المقياس: اللغة والحاسوب السنة الأولى والثانية ماستر الأستاذ: أنور طراد

المحاضرة الخامسة: التعرف الآلي على الكتابة العربية

1 **تاريخ حوسبة الخط العربي:** بدأت حوسبة الخطاطة العربية بشكل مقبول في نهاية الثمانينيات وبداية التسعينات على غرار الخط الإنجليزي، وسبب هذا التأخر يعود إلى جملة من الأسباب:

- المحدودية الشديدة لقدرات الحواسيب قبل هذه السنوات، فقد لا تستوعب هذه الحواسيب إلا خمسين ألف كلمة فقط، وهذا ما يدلّ على أن البرمجيات العاملة وبيانتها لا يمكن أن تتعدى هذه الذاكرة الضئيلة.
- إنغلاق أنظمة هذه الحواسيب أمام المطورين من خارج الشّركات المصنعة لها، مما جعل الأمر عسيرا على تعديل هذه البرمجيات.

بعد هذه الفترة، وفي بداية النّصف الثّاني من ثمانينيات القرن العشرين، وقعت تطورات مهمة في عالم الحاسبات الصغيرة ونشأت شركات مثل أبل، التي قامت بتطوير الحاسبات والبرمجيات، ثم طورت شركة إ ب م، الأمريكية حاسبات أخرى وبرمجيات أكثر دقة وأكثر سعة، وقامنت هذه الشركات بالفصل بين العتاد والبرمجيات مما سمح بسهولة تعديلها من طرف المطورين من خارج هذه الشركات، ففتح المجال واسعا أمام مطورين عرب في إنتاج برمجيات عربية وتطويرها، وتعريب بعض البرمجيات الشهيرة، ولقيت تلك المحاولات نجاحا كبيرا سمح بتطوير وحوسبة الخطاطة العربية.

2 – تحديات الخطاطة العربية التي تواجه التعرف عليها آليا:

❖ **اتصال الحروف:** كما هو معلوم فإن حروف اللغات الأوروبية يمكن كتاباتها متصلة ومنفصلة، أما اللّغات كالصينية، واليابانية، والكورية، فتكتب بالخط المنفصل، والخطاطة العربية والبنغالية والهندية، تكتب دائما بالحرف المتصل، ومن وجهة نظر أية آلية حاسوبية للتّعرف على الأنماط الرّسومية، **يكون _ التّعرف على الأنماط وهي منفصلة أيسر من التعرف عليها وهي متصلة**، بحيث يستوجب في حال الانفصال حلّ مسألة تعيين كلّ رمز رسومي (جرافيم)،

ويسمى **التقطيع**، إضافة إلى مسألة التعرف على الحرف الذي يرمز إليه كل جرافيم، وهذا يعني أنه لا مفرّ من حلّ كلتا المسألتين: (التّعرف+ التقطيع) آنيا، وهذا ما يضاعف التّحدي.

- **التداخل بين حدود الجرافيمات:** من بين الأمور التي تضاعف التحدي، نرى بعض التراكب والتداخل الطفيف بين حدود الحروف.
- **تغير رسم الحرف مع تغير موضعه في الكلمة:** وهذا ينتج عن الكتابة المتصلة، ويؤدي هذا التغيير إلى زيادة كبيرة في عدد الرموز الرسومية، التي يجب أن يتعامل معها أي نظام للتعرف الآلي على النص العربي المكتوب، بالمقارنة مع اللغات الأوربية (الحرف المنفصل) ومن وجهة نظر أية آلية حاسوبية للتعرف على الأنماط الرسوية، يكون التعرف أكثر دقة كلما قلت عدد الأنماط الرسومية، والعكس بالعكس. مثاله: (على- العربية- مع- القطاع)
- **الجرافيمات المركبة من أكثر من حرف واحد:** يتعلق الأمر ببعض الخطوط التي تحتوي على بعض الجرافيمات المركبة من حرفين أو ثلاثة (خاصة الخطوط المزخرفة)، والتي لا يمكن التعامل معها إلا كوحدة رسومية واحدة، وهذه الجرافيمات المركبة ترفع عدد الأنماط الرسومية التي يتوجب على أي نظام للتّعرف على النّص المكتوب التّعامل معها، مما يرفع درجة الصّعوبة، فإذا كانت الإنجليزية في هذا الصّدد قد تكتفي بحوالي ثمانين جرافيما، فإن العربية تحتاج إلى ما يزيد عن 190 جرافيما، مثال ذلك: (لا، **مم**، **حج**..).

3 - مراحل التّمييز الآلي للكتابة العربية آليا:

تمر عملية التمييز الآلي للكتابة العربية والتعرف على الحروف بمراحل أختصرها فيما يلي:

- **المرحلة الأولى**: تمييز السطور كلّ على حدة، وذلك بعملية مسح رأسي تحدد من خلاله الفواصل الرأسية ما بين السطور، وتفرّق بين هذه الفواصل، وتلك التي تفصل ما بين الحروف وتقاطعها، وعلامات تشكيلها.
- **المرحلة الثانية**: فك التشبيك لتحليل سلاسل الكتابة المتصلة إلى حروف منفصلة، ويتم ذلك بعملية مسح أفقي، لتمييز خانات الفراغ الفاصلة بين الكلمات، وتلك التي ترد

داخل الكلمات نفسها، مثل كلمة (زرع، مزروع...)، ثم بعد ذلك تمييز مقاطع التشبيك الأفقية التي تصل بين الحروف المتتالية وتتم عملية التمييز باستخلاص الملامح الرئيسة (خاصة في حالة الخط اليدوي.) للخط الجاري تمييزه، مثل: عدد النّقاط ومواضعها التي يحتويها الحرف، وجود حلقات مغلقة في شكل الحرف مثل الحرف: ط، ق،ف..، وطول الحرف وعرضه، وصعوده أو هبوطه، عن خط التشبيك الأفقي المتصل.

- **المرحلة الثالثة**: تحديد نمط الحرف، وذلك بـ: تقسيم الإطار المستطيل للحرف بعد تكبيره داخل ذاكرة الحاسوب إلى خانات وخلايا، مربعة أو مستطيلة ومقارنة الشكل المحصور في الخلايا الصغيرة بعدد محدود من النماذج النمطية التقريبية المخزنة قبل داخل ذاكرة الحاسوب، لتحديد أقربها إلى شكل الجزء المحصور بكل خلية.
- **المرحلة الرابعة**: مقارنة النمط الذي تم تحديده في المرحلة السابقة بشكل الحرف الجاري تمييزه مع أنماط الحروف المعمارية المخزنة في ذاكرة النظام الآلي، فبالنسبة للكتابة المطبوعة المنسوخة بالآلة الكاتبة تتم المقارنة بدقة عالية، أما في حالة الخط اليدوي فتتسم المقارنة بعد التحديد القاطع والدقيق، في ظل تداخلات كثيرة من حروف أخرى متشابهة، وترتكز عملية المقارنة على اساس الفرق بين نمط الحرف الجاري تمييزه، والأنماط المخزنة التي تنتمي إليه، وذلك باستخدام أساليب ومعادلات رياضية وإحصائية، تساعد على التقريب والتوفيق.
- **المرحلة الخامسة**: إذا تمت عملية المقارنة بنجاح اختتمت عملية تمييز الحرف، ليتقل نظام التمييز الآلي إلى الحرف الذي يليه، أما إذا فشلت عملية التمييز فيستلزم تقديم عدة بدائل له، باستخدام بيانات إحصائية وقرائن لغوية وسياقية لتخمين الحرف المبهم.

المراجع:

- نبيل علي: اللغة العربية والحاسوب.
- تطبيقات أساسية في المعالجة الآلية للغة العربية، محسن رشوان وآخرون.