رقمنة الوثائق العربية وتقنياتها

تحت إشراف:

الأستاذ الدكتور محسن عبدالرازق رشوان الخبير بالمجمع

بالاشتراك مع:

لجنة اللغة العربية والذكاء الاصطناعي بالمجمع

رقمنة الوثائق العربية وتقنياتها

ملخص البحث(*)

يرتكز البحث حول مفهوم رقمنة الوثائق العربية وما يتعلق بها ابتداءً من صورتها الورقية بمختلف أنواع الوثائق وصولًا إلى صورتها الرقمية. ويتناول البحث مختلف المجهودات لرقمنة الوثائق والخطوات الأساسية لبدء مشروع الرقمنة وما يتعلق بها من أجهزة وتقنيات مختلفة مثل التعرف الضوئي للحروف، والمشكل الآلي ومحول النصوص إلى نصوص مسموعة، وكذلك محركات البحث التي تساعد على الرقمنة بمختلف صورها. وفي النهاية سنستعرض بعضًا من نتائج العمل بمحركات التعرف الضوئي على بعض من الوثائق لإبراز الفرق بينها.

(*) شارك في عمل هذا البحث محمود سامي أبوالعزم.

مقدمة

نظرًا لما نشهده في عصرنا الحديث من كثرة المعلومات وأهميتها لمواصلة التقدم والمعرفة، فإن الوصول إلى المعلومة يُعتبر من أهم التحديات الموجودة حاليًا في مواجهة التقدم؛ حيث توجد المعلومات بشكل غير رقمي أو بمعنى آخر في صورة محفوظات ورقية بمختلف أنواعها؛ ولذلك تأتي أهمية رقمنة الوثائق لتحويل الوثائق من صورتها الورقية إلى صورتها الرقمية؛ حيث يتم تصوير الوثائق بأحدث تكنولوجيا التصوير، ومن ثم معالجتها باستخدام التكنولوجيا الحديثة عبر أحدث تقنيات الذكاء الاصطناعي ليعطي القدرة على استعادة المعلومات بأسهل الطرق وأسرعها.

أصبحت الرقمنة مصطلحًا يُستخدم كثيرًا في وقتنا الراهن؛ حيث إننا سنعتمد في كل تكنولوجياتنا على الرقمنة، وهذا يقودنا لتعريف الرقمنة. الرقمنة هي وسيلة باستخدام أنواع متعددة من التكنولوجيا، على سبيل المثال – وليس الحصر – الذكاء الاصطناعي.

الرقمنة هي تحويل المعلومات والوثائق إلى صورة تستطيع أجهزة الحاسب الآلي التعامل معها. هذا ليس بالشيء السهل؛ لأنك تريد تحويل الصور والوثائق إلى شكل يستطيع الحاسب الآلي التعامل معه وتخزينه وتحليله واسترجاعه في أي وقت. إن ذلك ليس سهلًا ولكنه يعد من أعقد مشاكل التقنية، وأصبح بالإمكان التعامل معها وحلها. وفي هذا البحث سوف نستعرض التحديات التي تواجه الرقمنة وخصوصًا الوثائق العربية.

أولًا: كيفية تحويل الوثائق من صورتها الورقية إلى صورتها الإلكترونية (الرقمية) عن طريق الماسح الضوئي، هذه الخطوة تتم باستخدام الماسحات الضوئية وكلما تغيرت خصائص مخرجات الورق تغيرت اختياراتنا في نوع الماسح الضوئي. كذلك حسب نوع الوثائق وشكلها وجودتها يتم تحديد الوجهة في نوع الماسح الضوئي ذاته.

ثانيًا: عندما تتحول أغلب الوثائق إلى النظام الرقمي فسنحتاج إلى وسائط تخزينية عملاقة السعة يكفي لكل ما نحتاج تخزينه. هذا التحدي ليس صعبًا، ولكن يجب التأكد من توافر السعة التخزينية؛ لأن أخذ النسخ الاحتياطية يتم بشكل دوري. أما بالنسبة للمعلومات المهمة فيجب أن نأخذ عدة نسخ احتياطية منها وتُخزن على أجهزة مختلفة. إن التحدي هنا يأتي بتحديد عدد النسخ

الاحتياطية حسب أهمية المعلومات والفترة الزمنية بين كل نسخة والتالية لها.

ثالثًا: يجب حماية المعلومات والحفاظ على سريتها من الاختراق أو السرقة . فالحفاظ على سرية المعلومات هو التحدي ذاته، وقد يتم ذلك عن طريق برامج أو عن طريق تخزين المعلومات بعيدًا عن الإنترنت.

رابعًا: حتى يتم الاستفادة من المعلومات المخزنة لابد من تحليلها واستخدامها على نحو صائب؛ ولذلك يعتبر تحليل كميات كبيرة من المعلومات في وقت قصير يحتاج أجهزة حاسب آلي فائقة السرعة وبرمجيات ذكاء اصطناعي متقدمة مثل التعرف الضوئي للحروف، وسوف نتحدث عنها بشكل مفصل في البحث.

وأخيرًا: بمجرد الانتهاء من الرقمنة فإنه يتعين علينا الحفاظ على قدرتنا على الاسترجاع أو البحث في كل المعلومات والاستفادة منها وتحليلها للمساعدة في اتخاذ القرارات المناسبة.

الرقمنة تعتبر من أهم الخطوات ولابد منها، لكن يجب أن نضع تلك التحديات والعمل على حلولها والبحث عن أفضل حلول الرقمنة بشكل مستمر.

يأتي هذا البحث لعرض مراحل رقمنة الوثائق العربية عبر مختلف السبل المستخدمة، وكذلك عرض المشكلات التي تواجه الرقمنة وخصوصًا الوثائق العربية وكيفية العمل على مشروع للرقمنة.

مجهودات رقمنة الوثائق وتقنياتها

تتصنف رقمنة الوثائق تبعًا لعدة محاور، ومن أهم تلك المحاور المجهودات المبذولة للوصول إلى النتائج المطلوبة من الرقمنة. وبناءً على ذلك سوف نستعرض تلك التصنيفات على النحو التالي:

١- المجهود البشري: يُعتبر من أقدم الوسائل للوصول إلى رقمنة الوثائق، والذي يعتمد كليًّا على العامل البشري في كتابة النصوص بشكل يدوي، ولكن مع التطور الحديث الذي نعاصره فإن وجوده بقي على أقل الحدود لما يتمتع به من مميزات وعيوب.

وتأتي مميزاته كما يلي:

أ. دقة النصوص المكتوبة وكفاءتها.

ولكن تعتبر عيوبه هي السبب الأساسي لقلة استخدامه حاليًّا، وهي كما يلي:

أ. البطء الشديد في عملية الكتابة والرقمنة بشكل عام.

ب. عدم القدرة على الحفاظ على الأشكال أو الرسومات أو أصل المخطوطات التي لها دلالة واضحة في سياق النصوص.

٢- المجهود التقني: يُعتبر من أهم المجهودات نظرًا لما يقدمه من حلول مختلفة لرقمنة الوثائق التي تعتمد بشكل كبير على سرعة الحاسوب وأجهزة الماسح الضوئي الحديثة وغيرها. وتُعتبر أهميته مبنية على مدى قبول نتائجه أو اعتماديتها في الرقمنة. وبالرغم من أهميته لكنه أيضًا يتميز ببعض المميزات والعيوب التي تجعله أحد الحلول المطروحة في الرقمنة.

وتتمثل مميزات المجهود التقني في التالي:

أ. سرعة إنجاز مهمة الرقمنة.

ب. القدرة على الحفاظ على الأشكال أو الرسومات أو أصل المخطوطات التي لها دلالة واضحة في سياق النصوص.

أما عيوبه وهي التي تقرر من مدى قبوله كحل للرقمنة تتمثل في التالي:

أ. تغير الدقة بتغير جودة الوثائق، وجودة تصوير الماسح الضوئي.

٣- المجهود البشري مع التقني: يعتبر هذا المجهود مزيجًا من المجهودين السابقين للقدرة على الدمج بين المميزات وتقليل العيوب على قدر المستطاع. كما يُعتبر التدخل البشري للتصحيح وزيادة جودة الناتج النهائي لرقمنة الوثائق. ولذلك تتمثل مميزاته في التالي:

أ. إنجاز مهمة الرقمنة بشكل أسرع من المجهود البشري وأقل سرعة من المجهود التقني منفردًا.

ب. القدرة على الحفاظ على الأشكال أو الرسومات أو أصل المخطوطات التي لها دلالة واضحة في سياق النصوص.

ج. دقة النصوص المكتوبة وكفاءتها.

ولكن تظهر عيوبه في:

أ. عملية الرقمنة أصبحت أكثر تعقيدًا لما تحتويه من اختلافات في طرق المزج بين العامل البشري والحاسوب،

ولهذا سوف نستعرض الخطوات الأساسية لمشاريع رقمنة الوثائق عن طريق

المجهود البشري مع التقني، وذلك لتوضيح عمليات الرقمنة وما تحتويه من عمليات معالجة معقدة متتابعة.

استعراض الخطوات الأساسية لمشاريع رقمنة الوثائق

لعمل مشروع رقمنة وثائق لابد من الاهتمام ببعض النقاط الأساسية وما قد تحتويه من نقاط فرعية تؤثر على المشروع بأكمله؛ ولذلك سوف أستعرض النقاط بشكل شامل، فمن الممكن الاستغناء عن بعضها إن لم تكن بالمشروع ذاته.

١- إعداد النظام:

وحدة إعداد النظام هي المكان الذي يتم فيه تكوين جميع معلمات النظام. يتم تكوين جميع معلمات سير العمل ومعلمات إدخال البيانات المختلفة باستخدام وحدة إعداد النظام. ويتم أيضًا تكوين المعلمات القياسية، مثل مخطط قاعدة البيانات، والوجهة التي سيتم استخدامها لتحميل الملفات في وحدة التحميل، وغيرها في وحدة إعداد النظام. وقد تدخل هذه البيانات بضبط سابق لمحركات المعالجة اللغوية التي سوف نستعرضها لاحقًا.

٢- إعداد الوثائق:

وحدة إعداد الوثائق هي المكان الذي يتم فيه تسجيل أنواع الوثائق في النظام والهيكل القياسي لكل نوع من أنواع الوثائق. هذه الوحدة مسؤولة عن إنشاء بنية المجلد، حيث تُوجد الصور بعد مسحها ضوئيًا. إضافةً إلى ذلك باستخدام وحدة إعداد الوثائق يقوم المستخدمون بتحديد الفهارس المطلوبة لكل نوع مستند يتم استخدامه لإنشاء قاعدة البيانات التي سيتم ملؤها أثناء مرحلة الفهرسة. يُعد إعداد الوثائق أيضًا المدخل الرئيسَ لوحدة الفهرسة لتطبيق عمليات التحقق المطلوبة في حقل الفهرسة.

٣- المسح الضوئي:

وحدة المسح الضوئي هي الوحدة التي تتعامل مع الماسح الضوئي حيث يتم إدراج الوثائق وتحويلها من شكلها الورقي إلى شكلها الرقمي كصور ونقلها إلى بنية المجلد المطلوب، وإجراء عمليات فحص الجودة والتحقق من عدد الصور الممسوحة ضوئيًا مقارنة بعدد صفحات كل مستند تم تسجيله في وحدة إعداد الوثائق. وباختلاف نوع الماسح الضوئي تختلف جودة الصور المنتجة.

٤- وحدة التعرف الضوئي للحروف:

تعتبر هذه الوحده أهم وحدات الرقمنة وتسمى في بعض الأوقات وحدة معالجة الصور؛ حيث يتم بها معالجات معقدة جدًّا بشكل متتابع للوصول لأفضل نتائج في كتابة النصوص، وكذلك الحفاظ على هوية الوثيقة. ولذلك فإن تطور هذه التقنية بالاضافة إلى جودة المسح الضوئي يمكن أن تصل تلك المرحلة إلى نتائج مبهرة في زمن قياسي. وعلى غرار تلك النتائج يتم مراقبتها بقسم مراقبة الجودة والعمل على تطوير النتائج وتحسينها عن طريق إعادة تكييف المحركات المعالجة للصور، وكذلك اللغة.

٥- الفهرسة:

وحدة الفهرسة هي المكان الذي يتم فيه إنشاء شاشات إدخال البيانات والتحقق من البيانات بناءً على المؤشرات المطلوبة المسجلة في وحدة إعداد الوثائق. تقوم وحدة الفهرسة أيضًا بإنشاء قواعد التحقق من صحة البيانات في كل حقل بناءً على نفس المصدر. في وحدة الفهرسة، يمكن عرض الصور الممسوحة ضوئيًّا لكل مستند بجوار شاشة إدخال البيانات لاستخراج البيانات من الصور، والتحقق من البيانات المُدخلة بالبيانات الحقيقية على الصور، كما يمكن مراقبة الجودة على جودة مسح الصور. مع بعض التكوينات، يمكن إعداد وحدة الفهرسة للسماح بمفهوم إدخال البيانات المزدوج في حالة الحاجة إلى مستويات جودة أعلى.

٦- مراقبة الجودة:

يتم تصميم وحدة مراقبة الجودة للسماح لأعضاء فريق الجودة والعملاء بمراجعة الوثائق الممسوحة ضوئيًّا والمتعرف عليها ضوئيًّا والمفهرسة؛ حتى يتم الموافقة عليها أو رفضها. في حالة رفض الوثائق يقوم المستخدمون بتسجيل سبب الرفض، وإرسالها مرة أخرى إلى فرق المسح أو المعالجه أو الفهرسة.

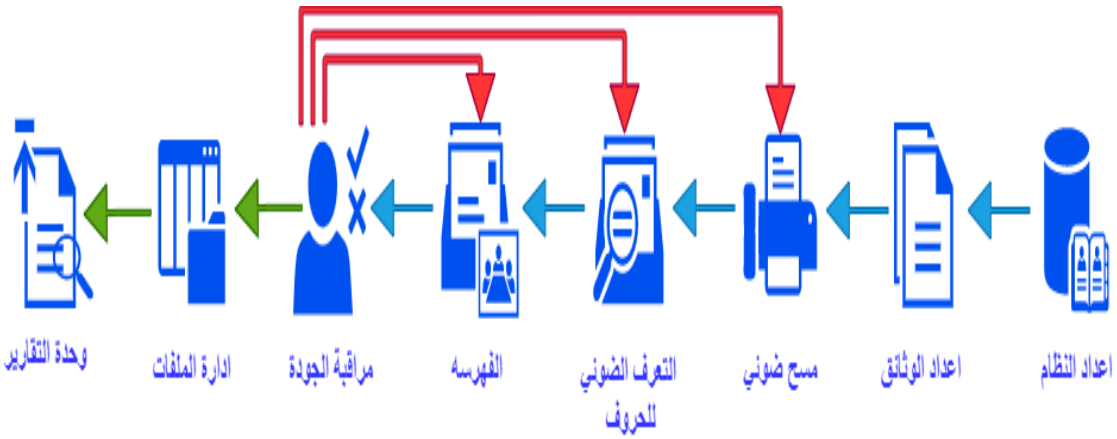
٧- إدارة الملفات:

تقوم هذه الإدارة بمتابعة الملفات وتاريخ إنشائها وتصديق عناصر الجودة. بينما تأتي أهميتها في القدرة على استعادة الوثائق حال البحث عنها باستخدام مفاتيح الفهرسة أو ما تحمله الوثيقة من نصوص. حيث إن حجم المعلومات والبيانات كبير جدًّا فإن هذه المرحلة أيضًا لها الكثير من المعالجات وخاصة بمحركات البحث.

٨- **وحدة التقارير:**

قد تقوم هذه الوحدة بإنشاء تقارير عامة مثل عدد الوثائق وتاريخ إنشائها أو تقارير خاصة مثل عمل استفسارات عن محتويات الوثائق من نصوص وتجميع المتشابهات وغيرها. وتأتي أهمية هذه الوحدة في المساعدة على أخذ القرار عن طريق تحليل البيانات.

وأخيرًا يوضح الشكل رقم (١) الصورة العامة والخطوات الأساسية لمشروع رقمنة الوثائق.



الشكل (١) – الصورة العامة والخطوات الأساسية لمشروع رقمنة الوثائق

وحتى يتسنى لنا عمل مشروع رقمنة الوثائق بأحدث التقنيات لابد من دراسة كل من أعمال المسح الضوئي إلى جانب تكنولوجيا التعرف الضوئي للحروف والتقنيات الأخرى التي تسهم في إثراء وتنوع مخرجات الرقمنة.

أجهزة الماسح الضوئي

سوف نستعرض الماسح الضوئي وأنواعه المتاحة وفقًا لمختلف أنواع الوثائق. حيث إنه باختلاف حجم الوثائق ونوعها وشكلها يتحدد نوع الجهاز المستخدم للمسح الضوئي.

١-ماسحات ضوئية مسطحة:

هذه هي الأنواع النموذجية التي كانت موجودة منذ الأيام الأولى من اختراع الماسحات الضوئية. لديهم سرير زجاجي حيث يتم الاحتفاظ بالمواد، في هذه الحالة، ورقة أو صورة فوتوغرافية، يجب مسحها ضوئيًا. تتعرض هذه الأجزاء للضوء البصري الذي يسقط على المادة ويلتقطها نقطة تلو الأخرى ويتم تخزينها على قرص التخزين لتكون نسخة رقمية. يجب على المستخدم وضع الوثيقة داخل الماسحة الضوئية

المسطحة كما يجب إغلاق الغطاء قبل المسح الضوئي بشكل يدوي، ولذلك هذه الطريقة تعتبر من أبطأ طرق المسح الضوئي؛ حيث إن إنتاجية الماسح الضوئي المسطح قد تصل إلى خمس صفحات في الدقيقة.

وتستخدم الماسحات الضوئية المسطحة بشكل أساسي لمسح أي شيء مسطح. المواد التي يتم مسحها ضوئيًّا باستخدام هذا الماسح هي (الصور – والوثائق – والكتب الرقيقة)

٢- ماسح ضوئي مزود بوحدة تغذية تلقائية للمستندات

يتم فيه إدخال المواد، ويسمى أيضًا وحدة التغذية، ويتم امتصاص هذه المواد بواسطة الماسح الضوئي ومسحها ضوئيًّا باستخدام مفهوم مشابه لمفهوم الضوء في الماسح الضوئي المسطح. الميزة الرئيسية لهذا الماسح الضوئي المسطح هي السرعة التي يتم بها مسح المواد ضوئيًّا. حيث إن سرعته قد تصل إلى مئة صفحة في الدقيقة؛ وذلك يساعد ويسهل عمليات المسح الضوئي للكميات الكبيرة من الوثائق.

٣-ماسح ضوئي علوي

تستخدم الماسحات الضوئية العلوية واحدة أو اثنتين من الكاميرات الرقمية، حوالي ١٠ إلى ١٢ بوصة أسفل الكاميرا (الكاميرات) ، سيتم فحص هذه المواد بواسطة الكاميرات الرقمية العلوية. في كل مرة يتسبب فيها شيء ما في تشغيل الكاميرات، يتم فحص المواد الموجودة تحتها وتخزينها في محرك الأقراص الثابتة بجهاز الكمبيوتر. لا تقتصر المواد التي يتم مسحها ضوئيًّا على الأشكال المسطحة، بل يمكن أن تكون بأي شكل. هذا يجعل من

السهل علينا مسح كل من (الكتب – والمخطوطات – والمنحوتات الحجرية – والوثائق)

٤-ماسحات محمولة

هو نوع قضيب صغير من الماسح الضوئي يمكن حمله باليد، ومن هنا جاء اسم الماسح الضوئي المحمول.

للتشغيل تحتاج فقط إلى التمرير فوق المواد التي تحتاج إلى مسح ضوئي، يتم التقاطها أثناء الضغط على الزر ويتم تخزينها في بطاقة الذاكرة التي يتم إدخالها في مكان مخصص لها.

بمجرد المسح الضوئي، يجب نسخ المحتوى الرقمي المخزن في بطاقة الذاكرة لإكمال عملية المسح. يمكن لهذا مرة أخرى مسح الأسطح المسطحة فقط مثل الماسح الضوئي المسطح. يتم استخدامه في الغالب لمسح بعض المستندات المتراكمة من اليوم السابق بسرعة.

٥-الماسحات الضوئية للكتب

هذا الماسح هو ماسح ضوئي علوي. على الرغم من حقيقة أنه ماسح ضوئي علوي، إلا أنه من الجدير بالذكر أن هذا الماسح الضوئي هو ماسح ضوئي منفصل نظرًا لطريقته الفريدة في العمل. هذا مخصص للكتب. مهد على شكل حرف V يتم حفظ الكتب عليه دون قطع العمود الفقري، وستقوم كاميرتان علويتان بالتقاط كل جانب من جوانب الصفحة واحدًا تلو الآخر. سيتم تخزين كل صفحة بترتيب رقم الصفحة المذكور في الكتاب. تتم معالجة هذه الصور المخزنة لاحقًا باستخدام أدوات التنظيف المجمعة وبعد ذلك يتم إنشاء ملف PDF من هذا. هذه طريقة مفيدة جدًا لمسح الكتب ضوئيًا دون قطع العمود الفقري للكتاب، ومن ثم يُطلق

على هذا الماسح الضوئي أيضًا الماسح الضوئي غير المدمر.

٦-ماسحات ضوئية كبيرة الحجم

هي الماسحات الضوئية المستخدمة لمسح الأوراق الكبيرة ضوئيًّا. يستفيد المقاولون المدنيون ومنشئو المباني والمهندسون المعماريون وقوائم الخرائط وما شابه من هذه الخدمات. سوف يتعاملون مع الخرائط أو التصميمات الكبيرة

تُعتبر مهمة المسح الضوئي هي تحويل الوثائق من الصورة الورقية إلى الصورة الرقمية، ولكن أيضًا تعتبر جودة الصور المنتجة من أهم عوامل نجاح عملية الرقمنة بالكامل. ولذلك يُوصَى بأن تكون دقة الصور ٣٠٠ نقطة في البوصة أو أعلى. وبمراجعة معظم المواسح الضوئية المذكورة سابقًا يتبين أنها جميعًا تدعم هذه الخاصية، بل من الممكن التصوير والحصول على دقة ألف ومئتي نقطة في البوصة. لكن يظل ثلاثمئة نقطة للبوصة الأنسب حتى لا يتضخم حجم الملفات وتوفي بالحد الأدنى لاحتياجات تقنيات الرقمنة.

أنواع الوثائق من حيث الشكل

هنا لا نتعرض للتمييز بين أنواع الوثائق على أساس تناولها للأفكار أو طرق إصدارها أو مراجعاتها، ولكن سنتناول التمييز بينها من حيث الشكل. فهناك تعدد في أشكال الوثائق وتنوع باختلاف حجم الصفحات وطرق إخراجها، وكذلك طرق حفظها. نرصد فيما يلي بعضًا من صور هذه الوثائق وأنواعها

عن بطلات العالم الحقيقيات

عن اللواتي لا يكتبُ عنهنّ أحد..
عن المستيقظات فجراً، المُصلّيات فرضاً، التاليات ذكراً، المُعدّات فُطوراً، المُجهّزات تلميذاً، الملبسات مريولاً، المسرّحات شعراً، المُراجعات برنامجاً، الحاليات صحوناً، المُوضّبات فراشاً، الحانيات ظهوراً، الكانسات أرضاً، الماسحات غباراً، المُرتّبات بيتاً، الطابخات غداءً، المُنتظرات أولاداً، المُطعمات حشداً، المُنظّفات قُدوراً، المُذاكرات دروساً، الحالاّت قروضاً، المصححات إملاءً، المُسمّعات استظهاراً، المُحفّظات قرآناً، المُعدّات للنوم أولاداً، المُفرقشات بعد ذلك أزواجاً!

١ – الكتب :

تعتبر الكتب هي أشهر أنواع الوثائق بسبب انتشارها مثل الكتب الدراسية والقصص الروائية. فإنها تتميز بتتابع صفحاتها بنسق قد يثبت بين الكتاب على عمود واحد بعرض الصفحات، ولكن قد يوجد أيضًا نسق به عمودان أو أكثر وهذه الكتب نادر وجودها.

١. المجلات :

تُعتبر المجلات أكثر تعقيدًا من الكتب باختلاف نسقها العام، وكذلك طرق الكتابة بها لا تشترط الالتزام بشكل معين يختلف من مجلة لأخرى وكذلك عدد الأعمدة وتداخلاتها.

٢. الصحف :

تُعتبر الصحف من أشهر الوثائق حيث كثرة إصداراتها التي قد تكون بشكل يومي. ولكنها تتميز بكثرة أعمدتها وصغر حجم الخط.

٣. القواميس والمعاجم

تتَّحِد معظم القواميس والمعاجم في طرق نسقها من حيث الأعمدة ولكن قد تتداخل أكثر من لغة وليست بالضرورة للغة واحدة.

٤. المخطوطات

تعتبر من أقدم الوثائق والتي تم تصويرها بأقدم وسائل التصوير التي تُدعى ميكروفيلم. وتتميز بقلة جودة الصور وصعوبة قراءتها.

٥. الملفات الشخصية

تتميز هذه الوثائق بثبات التنسيق العام بها، بثبات نوعها. قد تضم خلالها بطاقات تحقيق الشخصية، وجواز السفر، ورخص القيادة، وما شابه ذلك.

٦. الفواتير

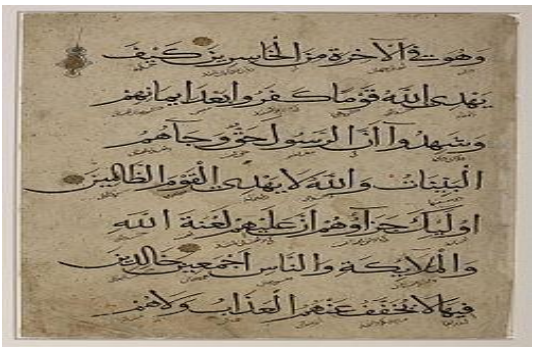
تضم تنوعًا كبيرًا في هذه الفئة من الوثائق باختلاف الجهة المصدرة لهذه الفواتير. فإنها تختلف في نسقها واللغة المكتوبة وحجم الخطوط وشكلها أيضًا.

أنواع الوثائق من حيث الكتابة

سنستعرض أنواع الوثائق باختلاف أنواع الكتابة فهناك تعدد في أنواع الكتابة بمرور العصور، حيث إنها بدأت من قديم الزمان حتى وقتنا الحاضر. نرصد فيما يلي بعضًا من صور كتابة الوثائق وأنواعها:

١. الكتابة اليدوية:

تعتبر هذه الكتابة هي أقدم الطرق لحفظ المعلومات وحتى عصرنا هذا، حيث يتم استخدام الخط اليدوي. وتتميز بأنها تختلف من شخص لآخر من حيث جودة الكتابة.



٢. الكتابة بالآلات الكاتبة

ظهرت هذه الكتابة في عام ١٧١٤ م وتطورت إلى أن أدخل عليها اللغة العربية عام ١٩١٤ م على يد المصريين فليب واكد وسليم شلبي. والتي تعتمد على الضغط على مجموعة من المفاتيح، لتطبع الكلمات والجمل والحروف بالحبر، وكان يتم استخدامها في الوثائق الرسمية، والمؤسسات الحكومية، ومكاتب الطباعة المختلفة. تتميز هذه الكتابة بثبات حجم الخط المستخدم وكذلك شكله.

٣. الكتابة بالكمبيوتر

قامت التكنولوجيا بتطوير شكل الكتابة ليتناسب مع طبيعتها التي تعتمد على ضغطة الزر في الكتابة على صفحة إلكترونية غير ملموسة، وهذا ما حدث في البداية على برامج الكتابة بجهاز الحاسب الآلي. ثم تطور الأمر ليشمل أجهزة المحمول، وأجهزة الحاسب الآلي المحمولة، والألواح الإلكترونية. لتصبح الكتابة الإلكترونية هي سمة الكتابة الرئيسية للعصر الحديث. وبما أنها على صفحة إلكترونية غير ملموسة وباستخدام البرامج الحاسوبية فإنها تتميز باختلاف حجم الكتابة وشكلها ولونها.

التقنيات المختلفة الخادمة لرقمنة الوثائق

حينما تتحول الوثائق من صورتها الورقية إلى صورة رقمية على شكل صور أو ملفات PDF فمن هنا تبدأ رحلة الرقمنة بمختلف أشكالها وأنواعها. وحتى يتسنى لنا الحصول على أفضل النتائج فلابد من الدخول بعدة مراحل وهي كالتالي:

١. معالجات مسبقة للصور

تبعًا لما قد يحدث من تشوهات في تصوير الوثائق من عوامل تاريخية أو ظروف تصوير أو غيرها فقد يصعب العمل عليها بهذا الشكل. لذا يتم العمل على تحسين جودة الصور عن طريق بعض المعالجات مثل:

أ. معالجة انحراف الصور أو دورانها .

ب.معالجة السطوع

ج.معالجة الضوضاء

٢. اكتشاف هيكل الوثيقة

يُعتبر اكتشاف هيكل الوثيقة هو الخطوة الأكثر أهميةً في عملية تحويل الصور النصية إلى ملفات نصية قابلة للبحث والتعديل.

يعمل مكتشف هيكل الوثيقة بمحرك مبني على أسس الذكاء الاصطناعي الذي تم تدريبه بمختلف أنواع الوثائق. يهدف مكتشف هيكل الوثيقة للعمل على ما يلي:

أ. تقسيم الصفحات وتمييز مناطق الفقرات النصية والصور الفوتوغرافية والجداول والأشكال الاسترشادية.

ب.تقسيم الفقرات إلى سطور.

ج. تقسيم السطور إلى كلمات.

د. تقسيم الكلمات إلى حروف.

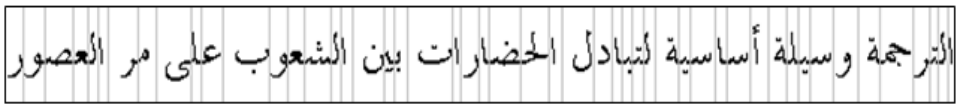
٣. التعرف الضوئي للحروف

محركات التعرف الضوئي للحروف قديمًا كانت مبنيةً على نماذج إحصائية وقواعد ثابتة، لكنها لم تدم طويلًا لضعف نتائجها بالإضافة إلى البطء الشديد في عمليات المعالجة. بينما تم بناء المحرك حديثًا على الذكاء الاصطناعي لما يتميز به من دقة وسرعة.

كما أن هناك الكثير من التحديات التي تواجه التعرف الضوئي وخصوصًا حروف اللغة العربية التي منها على سبيل المثال:

أ- اتصال الحروف:

إن الحروف العربية تترابط بعضها مع بعض لتكوِّن الكلمات وليست بشكل منفصل مثل اللغات الأخرى، وهذا ما يزيد من الأشكال التي يجب التعرف عليها من قبل الآلة وليست الحروف فقط.

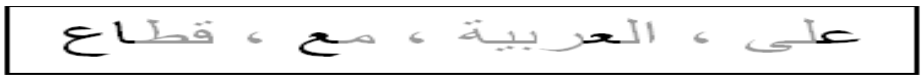


ب- التداخل بين حدود الجرافيمات:

تتميز اللغة العربية عن غيرها في الكتابة باختلاف الخطاطة والتي تعتمد معظمها على تداخل الحروف بعضها مع بعض، كما يُستمد منها الجمال والحلية في الخطوط؛ وذلك يزيد من تعقيد عملية التعرف الآلي عن التعرف من قبل الإنسان البشري. ومنها على سبيل المثال وليس الحصر الشكل التالي:

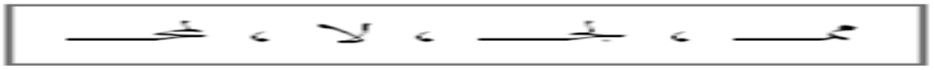
ج- تغير رسم الحرف مع تغير موضعه في الكلمة:

إن اللغة العربية تختلف عن غيرها من اللغات؛ حيث إن الحروف تختلف أشكالها باختلاف مواضعها في الكلمة سواء أكان الحرف في أول الكلمة، أم وسطها، أم نهايتها، أم منفردًا بالكلمة. فعلى سبيل المثال بالشكل التالي:

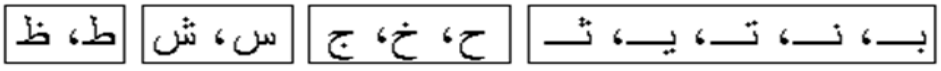


د- الجرافيمات المركبة من أكثر من حرف واحد

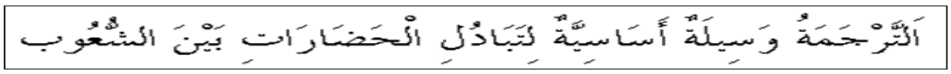
هناك بعض الحروف إذا تتابعت يتغير شكل كتابتها، وهذا ما يعرف بالجرافيمات المركبة كما هو مبين بالشكل التالي:



هـ- النَّقْط



و- التشكيل



ز- الحروف المطولة (الكاشيدا)

محمـــــــد رســـــول الله

هذه التحديات التي ذكرناها آنفًا تؤثر بشكل واضح على طريقة عمل محركات التعرف الضوئي وأيضًا علي النتائج المطلوبة بمدى صعوبتها.

هناك بعض محركات التعرف الضوئي التي تعمل على مستوى الحروف، وأخرى تعمل على مستوى الكلمات، وهناك نوع آخر يعمل على مستوى السطر كاملًا، ويُعتبر أدق المحركات لكنه أبطؤها.

ولذلك ينصح بعمل محرك التعرف الضوئي للحروف باستخدام الذكاء الاصطناعي على مستوى السطور.

ولعمل محرك مناسب للتعرف الضوئي للحروف فإنه يمر بعدة مراحل كما يلي:

أ. تجهيز بيانات:

تعتبر هذه المرحلة أهم المراحل؛ حيث إنها تؤثر بشكل قوي على المحرك بالكامل. وتعتبر البيانات في هذه الحالة عبارة عن صور بها سطور مكتوبة باللغة العربية يقابلها ملفات نصية أخرى مكتوبة بعناية لكل سطر على حدة. ويفضل في هذه البيانات الجمع بين اختلافات الخطوط التي قد يتعرض لها المحرك لاحقًا. ولابد من الوصول إلى كميات مناسبة من السطور قد تصل إلى ملايين السطور حتى تزداد قوة المحرك مقارنة باختلاف البيانات التي قد يتعرض لها. ونظرًا لأهمية هذه المرحلة فلابد من تدخل مراقبي الجودة عدة مرات للتأكد من سلامتها وجودتها على الشكل الأمثل.

ب. تقسيم البيانات:

يتم تقسيم البيانات إلى قسمين. الأول يكون القسم الأكبر وهو البيانات الخاصة بتدريب المحرك. والثاني هو البيانات الخاصة بتقييم واختبار المحرك وقياس كفاءته.

ج. زيادة البيانات:

تعمل هذه المرحلة على زيادة البيانات عن طريق عمل بعض المعالجات الحاسوبية على صور السطور الأصلية وتكرارها بشكل مختلف مع الحفاظ على النصوص نفسها دون القطع أو حذف كلمات منها. وتتم هذه المرحلة فقط على بيانات التدريب دون بيانات التقييم والاختبار.

د. تدريب محرك التعرف الضوئي:

يتم استخدام خوارزميات معقدة باستخدام بيانات التدريب فقط، الغرض منها استخراج خصائص الحروف وأشكالها والتدرب عليها وعمل إحصائيات واحتمالات بشكل متكرر للوصول إلى أفضل محرك قادر على التمييز بين الحروف والكلمات.

هـ. اختبار محرك التعرف الضوئي:

تتم هذه المرحلة باختبار المحرك لبيانات التقييم والاختبار حتى يتم استنباط ما هو مكتوب بالسطور.

و. اكتشاف آلي لأخطاء النصوص اللغوية:

يعتبر الاكتشاف الآلي للأخطاء هو المرحلة الأخيرة في محرك التعرف الضوئي، وتأتي أهميته في تدارك الأخطاء الناتجة من محرك التعرف الضوئي عن طريق استخدام بيانات وإحصاءات لغوية مبنية عن طريق تدريب مصحح لغوي بملفات نصية باللغة العربية بكميات هائلة تساعده في فهم النصوص وتتابع الكلمات. وبعد هذه المرحلة يتم اختبار المحرك الضوئي للحروف بالإضافة إلى مصحح اللغة معًا باستخدام بيانات التقييم وحساب الدقة بمراجعة الملفات النصية الأصلية، وحين الوصول إلى الدقة المطلوبة يتم إصدار محرك التعرف الضوئي بشكل نهائي، أما إذا لم تتحقق النتائج فيتم العودة لتدريب المحركات الخاصة بالتعرف الضوئي والمصحح اللغوي مرات ومرات.

٤. إصدار الملفات النصية:

في حالة استكمال عمل محرك التعرف الضوئي للحروف وتجهيز السطور كلّ على حدة تأتي هذه المرحلة، ويتم تجميع السطور والحفاظ على أماكنها من الملف الأصلي وإصدار نوع الملف المطلوب.

٥. المشكل الآلي:

التشكيل في اللغة العربية له أهمية لا تقل عن أصل الكلمة، فإذا تغير تشكيل حرف

واحد يتغير المعنى، ويتغير نطق الكلمات. وحتى يتم تشكيل الكلمات بأعداد كبيرة فذلك يتطلب مجهودًا لا يستهان به من قبل لغويين متخصصين باللغة، بالإضافة إلى أن الوقت المفقود في ذلك يكون كبيرًا جدًّا؛ لذا تبدأ فكرة المشكِّل الآلي الذي يعمل على فهم قواعد اللغة لتطوير محرك قائم على الذكاء الاصطناعي.

٦. تحويل النصوص إلى ملفات صوتية:

ومع الوصول إلى دقة عالية في كتابة النصوص العربية بالتشكيل يتسنى لنا أن نتحول من الصورة الكتابية للصورة السمعية حتى يتم الاستفادة من البيانات بالقدر المستطاع وزيادة عدد القراء والمستمعين وتوصيل المعلومات بأسهل الطرق التي يرغب فيها الإنسان. وقد تأتي أهمية نطق النصوص من أهمية نشر المعرفة مع ذوي الإعاقة البصرية لعدم القدرة على تحويل الكتب إلى طريقة "برايل".

إن الذكاء الاصطناعي أيضًا يطرق هذا الباب لعمل محرك لتحويل النصوص إلى ملفات صوتية ولكن باللغة العربية، إن التشكيل يلعب دورًا حساسًا في نطق الكلمات وبوجود مشكل آلي للغة العربية قوي يأتي محول النصوص إلى ملفات صوتية بشكل قوي أيضًا.

٧. محركات البحث في النصوص:

بتحول الصور إلى نصوص قابلة للتعديل ومع تزايد أعداد الملفات فإنه يلزم وجود محركات بحث للقدرة على استعادة الملفات المطلوبة. قد ينقسم البحث في النصوص إلى قسمين:

أ. البحث في الملفات عن طريق فهرس الملفات:

وهذا يُقصد به استعادة الملفات عن طريق البحث باسم الملف أو الكود الخاص به داخل فهرس المكتبة.

ب. البحث في الملفات عن طريق المحتوى:

وهنا يتم البحث عن طريق المحتوى والنصوص داخل الملفات. أو بمعنى آخر سيتم استرجاع كل الملفات التي تحتوي على نفس كلمات البحث وبدرجات متفاوتة على درجة التطابق بين هذه الكلمات المتوافرة داخل نصوص الملفات.

ومن هنا تزداد أهمية محركات البحث؛ حيث توفر الوقت والمجهود في استخراج المعلومات بجميع أشكالها. ومع ذلك تزداد صعوبة محركات البحث بزيادة حجم البيانات التي يبحث بها. ويتم الحكم على محركات البحث بكفاءتها في استرجاع البيانات المطلوبة في أقل وقت ممكن.

ولذلك فهناك طرق عدة خاصة بمحركات البحث في النصوص باستخدام الكلمات عن طريق ما يلي:

١. البحث باستخدام الكلمة ذاتها: هو البحث القائم على التطابق الكامل بين الكلمة المراد البحث بها ووجودها داخل النصوص بنفس هيئتها.

مثال ذلك : "راهب" ستسترجع "راهب" فقط.

٢. البحث باستخدام جذع الكلمة: هو البحث القائم على استرجاع جميع الكلمات التي تتوافق في جذع الكلمة نفسه، مثل الكلمات التي بها سوابق ولواحق.

مثال ذلك: "راهب " ستسترجع "راهب، فالراهب، والراهب، كالراهب، الراهبات، .....".

٣. البحث باستخدام جذر الكلمة: هو البحث القائم على استرجاع جميع الكلمات التي تتوافق في جذع الكلمة نفسه،

مثال ذلك: "راهب" ستسترجع جميع الكلمات ذات الصلة بالجذر "رهب" مثل "الراهبات، الرهبان، راهب، إرهاب، رهبة، رهيب، ....".

٤. البحث باستخدام الفرع (الليما): هو بحث يتبع ثلاث قواعد:

أ. هو الكلمة نفسها إذا كانت أداة.

ب. هو مفرد الكلمة في حالة الأسماء.

ج. هو الفعل الماضي المسند إلى الغائب في حالة الأفعال.

مثال ذلك: "مدرسة" ستسترجع جميع الكلمات ذات الصلة بالليما أو الفرع مثل "بمدرستهم، مدرستها، مدرسة، مدارس، ....".

٥. البحث الدلالي: هو طريقة للبحث عن طريق تصنيف الألفاظ في مجالات دلالية مثل الرفض والقبول، الإنجاز والفشل، الوفاق والخلاف، والسر والعلن فيرد في باب السر والعلن – علي سبيل المثال–

أفعال: يبطن، يكتم، يحفظ، أو يجهر، يفشي، يذيع، يشيع، ....

وأسماء: السر، اللغز، أو الإعلان، الإشاعة،....

مثال ذلك عند البحث عن "راهب " ستسترجع جميع الكلمات ذات الصلة بالمعنى مثل "الراهبات، الرهبان، راهب، عابد، ناسك، متبتل، ..... ".

نماذج مختلفة لرقمنة الوثائق العربية باستخدام محركات التعرف الضوئي

نستعرض خلال هذه الفقرة أهم المفارقات بين أهم محركات البحث الضوئي من حيث طرق التنصيب، وجودة المخرجات ودقتها حتى يتم إنجاز عملية الرقمنة.

ويعتبر أهم هذه المحركات التي تتميز بقدرتها على العمل بكفاءة عالية على وثائق اللغة العربية دون غيرها كالتالي:

١- Google أو " جوجل".

٢- ABBYY أو " آبي ".

٣- " Sotoor" أو "سطور"

ولكن لا يتوافر محرك Google في نسخة يتم تنصيبها في مكان العمل في حين أنه يعمل فقط على الإنترنت، ولكن يتوافر ABBYY و Sotoor في صورتي الإنترنت والتنصيب بمكان العمل.

وقياس الدقة سيكون بحساب معدل الكلمات الخطأ؛ حيث إننا سنقوم برصد الكلمات المستبدلة أو الكلمات المدخلة أو الكلمات المحذوفة نسبة إلى الكلمات بالنص الأصلي.

وفيما يلي نرصد بعض النتائج

الوثائق	عدد الكلمات الأصلية بالنص	معدل نسبة الكلمات الخاطئة بالنسبة للكلمات الأصلية ٪		
		Google	ABBYY	Sotoor
الكتب	٤٣٩	٣٣٪	٥٢٪	٢٨٪

وكلما قلت نسبة الكلمات الخطأ دل ذلك على قوة محرك التعرف الضوئي للحروف.

وفيما يلي بعض نماذج الوثائق المستخدمة للمقارنة بين محركات التعرف الضوئي ونتائجها.

نشأت المذاهب الأدبية القويمة عند العرب والفرنج إما مرحلة لطور مذهب تقدم به مبتدعوه ، وإما رد فعل لمذهب غلا فيه متبعوه. فمذهب عبد الحميد بن يحيى كان الطور الأول للأسلوب العربى الضيق الموجز دعت إليه مقتضيات المجتمع الجديد من تشعب اطراف الدولة وبدو ثمار الحضارة ودنو العربية من الفارسية . ومذهب ابن المقفع الذى ظهر فى فجر الحضارة العربية كان طوره الثانى دعا إليه اتساع الخلافة وتنوع الثقافة وشدة اختلاط العرب بالفرس .

ثم كان طوره الثالث مذهب الجاحظ الذى اقتضاه نقل العلوم الأجنبية وازدهار المدنية العباسية وانتشار المقالات الإسلامية وتعقد الحياة الإجتماعية واقتباس الآراء الفلسفية .

ثم أترف المسلمون وتقلبوا فى أعطاف النعمة وتأنقوا فى مظاهر العيش فظهر طوره الرابع فى مذهب ابن العميد المسجوع المنمق .

وإلى هنا كان التطور فى النثر الفنى تطورا طرديا يسير من الضيق إلى السعة ومن الجزالة إلى الرقة ومن الترسل المتوازن إلى الصنعة المطبوعة. فلما ضعفت الخلافة وصار الأمر إلى غير أهله جرت على الكتابة أعراض الفساد والوهن فكثر الزيف وانتشرت الصنعة وكان من ذلك مذهب القاضى الفاضل وهو الطور الخامس من أطوار الأسلوب العربى غلا فيه أصحابه حتى أفسدوا الفكرة وشوهوا الصورة. ومن هنا كان رد الفعل بظهور مذهب ابن خلدون إذ رغب عن السجع وزهد فى البديع وسار باللفظ وراء المعنى . ثم تطور هذا المذهب بتأثير الحضارة الأوربية ونقل الآداب الأجنبية إلى الأسلوب الذى يكتب به الكتاب الموهوبون اليوم .

. . .

كذلك كانت نشأة المذاهب الأدبية فى أوربا ، ففى فرنسا مثلا كانت اللغة وآدابها خاضعة لسلطان الإغريقية واللاتينية على أثر انبعاث الروائع اليونانية والرومانية فى إيطاليا ، فالألفاظ يكثر فيها العامى والدخيل، والتراكيب يطغى عليها هوميروس أو فرجيل . والنظم والنثر يجريان على المحاكاة الشكلية ، والموضوعات تؤخذ من الأساطير الوثنية . فكان لابد لهذه العبودية الأدبية من محرر يكفكف شرتها وينظم فوضاها ، فظهر المذهب الاتباعى فى أوائل القرن السابع عشر وهو يقوم على تقليد اليونان والرومان تقليد التلميذ لمعلمه لا تقليد المصور لمثاله . ثم طهروا اللغة من الدخيل والعامى، ووسعوا دائرتها بالوضع والاشتقاق ، وجعلوا للعقل السلطان المطلق على الأدب فلا يكادون يلقون بالهم إلى الخيال والعاطفة . ثم قيدوا الفنون الشعرية والقصصية بقيود من القواعد الصلبة وأخذوا أنفسهم بها حتى انفرجت الحال بينهم وبين العامة ، وانقطع السبب بينهم وبين الطبيعة . وكان رد الفعل الطبيعى لذلك ظهور المذهب الابتداعى الذى عزف عن المحاكاة الأجنبية ، واستمد موضوعاته من أسرار المسيحية وعهود الفروسية ، وقدم الخيال على العقل والشعور على المنطق والفردية على الجمعية والذاتية على الموضوعية . ثم أطلق الفن من القيود التى كبله بها الاتباعيون فأمن الشعراء

الصفحة الأصلية من كتاب "شرح شواهد الإيضاح".

المذاهب الأدبية المنحرفة

هذا المذهب وتأثر الحضارة الأوربية و نقل الآداب الأجنبية إلى الأسلوب الذي يكتب به الكتاب الموجودون اليوم .

و

تلك المذاهب الأدبية الغربية عند العرب والفرق إنما مرحلة تطور مذهب القديم به عن جاءوه ، وما ورد قبل المذهب خلافه

تبعوه. فمذهب عبد الحميد بن يحي كان الطور الأول للأسلوب العربي الضيق الموجز دعت إليه مقتضيات المجتمع الحديد من شعب اطراف الدولة و بدو ثمار الحضارة ودنو العربية من الفارسية . ومذهب ابن المقفع الذي ظهر في بفر الحضارة العربية كان طوره الثاني دعا إليه اتساع الخلافة وتنوع الثقافة وشدة اختلاط العرب با لفرس .

ثم كان طوره الثالث مذهب الجاحظ الذي اقتضاه قبل العلوم الأجنبية وازدهار الحياة العلمية وانتشار المقالات الإسلامية وتعدد الحياة الإجتماعية والقياس الآراء الفلسفية .

ثم أترف المحدثون وتأنقوا في أعطاف الفصة ولم تقفوا في مظاهر الجمال فظهر طوره الرابع في مذهب ابن العميد المسجوع المدبج .

ورأي هذا كان التطور في النثر الفني تطورا طبيعيا يسير من الدقيق إلى الفصة ومن الجزالة إلى الرقة ومن التربيل المتوازن إلى الصنعة المطبوعة . فلما مضت الخلافة وصار الأمر إلى غير أهله جرت على الكتابة أعراض الفساد والوهن فكثر التزيف وانتشرت الصنعة وكان من ذلك مذهب القاضي الفاضل وهو الطور الخامس من أطوار الأسلوب العربي فلا فيه لمذهب حتى أصدروا الفكرة وتوهموا الصورة. ومن هنا كان رد الفعل بظهور مذهب ابن خلدون إذ رغب عن السجع وزهد في الجمع وصار بالفقرة وراء المعنى . ثم تطور

كذلك كانت نشأة المذاهب الأدبية في أوربا ، ففي فرنسا مثلا كانت اللغة وآدابها خاضعة لسلطان الإغريقية واللاتينية على أثر اتباع الرومانية اليونانية والرومانية في إيطاليا ، فالألفاظ يكثر فيها العام والدخيل، والتراكيب يطغى عليها هوميروس أو فرجيل . والنظم والنثر يجريان على المحاكاة الشكلية ، والموضوعات تؤخذ من الأساطير الوثنية . فكان لابد لهذه العبودية الأدبية من محرر يكفكف شرتها وينظم فوضاها ، فظهر المذهب الاتباعي في أوائل القرن السابع عشر وهو يقوم على تقليد اليونان والرومان تقليد التلميذ لمعلمه لاتقليد المصور لمثاله . ثم طهروا اللغة من الدخيل والعامي ، و وسعوا دائرتها بالوضع والاشتقاق ، وجعلوا للعقل السلطان المطلق على الأدب فلا يكادون يلقون بالهم إلى

نتائج محرك Google

٨ المذا ٠بالاديةاكرءة

نتات المناهب الأدبية ا!قولاً عند الرب والنرغ إما /حلى | طور ٠ذهب "قدم به مبتدعوه ، و|٠ا رد فعل لمذهب غلايه .تبعوه ٠ فذهب عبد الجند بن يحي كان الطور الأول لللا صلوب الر نى الضرق المور دعت إ به .قتضيات التمع ا بمديد مزنتهب اطر١ف الدولة و بذو ممار الحنارة وذو المربية من ١١ذارمية . وذهب ا٠ن المقفع الذى ظهرفى بمر الحضارة الر بية كان طوره الثا٠لى دءا إ ليه 1فاع البلاقة وتنوع ا!ثقافة وثدة إخلاط العرب بالفرس.

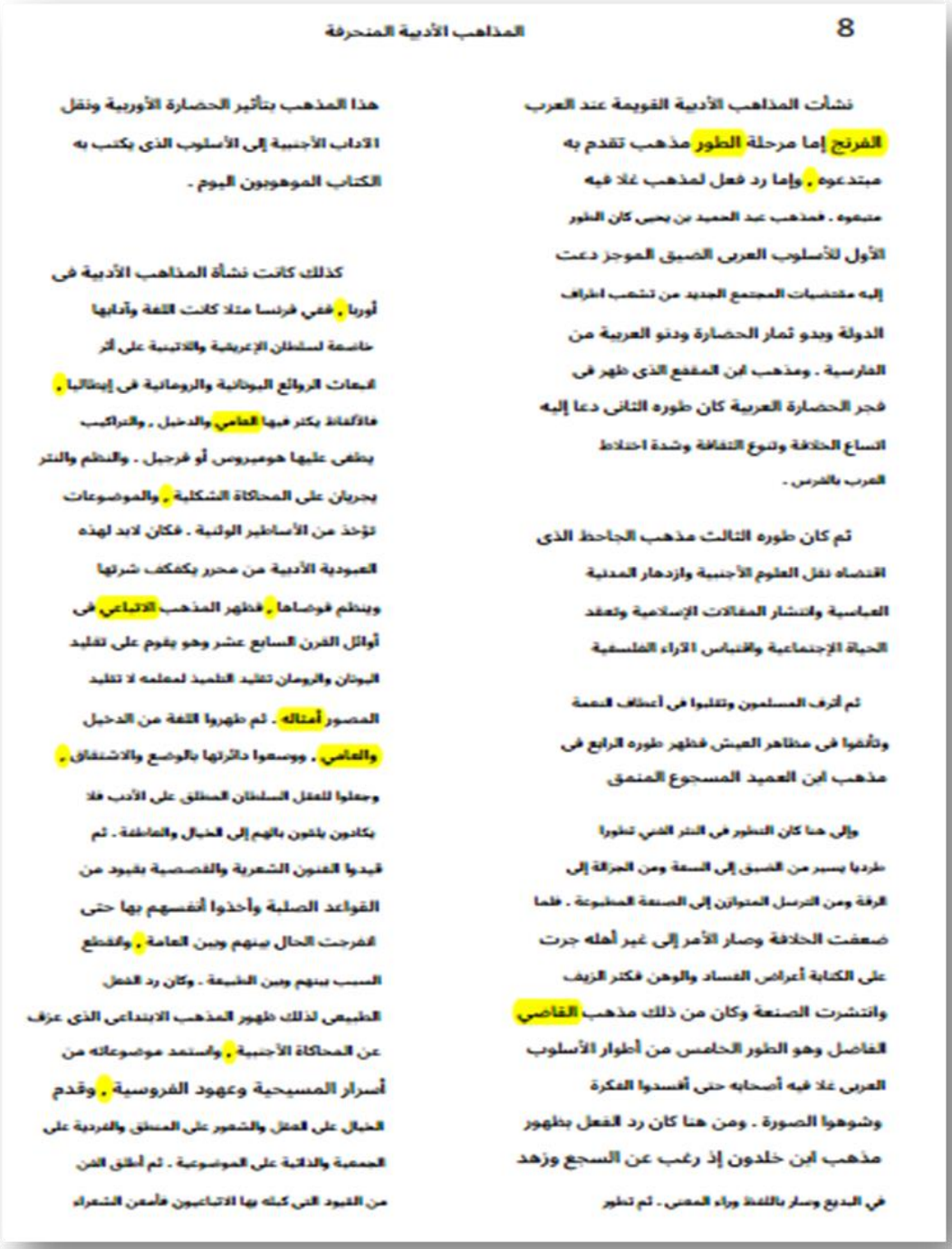
ثم كان طوره ازالت منهب ا٠١الحظ الذى اقتضاه تل العلوم الأجنرمة وازددار الدمية البامية وا٠١اثار المنالات الإسلا٠.ذ وتحقد الحناة |لإخ١.ة وا داس الأراء الفلسفية .

بم أرف المامون وقابرا فى أععذاف العمة وانقوا فى مظاهرال٠بش قطهر طوره الراع ق منهب ا بن المرد ا لممجوع المنق .

وذلى هنا كان الطور ف ايذالغنى طور| طرديا ي٠رساكتق إي الة وس اءلحزالا إلى ار5قو.نالارسلاذوازنالمالصعةالمطءوة.دا زفت ا.لحلاقة وبار؛الأص إلى غرأهله رت عل الكابة أعراض الفك والوز فكرالريف و؛تشرت اب كان مزنلك منهب القاغى افزل و٠والاورالحاى ض أطوارالأرب العر٠ل هلا به أمحا به حتى أفسدوا الغتي وشهوا الصورة. ومزهنا كانزدالغعلدظهور منهب ا ن خلدون اذ ر٠ب ز الم وز هد ف ابدح ودار باللفظ وراء المعتى ٠ ثم طور هذا المنهب ٠تأع٠ار ا لحضا رة الأوربية وقل الأداب الأجندة إلى الارلوب الذى كب به الكاب الموهو؛ون اوو؛ .

كذلك كاتذاة الذاهب الأدبنف أورا ، ض فرف مثلا كانت افة وآدابها ٠ذاضعة للعنان الإغريقية واللاميية هلى أر ابعاث ارواح ابوتاتية وارومابة ف المط١ ٧ ، فالألفاظ ئ؛با!اعاى والدخل، والراكب ٠طئ طيها هوميروس أوفرجيل. والطم والذ ربان على اباكاة الثكية، وا! ونوعا ت ٠وخذسالأماطرالوي ٠ فكازلابد لهذه البردية الأدبة ض محرر مكفكف ثرتما وينغلم فوزاها، فظهر المنب الاراش ف أوائل الترن اباع عأر وهو يقوم عل غايد ابولآن والروان تا٠٠د التنين لعله لاقيد المصرر لتالي . ثم طءروا اسة ٠ن الديحل واباى، ووبوا دارا بالوغح والاتتناق، وجعلوا ٧مقل الساطان المطلق هلى الأدب فلا يكادون ينقون باثمدلى الحيال واشاطغة. ثم ب وا !اذون !!ثعرية والقصصة بقيود ن !!قراءد الصبة وأخذوا أتفهم با حتى انفرجت ادال بنهم وس العامة، واتبع البب بنهم و ٠س ازبعة ٠ كان رد الفعل الطورلذوظءورالمذبالا٠تدالذنزف عن اشاكاة الأجنبية، واسدموزوعاته ٠ن أمرار اديحية وع٠ود الفروسية، وندم المال٤لسالعقلوالشعورلسالطقوالغرد١على ابمعزن والذاتية هلى المونوبة. ثم أطلق اش

نتائج محرك ABBYY (المعلم عليه بالأصفر خطأ)



8 المذاهب الأدبية المنحرفة

نشأت المذاهب الأدبية القويمة عند العرب الفرنج إما مرحلة الطور مذهب تقدم به مبتدعوه ، وإما رد فعل لمذهب غلا فيه متبعوه . فمذهب عبد الحميد بن يحيى كان الطور الأول للأسلوب العربي الضيق الموجز دعت إليه مقتضيات المجتمع الجديد من تشعب اطراف الدولة وبدو ثمار الحضارة ودنو العربية من الفارسية . ومذهب ابن المقفع الذي ظهر في فجر الحضارة العربية كان طوره الثاني دعا إليه اتساع الخلافة وتنوع الثقافة وشدة اختلاط العرب بالفرس .

ثم كان طوره الثالث مذهب الجاحظ الذي اقتضاه نقل العلوم الأجنبية وازدهار المدنية العباسية وانتشار المقالات الإسلامية وتعقد الحياة الإجتماعية وإقتباس الآراء الفلسفية

ثم أترف المسلمون وتقلبوا في أعطاف النعمة وتأنقوا في مظاهر العيش فظهر طوره الرابع في مذهب ابن العميد المسجوع المنمق

وإلى هنا كان التطور في النثر الفني تطورا مطردا يسير من الضيق إلى السعة ومن الجزالة إلى الرقة ومن الترسل المتوازن إلى الصنعة المطبوعة . فلما ضعفت الخلافة وصار الأمر إلى غير أهله جرت على الكتابة أعراض الفساد والوهن فكثر الزيف وانتشرت الصنعة وكان من ذلك مذهب القاضي الفاضل وهو الطور الخامس من أطوار الأسلوب العربي غلا فيه أصحابه حتى أفسدوا الفكرة وشوهوا الصورة . ومن هنا كان رد الفعل بظهور مذهب ابن خلدون إذ رغب عن السجع وزهد في البديع وسار باللفظ وراء المعنى . ثم تطور هذا المذهب بتأثير الحضارة الأوربية ونقل الآداب الأجنبية إلى الأسلوب الذي يكتب به الكتاب الموهوبون اليوم .

كذلك كانت نشأة المذاهب الأدبية في أوربا ، ففي فرنسا مثلا كانت اللغة وآدابها خاضعة لسلطان الإغريقية واللاتينية على أثر البعثات الرواقع اليونانية والرومانية في إيطاليا ، فالألفاظ يكثر فيها العامي والدخيل ، والتراكيب يطغى عليها هوميروس أو فرجيل . والنظم والنثر يجريان على المحاكاة الشكلية ، والموضوعات تؤخذ من الأساطير الوثنية . فكان لابد لهذه العبودية الأدبية من محرر يكفكف شرتها وينظم فوضاها ، فظهر المذهب الاتباعي في أوائل القرن السابع عشر وهو يقوم على تقليد اليونان والرومان تقليد التلميذ لمعلمه لا تقليد المصور أمثاله . ثم طهروا اللغة من الدخيل والعامي ، ووسعوا دائرتها بالوضع والاشتقاق ، وجعلوا للعقل السلطان المطلق على الأدب فلا يكادون يأذنون بالهم إلى الخيال والعاطفة . ثم قيدوا الفنون الشعرية والقصصية بقيود من القواعد الصلبة وأخذوا أنفسهم بها حتى افترجت الحال بينهم وبين العامة ، وانقطع السبب بينهم وبين الطبيعة . وكان رد الفعل الطبيعي لذلك ظهور المذهب الابتداعي الذي عزف عن المحاكاة الأجنبية ، واستمد موضوعاته من أسرار المسيحية وعهود الفروسية ، وقدم الخيال على العقل والشعور على المنطق والفردية على الجمعية والذاتية على الموضوعية . ثم أطلق الفن من القيود التي كبله بها الاتباعيون فأمعن الشعراء

نتائج محرك Sotoor

وبمراجعة النتائج بالإضافة إلى نماذج النتائج بين المحركات الثلاثة يتضح لنا أن محرك Google يتميز بمتوسط الدقة، ولكن يعيبه عدم قدرته على الحفاظ على النسق العام للصفحات. ويتميز محرك ABBYY بقدرته على الحفاظ على النسق العام للصفحات لكنه ضعيف من حيث الدقة. في حين أن محرك Sotoor يجمع بين الميزتين: حيث الدقة العالية والحفاظ على النسق العام بالصفحات.

دراسة عمل مشروع رقمنة الوثائق العربية

مما سبق ذكره فإن تحديد متطلبات مشروع رقمنة الوثائق العربية يتوقف على عدة نقاط:

١. عدد الوثائق ومدة عمل المشروع:

يتحدد من خلالهما عدد أجهزة الماسحات الضوئية وعدد أجهزة الحاسب ونسخ المحركات الخاصة بالتعرف الضوئي وكذلك الموارد البشرية المطلوبة لذلك.

٢. نوع الوثائق:

يتحدد من خلالها نوع الماسح الضوئي، إذا كان سيتعامل مع ورق أو كتب أو صحف وغيرها؛ حيث يختلف حجم الصفحات وأشكالها. وكذلك حساب قدرته الإنتاجية بعدد الصفحات ودقة التصوير بكل صفحة.

٣. الدقة المطلوبة:

تحدد شكل التدخل البشري في قياس الكفاءة والمراجعة.

٤. مدى أهمية سرية البيانات بالوثائق:

تحدد مكان العمل والجهات المعنية بعمل التصوير الضوئي للوثائق بالإضافة إلى مكان تنصيب محركات التعرف الضوئي ما إذا كانت على الإنترنت أو بمكان العمل ذاته.

٥. تكلفة المشروع:

هي المحرك الأساسي للمشروع؛ حيث القدرة على تنفيذ جميع جوانب المشروع بشكلها الصحيح. وبداية من الدراسات المبدئية للمشروع سيتم وضع تصور عام للتكلفة والمقارنة بعدد الوثائق لحساب تكلفة رقمنة الصفحة الواحدة من نموذج الوثائق.

٦. العائد المتوقع من عمل مشروع الرقمنة:

ينقسم العائد من عملية الرقمنة إلى شقين؛ أحدهما معنوي، وآخر مادي. إن الجانب المعنوي هو إثراء المعلومات والقدرة على الوصول إلى الكم الهائل من المعلومات

التي قد تفيد من يبحث عنها بشكل أسهل وأسرع. أما الجانب المادي فهو إعادة بيع المعلومات وخاصة إذا كانت ذات قيمة ونادرة الوجود.

وبشكل عام تتفاوت تكلفة رقمنة الصفحة الواحدة من الوثائق بمختلف مجهودات الرقمنة وتقنياتها السابق ذكرها. إن رقمنة الوثائق بالمجهود البشري فقط تعتبر من أغلى أنواع الرقمنة، كما أنها يعيبها البطء الشديد. أما المجهود التقني فيتميز بالسرعة العالية وقلة التكاليف. ولكن المجهود البشري والتقني معًا يتوسطهم في التكلفة والدقة والزمن.

المراجع

1. **الفجوة الرقمية**: رؤية عربية لمجتمع المعرفة، بالاشتراك مع د. نادية حجازي، سلسلة عالم المعرفة، رقم ٣١٨ أغسطس ٢٠٠٥.
https://drive.google.com/file/d/1ffWYkrE0GAEootsz9EzutJ6SmEHGS6tz/view
2. http://waelalghoolarabic.blogspot.com/2018/
3. https://www.mklat.com/%D8%A3%D8%AF%D9%88%D8%A7%D8%AA-%D8%A7%D9%84%D9%83%D8%AA%D8%A7%D8%A8%D8%A9/
4. https://www.scanjunction.com/what-are-the-different-types-of-scanners-and-their-uses/