إيسيسكو
ICESCO

مجلة الإيسيسكو للغة العربية

دوريَّة علميَّة محكَّمة تُصدرها

مُنظَّمة العالم الإسلامي للتربية والعلوم والثقافة

المجلد الأول - العدد الثاني
جمادى الآخرة 1446 / ديسمبر 2024

منشورات منظمة العالم الإسلامي للتربية والعلوم والثقافة
(إيسيسكو)

شارع الجيش الملكي، حي الرياض، ص. ب. 2275 ، ر. ب. 10104، الرباط، المملكة المغربية

المجلد الأول - العدد الثاني
جمادى الآخرة 1446 / ديسمبر 2024

© إيسيسكو
جميع حقوق إعادة الإنتاج والترجمة والاقتباس محفوظة

الرقم الدولي الموحد للدوريات الورقية (ISSN): 3007-5726
الرقم الدولي الموحد للدوريات الإلكترونية (E-ISSN) : 3007-5734

التصميم والطباعة في الإيسيسكو

+212537566052 | www.icesco.org | contact@icesco.org

هيئة التحرير

المشرف العام

د. سالم بن محمد المالك

المدير العامُّ لمنظَّمة العالم الإسلاميِّ
للتَّربية والعلوم والثَّقافة (إيسيسكو)

رئيس التَّحرير

أ.د. مجدي حاج إبراهيم

مدير التَّحرير

أ.م.د. أدهم محمد علي حموية

المحرِّر اللغوي

د. مهند عمر رنــة

الهيئة الاستشارية

- **أ.د. أحمد المتوكل**
  المملكة المغربية
- **أ.د. رمزي البعلبكي**
  الجمهورية اللبنانية
- **أ.د. سعد مصلوح**
  جمهورية مصر العربية
- **أ.د. عبد السلام المسدي**
  الجمهورية التونسية
- **أ.د. عبد العزيز الحربي**
  المملكة العربية السعودية
- **أ.د. محمد حسين آل ياسين**
  جمهورية العراق
- **أ.د. محمد عدنان البخيت**
  المملكة الأردنية الهاشمية
- **أ.د. مسعود صحراوي**
  الجمهورية الجزائرية الديمقراطية الشعبية
- **أ.د. وليد القصاب**
  الجمهورية العربية السورية
- **أ.د. أون يون كيونغ (نبيلة)**
  جمهورية كوريا
- **أ.د. رحمة أحمد الحاج عثمان**
  ماليزيا
- **أ.د. محمد طالب الحوري**
  الولايات المتحدة الأمريكية
- **أ.د. نيكولاس روزر نبوت**
  مملكة إسبانيا

"مجلّة الإيسيسكو للُّغة العربيَّة" دوريَّةٌ علميَّةٌ محكَّمةٌ للبحوث في اللُّغة العربيَّة وآدابها وعلومها، تُصدرها منظمَّة العالم الإسلاميّ للتَّربية والعلوم والثَّقافة (إيسيسكو)، في شهري يونيو وديسمبر (حزيران وكانون الأوَّل) من كلِّ عامٍ، ويشتمل نطاقُها على محورين لبحوث اللُّغة العربيَّة وآدابها وعلومها:

- المحور النَّظري، ويضمُّ البحوث اللِّسانيَّة والأدبيَّة والنَّقديَّة.
- المحور التَّطبيقي، ويضمُّ البحوث التَّعليميَّة والتَّرجميَّة والحوسبيَّة.

لا تمثِّل آراء الكتَّاب بالضرورة توجُّهات منظمة العالم الإسلامي للتربية والعلوم والثقافة (إيسيسكو)

مراسلة المجلة

مركز اللغة العربية للناطقين بغيرها
منظمة العالم الإسلامي للتربية والعلوم والثقافة
(إيسيسكو)
شارع الجيش الملكي، حي الرياض، ص.ب. 2275، ر.ب. 10104
الرباط، المملكة المغربية

www.ijal.icesco.org || ijal@icesco.org

ضوابط النشر

- أن يتَّسم البحث بالجدَّة والموضوعيَّة والرَّصانة العلميَّة.
- ألَّا يكون البحث منشورًا أو مقدَّمًا للنشر في أيِّ وعاءٍ علميٍّ آخر.
- ألا تتجاوز نسبة الاقتباس في البحث 30٪ (مع استثناء المصادر والمراجع).
- أن يكون عدد كلمات البحث ما بين 5000-7000 كلمة؛ إضافة إلى ملخص للبحث كلماتُه ما بين 200-300 كلمة، وترجمته إلى الإنجليزية.
- أن يكون التوثيق بطريقة الحواشي في كل صفحة، وتُدرج أرقامها بعد علامات الترقيم في المتن، والترقيم جديد لكل صفحة.
- أن يكون التوثيق وفق نظام شيكاغو Chicago.
- أن تُضاف قائمة للمصادر والمراجع مكتوبة بالحروف اللاتينية.
- أن تُرسل البحوث من خلال إنشاء حساب في موقع المجلة (ijal.icesco.org).

الأُنثى في الفكرِ النَّحويِّ العربيِّ: قراءةٌ معرفيَّةٌ

مريم بيد الله علي الحجازي، محمد صالح بوبركة 9

حصولُ الأمنِ الثَّقافيِّ رهينُ اعتمادِ اللُّغةِ الجامعةِ

أحمد العلوي العبدلاوي 37

العلمُ في اللُّغتينِ الإنكليزيَّة والعربيَّة ومُستقبَلُ التَّعليمِ والبحثِ العلميِّ والثَّقافةِ فيهِما

شفيق بنات، فؤاد عبد المطلب 53

نَسَقُ النَّقدِ المعرفيِّ: من النَّصِّيَّة إلى التَّأويليَّة

محمد سالم سعد الله 93

أبو جعفر الرُّعينيُّ الغرناطيُّ أوَّلُ مَنْ ألَّف في المثلَّثات اللُّغويَّة القرآنيَّة:

بحثٌ في معالِمِ الصَّنعةِ والتَّفرُّدِ والجِدَّةِ

محمد ابن سفاج 115

بنيةُ التَّكريرِ في الخطابِ النَّبويِّ: خطبةُ الوداعِ أُنموذجًا

مصطفى قنبر 167

شعريَّةُ التَّناوُبِ الأُسلوبيِّ في "القصيدةِ الدِّمشقيَّةِ" لنزار قبَّاني

محمد فلفل 197

إشكالاتُ التَّرجمةِ من التركيَّةِ إلى العربيَّةِ لدى الطُّلَّابِ الأتراكِ: دراسةٌ تقابليَّةٌ

هاني إسماعيل رمضان 217

تطبيقُ الذَّكاءِ الاصطناعيِّ في تعليمِ اللُّغاتِ وتعلُّمِها:

برنامجُ (QuestionWell.org) أُنموذجًا

محمد صبري شهرير 237

التَّعرُّفُ الضَّوئيُّ على المخطوطِ العربيِّ واقعُه وتحدِّياتُه: تقنيَّةُ (OCR-RDI) أُنموذجًا

حسين البسومي، محسن رشوان، ياسر السيد 259

© إيسيسكو ICESCO 2024

2024 - (2)1
Received اسـتُلم 2024-10-14
Accepted قُبل 2024-11-15

التَّعرُّفُ الضَّوئيُّ على المخطوطِ العربيّ: واقِعُه وتحدِّياتُه
تقنيَّةُ (OCR-RDI) أنموذجًا

حسين البسومي*، محسن رشوان، ياسر السيد*****

مُلخَّص

يناقش هذا البحث إمكانية توظيف تقنية التعرُّف الضوئي على النصوص (OCR)، في التعرُّف على المخطوط العربي القديم، وسُبُلِ استعمالها في محاكاة مهمة المحقق البشري في نسْخِ المخطوط، وتحويله إلى الصيغة الرقمية القابلة للتحكُّم فيها، مع بيان التحديات التي تواجه توظيف تلك التقنية، وقد توسَّل البحث المنهج الوصفي في التعريف بتقنية التعرُّف الضوئي على النصوص، وبرنامج "سطور"، وخصائص تحقيق المخطوط العربي، وأبرز ما يواجه خبيرَ التحقيق من صعوبات، وأيضًا المنهج التجريبي في عقْدِ تجربتين لقياس كفاءة برنامج "سطور"، ومقارنة نتائجه بنتائج برنامجين آخرين للتعرُّف الضوئي على المخطوط، وخَلَصَ البحث إلى أوجه تفاوُت البرامج الثلاثة في مستوى الدقة، والكشف عن مستوى التطور الذي حقَّقته تلك التقنية بعامة، مما يشجع على توظيفها في إنجاز مهمة قراءة المخطوط ونسْخِه إلكترونيًّا.

الكلمات الرئيسة: المخطوط العربي، تقنية OCR، سطور، زنكي، غوغل لنز

* أستاذ مشارك متخصِّص في علم اللغة الحاسوبي، كلية الآداب، جامعة الوادي الجديد، جمهورية مصر العربية، hosinmab_2010@yahoo.com.

** أستاذ الاتصالات والإلكترونيات، كلية الهندسة، جامعة القاهرة، جمهورية مصر العربية، mrashwan@rdi-eg.ai.

*** خبير التحول الرقمي وإدارة المشروعات، جمهورية مصر العربية، yasserfayez2@yahoo.com.

© إيسيسكو ICESCO 2024

ICESCO Journal of Arabic Language
ISSN: 3007-5726 | E-ISSN: 3007-5734
https://doi.org/10.70910/ijal1(2)10

Optical Character Recognition for Arabic Manuscripts Reality and Challenges, OCR-RDI Technology as a Model

Hussein Elbasoumy[*], Mohsen Rashwan[**], Yasser Elsayed[***]

Abstract

This study discusses the potential of using Optical Character Recognition (OCR) technology to read ancient Arabic manuscripts, and its role in simulating the textual editor's task of transcribing manuscripts into digitally manageable formats. The Study also aims to identify the challenges in using this technology. The study employed a descriptive approach to introduce Optical Character Recognition (OCR) technology, provide an overview of the *Sotoor* program, highlight the features of Arabic manuscript verification, and address the key challenges faced by experts in this field. Additionally, an experimental approach was adopted through two trials to evaluate the efficiency of the *Sotoor* program and compare its results with those of two other programs. The Study concluded by outlining the disparities in accuracy levels among the three programs, showcasing the overall advancements in this technology, and encouraging its adoption for the electronic reading and transcription of manuscripts.

Keywords: *Arabic manuscript, OCR, Sotoor, Zinki, Google Lens*

[*] Associate Professor of Computational Linguistics, Faculty of Arts, New Valley University, Egypt, hosinmab_2010@yahoo.com.
[**] Professor of Communication and Electronics, Faculty of Engineering, Cairo University, Egypt, mrashwan@rdi-eg.ai.
[***] Expert of Digital Transformation and Project Management, Egypt, yasserfayez2@yahoo.com.

© إيسيسكو ICESCO 2024

مُقدِّمة

يزخر التراث العربي بزادٍ وفير من المخطوطات التي تنتمي إلى شتَّى فروع العلم والمعرفة، وتحمل في طياتها نتاج قرائح علماء الحضارة الإسلامية وأدبائها ومؤرخيها وفلاسفتها بمختلف أعراقهم وأجناسهم، وقد ضاع كثيرٌ من تلك المخطوطات في النكبات التي مرَّت بالأمة الإسلامية عبْرَ تاريخها، وبقي لنا منها زادٌ يسيرٌ، نَشَطَتْ ثلَّة من العلماء لتحقيقه وإخراجه على نحْوٍ يسهُلُ على طلبة العلم التعامل معه، والإفادة منه، وقد أحسنوا بلاء فيما انتدبوا أنفسهم له، غيْرَ أن هذه الجهود لم تأتِ إلا على أقلِّ ذلك الموروث، ولا يزال أكثره مخزونًا في أرفف المكتبات، يحول دوْنَ تحقيقه وإخراجه للناس كثيرٌ من الجهود المضنية والسنوات الطوال.

وفي جانب آخر شهدت التقنيات الحاسوبية طفرات كبيرة متسارعة في تطوير قدراتها، وزيادة سعة تخزينها، وترقية نماذجها وأدواتها، فصارت أقدر من ذي قبْلُ على محاكاة الإنسان في كثيرٍ من مهامِّه في أغلب مجالات الحياة، ومن المجالات التي استقطبت جهود الحاسوبيين لتوظيف التقنيات الحاسوبية المتطورة؛ مجالُ تحقيق موروثنا العربي والإسلامي من المخطوطات.

ونظرًا إلى ما يكتنفه علْمُ التحقيق من صعوبات جمَّة، وإلى ما يحتاجه من معارف متنوعة وخبرات عملية متراكمة؛ للتعامل مع المخطوطات بتنوعاتها المتباينة؛ لا يزال إسهام التقنيات الحاسوبية – في إنجاز أغلب مهام التحقيق – إسهامًا جزئيًّا ومحدودًا، ومن المهام التي لقيت كثيرَ اهتمام من المحقِّقين والحاسوبيين لتوظيف التقنيات الحاسوبية في إنجازها؛ التعرُّف الآلي على المحتوى النَّصِّي للمخطوط على اختلاف أحواله، فتعدَّدت البرامج التي وظَّفت تقنية التعرُّف الضوئي على النص (OCR)، ووظَّفت فيها بعْضَ تقنيات الذكاء الاصطناعي لمحاكاة المحقِّق البشري في معارفه ومهاراته في قراءة المخطوط.

ويعنى هذا البحث بالتعريف بأحد تلك البرامج، وهو برنامج "سطور" (Sotoor)،[1] الذي طوَّرته الشركة الهندسية لتطوير نظم الحاسبات (RDI)، ولمزيد من التوضيح وبيان مدى

[1] انظر: برنامج "سطور".

© إيسيسكو ICESCO 2024

قدرته على أداء مهامِّه، لجأ البحث إلى مقارنة نتائجه بنتائج برنامج "زنكي" (Zinki)،[1] وبرنامج "غوغل لنز" (Google Lens).[2]

وتتمثَّل المشكلة البحثية في ضخامة الجهود المطلوبة من أجْلِ تحقيق موروثنا من المخطوطات العربية والإسلامية، وصعوبة تعامل التقنيات الحديثة مع نصوص تلك المخطوطات؛ نظرًا إلى أنها مكتوبة بخط اليد، بخطوط متنوعة، وقد قام على نسْخِها من يُحسن مهنة النساخة ومن لا يُحسنها، إضافة إلى تعرُّضها لسوء التخزين والحفظ، فأصابها التلف والطمس، وغيرها من المشكلات التي تصعِّب على التقنيات الحاسوبية التعرُّف على نصوصها، وتحويلها إلى نصوص رقمية قابلة للتحكُّم فيها، ونتيجة لذلك لا تزال تقنية التعرُّف الضوئي على نصوص المخطوطات تعاني مشكلات تحتاج إلى حلٍّ، وهو ما يحاول هذا البحث مناقشته من خلال التطبيق على برنامج "سطور".

وإن هذا البحث بينيٌّ، جمع بين جانبين؛ أحدهما المخطوطات العربية بما يختصُّ به نَسخُها من تنوُّع ونسبية، والآخر تقنية التعرُّف الضوئي على الحروف العربية بما يحكمها من منهج رياضي، وقد تناولت كلًّا من الجانبين في ضوء الآخر، والمطالع للنتاج البحثي اليوم لا يجد كثيرًا من البحوث التي تعالجهما معًا، وإن كثرت تلك التي تعنى بكلٍّ منهما بعيدًا من الآخر، ومن الدراسات التي تناولت الجانبين معًا ما يأتي دراسة بعنوان "استخدام تقنية (ICR) التعرُّف الذكي على الحروف المكتوبة بخطِّ اليد في قراءة الوثائق والمخطوطات العربية، وانعكاس ذلك على مؤسسات حفْظِ التراث"،[3] وفيها تفريق بين تقنية التعرُّف الذكي على الحروف المكتوبة بخط اليد (ICR)، وتقنية التعرُّف البصري على الحروف (OCR)، مع دراسة حالة لمشروعين من المشروعات التي طبقت تقنية (ICR)؛ أحدهما طُبِّق على نوعية معينة من

[1] انظر: برنامج "زنكي".

[2] انظر: برنامج "غوغل لنز".

[3] انظر: نسمة علي عبد الحميد، "استخدام تقنية (ICR) التعرف الذكي على الحروف المكتوبة بخط اليد في قراءة الوثائق والمخطوطات العربية، وانعكاس ذلك على مؤسسات حفظ التراث"، **الرزنامة**، العدد (20)، (2022).

© إيسيسكو ICESCO 2024

المخطوطات العربية، وهو المشروع التعاوني بين المكتبة البريطانية ومكتبة قطر الرقمية، والآخر مشروع مصحف مسقط الإلكتروني، الذي طُبِّق على مصحف عُمان المخطوط لهذا الغرض بخاصة، وقد خلصت الدراسة إلى أهم التحديات أمام الوصول إلى مستوى الجودة المطلوب في تطبيق هذه التقنية على المخطوطات والوثائق العربية المكتوبة بخط اليد.

وهذه الدراسة تتقاطع مع هذا البحث في الموضوع والمشكلة والأدوات، وتفترق عنه في أمرين؛ أحدهما أن موضوعها اقتصر على التعرُّف على الكلمات الدالة فقط، لا كامل المخطوط، وذلك في المشروع التعاوني بين المكتبة البريطانية ومكتبة قطر الرقمية، واقتصر في مشروع مصحف مسقط على نصوص القرآن الكريم، وهي مادة لغوية خاصة لا عامة، أما هذا البحث فلم يتوقف عند الكلمات الدالة، بل تعداها ليشمل كامل نصِّ المخطوط، ولم يتوقف عند مادة لغوية خاصة، بل توسع ليشمل نصوصًا لغوية عامة متنوعة، والأمر الآخر أن الدراسة السابقة أُنجزت عام (2022)، والفارق الزمني بينها وبين هذا البحث كافٍ لإحداث تطوير ملحوظ في التقنيات، فقد أشارت الدراسة إلى أن السوق يخلو من تطبيقات تعنى بالتعرُّف البصري على الخط العربي المكتوب بخطِّ اليد،[1] في حيْنِ أن في السوق اليوم أكثرَ من تطبيق، منها التطبيقات الثلاثة التي يقارن هذا البحث بين نتائجها، وكذا أشارت الدراسة إلى الضعف الملحوظ في نتائج التطبيق التي جرَّبته حينذاك،[2] مع أنه كان يعمل على الكلمات الدالة فقط، في حيْنِ كَشَفَ هذا البحث عن نتائج متقدمة للتطبيقات الثلاثة في التعامل مع عينة التجربة.

دراسة أُخرى كانت بعنوان "رقمنة الخط العربي ودوره في تحقيق المخطوطات: مشاريع تقنية وتحديات بشرية"،[3] وعَرَضَتْ لأنواع الخطوط العربية ومميزاتها، وما تمثِّله من

[1] انظر: المرجع السابق، ص381.

[2] انظر: المرجع السابق، ص404.

[3] انظر: جميلة روقاب، "رقمنة الخط العربي ودوره في تحقيق المخطوطات: مشاريع تقنية وتحديات بشرية"، **المجلة الجزائرية للأبحاث والدراسات**، 2(8)، (2019).

© إيسيسكو ICESCO 2024

صعوبة في تحقيق المخطوط العربي، ثم عَرَضَتْ لمشروع الذخيرة اللغوية العربية لعبد الرحمن الحاج صالح، من جهة فائدته في خدمة التراث المخطوط، ثم عَرَضَتْ لبعْضِ برامج تقنية الخط العربي وأدوات الرقمنة، وبيَّنت تبايُن الباحثين في رقمنة الخط العربي، وتختلف هذه الدراسة عن هذا البحث في أنها عَرَضَتْ لبعْضِ خصائص الخط العربي وإمكانات رقمنته، من دوْنِ العناية بغيْرِه من الجوانب التي تؤثر في نسْخِ المخطوط، ولم تقدِّم دراسة تجريبية لقياس كفاءة تقنية (OCR) في التعرُّف الضوئي على نصوص المخطوطات العربية.

وهناك دراسة ثالثة بعنوان "نحْوَ التحقيق الإلكتروني (الرقمي) للمخطوطات: استخدام تطبيقات تكنولوجيا المعلومات في تحقيق المخطوطات لدى الباحثين الجزائريين في مجال المخطوطات"،[1] هَدَفَتْ إلى استطلاع رأي الباحثين الجزائريين في استعمال التطبيقات التقنية الحديثة في تحقيق المخطوطات، وذلك بتوزيع استبانة وتحليل نتائجها، وقد عَرَضَتْ لمبادئ علْمِ تحقيق المخطوطات، والعلوم المساعدة إياه، وأقسامه، كتحقيق العنوان، واسم المؤلِّف، ونسبة الكتاب إليه، وتحقيق متْنِ الكتاب، ثم عَرَضَتْ لعُدة المحقِّق من العدسات المكبرة، والقارئة، وأشارت إلى صفات المحقِّق، وإجراءات عملية التحقيق وصعوباتها، وهذه الدراسة مع قرْبِ عنوانها من موضوع هذا البحث، جاءت معالجتها الموضوع مفارقة إياه في مشكلته البحثية، وأهدافه، ومنهجه.

تقنية التعرُّف الضوئي على الحروف العربية (OCR)

أولاً: مفهومها وأهميتها

تقنية التعرُّف الضوئي على الحروف (OCR) هي تقنية ذكية تُستعمل لتحويل النصوص في الصور إلى نصوص رقمية قابلة للقراءة والتحرير، وهي ذكية لأنها تحاكي طريقة عَمَلِ النظام العصبي عنْدَ الإنسان؛ إذ تقطِّع صُوَرَ الحرف المراد التعرُّف عليها، وتطابِقُها مع مقاطع مخزَّنة

[1] انظر: مولاي محمد، "نحو التحقيق الإلكتروني (الرقمي) للمخطوطات: استخدام تطبيقات تكنولوجيا المعلومات في تحقيق المخطوطات لدى الباحثين الجزائريين في مجال المخطوطات"، **مجلة الحوار المتوسطي**، 5(1)، (2014).

© إيسيسكو ICESCO 2024

مسبقًا، لذا يمكنها التعرُّف على الوثائق العادية، بل الوثائق التالفة التي تظهر فيها الحروف غيْرَ واضحة أو مطموسًا بعْضُ أجزائها.[1]

وتنقسم تلك التقنية بعامة إلى تقنيتين فرعيتين؛ إحداهما تقنية التعرُّف على الأحرف المطبوعة آليًّا (Optical Character Recognition - OCR)، والأخرى تقنية التعرُّف على الأحرف المكتوبة بخط اليد (Handwriting Recognition - HWR)،[2] ولكن "النوع الثاني يشكِّل تحديًا أكبر، وذلك للاختلافات الكبيرة في طرق الكتابة، وخصوصيات كل شخص في كتابة الحروف".[3]

وتتمثَّل أهمية هذه التقنية في تمكين الحاسوب من التعامل مع المحتوى النصي لتلك الوثائق التي حوَّلتها من صُوَرٍ إلى نصوص رقمية يمكن للحاسوب فهرستها، وتصنيفها، ومساعدة الباحثين في البحث والتنقيب عن مُرادهم فيها؛ إذ إن غزارة الأدبيات التراثية وتنوّعها، وعجْزَ وسائل التوثيق التقليدية عن وضْعِها وضعًا متكاملًا وافيًا في أيدي الباحثين والمؤرخين؛ قد أبرز تحديًا كبيرًا أمام نفْضِ الغبار عن ذلك التراث العربي الفريد في ضخامته وتنوُّعه وثرائه،[4] هذا مع إثراء المحتوى الرقمي العربي، وتمكين تقنيات الذكاء الاصطناعي من التعلُّم منها، ومن ثم أداء عمليات تكشف عما تحمِلُه تلك النصوص من خصائص مؤلِّفيها، وسمات عصورهم وبلدانهم، وتستخلص مضامينها العلمية والحضارية التي تمثِّل تاريخ الأمة وثقافتها، ومن شأنِ ذلك أن يؤدي إلى "توفير معلومات تراثية وإحصائيات متنوعة ذات أهمية في كتابة التاريخ، أو إعادة كتابته، من حيْثُ إنها تقدِّم إمكانات تحليلية لمضامين المخطوطات

---

[1] انظر: صديق بسو، "خوارزميات وبرامج التعرف الآلي على الخط العربي"، **ملتقى التكنولوجيات الجديدة ودورها في صناعة اللغة العربية واستعمالها**، المجلس الأعلى للغة العربية، جامعة الإخوة منتوري، قسنطينة، 17 مايو 2017، ص197.

[2] انظر: شذى عبد الرحيم خليفة الرضي، "التعرف على النصوص باستخدام تعلم الآلة"، **مجلة العلوم الإنسانية والطبيعية**، 3(11)، (2022)، ص29.

[3] بسو، **خوارزميات وبرامج التعرف الآلي على الخط العربي**، ص194.

[4] انظر: هاشم فرحات، **دور تكنولوجيا المعلومات في ضبط المخطوطات العربية وإتاحتها: مراجعة علمية للإنتاج الفكري** (الرياض: جامعة الملك سعود، 2001)، ص28.

© إيسيسكو ICESCO 2024

يمكن معها الحصول على دلالات تاريخية أوسع من تلك التي يمكن الحصول عليها بالوسائل التقليدية، إضافة إلى تميُّزها بالدقة والاستعداد إلى الإحصاء كلَّما أمكن ذلك".[1]

ونظرًا إلى تلك الأهمية، "خَطَتْ كثيرٌ من اللغات أشواطًا كبيرة في ميدان التعرُّف الآلي على الخط، ومنها التعرُّف على الأحرف اللاتينية، الصينية، اليابانية... حيْثُ عَرَفَتْ الكثير من اللغات التي تعتمد على الحرف اللاتيني نِسَبًا مرتفعة في التعرُّف تَصِلُ إلى (100٪)، بينما تظلُّ نِسَبُ التعرُّف على الحرف العربي منخفضة، رَغْمَ أن البحوث التي نُشِرَتْ حَوْلَ التعرُّف على الحرف العربي كانت سنة (1975)".[2]

ثانيًا: برنامج "سطور" كيفية عمَلِه ومكوّناته

يعنى هذا البحث بالتعريف ببرنامج "سطور" (Sotoor) الذي طوَّرته شركة (RDI) اعتمادًا على تقنية (OCR)،[3] ويعرض لمكوناته الأساس التي تمكِّنه من أداء مهامه؛ إذ يختصُّ كلُّ واحد منها بمرحلة من المراحل الثلاث التي تمرُّ بها القراءة الضوئية للصورة، وهذه المراحل هي:

- مرحلة اكتشاف الخطوط (Line Detection): يُنجزها أنموذجٌ دُرِّبَ لاكتشاف خطوط النصِّ داخل الصفحة (Line Detection Model).
- مرحلة التعرُّف على النص (Text Recognition): يُنجزها أنموذجُ التعرُّف على النص (Text Recognition Model) للتعرُّف على الحروف داخل الخطوط المكتشفة.
- مرحلة التجميع (Clustering): تُنجزها خوارزمية (Regioner) لإعادة تجميع الخطوط المكتشفة وتنظيمها في فقرات وأقسام.

---

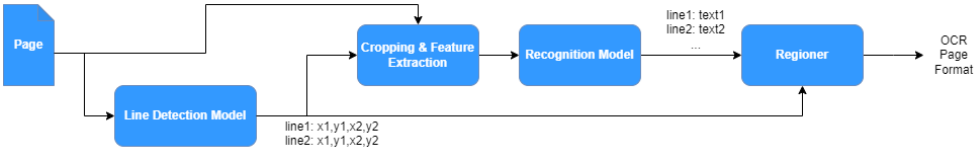
[1] السابق نفسه.

[2] بسو، **خوارزميات وبرامج التعرف الآلي على الخط العربي**، ص192-193.

[3] يمكن مراجعة الموقع الشابكي لبرنامج "سطور".

© إيسيسكو ICESCO 2024

وتَضِحُ تتابعية المراحل السابقة وما يُنجزها من نماذج من خلال المخطط الآتي:



شكل (1): نظرة عامة على نظام عمل برنامج "سطور"

وفيما يأتي توضيح المكونات الثلاثة الرئيسة للنظام:

1. أنموذج اكتشاف الخطوط (Line Detection Model):

يعمل باستخراج مربعات الإحاطة (Bounding Boxes) لخطوط النصِّ من صورة الصفحة، وهو أنموذج دُرِّب على الصفحات عن طريق الحركة التراجعية (Regression) لاكتشاف إحداثيات مربع الإحاطة،[1] ويستعمل النظامُ أحدَثَ خوارزميات تعلُّم الآلة (Machine Learning) لتدريب أنموذج الحركة التراجعية لتعلُّم خرائط هندسة (RBOX) لمربعات الإحاطة للنصِّ، وهذه الخرائط هي خَمْسُ خرائط تمثِّل الدوران والمسافات الأربعة من الأعلى والأسفل واليسار واليمين بيْنَ حدود الصورة وبكسلات مربع الإحاطة، ويولِّد الأنموذج مخرجين أساس؛ خريطة هندسة (RBOX)، وخريطة الاحتمالات لموقع النصوص.[2]

ولا يقتصر الأنموذج على التعرُّف على الخطوط التي تتألَّف منها النصوص العربية فقط، بل يمكنه أن يتعرَّف على اللغات الأخرى كذلك، فقد دُرِّب على مجموعة بيانات من صُوَرِ المستندات المطبوعة التي تحتوي على نصوص عربية وأجنبية (إنكليزية وإيطالية)؛ لذا يعمل الأنموذج كاشفًا للخطوط المطبوعة مهما كانت اللغة، وفيما يأتي بيان توزيع صُوَرِ النصوص التي تدرَّب عليها، واختُبر بها، موضَّح في الجدول (1).

[1] See: Raymond W. Smith, "An Overview of the Tesseract OCR Engine", *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Curitiba, Brazil, (2007), pp. 629-633.

[2] See: Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng & Xiangyang Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals", *IEEE Transactions on Multimedia*, 20(11), (2018), pp. 3111-3122.

© إيسيسكو ICESCO 2024

توزيع البيانات	التدريب	الاختبار
عدد الصور	18299	1976

جدول (1): توزيع بيانات التدريب والاختبار لاكتشاف الخطوط

وقد رُوعي في تدريب الأنموذج ما يأتي:

(أ) **استعمال ثلاثة أنواع من دوال الخسارة (Loss Function)**، ودالة الخسارة مقياس كمِّي يُستعمل في تعلُّم الآلة لقياس مدى اختلاف النتائج التي يتنبَّأ بها الأنموذج عن النتائج التي توصَّل إليها، فهي مقياس لمدى (سوء) أداء الأنموذج في مهمَّته، وهي أنواع:[1]

- دالة خسارة التصنيف: خسارة (Dice) المطبَّقة على خريطة الاحتمالات (Score Map) المولَّدة من الأنموذج، وخريطة موقع النصوص حسب البيانات نفسها (GT Text Location Binary Mask).
- دالة خسارة الهندسة: خسارة (Log IOU) المطبَّقة باستعمال خرائط الهندسة.
- دالة خسارة الدوران: المطبَّقة على خريطة الدوران.

والخسارة الإجمالية للتدريب هي مزيج من الخسائر الثلاث.

(ب) **تطبيق التحسينات (Augmentations)** على الصور في أثناء التدريب، وتُستعمل التحسينات لتوسيع مجموعة البيانات التدريبية اصطناعيًّا، وذلك بتطبيق مجموعة من التعديلات العشوائية على الصور التي يتدرَّب عليها، مما يؤدي إلى إنشاء صُوَرٍ جديدة تختلف قليلاً عن الصور الأصلية، ومن هذه التحسينات:[2]

- الدوران (Rotation)؛ إذ تدوَّر الصورة بزاوية عشوائية حول مركزها.
- الاقتصاص العشوائي (Random Cropping with Aspect Ratio)، ويُراعى ضبْطُ الخطوط وفْقَ عملية الاقتصاص.

[1] See: Zhongheng Zhang, "Introduction to Machine Learning: K-Nearest Neighbors, Regression, and Beyond", *Annals of Translational Medicine*, 7(7), (2019), pp. 1-20.
[2] See: Connor Shorten & Taghi M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 6(1), (2019), p. 60.

© إيسيسكو ICESCO 2024

- تحسينات الألوان (Color Augmentations)، بما يشمل تغيير التشبُّع، والتلوين، والتحويل إلى اللونين الأسود والأبيض (Binarization)، وغيرها من التحسينات.

(ج) تطبيق أنموذج التدهور (BSRGAN) على الصورة، وهو أنموذج عميق للتعلُّم الآلي مصمَّم لتدهور الصور واقعيًّا، أي إنه يحوِّل الصورة ذات الجودة العالية إلى صورة ذات جودة أقل، ليحاكي بذلك ما يمكن أن تتعرَّض له الصورة من تَلَفٍ أو تدهور في أثناء عملية التقاطها أو نقْلِها.[1]

يدرَّب الأنموذج في خمسين جولة (Epoch)، بهدف تعديل أوزانه وتحسين قدرته على التنبؤ، وذلك بمعدَّل تعلُّم (Learning Rate) ابتدائي (3-e3)، وتقليل معدَّل التعلُّم أسيًّا (Exponential Decay) لكل جولة بمعامل تقليل (0.1-)، مع تقليل معدَّل التعلُّم على استقرار الخسارة بمعامل (0.1) (LR Reduction on Plateau).

وقد يُقاس أداء الأنموذج بمقياس نجاح (PRIMA)، وهو مقياس عامٌّ يشتمل على مقاييس فرعية يتضمنها الجدول الآتي، وقد ضُمِّنت النتائج الجدول (2)، وكلُّ الأرقام المذكورة في الجدول هي معدلات نجاح، وكلما زادت أشار ذلك إلى جودة الأنموذج.

الدمج	الانقسام	الفقد	الفقد الجزئي	الكشف الكاذب	الإجمالي
٪97.31	٪99.05	٪98.73	٪93.85	٪97.99	٪96.32

جدول (2): أداء الأنموذج على عينة الاختبار

وفي أثناء التنبُّؤ يتطلَّب أنموذج اكتشاف الخطوط تحديد موقع مربع الإحاطة لكلِّ خطِّ نصٍّ، ويُنتِج الأنموذج مسافات (AABB) (خريطة الهندسة)، وخريطة الزاوية، وخريطة الاحتمالات، وببعْضِ المعالجة اللاحقة وإزالة التشويش غير المحدودة (Non-Maximum Suppression)، يمكن الحصول على مواقع مربعات الإحاطة النهائية، ويمكن رؤية الشكل (2) لبعْضِ العينات المرئية من مخرجات الأنموذج.

[1] See: Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao & Xiaoou Tang, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), pp. 63-79.

© إيسيسكو ICESCO 2024

شكل (2): اكتشاف الخطوط المرئية من الأنموذج، إلى اليسار جريدة إيطالية، وإلى اليمين كتاب عربي

2. أنموذج التعرُّف على النصِّ (Text Recognition Model):

هو جوهر نظام (OCR)؛ إذ يتعرَّف على النصِّ في الخطوط المكتشفة من الأنموذج السابق، ومهمته الأساس فكُّ ترميز الصورة التي تمثِّل خطًّا من النصِّ، وتحويلها إلى نصٍّ قابل للقراءة، ومدخلات الأنموذج صورة تمثِّل خطَّ نصٍّ واحد، والمخرَج نصٌّ قابل للقراءة، ويُطابق النصَّ في الصورة المدخلة.

ويستعمل النظام أنموذج (seq2seq)، وهو يعامل كلَّ صورة بوصفها سلسلةً من الإطارات الخاصة (Feature Frames)، ثم يُنتِج سلسلةً من رموز النصِّ التي تتفق مع تلك السلسلة من المميزات (Features)، ويستعمل أيضًا هيكلية (Encoder-Decoder)، والمرمِّز (Encoder) نوعٌ من التوافق (Conformer)، ويُركِّز على تعلُّم التمثيلات البصرية، ومفكُّ الترميز (Decoder) أنموذج تسلسُلٍ يعتمد (conv-1d)، ويُركِّز على تعلُّم اللغة، وكلاهما – المرمِّز ومفكُّ الترميز – يُدمَج لصوغ أنموذج ناقل (Transducer) دُرِّب بخسارة (RNN-T) المختصرة.

وقد اعتُمدت مصادر مختلفة من البيانات المطبوعة لتزويد الأنموذج بخطوط متنوعة، وأحجام خطوط متباينة، ومجالات لغوية ومعرفية مختلفة، وفيما يأتي توضيح توزيع البيانات الرئيسة التي تدرَّب عليها النظام:

اسم البيانات	المجال	عدد الخطوط	النسبة المئوية
Sotoor	الكتب	217216	٪12.76
WIPO	براءات الاختراع	114634	٪6.74
Book Titles	الكتب	174971	٪10.28
Magz	المجلات والصحف	1194948	٪70.22

جدول (3): توزيع بيانات تدريب أنموذج الخطوط العربية المطبوعة

ويستخرج النظام الخصائص (Features) من صُوَرِ الخطوط الرمادية (Grayscale) عن طريق تغيير الحجم إلى ارتفاع ثابت، وقلْبِ الصور حتى يتطابق ترتيب النصِّ مع ترتيب الأطر (Frames)، ليمثِّل كلُّ عمود في خريطة الخصائص إطارًا، وكلُّ مجموعة من الخصائص تمثِّل حرفًا، ويعتمد النظام أنموذج (seq2seq) في تدريب الأنموذج على تعلُّم العلاقة بين تسلسُل الإطارات البصرية (المدخل)، وتسلسُل الحروف (المخرج)، وكيفية تحويل الإطارات البصرية إلى نصٍّ.

ويعتمد النظام أيضًا على التحسينات (Augmentations) لتحسين أداء الأنموذج، وهذه بعض التحسينات البارزة:

- الهندسية: القص (Shear) والدوران.
- الألوان: تغيير تحسينات المقدمة والخلفية.
- إدخال صورة في الخلفية.
- الضوضاء (Noise)، من مثل (Salt and Pepper, Gaussain Noise).
- ضغط (JPEG).

وتُستعمل سياسة التحسين تلك في أثناء التدريب (Online)، وتطبق هذه التحسينات عشوائيًّا على الصور المدخَلة، وفيما يأتي عينة من هذه التحسينات في الشكل (3).

© إيسيسكو ICESCO 2024

شكل (3): عينات صور الخط المحسنة

يستعمل النظام دالة خسارة (RNN-T) المختصرة على ناتج المرمِّز ومفكِّ الترميز (Encoder-Decoder) المشترك، للعثور على أفضل مسار محتمل في أثناء التدريب، وقد دُرِّب الأنموذج في ثمان وعشرين جولة على جميع البيانات المذكورة أعلاه، بمعدل تعلُّم ابتدائي (3-e1)، والتسخين الأولى (Warmup) للأنموذج لأول ثلاثة آلاف خطوة، وتقليل معدل التعلُّم، وقد كانت نتائج الأنموذج على مجموعات الاختبار المستخرجة لكلِّ مصدر تدريب كما هي موضَّحة في الجدول (4).

اسم الاختبار	نسبة **WER** (المصححة)
Sotoor	٪8.59
WIPO	٪28.18
Books	٪6.53
Magz	٪9.45

جدول (4): نتائج أنموذج الخطوط العربية المطبوعة

ويُلحظ أن النتائج جيدة لجميع مجموعات الاختبار، باستثناء مجموعة (WIPO)، لأنها الأقل تمثيلًا، وتحتوي بعض التحويلات السياقية بسبب طبيعتها العلمية.

3. خوارزمية التجميع (Clustering Algorithm):

هي خطوة معالجة لاحقة لدمْجِ الخطوط في فقرات (K)، ثم دمْجِ تلك الفقرات في مناطق لإنشاء هيكل صفحة (OCR) المقروءة التي تمثِّل المخرج النهائي، وذلك بإجراءات هي:[1]

[1] See: Palaiahnakote Shivakumara, Partha Pratim Roy & Umapada Pal, "Multi-Oriented Scene Text Detection in Video Based on Clustering and Distance Metric", *Pattern Recognition 76*, (2018), pp. 37-54.

© إيسيسكو ICESCO 2024

- **تحليل المسافات بين الخطوط**: لتحديد ما إذا كانت هذه الخطوط تنتمي إلى الفقرة نفْسِها أو لا تنتمي.
- **تحليل حجم الخط**: لتحديد ما إذا كانت هناك تغييرات في حجْمِ الخط تدلُّ على بداية فقرة جديدة.
- **تحليل محاذاة الخطوط**: لتحديد ما إذا كانت الخطوط متساوية في المحاذاة، وتنتمي إلى الفقرة نفْسِها.
- **تحليل المحتوى**: في بعض الحالات المتقدمة، يمكن تحليل المحتوى اللغوي لتحديد العلاقات بين الجمل والفقرات.

وتنفِّذ خوارزمية التجميع بالإجراءات السابقة عمليتي دمْجٍ؛ إحداهما تدمج الخطوط معًا لتكوين فقرات، والأخرى تدمج فقرات لتكوين مناطق أو صفحة كاملة، ويمكن رؤية مخرجات التجميع في الشكل (4).

شكل (4): من اليسار إلى اليمين؛ مخرجات اكتشاف الخطوط، دمج الفقرات من الخطوط (المرحلة 1)، دمج المناطق من الفقرات (المرحلة 2)

مكوِّنات برنامج "سطور":

نوجز هنا أبرز مكوِّنات برنامج "سطور"، وأهم الإجراءات التي تُتَّبع في أثناء القراءة الضوئية للصورة:

1. واجهة المستخدم (User Interface - UI):

واجهة بسيطة سهلة الاستعمال، تتيح تحميل الصور، وتحديد إعدادات التحويل، ثم عرْضَ النتائج.

2. وحدة استيراد الصور (Image Acquisition):

تتيح تحميل الصور من مصادر مختلفة، من مثْلِ الماسحات الضوئية، والكاميرات، أو الملفات المخزنة محليًّا.

3. المعالجة الأولية للصورة (Image Preprocessing):

تحسين جودة الصور وإعدادها لعملية التعرُّف، ويشمل ذلك:

- إزالة الضوضاء.
- ضبط التباين والإضاءة.
- تصحيح الانحرافات والتشوهات.
- تحويل الصور إلى الأبيض والأسود (Binarization).

4. تقسيم الصورة (Segmentation):

تقسيم الصورة إلى أجزاء صغيرة تشمل الأحرف، والكلمات، والأسطر، لتسهيل عملية التعرُّف.

5. وحدة التعرُّف على الأحرف (Character Recognition Engine):

النواة الأساس لبرنامج (OCR)، حيث تُطبَّق الخوارزميات للتعرُّف على النصوص، ويشمل ذلك:

- التعرُّف على الأنماط (Pattern Recognition).
- خوارزميات التعلم الآلي (Machine Learning Algorithms).
- الشبكات العصبية العميقة (Deep Neural Networks).

© إيسيسكو ICESCO 2024

6. وحدة ما بعد المعالجة (Post-Processing):

تحسين النص المستخرج وتصويب الأخطاء، ويشمل ذلك:

- المعاجم لتصويب الأخطاء الإملائية.
- نماذج لغوية لتحسين دقة النص.

7. التصدير والتخزين (Export and Storage):

حفْظُ النصوص المستخرجة في تنسيقات مختلفة، من مثل (PDF, TXT, DOCX)، أو تخزينها في قواعد بيانات.

المخطوط العربي وتحديات التعرُّف عليه ضوئيًّا

يعتمد التعرُّف الآلي على النصوص المكتوبة (OCR) على تقنيات الذكاء الاصطناعي التي تحاول محاكاة قدرات العقل البشري في قراءته النصوص التي تشتمل عليها الصور، ونسْخِها إلى صيغة نصية رقمية مقروءة، ويمرُّ ذلك بخطوات وإجراءات تماثل ما يقوم به المحقِّق البشري في معالجة المخطوط، ونَسْخِه، والتغلُّب على كل ما يواجهه من تحديات، وبذلك يَضِحُ أن أغلب الصعوبات التي تواجه المحقِّق البشري في أثناء نسْخِ المخطوط هي نفْسُها التي تواجه التقنية في أثناء عمَلِها، مع وجود ما تختصُّ به التقنية مقارنة بجهْدِ المحقِّق البشري.

ويعنى هذا المبحث ببعض تلك التحديات التي رَصَدَها خبراء التحقيق في أثناء معالجتهم أنواع المخطوطات المختلفة، وذلك لأهميتها في تطوير التقنية، وقياس أدائها.

إن تحقيق النصوص علْمٌ وصناعةٌ، وله جانبان؛ "جانبٌ نظري علمي أُلِّفت فيه كُتُبٌ، وشُرِحت تجاربُ، مصحوبة ببعْضَ النماذج، محتوية على بعْضِ القواعد أو النصائح التي تعين الراغب في ممارسة هذه المهمة الصعبة الدقيقة، وجانبٌ آخر عملي فني، لا بدَّ فيه من الممارسة والتجارب الشخصية، حتى يتهيَّأ للمرء قدْرٌ من الخبرة والمهارة والقدرة على حلِّ الصعوبات العملية التي تتنوَّع من حالةٍ إلى حالةٍ، ولا تغني فيها القواعد أو النصائح المدونة"،[1] لذا تتطلَّب هذه المهنة في من يتصدَّى لها أن يكون مزوَّدًا بعلْمٍ وخبرةٍ يمكنانه من

[1] حسن الشافعي، "بعض صعوبات تحقيق المخطوطات العربية"، **مجلة كلية دار العلوم**، العدد (10)، (1980)، ص231.

© إيسيسكو ICESCO 2024

نقْلِ ما يريده المؤلِّف نقلاً أمينًا دقيقًا، وأن يضيئه بتعليقاته وتوجيهاته، ومما اشترطه العلماء في المحقِّق أن يكون مجيدًا اللغة العربية بفروعها ومستوياتها المختلفة وعصورها المتعاقبة، وأن يكون ملمًّا بالعلم أو الموضوع الذي يتناوله المخطوط، وأن يكون ملمًّا بالتحقيق علمًا وفنًّا،[1] وقد تتَّسع الخبرات التي يحتاجها المحقِّق من أجْلِ فحْصِ المخطوط، لتشمل معرفته وخبرته العملية "أنواع الورق، وأنواع الخط، وقواعد الكتابة، وطُرُقَ التجليد، ونحْوَها مما يعين على تقدير عُمُرِ النسخة إن لم يحدَّد تاريخ النسخ".[2]

وقد تنوَّعت التحديات التي تكتنف تحقيق المخطوط، وفَرَضَتْ على المحقِّق التزوُّد بالعلم والخبرة، وبذْلَ مزيد من الجهد في دراسة المخطوط من جميع جوانبه، وفيما يأتي إشارة إلى أهمِّ تلك التحديات التي يمكن أن تواجه المحقِّق:

1. التصحيف والتحريف:

يندرج تحْتَ هذه الظاهرة كلُّ اضطراب أو خلْطٍ في نقْطِ الحروف أو رسْمِها، ترتَّب عليه تبادُل الحروف فيما بيْنَها، ومن ثَمَّ تداخُل المعاني والتباسُها، وقد رجَّح محمود الطناحي تعريف التصحيف بأنه "تغيُّر في نقْطِ الحروف أو حركاتها، مع بقاء صورة الخط، كالذي تراه في كلمات مثْلِ (نَمَتْ) و(نِمْتُ)"،[3] وعرَّفَ التحريف بأنَّه "العدول بالشيء عن جهته... فقد يكون بالزيادة في الكلام، أو النقص منه، وقد يكون بتبديل بعْضِ كلماته، وقد يكون بحمْلِه على غيْرِ المراد منه".[4]

وتعدُّ "قضية التصحيف والتحريف من أخطر قضايا تحقيق النصوص، لأنَّها تتصل بسلامة النصِّ، وتأديته على الوجه الذي تَرَكَهُ عليه مؤلِّفه، وهي الغاية التي ليس وراءها غاية، من تحقيق النصوص وإذاعتها، وقد يُتسامح في بعْضِ جوانب التحقيق الأخرى، مع أهميتها،

---

[1] انظر: السابق نفسه.

[2] المرجع السابق، ص237.

[3] محمود محمد الطناحي، **مدخل إلى تاريخ نشر التراث العربي** (القاهرة: مكتبة الخانجي، 1984)، ص286.

[4] المرجع السابق، ص287.

© إيسيسكو ICESCO 2024

كتوثيق النُّقول، وتخريج الشواهد، وصُنْعِ الفهارس الفنية، ولكن أن يُترك اللفظ مصحفًا أو مُزالاً عن جهَتِه، فهذا مما لا يُتسامح فيه، ولا يُعفى عنه".[1]

وقد أرجع الأصفهاني التداخل بيْنَ حروف العربية إلى "أن الذي أبدع صُوَرَ حروفها لم يضَعْها على حكمة، ولا احتاط لمن يجيء بعْدَه، وذلك أنه وَضَعَ لخمسة أحرف صورة واحدة، وهي الباء، والتاء، والثاء، والياء، والنون، وكان وجْهُ الحكمة فيه أن يَضَعَ لكلِّ حرفٍ صورة مُباينة للأخرى، حتى يُؤمَن عليه التبديل"،[2] وكثيرًا ما أوقع هذا الالتباسُ النُّسَّاخَ في الخطأ سواء بالتحريف أم التصحيف أم التكرار أم الإضافة أم السقط،[3] فيُخطئ "الناسخ في قراءة ما هو مكتوب في الأصل، ويكتب غيْرَه، وهذا الجنس من التحريف لا تحصى أنواعه".[4]

ونظرًا إلى خطورة هذه الظاهرة وشيوعها، نَشَطَ علماء التراث في رَصْدِ الكلمات المشْكِلة التي يزيد فيها احتمال وقوع الخطأ عنْدَ نسْخِها، فقد ألَّف فيها العسكري (ت382هـ) كتابه "التصحيف والتحريف"، وشَرَحَ ما يَقَعُ فيه، ويقول في صدْرِه: "شرحْتُ في كتابي هذا الألفاظ والأسماء المشْكِلة التي تتشابه في صورة الخط، فيَقَعُ فيها التصحيف، ويدخلها التحريف"،[5] ومن المصنَّفات المهمَّة في هذا الباب كتاب "تصحيح التصحيف وتحرير التحريف"، للصفدي (ت764هـ)، فقد شكا في مستهلِّه من شيوع التصحيف والتحريف، فقال: "إن التصحيف والتحريف قلَّما سَلِمَ منهما كبير، أو نجا منهما ذو إتقان"،[6] وقد جَمَعَ الصفدي الكلمات التي يُحتمل أن يقع فيها التصحيف والتحريف، وصنَّفها وفْقَ احتمالاتها، فقد تأتي على:[7]

---

[1] المرجع السابق، ص285.

[2] الأصفهاني، **التنبيه على حدوث التصحيف**، تحقيق: محمد أسعد طلس (بيروت: دار صادر، ط2، 1992)، ص27.

[3] انظر: الشافعي، **بعض صعوبات تحقيق المخطوطات العربية**، ص241.

[4] برجستراسر، **أصول نقد النصوص ونشر الكتب**، إعداد وتقديم: محمد حمدي البكري (الرياض: دار المريخ، 1982)، ص81.

[5] العسكري، **التصحيف والتحريف وشرح ما يقع فيه** (القاهرة: مطبعة الظاهر، 1908)، ص3.

[6] الصفدي، **تصحيح التصحيف وتحرير التحريف**، تحقيق: السيد الشرقاوي (القاهرة: مكتبة الخانجي، ط1، 1987)، ص4.

[7] انظر: جلاط مريم، "التصحيف والتحريف في النصوص المحققة"، **المدونة**، 7(1)، (2020)، ص502.

© إيسيسكو ICESCO 2024

- حرفين، كما في (جابر، حائر).
- ثلاثة أحرف، كما في (يَغيبُ، يعتب، يعبث).
- أربعة أحرف، كما في (قلْبٌ، قلَب، قلْبَ، قُلِّب).
- خمسة أحرف، كما في (خبَرٌ، جبَر، خير، حبْر، حبّر، جبرَ).
- ستة أحرف، كما في (بشير، تُستر... إلخ).
- سبعة أحرف.
- ثمانية أحرف.

ونظرًا إلى غلبة وقوع النسَّاخ في مثْلِ هذا التصحيف، وخطورة ما قد يترتَّب عليه، فقد فَرَضَ ذلك على المحقِّق أن يتزوَّد بما يمكِّنه من أن يتفادى ذلك من العلم والخبرة، "فالاحتراس من التصحيف لا يُدرك إلا بعلمٍ غزيرٍ، وروايةٍ كثيرةٍ، وفهمٍ كبيرٍ"،[1] و"علاج هذه الظاهرة الخطيرة لا يكون إلا بمعرفةٍ دقيقةٍ بأسرار اللغة، وخصائص مفرداتها وتراكيبها، وتصرُّف هذه المفردات والتراكيب في كلام العرب، ثم إلمامٍ كاشفٍ بتاريخ هذه الأمة العربية، وأحوال رجالها وكُتُبِها ومصطلحات علومها، وكلِّ ما يمتُّ إليها بسَبَبٍ، وهذا لازمٌ لكلِّ من يشتغل بتراث الأمَّة".[2]

ومع كلِّ ذلك لم يسلم كبار المحقِّقين من الوقوع في هذه الظاهرة، فتسلَّلت منهم هفوات إلى ما حقَّقوه من أعمال، وقد أورد محمود الطناحي من تلك الهفوات في كتابه "مدخل إلى نشْرِ التراث العربي"، منها قوْلُه: "ومن أطرف ما رأيته من ذلك التصحيف الناشئ عن قُرْبِ الحروف وبُعْدِها، ما جاء في كتاب محقَّق، قال: (وأنشدَ شِعْرًا بَيَّنَ سُكْرَهُ)، بهذا الضبط الذي تسمعون، وقد تأمَّلت ذلك الكلام فلم أجد شيئًا، وإذا صحَّته: (وأنشدَ شِعْرَ ابْنِ سُكَّرَةَ)".[3]

---

[1] العسكري، **التصحيف والتحريف وشرح ما يقع فيه**، ص3.

[2] الطناحي، **مدخل إلى تاريخ نشر التراث العربي**، ص304.

[3] السابق نفسه.

© إيسيسكو ICESCO 2024

2. فرادة المؤلِّف أو النَّاسخ التي تميِّز المخطوط من غيْرِه، وإن ناظَرَهُ غيْرُه في الزمان أو المكان أو الموضوع:

قد يتميَّز المؤلِّف باستعمال معجمٍ لغويٍّ (كلمات وتراكيب) مختلفٍ عن معاصريه من المؤلِّفين، وكذلك النَّاسخ قد يُخالِف معاصريه في الخطِّ الذي ينسخ به، أو في طريقة استعماله، كأنَّ تكون الحروف عنده متقارِبة أو متداخِلة أو مزخرَفة على نحْوٍ يختلف عن غيْرِه من النُّسَّاخ؛ إذ "يختلف كلُّ خطٍّ حسْبَ كاتِبِه، فهناك من يُقارب بيْنَ حرْفِ الدال واللام، أو الغين والفاء، رَغْمَ الفرق الموجود بينها"،[1] وتتجلَّى تلك الفرادة في الرموز التي يستعملها وأشكالها، بالإضافة إلى تفاوُتِ النُّسَّاخ في مستوى الدقة، فبعْضُهم يسهو، فتسقط منه كلمة، أو جملة، أو سطر، فقد يسقط "سطْرٌ عند نسْخِ الكتاب، لأنَّ النَّاسخ بعْدَ إتمام السطر لا يبدأ بما بعْدَه، بل يجاوز سطرًا كاملاً، ويبتدئ بالثالث".[2]

ومما يدخل في إطار فرادة المؤلِّف أو النَّاسخ ما يقع من أحدهما أو كليهما من أخطاء نحويَّة، فهذا النوع من التحريف يَقَعُ من بعْضِ النُّسَّاخ إذا "لم ينتبهوا إلى ما هو مكتوب في النُّسخ، فكثيرًا ما أبدلوا الصحيح في الأصل بالدارج في لغَتِهم، فأبدلوا النصب والجزم بالرفع، وأبدلوا المؤنَّث بالمذكَّر، والفاء بالواو، إلى غيْرِ ذلك، وكان أكثر خطَئِهم في الأعداد، لأنَّ العادة كانت جارية على أن ينطقوا بالأعداد طبْقَ اللغة الدارجة، ولهذا السبب فإن النُّسخ التي لا خطأ فيها في الأعداد نادرة".[3]

ويترتب على تلك الفرادة أن يجتمع أحيانًا للمخطوط الواحد أكثرُ من نسخة متفاوتة في خصائصها ودقَّتها، ومتباينة في أكثر من موضع، وهو ما يقتضي من المحقِّق أن يقابل بيْنَها، وأن يوازن، ثم يرجِّح الأقرب إلى مراد المؤلِّف، وهذا يتطلَّب منه:

---

[1] المرجع السابق، ص62.

[2] برجستراسر، **أصول نقد النصوص ونشر الكتب**، ص24.

[3] المرجع السابق، ص84.

© إيسيسكو ICESCO 2024

- أن يرجع إلى "ترجمة المؤلِّف وسيرته وأخباره ومؤلَّفاته، وأن يتمرَّس في لغَتِه وأساليبه، فقد يضيء ذلك النصَّ، ويساعد على مقابلته بالقاموس اللغوي للمؤلِّف، فلكلِّ كاتبٍ أو مؤلِّفٍ مفرداته ومصطلحاته وأساليبه التي تتكرر لديه وتشيع في كتاباته".[1] – أن يتثبَّت من الكلمات الغامضة بأن "يقوم المحقِّق بالبحث عن صورتها في مواضع أخرى في صفحات المخطوط، فيختار أقرب صورة من صُوَرِ الكلمات الواضحة لهذه الكلمة الغامضة"،[2] وأن يتعرَّف "الصورة التي يستخدمها النَّاسخ لكلِّ حرْفٍ من الحروف في حالِ اتصالها أو انفصالها أو البدء بها أو الانتهاء بها، ومقارنة أحرف الكلمة الغامضة بصُوَرِ الأحرف عنْدَ الناسخ".[3]

3. توالي نسْخِ المخطوط:

تتباين الثقة في كتابة المخطوط ونسْخِه بالنظر إلى من قام بذلك، فأعلاها أن ينهض المؤلِّف نفْسُه بكتابته ونسْخِه، فهو أدرى بمقصوده منها، وأحرص على إخراجه على الوجه الصحيح، والأقل منه أن ينقل عن المؤلِّف أحدُ طلابه سماعًا أو استملاءً أو استنساخًا، مع أهمية أن يبذل المؤلِّف جهْدَه في المراجعة والتصحيح، ومن دوْنِ هاتين المرتبتين السابقتين مرتبة النُّسَّاخ، وهي أقلُّ منهما بكثير، فقد كان هؤلاء النُّسَّاخ يتكسبون معاشهم من نسْخِ الكتب، وكان كثير منهم يعنى بسرعة الانتهاء من النَّسخ، وحُسْنِ منظره أكثر من عنايته بدقة نقْلِه على نحْوِ ما أراد المؤلف،[4] حتى كثرت الشكاية من تساهُل هؤلاء النُّسَّاخ في عَمَلِهم، وكثرة ما يقعون فيه من تصحيف، يقول العسكري: "قد كان الناس فيما مضى يغلطون في اليسير دوْنَ الكثير، ويصحِّفون في الدقيق دوْنَ الجليل، لكثرة العلماء وعناية المتعلمين، فذَهَبَتِ العلماء وقلَّتِ العناية، فصار ما يصحِّفون أكثر مما يصحِّحون، وما يسقطون أكثر مما يضبطون".[5]

---

[1] المرجع السابق، ص252.

[2] السابق نفسه.

[3] السابق نفسه.

[4] انظر: المرجع السابق، ص17-18.

[5] العسكري، **التصحيف والتحريف وشرح ما يقع فيه**، ص5.

© إيسيسكو ICESCO 2024

ومع تعاقُبِ الأزمنة وتباعُدِ الأمصار، قد يُعاد نسْخُ المخطوط المنسوخ من قبْلُ على يَدِ ناسِخٍ جديد، فتتراكَمُ بذلك صُوَرٌ متنوعة من التصحيف والتحريف، وتزيد التعليقات والرموز، فتأتي تلك النسخة الأخيرة أكثر غموضًا، وأبعد عن مراد مؤلِّفها الأول، "ولما كان المؤلِّفون لا يطَّلعون على كلِّ ما يُنسخ من كُتُبِهم، فقد كَثُرَ عَدَدُ الإبرازات [الطبعات، النسخ]، وزاد احتمال وقوع الفرق بينها"،[1] "وهو ما يزيد من صعوبة فهْمِ النص بشكْلٍ دقيق، وتقديمه في صورته الأولى"،[2] ويزيد الأمر صعوبة أن تُصاب إحدى هذه النُّسَخ المتأخِّرة بطمْسِ معلومات ناسِخِها، أو أن تُترك سهوًا أو قصدًا منه، ثم نجِدُها تُزاحِم النُّسخ الأولى بين يَدَي المحقِّقين، وهو ما يفرض عليهم تحديًا كبيرًا لتحديد النسخة الأساس التي تُقابَل عليها سائر النُّسَخ، مع أهمية الأخذ في الحسبان أن "قِدَمَ التاريخ للنسخة ليس وحْدَه مبرِّرًا لتفضيلها، ولهذا نحتاج إلى حُجَجٍ أقوم وأثبت من تاريخ النسخة"،[3] ويدخل في هذه الحجج معرفة كاتب النُّسخة، أهو مؤلِّفها، أم تلميذ المؤلِّف، أم ناسِخٌ يتكسَّب من نساخته؟ وإذا كان الكاتب ممن يتكسَّبون من نساختهم، فما مدى علْمِه بموضوع المخطوط؟ وماذا يُعرف به من كفاءة وأمانة في النقل؟ وغيْرُ ذلك من مرجِّحاتٍ يمكن للمحقِّق أن يعتمدها.[4]

4. الخطُّ المكتوب به المخطوط:

تعود بعْضُ التحديات التي يجِدُها المحقِّق إلى "اختلاف أنواع الخطوط، وطُرُقِ الإعجام والتشكيل، وطُرُقِ كتابة الأرقام، وبعْضِ علامات الترقيم القديمة، وهذا بابٌ واسِعٌ لا بدَّ فيه من ثقافة نظرية عملية للتحقيق"،[5] وكذا يعود بعْضُ تلك التحديات إلى تبايُن خصائص الخطِّ العربي ما بين مشرقيٍّ ومغربيٍّ، وقديمٍ وحديثٍ، فلتبايُن الخطوط أثرٌ بالغٌ في وقوع

[1] برجستراسر، **أصول نقد النصوص ونشر الكتب**، ص27.

[2] حميد قريتلي، "صعوبات تحقيق المخطوط: قراءة الخطوط نموذجًا"، **مجلة مالك بن نبي للبحوث والدراسات**، 2(2)، (2020)، ص58.

[3] برجستراسر، **أصول نقد النصوص ونشر الكتب**، ص16.

[4] انظر: المرجع السابق، ص16-17.

[5] الشافعي، **بعض صعوبات تحقيق المخطوطات العربية**، ص241.

© إيسيسكو ICESCO 2024

التصحيف والتحريف في نُسخ المخطوط، فإذا كان "الكتاب قد كُتِبَ أولاً بالكوفي، ثم نُسِخَ بالخط النَّسخي، ثم بالمغربي، ثم أُعيدت كتابته بالنَّسخي، ثم كُتِبَ بالفارسي أو الرقعة التركي، فلا نهاية لاحتمال وقوع التحريف في مثْلِ هذا الكتاب، وأكثرُ ذلك يحدث عنْدَ النقل من خطٍّ لخطٍّ، وعنْدَ النَّسخ من أصلٍ قديمٍ؛ لأن الناسخ في هذه الحالات لا يعرف خطَّ الأصل معرفةً كافيةً في كثير من الأحيان".[1]

5. خوارج النص:

من التحديات التي تواجه المحقِّق في قراءة المخطوط، وتزيد من نسبة وقوع التصحيف والتحريف اختلاطُ الحواشي بمتْنِ المخطوط؛ إذ "تزخر بعْضُ المخطوطات العربية بكثيرٍ من الحواشي والتعليقات الجانبية، وقد تقتحم بعْضُ هذه التعليقات متْنَ المخطوط، فتدخل بيْنَ سطوره، أو بيْنَ جُمَلِه، فتختلط بالمتن، فلا يعرف القارئ إن كان الكلام كلُّه للمؤلِّف، أو أن بعْضَه للناسخ، وذلك أن بعضًا ممن يمتلكون المخطوط يضيفون إليه بعْضَ تعليقاتهم، للتوضيح أو الشرح أو التفسير، أو غيْرِ ذلك، وقد يفعل الناسخ مثْلَ ذلك، فيتدارك جملةً أو كلمةً سقطت في أثناء النَّسخ، أو قد يرغب في إيضاح شيء من النص، وقد يكون بعْضُ هذه الحواشي من المؤلِّف نفْسِه إذا كان المخطوط بخطِّه".[2]

6. حالة المخطوط:

سوء الحالة التي عليها المخطوط نتيجة سوء التخزين، أو تعرُّضه لما يُفسد أوراقه، أو يزيل أحباره؛ يمثِّل تحديًا كبيرًا أمام المحقِّق، وينبغي له أن يجتهد في تقدير ما ترتَّب على ذلك من سقْطٍ أو طمْسٍ، فإن "وجود طمْسٍ في المخطوط، أو تَلَفٍ نتيجة الرطوبة أو الإصلاح الجائر، أو غيْرِ ذلك؛ مما يقطع التواصل بين الجمل، ويضرُّ بالمعنى والسياق، وتَقَعُ مثْلُ هذه الصعوبة كثيرًا في المخطوطات كلما تقادَمَ تاريخ نسْخِها، أو لم تحتفظ بصورة مناسبة".[3]

---

[1] برجستراسر، **أصول نقد النصوص ونشر الكتب**، ص81.

[2] المرجع السابق، ص260.

[3] المرجع السابق، ص253.

© إيسيسكو ICESCO 2024

7. الخلل في ترتيب صفحات المخطوط:

من التحديات التي تواجه المحقِّق في نُسَخ مخطوطه "الخلل في ترتيب الصفحات، فكثير من الكتب القديمة جُمِعَتْ من ملازم أو أوراق مفردة يسهل وقوع الخلل في ترتيبها، وتتعقَّد هذه المشكلة عندما يأتي ناسِخٌ آخر، فينقل عن أصلٍ مختلِّ الترتيب، مع اختلاف في أوائل الصفحات ونهاياتها، فيصعب اكتشاف الخلل، ويحتاج لمهارة شديدة"،[1] وتلحق بهذه المشكلة مشكلة السقط، "أي النقص في ثنايا الكتاب، وأحيانًا يكون كلمة أو عبارة أو سطرًا أو عدة سطور، أو أكثر، ويفطن إليه المحقِّق بالسياق أو بعلامات معينة".[2]

الجزء التطبيقي

يتمثَّل الهدف من هذا الجزء التطبيقي في رصْدِ التحديات التي لم يستطع برنامج "سطور" – محل البحث – أن يتغلب عليها حتى اليوم، والمقارنة بيْنَه وبيْنَ برنامجين آخرين من أفضل البرامج التي تنافِسُه في مجال القراءة الضوئية للمخطوطات العربية، هما "زنكي" و"غوغل لنز".

وتأتي هذه المقارنة بهَدَفِ بيان تفاوُتِ البرامج الثلاثة في قدرتها على التعامل مع التحديات التي تواجه تحويل النصِّ من صيغة الصورة إلى صيغة النصِّ القابل للتحكُّم فيه، ويمكن الإفادة في رصْدِ تلك التحديات وتفسير أوجه القصور التي وَقَعَتْ فيها البرامج الثلاثة، مما ورد من تحديات أشار إليها المبحث السابق من البحث.

ولتحقيق ذلك الهدف عَقَدَ البحث تجربتين للتحقُّق من قدرة البرامج الثلاثة على التعامل مع تلك التحديات؛ أولى التجربتين جزئية، والثانية عامة، وفيما يأتي عرْضُ التجربتين ونتائجهما.

---

[1] الشافعي، **بعض صعوبات تحقيق المخطوطات العربية**، ص240.

[2] السابق نفسه.

© إيسيسكو ICESCO 2024

التجربة الأولى الجزئية:

قراءةٌ ضوئية لصفحة من مخطوط "كتاب السُّنن للإمام الشافعي رضي الله عنه"؛[1] أربَعَ قراءاتٍ؛ إحداها قراءة بشرية من خبير في التحقيق، والثلاث الأُخَرُ قراءات آليَّة من البرامج الثلاثة، ثم قُورنت كلُّ قراءة آليَّة بقراءة الخبير مقارنة بشريَّة، وتمثَّل معيار الصواب والدقة في هذه التجربة في مطابقة القراءات الأربع ما وَرَدَ في المخطوط، لأن الهدف من هذه الخطوة ليس التصحيح والتحقيق، بل قياس قدرة الأنظمة الثلاثة على نَسْخِ المخطوط على ما هو عليه بدقة، وفيما يأتي بيان إجراءات التجربة.

أولًا: صفحة المخطوط[2]

شكل (5): صورة صفحة المخطوط المختارة للتجربة الأولى

[1] المخطوط من مقتنيات مكتبة الجامع الأزهر، برقم (22374 – حديث)، وهو بعنوان "كتاب السنن للإمام الشافعي"، برواية أبي إبراهيم بن إسماعيل بن يحيى المزني، عنه رحمه الله، ثم رواية أبي جعفر أحمد بن محمد سلامة الأزدي الطحاوي، عنه رحمه الله، ثم رواية الشريف أبي القاسم الميمون بن حمزة بن الحسين الحسيني، رحمه الله، وقد كُتِبَ هذا المخطوط بقلم طه بن يوسف بن طه بن يوسف بن عامر، وأتمَّه في رجب (1311هـ).

[2] الصفحة المختارة برقم (62) من إجمالي عدد صفحات المخطوط البالغ (144) صفحة.

© إيسيسكو ICESCO 2024

ثانيًا: الخصائص الفنية للصفحة

استعمل الناسخ خطَّ الرقعة، وتميز بما يأتي:

- تسهيل الهمزة في أول الكلمة، سواء أكانت أول عَلَمٍ، كما في (اسيد)، أم أول الفعل، كما في (ادركت)، أو اسم، كما في (ابى)، أي (أبِي)، أو حرف، كما في (ان)، أي (أنَّ)، وقد تُسبق الهمزة بالفاء أو الواو، كما في (فاذنت، واخبرنى).
- إهمال وضع نقطتين تحت الياء المتطرفة (الياء الشامية)، كما في (واخبرنى، ابى، المزنى).
- الوقوع في السقط، فقد أغفل كلمة (فصليت)، وأضافها محقِّق الكتاب اعتمادًا على نسخة أخرى للمخطوط.
- زيادة بعْضِ الأحرف في الكلمة زيادةً أدت إلى تغيُّر صورة الكلمة، ومن ثمَّ معناها، كما وقع في كلمة (أنا) في قوله: "قال انا المزنى"، وصوابها: "قال إن المزنى".
- تداخُل بعض الحروف عند الناسخ لضيْقِ المساحة المتاحة لتكملة الكلمة، كما في (محذورة) في السطر الخامس، أو لمحاولة تجويد النسخ، وتوفير المساحة، كما في أغلب الكلمات التي وَقَعَ فيها خطأٌ في النسخ.
- إهمال كتابة ياء كلمة (شيء)، وكتابة الهمزة على ياء غير منقوطة (شئ).
- ضبْطُ بعْضِ الأحرف التي يرى الناسخ احتمالية الخلط فيها، كما في ضبْطِه تاء الفاعل في الكلمات (صَليتُ، قمتُ، فعلتَ، صليتَ)، أو ضبْطِ بعْضِ الأعلام، كما في (مُحَيْريز، رَوَّاد، جُبَيْر، يَزِيدَ، مُعَاوية)، وقد يضبط الكلمة ضبطًا تامًّا بنيةً وإعرابًا، كما في (لِرَسُولِ، بْنِ، يَزِيدَ، تَعُدْ)، وقد يأتي الضبط غير ذي جدوى في منْعِ التباس فهْمِ للكلمة، كما في (لِرَسُولِ، وسَلم، بالصَّلوة)، في حيْنِ تَرَكَ كلمات أُخرى كانت تحتاج إلى ضبْطٍ، ولا سيما الأعلام، كما في (عتاب، اسيد، جريج).

ثالثًا: قراءات المخطوط

1. قراءة الخبير البشري:

"لِرَسُولِ الله صَلى الله عليه وسَلم فقدمت على عتاب بْنِ اسيد عامل رسول الله صلى الله عليه وسَلم بمكة فاذنت معه بالصَّلوة عن امر رسول الله صلى الله عليه وسَلم قال ابن

© إيسيسكو ICESCO 2024

جريج: واخبرنى ذلك من ادركت من آل ابى محذورة فمن ادرك ابا محذورة على نحو مما اخبر ابن مُحَيْريز اخبرنا احمد قال انا المزنى قال اخبرنا الشافعى رحمه الله عن عبد المجيد عن عبد العزيز بن ابى رَوّاد قال: اخبرنا ابن جريج قال اخبرنى عمر بن عطاء بن ابى الخَوّاران نافع بن جُبَيْر ارسَله الى السَّائب بن يَزِيدَ بن اخت نمر يَسْاله عن شئ رآه من مُعَاوية في الصَّلوة فقال نعم صَليتُ مع معاوية الجمعة في المقصُورة فلما سلم قمتُ في مقامى فصليت فلما دخل ارسل الى فقال لا تَعُدْ لما فعلتَ اذا صَليتَ الجمعة فلا تصلها بصَلوة حتى تكلم او تخرج فان نبيَّ الله صَلى اللهُ عليه وسَلم امر بذلك لا توصل صَلوةً بِصَلوة حتى تخرج او تكلم".

2. قراءة برنامج "سطور":

"لرسول الله صلى الله عليه وسلم فقدمت على عتاب بن اسيد عاملو بمكة فاذنت معه بالصلوة عن او رسول الله صلى الله عليه وسلم قال ابن جريح واخبرني ذلك من ادركت من ال ابي محذوف فمن ادرك ابا محذورة على نحو مما اخبر ابن محير بيرز اخبرنا احمد قال انا المزنى قال اخبرنا الشافعى رحمه الله عن عبد المجيد عبد العزيز بن ابي رواد قال اخبرنا ابن جريح قال اخبرني عمر بن عطاء بن ابى الخوار ان نافع بن جبير ارسله الي السائب بن يزيد بن اخت تمر يساله عن شئ رآه من معاوية في الصلوة فقال نع صليت مع معاوية الجمعة فى المقصورة فلما سلمى قمت فى مقامى فلما دخل ارسل الى فقال لا تعد لما فعلت اذا صلت الجمعة فلا تصلها بصلوة حتى تكلم اوتخرج فان نبى الله صلى الله عليه وسلم امر بذلك لا نوصل صلوة بصلوة حتى تخرج وتكلم".

3. قراءة برنامج "زنكي":

"لرسول الله صلى الله عليه وسلم فقدمت على عتاب بن أسيد عامل رسول الله صلى الله عليه وسلم بملة فاذنت معه بالصلوة عن امر رسول اللاه صلى الله علياه وسلم قال ابن جزيج واخبرني ذلك من ادركت من ال ابي محدو فمن أررك ايا محدورة على نحو مما اخير ابن محيرين أخبرنا أحمد قال انا المزى قال أخبرنا الشافعي رحمه الله عن عيد المجيد عن عبار الغريز بن ابي رواد قال اخبرنا ابن خريج قال اخبرني عم بن عطاء ين الي الخواران نافع بن جبير ارسله الى

© إيسيسكو ICESCO 2024

الايث بن يزيدين أخت من لبيا له عن شيء راه من معاوية في الصلاة فقال نعم صليت مع معاونن الجمعر في المقصورة فلما سلم ثمت في مقامى فلما رخل ارسل الى فقال لا تقد لما فعلت إذا صلبت الجمعة فل نصبها بصلوة حتى تكلم أو تخرج فان بني الله صلى الله عليه وسلم امر بذلك لا تفصل صلوة بصالوة حتى تخرج او تكلم".

4. قراءة برنامج "غوغل لنز":

"لرسول الله صلى الله عليه وسلم فقدمت على عناب بنِ اسيد عامل رسول الله صلى الله عليه وسلم مكة فاذنت معه بالصلوة عزام رسول الله صلى الله عليه وسلم قال ابن جريج واخبرنى ذلك من ادركت من الرابى محدودة في ادرك ايا محذورة على نحو مما اخبر ابن مخيرين اخبرنا احمد قال انا المزى قال أخبرنا الشافعي رحمه الله عن عبد المجيد عن عبد العزيز بن ابى رواد قال أخبرنا ابن جريج قال اخبرتي عمر بن عطاء بن اے الخواران نافع بن جبير ارساء إلى السائب بن يزيد بن اخت غرينا له عن متى رآه من مُعَاوِية في الصلاة فقال نع صليت مع معاوية الجمعة في المقصورة فلما سلم قمت في مقامي فلما دخل ارسل الخفقال لا تعد لما فعلت إذا صليت الجمعة فلا تصلها بصلوة حتى تكلم او تخرج فان نبي الله صلى الله عليه وسلم امر بذلك لا توصل صلوة يصلوة حتى تخرج او تكلم".

رابعًا: تحليل القراءات الآلية

1. الخصائص العامة للقراءات الآلية:

(أ) تفاوت الالتزام بنقْلِ فرادة الناسخ في بعْضِ الظواهر الكتابية، وقد كانت لذلك صُوَرٌ، منها:

- التفاوت في تسهيل الهمزة أو تحقيقها، فالمخطوط التزم التسهيل، والتزم بذلك برنامجا "زنكي" و"غوغل لنز"، وخالفه برنامج "زنكي" في إثبات الهمزة في قراءة الكلمات (أسيد، أررك [أدرك]، أخبرنا، أحمد، أخت)، وأثبت برنامج "غوغل لنز" الهمزة في (أخبرنا، إلى).
- تصحيح بعْضِ الأخطاء الإملائية، فقد أسقط المخطوط الياء من كلمة (شيء)، وهي فيه (شئ)، ولكن حافظ برنامج "سطور" على صورة الكلمة الخطأ، وصحَّحها برنامج "زنكي".

© إيسيسكو ICESCO 2024

- التزم المخطوط تَرْكَ نقْطِ الياء المتطرفة (الياء الشامية)، وقد تفاوتت البرامج في الحفاظ على ذلك، فكلمة (واخبرنى) وردت في المخطوط غير منقوطة، وجاءت في برنامجي "سطور" و"زنكي" منقوطة (واخبرني)، وحافظ عليها برنامج "غوغل لنز" غيْرَ منقوطة، وكلمة (الشافعي) وَرَدَتْ في المخطوط غيْرَ منقوطة، وكذلك في برنامجي "زنكي" و"غوغل لنز"، ونقَطَها برنامج "سطور"، ووَرَدَتْ كلمة (أبي) في المخطوط غيْرَ منقوطة، وكذلك في برنامج "غوغل لنز"، ونقَطَها برنامجا "سطور" و"زنكي".
- تغيير بعْضِ اللازمات الكتابية العائدة إلى نوْعِ الخطِّ، من مثْلِ كلمة (صلاة)، فقد أثبتها المخطوط (صلوة) بكتابة الألف واوًا، إشارة إلى أصلها، كما هي في الخط المصحفي العثماني، أما البرامج الثلاثة فقد غيَّرتها إلى الكتابة العادية، وكذلك كلمة (أبى)، فقد جاءت الياء في المخطوط مكتوبة بطريقة عكسية، في حيْنِ كَتَبَها برنامجا "سطور" و"زنكي" بالطريقة التقليدية، وحافظ على طريقتها برنامج "غوغل لنز".

(ب) تفاوَتَتْ الأنظمة الثلاثة في دقة قراءة الصورة على نحْوٍ يطابق المخطوط، فقد بَلَغَ إجمالي كلمات النص (151) كلمة، أخطأ برنامج "سطور" في (13) كلمة، أي بنسبة دقة (91.56٪)، وأخطأ برنامج "زنكي" في (31) كلمة، أي بنسبة دقة (78.58٪)، وأخطأ برنامج "غوغل لنز" في (19) كلمة، أي بنسبة دقة (88.97٪).

2. تحليل أخطاء القراءات الآلية:

(أ) أربعة أخطاء تفرَّد بها برنامج "سطور"، هي:

- (عاملو)، بدمْجِ كلمتين، وقلْبِ الراء واوًا، والصواب (عامل رسول).
- (سلمى)، بزيادة ألِفٍ آخِرَ الكلمة، والصواب (سلم).
- (اوتخرج)، بدمْجِ كلمتين، والصواب (أو تخرج).
- (وتكلم) بدمْجِ كلمتين، وإسقاط حرْفٍ من الكلمة الأولى، والصواب (أو تكلم).

(ب) (18) خطأ تفرَّد بها برنامج "زنكي"، هي:

- (علياه)، بزيادة الألف، والصواب (عليه).
- (أررك)، بتصحيف الدال راءً، والصواب (أدرك).

© إيسيسكو ICESCO 2024

- (محدورة)، بتصحيف الذال دالاً، والصواب (محذورة).
- (اخير)، بتصحيف الباء ياءً، والصواب (أخبر).
- (عيد)، بتصحيف الباء ياءً، والصواب (عبد).
- (عبار)، بزيادة الألف، وتصحيف الدال راءً، والصواب (عبد).
- (الغزيز)، بتصحيف العين غينًا، والصواب (العزيز).
- (عم)، بإسقاط حرْفٍ من الكلمة، والصواب (عمر).
- (ين)، بتصحيف الباء ياءً، والصواب (بن).
- (الايث)، بإسقاط السين، وتصحيف الهمزة المنبورة ياءً، وتصحيف الباء ثاءً، والصواب (السائب).
- (يزيدين)، بدمْجِ كلمتين (يزيد، بن)، ثم تصحيف الباء ياءً، والصواب (يزيد بن).
- (معاونن)، بتصحيف الياء نونًا، والتاء المربوطة نونًا، والصواب (معاوية).
- (الجمعر)، بتصحيف التاء المربوطة راءً، والصواب (الجمعة).
- (ثمت)، بتصحيف القاف ثاءً، والصواب (قمت).
- (رخل)، بتصحيف الدال راءً، والصواب (دخل).
- (تقد)، بتصحيف العين قافًا، والصواب (تعد).
- (فل)، بإسقاط الألف، والصواب (فلا).
- (بني)، بقلْبِ حرْفٍ مكان حرْفٍ، والصواب (نبي).

(ج) ثمانية أخطاء تفرَّد بها برنامج "غوغل لنز"، هي:

- (عناب)، بتصحيف التاء نونًا، والصواب (عتاب).
- (عزام)، بدمْجِ كلمتين (عن، أمر).
- (الرابي)، بدمج كلمتين (آل، أبي).
- (في)، بإسقاط الميم، وتصحيف النون ياءً.
- (اخبرتي)، بتصحيف النون تاءً.

© إيسيسكو ICESCO 2024

- (إرساء)، بتصحيف اللام ألفًا، والهاء همزةً مفردة على السطر، والصواب (أرسله).
- (متى)، بتصحيف الشين ميمًا وتاءً، والهمزة المنبورة آخر الكلمة ألفًا لينة، والصواب كما جاء في المخطوط (شئ).
- (الخفقال)، بدمْجِ كلمتين (إلي، فقال)، مع تصحيف الياء المتطرفة غير المنقوطة خاءً.

(د) أربعة أخطاء اشترك فيها برنامجا "سطور" و"زنكي"، هي:

- (جريج)، في الموضع الأول من النصِّ، قرأها برنامج "سطور" (جريح)، بتصحيف الجيم الثانية حاءً، وقرأها برنامج "زنكي" (جزيج)، بتصحيف الراء زايًا.
- (جريج)، في الموضع الثاني من النصِّ، قرأها برنامج "سطور" (جريح)، بتصحيف الجيم الثاني حاءً، وقرأها برنامج "زنكي" (خريج)، بتصحيف الجيم الأولى خاءً.
- (صليت)، قرأها برنامج "سطور" (صلت)، بإسقاط الياء، وقرأها برنامج "زنكي" (صلبت)، بتصحيف الياء باءً.
- (توصل)، قرأها برنامج "سطور" (نوصل)، بتصحيف التاء نونًا، وقرأها برنامج "زنكي" (تفصل)، بتصحيف الواو فاءً.

(هـ) خطآن اشترك فيهما برنامجا "سطور" و"غوغل لنز"، هما:

- (امر)، قرأها برنامج "سطور" (او)، بإسقاط الميم، وتصحيف الراء واوًا، وقرأها برنامج "غوغل لنز" (عزام)، بدمْجِها مع الكلمة التي قبْلَها (عن)، وإسقاط الراء.
- (نعم)، قرأها البرنامجان (نع)، بإسقاط الميم.

(و) ستة أخطاء اشترك فيها برنامجا "زنكي" و"غوغل لنز"، هي:

- (بمكة)، قرأها برنامج "زنكي" (بملة)، بتصحيف الكاف لامًا، وقرأها برنامج "غوغل لنز" (مكة)، بإسقاط الباء.
- (أبا)، قرأها البرنامجان (أيا)، بتصحيف الياء باءً.
- (أبي)، قرأها برنامج "زنكي" (إلي)، بتصحيف الباء لامًا، وقرأها برنامج "غوغل لنز" (اے)، محافظًا على الكيفية التي جاءت بها في المخطوط.

© إيسيسكو ICESCO 2024

- (الخواران)، بدمْجِ البرنامجين كلمتين في كلمة واحدة، والصواب (الخوران إن).
- (يسأله)، بفصْلِ برنامج "زنكي" بين جزأي الكلمة، وجعْلِهما كلمتين منفصلتين (لبيا، له) بتصحيف الياء لامًا، وتصحيف السين باءً وياءً، وجعْلِ سائر الكلمة كلمةً مستقلةً (له)، وقد أسقط برنامج "غوغل لنز" الجزء الأول من الكلمة (يسا).
- (بصلوة)، قرأها برنامج "زنكي" (بصالوة)، بزيادة ألِفٍ في الكلمة، وقرأها برنامج "غوغل لنز" (يصلوة)، بتصحيف الباء ياءً.

(ز) ثلاثة أخطاء اشتركت فيها البرامج الثلاثة، هي:

- (محذورة)، في الموضع الأول من النصِّ، قرأها برنامج "سطور" (محذوف)، بتصحيف الراء والتاء المربوطة فاءً، وقرأها برنامج "زنكي" (محدو)، بتصحيف الذال دالاً، وإسقاط الراء والتاء، وقرأها برنامج "غوغل لنز" (محدودة)، بتصحيف الذال دالاً، وتصحيف الراء دالاً.
- (مُحَيْريز)، قسَّمها برنامج "سطور" إلى قسمين (محير، بيرز)، بزيادة باءٍ، وقرأها برنامج "زنكي" (محيرين)، بتصحيف الزاي نونًا، وقرأها برنامج "غوغل لنز" (مخيرين)، بتصحيف الحاء خاءً، وتصحيف الزاي نونًا.
- (نمر)، قرأها برنامج "سطور" (تمر)، بتصحيف النون تاءً، وقرأها برنامج "زنكي" (من)، بإسقاط النون، وبتصحيف الراء نونًا، وقرأها برنامج "غوغل لنز" (غرينا)، بتصحيف النون والميم غينًا (غر)، ثم دمْجِها في نصْفِ الكلمة اللاحقة (يساله)، بتصحيف السين نونًا.

التجربة الثانية العامة:

قراءة (15) صورة من (15) مخطوطًا، وذلك وفق الصوابط الآتية:

- الاختيار العشوائي للمخطوطات والصفحات منها.
- القراءة الضوئية لكل صورة بالبرامج الثلاثة، مع قراءة من خبير في التحقيق.
- مراعاة أن تكون القراءة البشرية لخبير التحقيق مطابقةً المخطوط.

© إيسيسكو ICESCO 2024

- مضاهاة قراءة كلِّ برنامج مع القراءة البشرية لخبير التحقيق، وعدِّ أيِّ مخالفة لقراءة الخبير خطأً يُحسب على البرنامج، ويقلِّل من نسبة دقَّته.
- المقارنة بين مستوى دقَّة البرامج الثلاثة لكلِّ مخطوط على حدة.

وكانت نتائج التجربة وفق ما يأتي:

م.	اسم المخطوط	نسبة الدقة		
		سطور	زنكي	غوغل لنز
1	آكام المرجان في أحكام الجان	٪79	٪76	٪64
2	رسالة أحكام الفرس	٪78	٪78	٪80
3	كتاب متن الألفية في عالم العربية	٪67	٪49	٪40
4	نخبة الفوائد من عقد القلائد في حل قيد الشرائد ونظم الفرائد	٪86	٪79	٪75
5	كتاب الرسالة في الفقه	٪49	٪37	٪18
6	الشرح الصغير على أقرب المسالك إلى مذهب مالك	٪70	٪55	٪63
7	الفوائد الباهرة في بيان حكم شوارع القاهرة	٪71	٪62	٪60
8	المنقذ من الضلال	٪61	٪50	٪58
9	منهاج الطالبين ومسالك الراغبين	٪79	٪62	٪66
10	بحوث في التشريع الإسلامي وأسانيد قانون الزواج والطلاق	٪36	٪28	٪39
11	تحفة الأطفال بشرح تحرير تنقيح اللباب	٪70	٪56	٪60
12	حاشية الجوهري على ابن قاسم	٪68	٪61	٪62
13	خزانة الفقه	٪84	٪80	٪48
14	شرح العمدة في الأحكام في الحديث عن خير الأنام	٪55	٪49	٪32
15	عطية الرحمن في إرصاد الجوامك والأطيان	٪79	٪54	٪74

جدول (5): عناوين المخطوطات المختارة للتجربة الثانية ونسب الدقة في قراءتها

النسب	سطور	زنكي	غوغل لنز
المعدل	٪69	٪56	٪55
نسبة الدقة العظمى	٪86	٪80	٪80
نسبة الدقة الدنيا	٪36	٪28	٪18
الانحراف المعياري	٪13	٪14	٪16

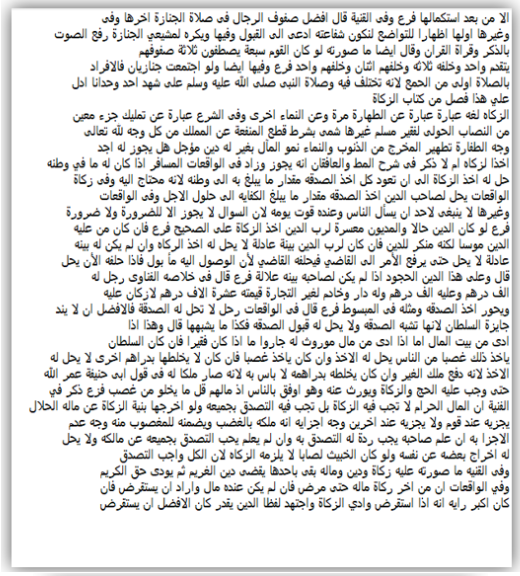
جدول (6): نسب الدقة لكل من البرامج الثلاثة

© إيسيسكو ICESCO 2024

وفيما يأتي عرْضٌ لقراءة برنامج "سطور" صفحات بعْضِ المخطوطات، متباينة في نِسَبِ دقَّتها:

1. نتائج متقدِّمة (85٪ وأكثر):

من مخطوط "نخبة الفوائد من عقد القلائد في حل قيد الشرائد ونظم الفرائد".



الا من بعد استكمالها فرع وفى القنية قال افضل صفوف الرجال فى صلاة الجنازة اخرها وفى
وغيرها اولها اظهارا للتواضع لئكون شفاعته ادعى الى القبول وفيها ويكره لمضيعي الجنازة رفع الصوت
بالذكر وقراة القران وقال ايضا ما صورته لو كان القوم سبعة يصطفون ثلاثة صفوفهم
يقتدم واحد وخلفه ثلاثة وخلفهم اثنان وخلفهم واحد فرع وفيها ايضا ولو اجتمعت جنازتان فالافراد
بالصلاة اولى من الجمع لانه تختلف فيه وصلاة النبى صلى الله عليه وسلم على شهد احد وحدانا اول
على هذا فصل من كتاب الزكاة
الزكاه لغه عبارة عبارة عن الطهارة مرة وعن النماء اخرى وفى الشرع عبارة عن تمليك جزء معين
من النصاب الحولى لفقير مسلم غيرها شمى بشرط قطع المنفعة عن المملك من كل وجه لله تعالى
وجه الطقارة تطهير المخرج من الذنوب والنماء نمو المال بقبر له دين مؤجل هل يجوز له اجد
اخذا لزكاه ام لا ذكر فى شرح المط والعافقان انه يجوز وزاد فى الواقعات المصافر اذا كان له ما فى وطنه
حل له اخذ الزكاة الى ان تعود كل اخذ الصدقه مقدار ما يبلغ به الى وطنه لانه محتاج اليه وفى زكاة
الواقعات يحل لصاحب الدين اخذ الصدقه مقدار ما يبلغ الكفايه الى حلول الاجل وفى الواقعات
وغيرها لا ينبغى لاحد ان يسأل الناس وعنده قوت يومه لان السوال لا يجوز الا للضرورة ولا ضرورة
فرع لو كان الدين حالا والمديون معسرة لرب الدين اخذ الزكاة على الصحيح فرع فان كان من عليه
الدين موسا لكنه منكر للدين فان كان لرب الدين بينة عادلة لا يحل له اخذ الزكاه وان لم يكن له بينه
عادلة لا يحل حتى يرفع الأمر الى القاضى فيحلفه القاضى لأن الوصول اليه ما يول فاذا حلفه الأن يحل
قال وعلى هذا الدين المجحود اذا لم يكن لصاحبه بينه عادلة فرع قال فى خلاصه الفتاوى رجل له
الف درهم وعليه الف درهم وله دار وخادم لغير التجارة قيمته عشرة الاف درهم لازكان عليه
ويجوز اخذ الصدقه ومثله فى المبسوط فرع قال فى الواقعات رجل لا تحل له الصدقه فالافضل ان لا يند
جابزة السلطان لانها تشبه الصدقه ولا يحل له قبول الصدقه فكذا ما يشبهها قال وهذا اذا
ادى من بيت المال اما اذا ادى من مال موروث له جاز وا ما اذا كان فقيرا فان كان السلطان
ياخذ ذلك غصبا من الناس يحل له الاخذ وان كان ياخذ غصبا فان كان لا يخلطها بدراهم اخرى لا يحل له
الاخذ لانه دفع ملك الغير وان كان يخلطه بدراهمه لا باس به لانه صار ملكا له فى قول ابى حنيفة عمر الله
حتى وجب عليه الحج والزكاة ويورث عنه وهو اوفق بالناس اذ مالهم قل ما يخلو من غصب فرع ذكر فى
القنية ان المال الحرام لا تجب فيه الزكاة بل تجب فيه التصدق بجميعه ولو اخرجها بنية الزكاة عن ماله الحلال
يجزيه عند قوم ولا يجزيه عند اخرين وجه اجزايه انه ملكه بالغضب وبضمنه للمغصوب منه وجه عدم
الاجزا به ان علم صاحبه يجب ردة له التصدق به وان لم يعلم يجب التصدق بجميعه عن مالكه ولا يحل
له اخراج بعضه عن نفسه ولو كان الخبيث لصبا لا يلزمه الزكاه لان الكل واجب التصدق
وفى القنيه ما صورته عليه زكاة ودين وماله يفى باحدهما يقضى دين القريم ثم يودى حق الكريم
وفى الواقعات ان من اخر ركاة ماله حتى مرض فان لم يكن عنده مال وارد ان يستقرض فان
كان اكبر رايه انه اذا استقرض وادى الزكاة واجتهد لفظا الدين يقدر كان الافضل ان يستقرض

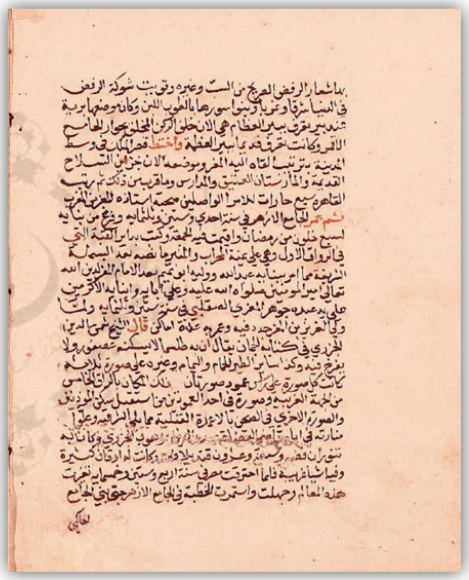
2. نتائج متوسطة (75–85٪):

من مخطوط "آكام المرجان في أحكام الجان".

عن معاذ وساقه الحافظ ابو نعيم عن الليث بن سعد عن عبد العرس
عن عمه عن معاذ كما رواه ابن ابي الدنيا حدثنا محمد بن الحسين
حدثنا ابو الوليد الكندى ثنا كثير بن عبد الله ابوهاشم
التامى قال دخلنا على ابي رجا العطاردى فسالناه هل
عندك علم من الجن ممن تابع النبى صلى الله عليه وسلم فتبسم
وقال اخبركم الذي رايت وبالذى سمعت كنا فى سفر حتى
اذا نزلنا على الما وضربنا اخبيتنا وذهبت اقبل فاذا انا بحية
دخلت الخبا وهى تضطرب فعهدت الى اداوتى فنضحت عليها
من الما فسكنت حتى اذن مرتن بالرجل فقلت لاصحابي
انتظروني حتى اعلم حال هذه الحية الى ما تصير فلما صلينا
العصر ماتت فعمدت الى عيدى فاخرجت منها خرقة بيضا
فلففتها وحفرت لها ودفنتها وسرنا بقية يومنا وليلتنا
حتى اذا اصبحنا ونزلنا على الما وضربنا اخبيتنا وذهبت اقل
فاذا انا باصوات سلام عليكم مرتين لا واحد ولا عشرة
الاماية واكثر فقلت من انتم قالوا نحن الجن بارك الله عليك
فما اصطعمت البنا ما لا تستطيع ان تجازيك قلت ما استطعت
اليكم قالوا ان الحية التى ماتت عندك كان ذلك اخر من بلغ
بندى لعمى بايع النبى صلى الله عليه وسلم من الجن قلت ورواه
الحافظ ابو نعيم فقال حدثنا عبد الله بن محمد بن جعفر انا احمد
ابن الحسن بن عبد الجبار ثنا بشر بن الوليد الكندى وقال
فيه كل واحد ولا عشرة الاماية ولا الفا اكثر من ذلك ن
قلت وقد تقدم رصل اسماهم ما ذكره ابن دريد شاصره
وماصره ومنشى * وماشى * والا خطب وساق الحافظ ابوهم
بسنده عن ابن اسحاق قال واسما وهم فيما ذكر لي حسا وسما
وشاصر وماصر وناصر واننا الازى واتين والاخصم

3. نتائج ضعيفة (65–75٪):

من مخطوط "الفوائد الباهرة في بيان حكم شوارع القاهرة".



بها شعار الرفض الصريح من الست وغيره وقويت شوكة الرفض
فى الدنيا شر قا وغربان بنو اسويرها بالطوب اللين وكان موضعها برية
وتدبير تعرف يسير العظام هي الان خلق الركن المخلق بجوار الجامع
الاقمر وكانت تعرق فدى ابي العظمة واختط قعر الملك فى وسط
المدينة ترتب القاه اليه المقر وبوضعه الان خزاين السلاح
القديمة والمارستان العتيق والمدارس وما قرب من ذلك ثم رتبه
القاهرة سبع حارات للام الواصلين صحبه استاذه المغربي المغرب
ثم عمر الجامع الازهر فى سنة احدى وستين وثلثمايه وفرغ من بنايه
لسبع خلون من رمضان واقيمت فيه الجمعة وكتب داير القبة التى
فى الرواق الاول وهى على عنة المحراب والمنبر ما نصه بعد البسملة
الشريفة مما امر شا به عبد الله وليه ابو تميم بعد الامام المغفر لدين الله
تعالى امير المومنين قبلوله الله عليه وعلى ابايه وابنا به الاكرمين
على يد عبده جوهر المغري الصقلي في سنوبتين وثلثمايه ولما
وكى العزيز بن المعز جدد فيه وعمر به عدة اماكن قال الشيخ شمن الدين
الجزري فى كتابه الحمان يقال ان به طلسما لا يسكنه عصفور ولا
يفرخ قيد وكذا ساير الطير الحمام واليمام وغيره على صورة للانه
كانت كل صورة على راس عمود صورتان ذلك المكان بالراق الخاص
من الجهة الغربية وصورة فى احد العمودين من استقبل سكن المؤذن
والصورة الاخرى فى الصحن بالاعمدة القتلية مما يلى الشرقيه وعلوا
منارته في ايام الضاعف والحضر ادعون الخرزى وكان له
تنور ان قصه وسمين وعرون قنديلا وكانت له اوقاف كثيرة
وفيه اشيا غريبة فانما احترقت مصر فى سنة اربع وستين وخمسما نه تغرت
هذه المعالم وجهلت واستمرت الخطبة فى الجامع الازهر حتى بنى الجامع

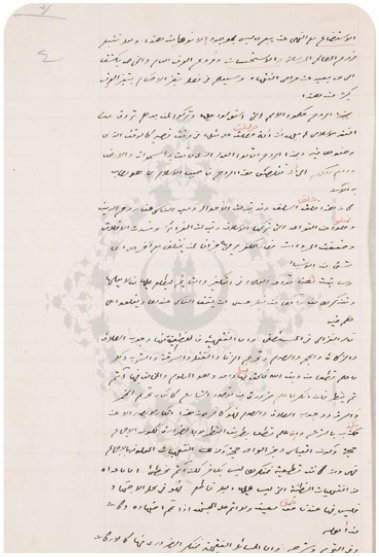
4. نتائج مرفوضة (0–50٪):

من مخطوط "كتاب الرسالة في الفقه"، بنسبة (49٪).

من مخطوط "مختصر خليل"، بنسبة (39٪).

© إيسيسكو ICESCO 2024

من مخطوط "بحوث في التشريع الإسلامي وأسانيد قانون الزواج والطلاق"، بنسبة (36٪).



خاتمة

إثر مناقشته جوانب مشكلته البحثية، والتجارب التي أُجريت على صفحات من عدة مخطوطات عربية، خَلَصَ هذا البحث إلى ما يأتي:

1. شهدت تقنية التعرُّف الضوئي على الحروف العربية المكتوبة بخطِّ اليد تطورًا ملحوظًا، يشجع على اعتمادها في المشروعات القومية لرقمنة تراثنا الممتدِّ من الوثائق والمخطوطات العربية، ويدعو إلى مواصلة الاستثمار فيه من حيْثُ إثراء قاعدة بياناته من التجارب الواقعية، وتمكين تقنيات الذكاء الاصطناعي من الإسهام في حلِّ مشكلاتها.
2. تفاوَتَت البرامج الثلاثة في مستوى دقَّتها في التعرُّف على المحتوى النصِّي لعينة التجربة، وفي طبيعة العثرات التي وَقَعَتْ فيها.
3. يمكن الإفادة من تطبيقات الذكاء الاصطناعي – من مثْلِ (Chatgpt4o) – في ترميم النصوص المقروءة آليًّا بما يحسِّن من نتائج تقنية (OCR)، ويزيد من فُرَصِ اعتمادها.
4. الإفادة من المخطوطات المحقَّقة في إثراء قاعدة بيانات الحروف المكتوبة بخطِّ اليد، وتصنيفها وفْقَ خصائصها المختلفة، من مثْلِ نوْعِ الخطِّ، والعصر الذي تنتمي إليه، والموضوع الذي تعالجُه، والناسخ، وغيْرِها من المتغيرات التي يمكن أن تفيد في ذلك، بالربط بين النصِّ المحقَّق ونُسَخِ مخطوطه، ليتعلَّم البرنامج منها، ويربط بيْنَ جميع أشكال الحرف وصورته الرقمية.

© إيسيسكو ICESCO 2024

المصادر والمراجع

الأصفهاني، **التنبيه على حدوث التصحيف**، تحقيق: محمد أسعد طلس (بيروت: دار صادر، ط2، 1992).

برجستراسر، **أصول نقد النصوص ونشر الكتب**، إعداد وتقديم: محمد حمدي البكري (الرياض: دار المريخ، 1982).

جميلة روقاب، "رقمنة الخط العربي ودوره في تحقيق المخطوطات: مشاريع تقنية وتحديات بشرية"، **المجلة الجزائرية للأبحاث والدراسات**، 2(8)، (2019).

حسن الشافعي، "بعض صعوبات تحقيق المخطوطات العربية"، **مجلة كلية دار العلوم**، العدد (10)، (1980).

حميد قريتلي، "صعوبات تحقيق المخطوط: قراءة الخطوط نموذجًا"، **مجلة مالك بن نبي للبحوث والدراسات**، 2(2)، (2020).

شذى عبد الرحيم خليفة الرضي، "التعرف على النصوص باستخدام تعلم الآلة"، **مجلة العلوم الإنسانية والطبيعية**، 3(11)، (2022).

صديق بسو، "خوارزميات وبرامج التعرف الآلي على الخط العربي"، **ملتقى التكنولوجيات الجديدة ودورها في صناعة اللغة العربية واستعمالها**، المجلس الأعلى للغة العربية، جامعة الإخوة منتوري، قسنطينة، 17 مايو 2017.

الصفدي، **تصحيح التصحيف وتحرير التحريف**، تحقيق: السيد الشرقاوي (القاهرة: مكتبة الخانجي، ط1، 1987).

العسكري، **التصحيف والتحريف وشرح ما يقع فيه** (القاهرة: مطبعة الظاهر، 1908).

محمود محمد الطناحي، **مدخل إلى تاريخ نشر التراث العربي** (القاهرة: مكتبة الخانجي، 1984).

مريم جلاط، "التصحيف والتحريف في النصوص المحققة"، **المدونة**، 7(1)، (2020).

مولاي محمد، "نحو التحقيق الإلكتروني (الرقمي) للمخطوطات: استخدام تطبيقات تكنولوجيا المعلومات في تحقيق المخطوطات لدى الباحثين الجزائريين في مجال المخطوطات"، **مجلة الحوار المتوسطي**، 5(1)، (2014).

© إيسيسكو ICESCO 2024

نسمة علي عبد الحميد، "استخدام تقنية (ICR) التعرف الذكي على الحروف المكتوبة بخط اليد في قراءة الوثائق والمخطوطات العربية، وانعكاس ذلك على مؤسسات حفظ التراث"، **الرزنامة**، العدد (20)، (2022).

هاشم فرحات، **دور تكنولوجيا المعلومات في ضبط المخطوطات العربية وإتاحتها: مراجعة علمية للإنتاج الفكري** (الرياض: جامعة الملك سعود، 2001).

برنامج "زنكي".

برنامج "سطور".

برنامج "غوغل لنز".

Connor Shorten & Taghi M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 6(1), (2019).

Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng & Xiangyang Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals", *IEEE Transactions on Multimedia*, 20(11), (2018).

Palaiahnakote Shivakumara, Partha Pratim Roy & Umapada Pal, "Multi-Oriented Scene Text Detection in Video Based on Clustering and Distance Metric", *Pattern Recognition 76*, (2018).

Raymond W. Smith, "An Overview of the Tesseract OCR Engine", *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Curitiba, Brazil, 2007.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao & Xiaoou Tang, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

Zhongheng Zhang, "Introduction to Machine Learning: K-Nearest Neighbors, Regression, and Beyond", *Annals of Translational Medicine*, 7(7), (2019).

© إيسيسكو ICESCO 2024

References

Al-Aṣfahānī, *al-Tanbīh ʻalā Ḥudūth al-Taṣḥīf,* Muḥammad Asʻad Ṭals (Ed.) (Beirut: Dār Ṣādir, 2nd Ed., 1992).

Al-ʻAskarī, *al-Taṣḥīf wal-Taḥrīf wa-Sharḥ mā Yaqaʻu fīh* (Cairo: Maṭbaʻah al-Ẓāhir, 1908).

Al-Ṣafadī, *Taṣḥīḥ al-Taṣḥīf wa-Taḥrīr al-Taḥrīf*, al-Sayyid al-Sharqāwī (Ed.) (Cairo: Maktabah al-Khānjī, 1st Ed., 1987).

Bergsträsser, *Uṣul Naqd al-Nuṣūṣ wa-Nashr al-Kutub*, Muḥammad Ḥamdī al-Bakrī (Ed.) (Riyadh: Dar al-Mirrīkh, 1982).

Connor Shorten & Taghi M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 6(1), (2019).

Ḥamīd Qrītlī, "Ṣuʻūbāt Taḥqīq al-Makhṭūt: Qirāʼah al-Khuṭūṭ Namudhajan", *Majallah Mālik bin Nabī lil-Buḥūth wal-Dirāsāt*, 2(2), (2020).

Ḥasan al-Shāfiʻī, "Baʻḍ Ṣuʻūbāt Taḥqīq al-Makhṭūṭāt al-ʻArabiyyah", *Majallah Kulliyyah Dar al-ʻUlūm*, (10), (1980).

Hāshim Farḥāt, *Dawr Tiknulūjyā al-Maʻlūmāt fī Ḍabṭ al-Makhṭūṭāt al-ʻArabiyyah wa-ʼItāḥatihā: Murājaʻah ʻIlmiyyah lil-ʼIntāj al-Fikrī* (Riyadh: King Saud University, 2001).

Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng & Xiangyang Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals", *IEEE Transactions on Multimedia*, 20(11), (2018).

Maḥmūd Muḥammad al-Ṭanāḥī, *Madkhal ʼilā Tārīkh Nashr al-Turāth al-ʻArabī* (Cairo: Maktabah al-Khānjī, 1984).

Maryam Jalāṭ, "al-Taṣḥīf wal-Taḥrīf fī al-Nuṣūṣ al-Muḥaqqaqah", *al-Mudawwanah*, 7(1), (2020).

Nasmah Alī Abdulḥamīd, "Istikhdām Tiqnyyah (ICR) al-Taʻarruf al-Thakī ʻalā al-Ḥurūf al-Maktubah bi-Khaṭ al-Yad fī Qirāʼah al-Wathāʼiq wal-Makhṭūṭāt al-ʻArabiyyah wa-INʻikās thālik ʻalā Muʼassasāt Ḥifẓ al-Turāth", *al-Ruznāmah*, No. (20), (2022).

ICESCO 2024 إيسيسكو ©

Palaiahnakote Shivakumara, Partha Pratim Roy & Umapada Pal, "Multi-Oriented Scene Text Detection in Video Based on Clustering and Distance Metric", *Pattern Recognition 76*, (2018).

Raymond W. Smith, "An Overview of the Tesseract OCR Engine", *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, Curitiba, Brazil, 2007.

Ṣaddīq Bisū, "Khuwārizmiyyah wa-Barnāmaj al-Ta'arruf al-'Ālī 'alā al-Khaṭṭ al-'Arabī", *Multaqā al-Tiknulujiyyāt al-Jadidah wa-Dawruha fī Sinā'ah al-Lughah al-'Arabiyyah wa-Isti'maliha*, al-Majlis al-'A'lā li-Lughah al-'Arabiyyah, Freres Mentouri University, Constantine, May 17, 2017.

Shatha Abdulraḥīm Khalīfah al-Raḍī, "al-Ta'arruf 'alā al-Nuṣūṣ bi-Istikhdām Ta'allum Al-'Ālah", *Majallah al-'Ulūm al-'Insāniyyah wal-Ṭabī'iyyah*, 3(11), 2022.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao & Xiaoou Tang, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

Zhongheng Zhang, "Introduction to Machine Learning: K-Nearest Neighbors, Regression, and Beyond", *Annals of Translational Medicine*, 7(7), (2019).

Google Lens.

Sotoor.

Zinki.

ICESCO 2024 إيسيسكو ©