومن الملاحظ أن عملية الترتيب تكون حسب المحور العمودي (Y-axis) لاعتماد عملية التقطيع على هذا المحور .

Num Point	Peak Values	Peak Locations	
		x-axis	y-axis
1	5.4341	11	1
2	5.3554	12	2
3	6.8350	5	4
4	6.6242	6	21
5	6.8486	5	25
6	6.8132	7	40

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

جدول (2) النقاط مرتبة تبعا إلى المحور العمودي (y-axis)

3-عملية تقطيع السطر النصي إلى مفردات الحروف المؤلفة له

يتم اختيار كل نقطتين متتاليتين في قائمة النقاط واعتبارهما نقطتين مرشحتين للقطع ولكي تتفق عملية التقطيع مع طبيعة الكتابة العربية التي تكون من اليمين إلى اليسار نبدأ بآخر نقطة (التي تمثل عندها أعلى قيمة للمحور العمودي في قائمة النقاط) والنقطة التي تسبقها من العلاقة الآتية:-

المسافة = دليل العمود للنقطة الحالية – دليل العمود للنقطة التي تسبقها .....(4)

ويتم إقرار القطع بين النقطتين المحددتين حسب الحالات الآتية:-

الحالة الأولى:-

إذا كانت المسافة بين النقطتين المرشحتين أكبر من عتبة Threshold (حيث تم تحديد هذه العتبة بمساعدة المميزات التي توفرها سعة الانتشار العرضي للحروف) يتم إقرار القطع بين النقطتين ،فمثلا المقطع "بسم" , النقطتين (40,25) من نقاط المحور العمودي يتم هنا إقرار القطع لانه المسافة الموجودة بين النقطتين تجاوزت حد Threshold أي يتم الحصول على حرف "ب" من القطع بين النقطتين .

الحالة الثانية:-

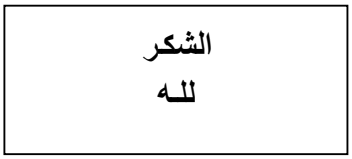
أما إذا كانت المسافة المحسوبة اقل من العتبة المحددة فيتم تجميع المسافات بين كل نقطتين متتاليتين بحيث تستمر عملية التجميع حتى تصل إلى حد معين تتوقف عملية التجميع وبعدها تحدد نقطتين للقطع من سلسلة النقاط المجمعة وهما النقطة الأولى والنقطة الأخيرة من السلسلة, وهذا ما يوضحه تقطيع الحرف "س" في المقطع "بسم" ,حيث يتم تجميع النقاط (25,21,4) ومن ثم يتم القطع بين النقطتين (25,4)وكذلك بالنسبة للحرف الأخير من المقطع "بسم" ,حيث يتم تجميع النقاط (1,2,4) ومن ثم يتم القطع بين النقطتين(1,4).

4-عملية التصنيف

قبل الشروع بتصنيف الحرف المقطوع يتم أجراء بعض العمليات الثانوية على الحرف المقطوع لغرض تهيئته إلى عملية التصنيف ومن هذه العمليات عملية إزاحة النقطتين المرشحتين للقطع إلى اليمين وذلك لتحديد بداية القمة للحرف المقطوع وكذلك إزالة الأجزاء المتصلة بالحرف(خطوط الاتصال) من نهايتي الحرف وذلك للحصول على الحرف بشكل تام وكذلك عملية نقل الحرف إلى إطار وتقيس الموقع وهي من العمليات التي تساعد على تقليل كل من الخزن وزمن المعالجة للحرف. بعد هذه العمليات يتم استخراج صفات الحرف المقطوع ،حيث يستخرج موقع وقمة الحرف المقطوع من خلال حساب التأثير الذي يحدثه الحرف في أيونات الصوديوم الموجودة في السائل المحيطي ، ثم يتم استخراج أطوال أعمدة الصوديوم النافذة لهذا الحرف في طبقة CBC والذي يمثل قمة الحرف ولغرض التصنيف يتم مقارنة موقع قمة الحرف مع مواقع القمم لحروف والمستخرجة نتيجة عمليات تعلم سابقة لهذه الحروف تم تعليمها إلى النموذج في طور التعلم المشرف عليه.

المناقشة والنتائج والاستنتاجات

تم اختيار الوثيقة الموضحة في الشكل (10) لمتابعة سير تنفيذ النظام.



شكل (10)الوثيقة المدخلة

ولغرض السيطرة على الكم البياني الهائل عند القيام بعملية طباعة النتائج من كل مرحلة من مراحل النظام المتتالية تم اختيار وثيقة بالحجم الموضح بالشكل (10)،وفيما يأتي النتائج المتولدة لكل مرحلة في النظام :-

1-مرحلة التحويل

يتم إدخال اسم ملف الوثيقة (InputFile.TIF) ،ناتج هذه المرحلة وهو موضح في الشكل (11) .

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000001100000000000000000110000000100000
0000
0000000000000000000000000001100000000000000000110000000000000000111000000010000
0000
0000000000000000000000011111000000000000000001111000000000000011100000110000
0000
0000000000000000000000011110000000000000000001110000000000000001000000110000
0000
0000000000000000000000011100000000000000000001110000000000000001000000011000
0000
0000000011000000000000011110000000000000000000010000000000000001000000011000
0000
0000000011000000000000001111000000000000001100111100000000000011000000011000
0000
0000000011000000000000000111000000000000001110010110000000000011000000011000
0000
0000000001111111111111111111111111111111111111111111111111111111100000011000
0000
0000000001111111111111111111111111111111111111111111111111111111100000011000
0000
0000000001111111111111111111111111111111111111111111111111111111100000011000
0000
0000000011000000000000000000000000000000000000000000000000000000000000010000
0000
0001111111000000000000000000000000000000000000000000000000000000000000000000
0000
0001111110000000000000000000000000000000000000000000000000000000000000000000
0000
0000011100000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000
0000

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000100000000000000010000000
0000
0000000000000000000000000000000000000000000000000000000000000011100000000000001110000000
0000
0000000000000000000000000000000000000000000000000000000000001000000000000000111100000000
0000
0000000000000000000000000000000000000000000000000000000000011000000000000001100000000000
0000
0000000000000000000000000000000000000000000000000000000000011000000000000001100000000000
0000
0000000000000000000000000000000000000000000000000000000000011100000000000001100000000000
0000
0000000000000000000000000000000000000000000000000000000000011110000000000000100000000000
0000
0000000000000000000000000000000000000000000000000000000000011111000000000000011000000000
0000
0000000000000000000000000000000000000000000000000000000000011111100000000000011000000000
0000
0000000000000000000000000000000000000000000000000000000000010111111111111111111111111111
0000
0000000000000000000000000000000000000000000000000000000000011111111111111111111111111111
0000
0000000000000000000000000000000000000000000000000000000000011111111111111111111111111111
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000

OTI 01RID0000000000000000000000000000000000000000000000000000000000000000
0000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000

شكل (11) ناتج مرحلة التحويل للوثيقة

2 –مرحلة اختيار درجة الانتباه

يتم في هذه المرحلة إدخال درجة الانتباه التي تم الحصول عليها بالطريقة الموضحة في المرحلة الثانية من النظام ،حيث كانت الدرجة المناسبة (0.5) .

3–مرحلة حساب تأثير النص المدخل في طبقة CSF

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

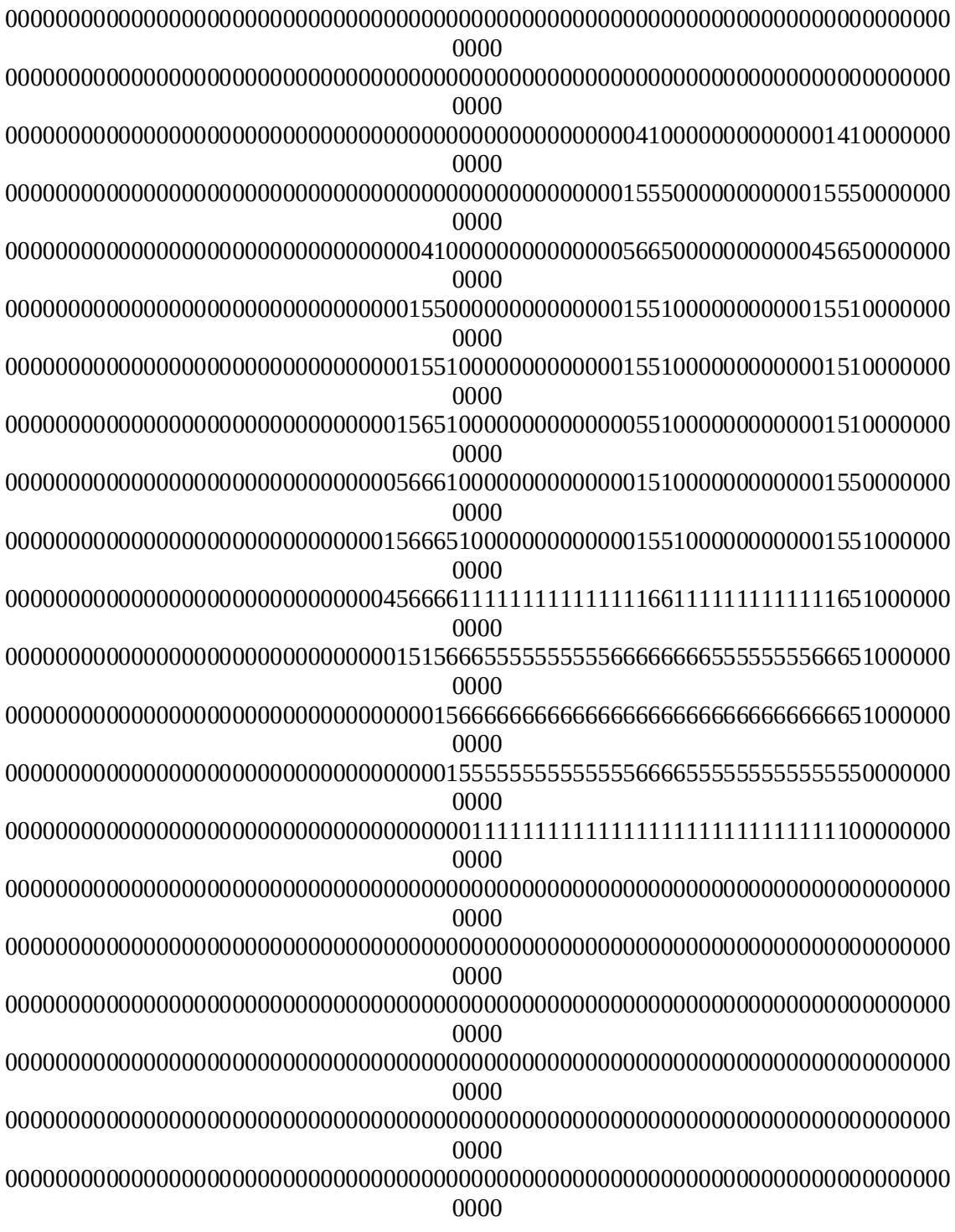
في هذه المرحلة يتم تطبيق العلاقة (1) على بيانات الملف (InputFile.TIF) وبدرجة الانتباه المحددة من المرحلة الثانية يكون ناتج هذه المرحلة وهو موضح في الشكل (12).

00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000055000000000000000551000000400000
0000
00000000000000000000001155000000000000000015510000000000000565000001500000
0000

OTI 01RID00000000000000000000555650000000000000001

000000555000055000000000
00000000000000000000156651000000000000000015651000000000000151000005510000
0000
00000000000000000000156620000000000000000015551000000000000151000001550000
0000
00000005500000000000156651000000000000000101125100000000000151000001550000
0000
00000005510000000000156651000000000000015511556510000000000155000001551000
0000
00000005511111111111112666111111111111156611526611111111111155100001551000
0000
00000001566555555566666666666555555566666666666666555555556665100001551000
0000
00000001566666666666666666666666666666666666666666666666666665100000550000
0000
00000001566555555555666666665555555555666666666666555555555555000000550000
0000
00001115511111111111111111111111111111111111111111111111111110000000410000
0000
00555556510000000000000000000000000000000000000000000000000000000000000000
0000
00556665100000000000000000000000000000000000000000000000000000000000000000
0000
00015551000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000
0000
00000000000000000000000000000000000000000000000000000000000000000000000000
0000

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com



شكل (12) طبقة CSF بعد مرحلة الحساب

4-مرحلة استخراج القمم للحروف المكونة لوثيقة النص

يكون ناتج هذه المرحلة قائمة بقيم القمم للحروف المنتشرة على سطح وثيقة النص وهو ما يوضحه الجدول (3) .

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

جدول (3) قائمة بالقمم مع دلائلها على سطح وثيقة النص المدخل

رقم القمة	القمة	المحور الأفقي	المحور العمودي
1	6.0178	7	71
2	6.0864	8	34
3	6.3964	9	54
4	6.5884	10	32
5	6.6416	12	33
6	5.7050	12	79
7	6.8571	14	34
8	6.7528	14	50
9	6.5473	15	20
10	6.7438	15	56
11	6.5443	15	69
12	6.134	18	17
13	6.3560	19	15
14	6.1921	28	61
15	6.1332	28	76
16	6.6560	33	43
17	6.2817	35	45
18	6.6801	36	61
19	6.5363	36	74

5- مرحلة تقطيع النص المدخل إلى الحروف المكونة له

إن هذه المعالجة تعتمد على دلائل القمم المستخرجة في المرحلة السابقة و تمت الإشارة إليها إذ تتكون هذه المرحلة من أربع عمليات ثانوية تكرر على كل سطر نصي في الوثيقة وهي :-

1-عملية تحديد دلائل القمم الخاصة بكل سطر نصي في الوثيقة

يتم تحديد دلائل القمم الخاصة بالسطر النصي الأول التي يكون فيها قيم المحور الأفقي واقع ضمن المدى (1,20) وكما في الجدول (4) .

جدول (4) قائمة بالقمم مع دلائلها الخاصة بالسطر الأول للوثيقة

رقم القمة	القمة	المحور الأفقي	المحور العمودي
1	6.017	7	71
2	6.0864	8	34
3	6.3964	9	54
4	6.5884	10	32
5	6.6416	12	33
6	5.7050	12	79
7	6.8571	14	34
8	6.7528	14	50
9	6.5473	15	20
10	6.7438	15	56
11	6.5443	15	69
12	6.1341	18	17
13	6.3560	19	15

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

2- عملية ترتيب دلائل القمم الخاصة بكل سطر نصي في الوثيقة

ترتب دلائل القمم للسطر الأول بحسب المحور العمودي لها و هو موضح في الجدول (5) .

جدول (5) قائمة بالقمم مع دلائلها بعد عملية الترتيب

رقم القمة	القمة	المحور العمودي	المحور الأفقي
1	6.3560	19	15
2	6.1341	18	17
3	6.5473	15	20
4	6.5884	10	32
5	6.6416	12	33
6	6.8571	14	34
7	6.0864	8	34
8	6.7528	14	50
9	6.3964	9	54
10	6.7438	15	56
11	6.5443	15	69
12	6.0178	7	71
13	5.7050	12	79

3- عملية تقطيع السطر النصي إلى الحروف المؤلفة له

اعتماداً على دلائل القمم الناتجة من العملية السابقة تحدد في كل مرة سلسلة من النقاط المتتالية التي تمثل نقاط القمم للحرف، مثلا للحرف "شـ" تحدد النقاط (56,54,50,34) وفي النهاية تحدد نقطتا القطع وهما النقطتان الأولى والأخيرة من سلسلة النقاط المتتالية (56,34).

4-عملية التصنيف :

تتكون هذه العملية من أربع خطوات وقد سبقت الإشارة إليها وهي :-

أ- إزاحة نقطتي القطع إلى اليمين . ج- نقل الحرف إلى إطار ثابت وتقيس الموقع

ب- تعديل نهايتي الحرف . د- استخراج الصفات للحرف .

ولغرض اختصار عمليات الطباعة لكل أحرف السطر تم اختيار حرف " شـ" في كلمــة " الشكر"ولتوضيح هذه الخطوات على هذا الحرف وكما في النتائج الموضحة في الشكل (13) .

الحرف قبل إزاحة نقطتي القطع	الحرف بعد إزاحة نقطتي القط	تعديل بداية الحرف
00000000000000000000000	00000000000000000000000	00000000000000000000000
00000000000000000000000	00000000000000000000000	00000000000000000000000
00000000000000000000000	00000000000000000000000	00000000000000000000000
00000000000000000000000	00000000000000000000000	00000000000000000000000
00000000000000000000000	00000000000000000000000	00000000000000000000000
00000000000000000001100	00000000000000000000000	00000000000000000000000
11000000000000000001100	00000000000000000000000	00000000000000000000000
11000000000000000001111	00000000000000000000000	00000000000000000000000
10000000000000000001110	00000000000000000110000	00000000000000000000011
00000000000000000001110	00000000000000000110000	00000000000000000000011
10000000000000000000001	00000000000000000111100	00000000000000000000011
11000000000000001100111	00000000000000000111000	00000000000000000000011
11000000000000011100101	00000000000000000111000	00000000000000000000011
11111111111111111111111		
11111111111111111111111		
11111111111111111111111		
00000000000000000000000		
00000000000000000000000		
00000000000000000000000		

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

00000000000000000000000

تعديل نها ية الحرف	عملية نقل الحرف إلى إطار ثابت وتقيس الموقع	حساب تأثير الحرف على طبقةCSF
0000000000000000000	00000000000000001	000000000000001551
0000000000000000000	00000000000001000	000000000000001662
0000000000000000000	00000000000000001	000000000000001666
0000000000000000000	00000000000001000	000000000000001666
0000000000000000000	00000000000000001	000000000000001666
0000000000000011000	00000000000001110	000000000000011566
0000000000000011000	00000000000000001	
0000000000000011110	00000000000001100	
0000000000000011100	00000000000000001	
0000000000000011100	00000000000001100	
0000000000000000010	00000000000000001	
0000000000011001111		
0000000000111001011		
1111111111111111111		
1111111111111111111		
1111111111111111111		
0000000000000000000		
0000000000000000000		
0000000000000000000		
0000000000000000000		
0000000000000000000		
0000000000000000000		
0000000000000000000		

شكل (13) الخطوات المختلفة التي تجرى على حرف "شـ" في عملية التصنيف

في نهاية هذه العملية يتم إيجاد موقع وقمة الحرف المقطوع وذلك لغرض تصنيف الحرف .

- موقع القمة =(9,13)
- القمة للحرف =(6.7)

تكرر خطوات عملية التقطيع على حروف السطر المحدد كلها إلى نهاية السطر ،بعد انتهاء عملية تقطيع السطر إلى حروف تستأنف مرحلة تقطيع الوثيقة إلى الحروف المكونة لها وذلك من خلال إعادة العمليات نفسها على السطر الثاني وهكذا إلى أن تنتهي الوثيقة ،ويكون ناتج تمييز الوثيقة المدخلة الحروف الآتية :- " ا ل ش ك ر ل ل ه " وقد تم اختبار النظام على مجموعة من الوثائق كانت تحتوي الواحدة منها بحدود (160)كلمة إذا كانت نسبة نجاح عملية التقطيع في النظام اكثر من % 95 ونسبة التمييز 85%ومن متابعة سير تنفيذ البرنامج أمكن تحديد مسببات الفشل لعملية التقطيع والتمييز ,حيث كانت تحدث هذه عند الحروف ذات سعة الانتشار العرضي الكبير مثل حرف (ضـ ،س) ويرجع السبب الرئيس لوجود بعض المواقع الفارغة من البيانات في ضمن النمط التي من المفترض أن تحتوي تلك المواقع على البيانات ,كذلك أن زيادة عملية التعلم للدماغ يعزز من عملية تمييز النظام ،أي أن تكرار عملية إدخال كلمات مختلفة تحتوي على الحرف نفسه يعزز من عملية التمييز مثلا عند تكرار عرض أربع كلمات تحتوي على حرف الباء مثل (بسم ، بسمه ، بسام ،بيان) سوف يعزز من تمييز النظام للحرف "بـ" ، وكذلك **من** الجدير بالذكر أن النظام في تمييز للوثائق العربية المطبوعة يأخذ فقط الوثائق المطبوعة بدون حركات.

المصادر :

[1] Saleh S.F. "An On- Line Automatic Arabic Document Reader ",M.Sc. thesis ,Basrah Univ. 1998.

[2] فاتن حسين عبد الله ،"تطوير نظام تمييز الحروف العربية باستخدام سلسلة فريمان" ,اطروحة ماجستير ،جامعة البصرة،العلوم، 2000.

[3] Al-Amiry ,M.J.A , "A Real Time Universal Character Recognizer[UCR] with Writing Tablet Design ",M.Sc. thesis ,Basrah ,Univ October 2000.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com

مهدي صالح جالي، "دراسة في معالجة الصور الرقمية بواسطة المعالج الشبكي ",اطروحة ماجستير ,جامعة [4] البصرة،1985.

[5] AL-Yousefi ,H and Udpa ,S.S, "Recognition of Arabic Characters", IEEE Transaction on PAMI vol.14,No.8,pp.853-857, August 1992.

[6] EL-Desouky ,A.L., Mofreh, M.S., Aida,O.A. and Hesham,A. "A Handwritten Arabic Character Recognition Technique for machine Reader ",IEEE Transaction on computer , 9 –pp.212-216 ,1991.

Pattern [7] Rasheed ,W.A.M ,"A Medium Scale Based Brain Model with Recognition Implementation "PH.D.Thesis Basrah Univ. 1996.

[8] Aurthur, G.G. "Text Book of Medical Physiology ", Saunder ,Canada ,1971.

Independent the Brain Modeling for Arabic Characters Recognition

A. Fathal And S. A.Sead

college of science /computer science department

Abstract

The research introduces a system to recognize the printed Arabic characters ,the system algorithm has been concluded from the modeling of the brain human behaviors .The basis of this system depends on the interpretation of human ability in recognizing and reading the characters in connection with failure in writing them .The design of the suggested system including five principal stages. The first stage reads the printed Arabic documents making use of the HP Scanner Jet which generates TIF (Tagged Image File)file .The second one determing the proper degree of attention for the model operation. The third one processes the effects of input text on activating of some brain components .The fourth one extracts the tops of the characters from the surface of the text document The last divides the whole text into separate characters (segmentation). A number of new ideas and methods are proposed for the purpose of segmentation ,The application of these proposals have given good and satisfactory results ,by which the adopted process of segmentation can be simplified in other well–known methods then ,the segmented characters can be classified depending on the abstracting experience as a result of the previous work .As the well as the system is supplied with the phase of learning according to its supervisor .The present work introduces a design of a system for recognizing the characters is able to overcome the most attributes of Arabic writing .The new direction of this work can be achieved throughout mixing the human behavior in building such systems.

PDF Created with deskPDF PDF Writer - Trial :: http://www.docudesk.com