تطوير مُمارسات إنشاء وإدارة السجلات الببليوجرافية لمصادر المعلومات بالمكتبات ومؤسسات المعلومات: دراسة تطبيقية لبناء منظومة تقنية ذكية

إعداد

د/ محمود محمد عبد العليم عبد الصمد

دكتوراه المكتبات والمعلومات

كلية الآداب – جامعة المنيا

mahmoudmm55432@gmail.com

- **المستخلص :**

تهدف هذه الدراسة إلى تقديم مُقترح عملي لمنظومة تقنية ذكية، مُستندة في بنائها على تطبيقات الذكاء الاصطناعي AI، وتقنيات التعلُّم الآلي ML، والتعلُّم العميق DL، والتي يُمكن توظيفها بالمكتبات ومراكز المعلومات للقيام بأحد أهم العمليات الفنية في مجال علوم المكتبات والمعلومات، وهي إدارة واستخراج المعلومات الببليوجرافية من الصور الرقمية لمصادر المعلومات، من أجل بناء سجل ببليوجرافي معياري لمصادر المعلومات الضخمة التي تقتنيها المكتبات.

وتتناول الدراسة في إطارها النظري التعريف بتقنيات الذكاء الاصطناعي AI، ويأتي الإطار التطبيقي ليُقدم مُقترحًا بمنظومة ذكية مُكونة من مجموعة من التقنيات المُتكاملة فيما بينها وظيفيًا لتحقيق مجموعة من الأهداف وهي : التعرُّف الذكي ICR على الصور الرقمية لأوعية ومصادر المعلومات، استخراج المعلومات الببليوجرافية

منها، تصدير المعلومات الببليوجرافية بملف إكسيل Excel طبقًا للحقول المعيارية المُتفق عليها، وتُستخدم الدراسة في سبيل ذلك المنهج التجريبي التطبيقي، لتنتهي الدراسة بتقديم مجموعة من النتائج والتوصيات التي من شأنها أن تُساهم في تطوير عمليات ومُمارسات التعامل مع مصادر المعلومات بالمكتبات ومراكز المعلومات.

- **الكلمات المُفتاحية :** البيانات الببليوجرافية – التعرُّف الذكي على الحروف – التعلم العميق – العمليات الفنية – السجلات الببليوجرافية.

0- التمهيد :

يشهد العصر الحالي ثورة رقمية في تطبيقات الذكاء الاصطناعي فأصبحت الحاجة مُلِحة لتبني تقنيات مُتطورة تعمل على تسريع وتنظيم عمليات مُعالجة مصادر المعلومات الضخمة التي تقتنيها المكتبات، بما يُعزز من قدرة المكتبات على تلبية احتياجات المُستفيدين بكفاءة أعلى.

تأتي هذه الدراسة لتُسلط الضوء على أهمية دمج تقنيات الذكاء الاصطناعي في إدارة المعلومات الببليوجرافية، وتسليط الضوء على كيفية استفادة المكتبات ومراكز المعلومات من تقنيات الذكاء الاصطناعي لتحسين كفاءة عمليات الفهرسة والتكشيف، ولتحقيق ذلك تستخدم الدراسة المنهج التجريبي التحليلي، ليُقدم مُقترحًا بمنظومة ذكية مُكونة من مجموعة من التقنيات المُتكاملة فيما بينها وظيفيًا لتحقيق مجموعة من الأهداف وهي : التعرُّف الذكي ICR على الصور الرقمية لأوعية ومصادر المعلومات، استخراج المعلومات الببليوجرافية منها، تصدير المعلومات الببليوجرافية بملف إكسيل Excel طبقًا للحقول المعيارية المُتفق عليها، وفي الختام جاءت الدراسة بمجموعة من النتائج والتوصيات [1].

1- أهمية الدراسة :

تُواجه المكتبات ومؤسسات المعلومات خاصة الكبيرة منها مثل [المكتبات الوطنية، الجامعية، العامة] تحديات أساسية تتعلَّق بإدارة المعلومات الببليوجرافية لمُقتنياتها من حيث الإنشاء والضبط، أبرزها تحديات الوقت والجهد المُستغرق في عمليات التكشيف والفهرسة الوصفية المطلوبة لإدخال البيانات، وبناء السجلات الببليوجرافية اللازمة لضبط مُقتنياتها، ومن هُنا جاءت هذه الدراسة كنقطة تحول هامة في مجال إدارة المعلومات الببليوجرافية بالمكتبات ومراكز المعلومات، لما لها من تأثيرات إيجابية مُتُعددة، تُحسِن من إدارة المعلومات في المكتبات ومراكز المعلومات، بإدخال تقنيات مُتقدمة تعمل على تعزيز الأداء والدقة والكفاءة، من خلال :

- تُقدم الدراسة نموذج تطبيقي للذكاء الاصطناعي في المكتبات – تطبيق تكنولوجي جديد – قادرًا على تطوير نُظم معلوماتية مُتقدمة في المُستقبل، ويستفيد من بنيته الباحثون والمُمارسون في مجال إدارة المعلومات بالمكتبات ومؤسسات المعلومات.
- تُقدم الدراسة مجموعة من التوصيات التي تُساهم في تحسين المُمارسات الحالية للفهرسة الوصفية، وتوجيه الجهود المُستقبلية نحو تحسين إدارة المعلومات الببليوجرافية.

2- أهداف الدراسة :

تهدف هذه الدراسة إلى تقديم مُقترح عملي لمنظومة تقنية ذكية، مُستندة في بنائها على تطبيقات الذكاء الاصطناعي AI، وتقنيات التعلُّم الآلي ML، والتعلُّم العميق DL، والتي يُمكن توظيفها بالمكتبات ومراكز المعلومات للقيام بأحد أهم العمليات الفنية في مجال علوم المكتبات والمعلومات، وهي إدارة واستخراج المعلومات الببليوجرافية من الصور الرقمية لمصادر المعلومات، من أجل بناء سجل ببليوجرافي معياري لمصادر المعلومات الضخمة التي تقتنيها المكتبات، وتحتاج إلى التعرُّف السريع على محتوياتها، ومضامينها، ليُستخدم هذا السجل فيما بعد في عمليات التكشيف أو المُقابلة بما هو موجود لدى المكتبة من مُقتنيات سابقة، أو كخطوة تمهيدية لعمليات الفهرسة الوصفية لمقتنيات المكتبات.

3- مشكلة الدراسة :

تتَّضح مشكلة الدراسة من خلال تعرُّضها لتحديات أساسية تتعلَّق بإدارة المعلومات الببليوجرافية في المكتبات ومراكز المعلومات، خاصة الكبيرة منها مثل [المكتبات الوطنية، الجامعية، العامة]، وهي تحديات أساسية تتعلَّق بإدارة المعلومات الببليوجرافية لمُقتنياتها من حيث الإنشاء والضبط، أبرزها تحديات الوقت والجهد المُستغرق في عمليات التكشيف والفهرسة الوصفية المطلوبة لإدخال البيانات، وبناء السجلات الببليوجرافية اللازمة لضبط مُقتنياتها.

4- تساؤلات الدراسة :

تسعى الدراسة الحالية للإجابة على التساؤلات الرئيسية التالية :

- كيف يُمكن توظيف تقنيات الذكاء الاصطناعي لبناء منظومة تقنية قادرة على التعرُّف بدقة على المصادر الرقمية واستخراج بياناتها الببليوجرافية وفقًا لمعايير محددة؟
- كيف يُمكن أن تُساهم هذه المنظومة في تحسين مُمارسات بناء السجلات الببليوجرافية لكمية كبيرة ومتنوعة من المصادر الرقمية ؟

5- حدود الدراسة :

1- **الحدود الزمنية :** أجريت الدراسة في الفترة من شهر يناير حتى ديسمبر لعام 2024م.

2- **الحدود الموضوعية :** تناولت الدراسة موضوع توظيف تقنيات الذكاء الاصطناعي لبناء منظومة تقنية قادرة على إنشاء وإدارة المعلومات الببليوجرافية من مصادر المعلومات الرقمية.

6- منهج الدراسة وأدواته :

- **منهج البحث التجريبي التحليلي :** وقد استخدمت الدراسة في هذا السبيل عددًا من الأدوات :

1. المصادر المعلوماتية ذات الصلة – مع عينة من الصور الرقمية لمصادر المعلومات.
2. التعامل مع مُحركات التعرُّف الذكي على النصوص، ومنها : Tesseract OCR, Easy OCR, Paddle OCR, Zinki OCR, OCR-D, Google Cloud Vision OCR، وهي من أبرز المُحركات التي يُمكن استخدامها ودمجها بالمنصة لخمة أغراض المنظومة، والمنصة الببليوجرافية المُقترحة لمرونتها واعتمادها على تقنيات التعلُّم العميق.
3. بيئة فيجوال استديو كود Visual Studio Code – منصة ريبليت Replit.
4. لغة البرمجة بايثون Python، ومكتباتها، ومنها:
   - مكتبة pytesseract, OpenCV, Pandas, Farasa, spacy .
   - تقنيات التعلُّم الآلي ML، والعميق DL، والشبكات العصبية من نمط LSTM.
   - مجموعة من الكلمات المُفتاحية المُنتقاة .
   - تصميم مكتبات برمجية بلغة البايثون ذاتية التحديث، للاستفادة منها بالمشروع.

7- الدراسات السابقة :

تمَّ تحديد استراتيجية البحث في الإنتاج الفكري العربي، والأجنبي الذي تناول موضوع "توظيف التقنيات الحديثة والذكية في استخراج وإدارة المعلومات الببليوجرافية من مصادر المعلومات بالمكتبات ومراكز المعلومات" بتحديد الكلمات المفتاحية : [المُعالجة الفنية – المُعالجة الذكية – الذكاء الاصطناعي – التعلُّم الآلي – البيانات الببليوجرافية]، وطبقًا للنتائج المُسترجعة انتقى الباحث أقربُها إلى موضوع الدراسة، من خلال البحث داخل المصادر التالية :

- بنك المعرفة بقواعد البيانات المُتاحة به العربية والإنجليزية : URL https://www.ekb.eg
- اتحاد المكتبات الجامعية المصرية : www.eulc.edu.eg : URL
- محركات البحث الإلكترونية المُستخدمة :

- الباحث العلمي Google Scholar : URL https://scholar.google.ae :
- الباحث الدلالي Semantic Scholar : : URL www.semanticscholar.org

فكانت الدراسات كالتالي :

- **الدراسات العربية :**

تناولت دراسة **(أحمد، محمد حسين 2021)** موضوع استخراج البيانات الببليوجرافية من مصادر المعلومات خاصة المواد النصية التي تتضمن [الكتب والمقالات العلمية]، حيث قدَّم البحث توضيح لماهية الذكاء الاصطناعي وتطبيقاته، والتعريف بماهية الفهرسة الوصفية، وتوضيح دور الناشر في عملية إنشاء التسجيلات الببليوجرافية، والاستفادة من إمكانيات التعلم الآلي في استخراج البيانات الببليوجرافية من مصادر المعلومات النصية، وانتهت الدراسة بتقديم نموذج مُقترح يُمثل إطار عام لاستخراج البيانات الببليوجرافية من مصادر المعلومات – النصية – العربية، مُستند على نظام Grobid, Germine machine learning library ، مع عرض لبنية النموذج المُقترح لاستخراج البيانات الببليوجرافية.

تم الخروج بعدد من النتائج والتوصيات أهمها أن الفهرسة الوصفية عملية تكرارية لابد من التفكير للحد منها، وأنه بوجد تنوع في مُخططات الكُتب مما يُمثل تحديًا في تدريب النماذج، ومن أهم توصياتها أن تتعاون مؤسسات النشر في إمداد النموذج مجموعات البيانات Dataset، وتوحيد مُخططات layouts للكُتب، وأن يتم عمل هاكاثون Hacathon على هوامش المؤتمرات لتشجيع الابتكار في توظيف تقيات الذكاء الاصطناعي في مجال علوم المكتبات والمعلومات [2].

وجاءت دراسة **(الهادي، محمد محمد 2023)** وضَّحت مفهوم المكتبة وخدمات المعلومات الذكية، وميَّزت بين أنواع المكتبات الذكية المُتشابهة، استخدمت الدراسة المنهج الوصفي التحليلي لمُراجعة الأدبيات المنشورة، لتحديد النظريات المُناسبة لتفسير أبعاد المكتبات الذكية، وتحديد خصائصها، وقد قدمت الدراسة مُقترح بنهج شامل من أجل بناء وتطوير المكتبات الذكية في إطار ثلاثة أبعاد رئيسية تتمثل في التكنولوجيا، الخدمة، البشر، كما وضَّحت الدراسة ضرورة تطبيق التكنولوجيا الحديثة مثل إنترنت الأشياء، الذكاء الاصطناعي، أدوات التنقيب عن البيانات، الحوسبة السحابية [3].

- **الدراسات الأجنبية :**

تهدف دراسة **(Tkaczyk, D., et al, 2021)** إلى تقييم ومقارنة عشرة أدوات لاستخراج البيانات الببليوجرافية من مصادر المعلومات وهي : Anystyle – Parser، Biblio، CERMINE، Citation، Citation –Parser، GROBID، ParsCit، PDFSSA4MET، Reference Tagger، Science Parse، تمت مقارنة الأدوات في نسختين : النسخة الجاهزة من الأداة، ونسخة تمت إعادة تدريبها لتناسب بيانات المشروع المُحددة، وتوصلت الدراسة إلى عدد من النتائج، وهي : الأداة الأفضل من حيث الأداء الجاهز هي GROBID، تليها CERMINE ثم أداة ParsCit، وأن الأدوات القائمة على التعلم الآلي تحقق دقة متوسطة مُشابهة لتلك التي تعتمد على القواعد أو التعبيرات العادية، لكن أدوات التعلم الآلي تُحقق استرجاعًا أعلى ثلاث مرات، تُظهر الدراسة أن تحسين النماذج لتناسب البيانات المحددة يُعزز الجودة، والنسخ المُعاد تدريبها من أدوات تحليل المراجع أفضل من الجاهزة [4].

وتناولت دراسة **(Kumar S, Vijay. Sheshadri, Kn 2019)** تطبيقات الذكاء الاصطناعي في المكتبات الأكاديمية، كيفية استخدام تكنولوجيا الذكاء الاصطناعي لتحسين خدمات المكتبات الأكاديمية، وقد هدفت الدراسة إلى استكشاف تطبيقات الذكاء الاصطناعي مثل الأنظمة الخبيرة، الشبكات العصبية الاصطناعية، مُعالجة الصور، مُعالجة اللغة الطبيعية، التعرُّف على الكلام، والروبوتات، اعتمدت الدراسة في منهجها على استعراض الأدبيات الحالية، وتفحص التطبيقات المختلفة للذكاء الاصطناعي في المكتبات، كما استعرضت كيف يمكن دمج تقنيات مُتُعددة لتقديم خدمات مكتبية أفضل، وتم تقديم أمثلة وتوضيحات حول كيفية تنفيذ هذه التقنيات لتحسين جودة الخدمة، وتوصلت الدراسة إلى مجموعة من النتائج من أهمها أن تقنيات الذكاء الاصطناعي مثل أنظمة التوصية، والمساعدين الافتراضيين يمكن أن تساعد في تحسين تجربة المستخدم، وجعل الخدمات أكثر تخصيصًا وفعالية، وأن الذكاء الاصطناعي يمكن أن يسهم في أتمتة مهام مثل فهرسة الكتب، وإدارة الاستعارات، مما يوفر الوقت ويزيد من كفاءة العمليات [5].

وقدمت دراسة **(Myanak & et al 2016)** إطار عمل تقني مُقترح لبرنامج تقني يُسمى (OCR++) وهو نظام مفتوح المصدر، مُصمم بلغة برمجة البايثون Python، وباستخدام نموذج Conditional random fields : CRF، يهدف إطار العمل إلى تحسين استخراج المعلومات من المقالات العلمية، تناول الإطار ثلاث فئات رئيسية من المعلومات : البيانات الوصفية (مثل العنوان، أسماء المؤلفين، التبعية، والبريد الإلكتروني)، والهيكل (مثل رؤوس موضوعات الأقسام، والنصوص الأساسية، رؤوس موضوعات الجداول والأشكال، الروابط والهوامش)، والببليوجرافيا

(مثل حالات الاستشهاد والمراجع). ولتحقيق أهداف الدراسة، قام الباحث بتحليل مجموعة متنوعة من المقالات العلمية باللغة الإنجليزية لفهم الأنماط الكتابية العامة، تم تقييم أداء الإطار باستخدام مجموعة من المقالات العلمية، وقورنت نتائجه مع الأدوات الحالية الرائدة، أظهرت التقييمات أن OCR++ يتفوق بشكل كبير على الأدوات الأخرى في استخراج المعلومات الهيكلية، مع تحقيق تحسين ملحوظ في الدقة (زيادة بنسبة حوالي 50%) ووقت المعالجة (تحسين بنسبة حوالي 52%)، بالإضافة إلى ذلك، تم تقديم حالتين جديدتين لاستخدام النظام في استخراج الروابط إلى مجموعات البيانات العامة من المنشورات، مما يعزز إجراءات المُعالجة الفنية في المكتبات الرقمية، ويمكن تصدير نتائج الإطار إلى مستندات مشفرة بتنسيق TEI [6].

وجاءت دراسة **(Tkaczyk, Dominika, et al 2015)** بدراسة وتحليل نظام CERMINE كحل شامل مفتوح المصدر لاستخراج البيانات الوصفية المُنظمة من المقالات العلمية الرقمية، يعتمد النظام على سير عمل معياري بفضل هيكله التقني باستخدام لغة البرمجة جافا JAVA ومكتباتها، كما يسمح بإمكانية توسيع الهيكلية المستقبلية، تناولت الدراسة تحليل بنية النظام المُكونة من تقنيات التعلم الآلي الخاضعة للإشراف وغير الخاضعة للإشراف، كما تضمنت الدراسة مُقارنة شاملة لنظام CERMINE مع حلول مُماثلة أخرى، ويصف منهجية التقييم، ويعرض النتائج المُحققة، يُتاح نظام CERMINE بموجب ترخيص مفتوح المصدر [7]، ويمكن الوصول إليه عبر الرابط : http://cermine.ceon.pl

8- الإطار النظري :

9- الذكاء الاصطناعي Artificial Intelligence :

1.10. ماهية الذكاء الاصطناعي Artificial Intelligence :

عرَّف قاموس جامعة كامبريدج البريطانية " الذكاء الاصطناعي Artificial Intelligence" بأنه المجال المعني بدراسة كيفية انتاج برامج حاسوبية وآلات تقنية تتمتع ببعض صفات العقل البشري، مثل القدرة على تفسير وإنتاج اللغة بطريقة تبدو بشرية، والتعرُّف على الصور أو إنشائها ، وحل المشكلات، والتعلم من البيانات الموردة لهم [8]، وقد ظهر مفهوم الذكاء الاصطناعي عندما قدم عالم الرياضيات والحاسوب والمنطق بجامعة مانشستر آلان ماتيسون تورنغ Alan Mathison Turing في عام 1950 م ورقته البحثية التي بعنوان "آلات الحوسبة والذكاء"، صاغ تورنغ لأول مرة مُصطلح الذكاء الاصطناعي Artificial Intelligence" وعرضه كمفهوم نظري وفلسفي.

2.10. البنية الطبقية لتقنيات الذكاء الاصطناعي :

تتكون تقنيات الذكاء الاصطناعي من مجموعة من الطبقات، وهي [9]:

- الطبقة الأولى : [طبقة البيانات] وهي طبقة تجميع البيانات، من مصادر مُتُعددة، وتجهيزها لتطبيقات الذكاء الاصطناعي، حيث تُوفر هذه الطبقة البنية التحتية الأساسية لتدريب نماذج الذكاء الاصطناعي، ويتم الاستفادة من هذه

البيانات باستخدام تقنيات مُختلفة أهمها : تعلّم الآلة، التعلم العميق، ومعالجة اللغة الطبيعية، والتعرُّف على الصور.

- الطبقة الثانية : [أطر تعلّم الآلة وطبقة الخوارزمية] وهي الطبقة المعنية بأداء الوظائف التي تقوم بمُعالجة البيانات، وتجسيدها بصورة تقنية في نماذج تدريب تقنية، ومن أمثلة هذه الأطر Tensor Flow –PyTorch –scikit–learn.
- الطبقة الثالثة : [طبقة النموذج] تتكون هذه الطبقة من [بنية النموذج – مُعلمات النموذج – أداة تحسين النموذج].
- الطبقة الرابعة : [طبقة التطبيق] يُقصد بها الواجهة التفاعلية التي تُمثل حلقة الوصل بين التقنية، والمُستخدمين لهذه التقنية، تسمح طبقة التطبيق للمستخدمين النهائيين بالتفاعل مع الأنظمة القائمة على تقنيات الذكاء الاصطناعي بكُل مُشتملاتها.

10- المكتبات ومؤسسات المعلومات في عصر الذكاء الاصطناعي :

تُعرف المكتبات ومؤسسات المعلومات بأنها مؤسسات علمية وثقافية تهدف إلى جمع وتنظيم واسترجاع وبث مصادر المعلومات بكُل أشكالها، ثُم تيسير وصول الباحثين والمُستفيدين إلى هذه المصادر بأسرع وقت، وبأقل جُهد، وبأكبر دقة مُمكنة [10]، وتتجه مكتبات ومراكز المعلومات نحو تبني تقنيات مُتقدمة، واستخدام تقنيات الذكاء الاصطناعي لتطوير مُمارسات عمليات المُعالجة الفنية أو المعلوماتية التي تقوم بها، لتحسين خدماتها، وتوفير معلومات ذات جودة عالية بشكل أكثر كفاءة وفعالية للمستفيدين.

1.11. أهمية المكتبات ومؤسسات المعلومات في عصر الذكاء الاصطناعي :

تزداد أهمية المكتبات ومؤسسات المعلومات في عصر الذكاء الاصطناعي، ويرجع ذلك لأهمية وقيمة مُقتنياتها، والعمليات المعلوماتية التي تتم بها، والخدمات التي تُقدمها لمُجتمعاتها، ففي مقالةٍ نُشرت حول أهمية البيانات في عصر الذكاء الاصطناعي بيَّنت ظهور تحدٍ تواجهه تقنيات الذكاء الاصطناعي، وهي مُعضلة انتهاء البيانات، وهي تلك المُعضلة التي تتمثل في تسارع وتيرة مُعالجة وتحليل المعلومات بالقدر الذي يفوق عملية نتاج المعلومات نفسها في العالم، لتجد فيها نماذج الذكاء الاصطناعي نفسها قد استهلكت واستوعبت كافة المعلومات المُتاحة، وحاجتها إلى معلومات أُخرى لعمليات التدريب والتحسين المُستقبلية.

ومن جانب آخر وضَّحت الدراسات أن من أهم وأجود المعلومات التي يُمكن أن تُستخدم لتدريب نماذج الذكاء الاصطناعي عليها للقيام بمهامها بجودة عالية، هي تلك المعلومات التي تُمثل الكُتب، والمقالات، والأبحاث، والدراسات التي كتبها أشخاص مُتخصصون في مجالاتهم [11]، تُوضح هذه الفكرة قيمة المُقتنيات التي تقتنيها المكتبات ومؤسسات المعلومات في عصر الذكاء الاصطناعي.

2.11. المُعالجة الفنية والتكشيف بالمكتبات ومراكز المعلومات:

1.3.10. ماهية المُعالجة الفنية :

تُعرفها الدراسة تعريفًا إجرائيًا بأنها مجموعة من العمليات والمعالجات التي تهدف إلى جمع المعلومات Acquisition وتنظيمها Organization، ووصفها ببليوجرافيًا Description، وفهرستها Cataloging، وجعلها مُتاحة للمُستفيدين Preservation and Retrieval بشكل فعّال، وهي تُعد مظلة تضم بين طياتها العمليات الإجرائية المعنية بتهيئة المعلومات، ومصادرها للاستعمال والاستفادة منها بكافة السُبل التي تُمكن من تعظيم الاستفادة من المحتوى المعلوماتي بكافة أشكاله، وصوره، وتُودي إلى تلبية الاحتياجات المعرفية، والبحثية للمُستفيدين والمُجتمع.

2.3.10. ماهية السجلات الببليوجرافية :

تُعد السجلات الببليوجرافية في المكتبات ومؤسسات المعلومات الأداة أساسية لإدارة المعلومات وتوثيق مصادر المعلومات، وهي تُساهم في توفير الوصول الفعال إلى المعرفة، والبحث الأكاديمي والعلمي، حيثُ تُشير السجلات الببليوجرافية إلى "سجلات أو قوائم تحتوي على معلومات مُفصلة عن المصادر المنشورة والمواد المعرفية المُختلفة، هذه السجلات تُستخدم في المكتبات، ومراكز المعلومات والأبحاث لتنظيم، وتوثيق الموارد المعرفية المتاحة، وتسهيل عمليات البحث والوصول إلى مصادر المعلومات ومُحتواها"، حيثُ تشمل السجلات الببليوجرافية عادةً معلومات مثل: العنوان، المؤلف، المؤلف المُشارك، المجلة أو الناشر، التاريخ، المكان: المكان الذي تم فيه النشر أو الإصدار، ملخص المحتوى والموضوع الرئيسي للمادة، الكلمات المفتاحية التكشيفية التي تُساعد في التصنيف، والعثور على المادة أثناء عمليات البحث.

3.3.10. الخطوات الإجرائية لمراحل إعداد السجلات الببليوجرافية :

تشمل عدة خطوات أساسية لضمان جودة وفعالية السجلات، تتمثل هذه الخطوات في:

1. تحديد المصادر والمواد : والتي سيتم توثيقها في السجلات الببليوجرافية.
2. جمع المعلومات الأساسية : لكل مصدر معرفي.
3. وصف وفهرسة المصادر : باستخدام المعايير المُعتمدة للوصف الببليوجرافي.
4. تنظيم وتصنيف السجلات بتنظيم السجلات، وتصنيفها بحسب نظام التصنيف المعمول به.
5. إدخال البيانات : إلى نظام قاعدة البيانات الخاص بالسجلات الببليوجرافية.

6. مراجعة وتدقيق : للتأكد من دقة واكتمال المعلومات المُدخلة.
7. إنشاء الفهارس والقوائم الببليوجرافية : للوصول إلى المصادر المعرفية بشكلٍ مُرتب ومُنظم.

باستخدام هذه الخطوات الإجرائية تضمن المكتبات ومؤسسات المعلومات توفير مصادر معرفية مُرتبة ومُنظمة بشكلٍ فعال، مما يُسهل على المُستفيدين استخدامها.

4.3.10. السجل الببليوجرافي الذكي (مدخل لعمليات الفهرسة والتحليل الموضوعي):

السجل الببليوجرافي يشكل جُزءًا أساسيًا من عمليات الفهرسة والتحليل الموضوعي، حيث يوفر معلومات دقيقة وشاملة عن كل مصدر معرفي داخل المكتبة أو مؤسسة المعلومات، مما يُسهل من استخدام هذه المعلومات في عمليات الفهرسة والتحليل الموضوعي.

يُعرِّف الباحث عملية الفهرسة والتحليل الموضوعي بأنها [عملية الإعداد الفني لأوعية ومصادر المعلومات من كُتب، ودوريات، ومواد سمعية، وبصرية، وغيرها من مصادر المعلومات، بهدف جعلها في متناول المُستفيدين بأيسر السُبل، وبأقل وقت وجُهد مُمكن]، فهي أداة للضبط الببليوجرافي، واسترجاع المعلومات، وتقييم المجموعات وفقًا لموضوعاتها، وإن نجاح أي مكتبة أو مؤسسة معلومات على مدى قدرتها على وصف وفهرسة مُقتنياتها، بصورة علمية قعَّالة، ويتوقف ذلك على مدى اتباعها للمعايير العالمية، واستخدامها للطُرق التقنية المعنية بإعداد الفهارس، وقوائم الضبط الببليوجرافي بصورة ذكية.

11- الإطار التطبيقي :

12- التعريف بمشروع المنظومة الذكية لإنشاء السجلات الببليوجرافية وأهدافه :

المشروع عبارة عن بناء وتصميم منظومة ذكية لإنشاء السجلات الببليوجرافية بأسلوب مُبتكر يعتمد على تقنيات الذكاء الاصطناعي، والتعلُّم الآلي ML، والتعلم العميق DL، والشبكات العصبية الاصطناعية ذات الذاكرة طويلة المدى LSTM، مع مجموعة من الأدوات والبرمجيات التقنية المُتقدمة مفتوحة المصدر Open Source، وذلك لتحقيق الأهداف التالية :

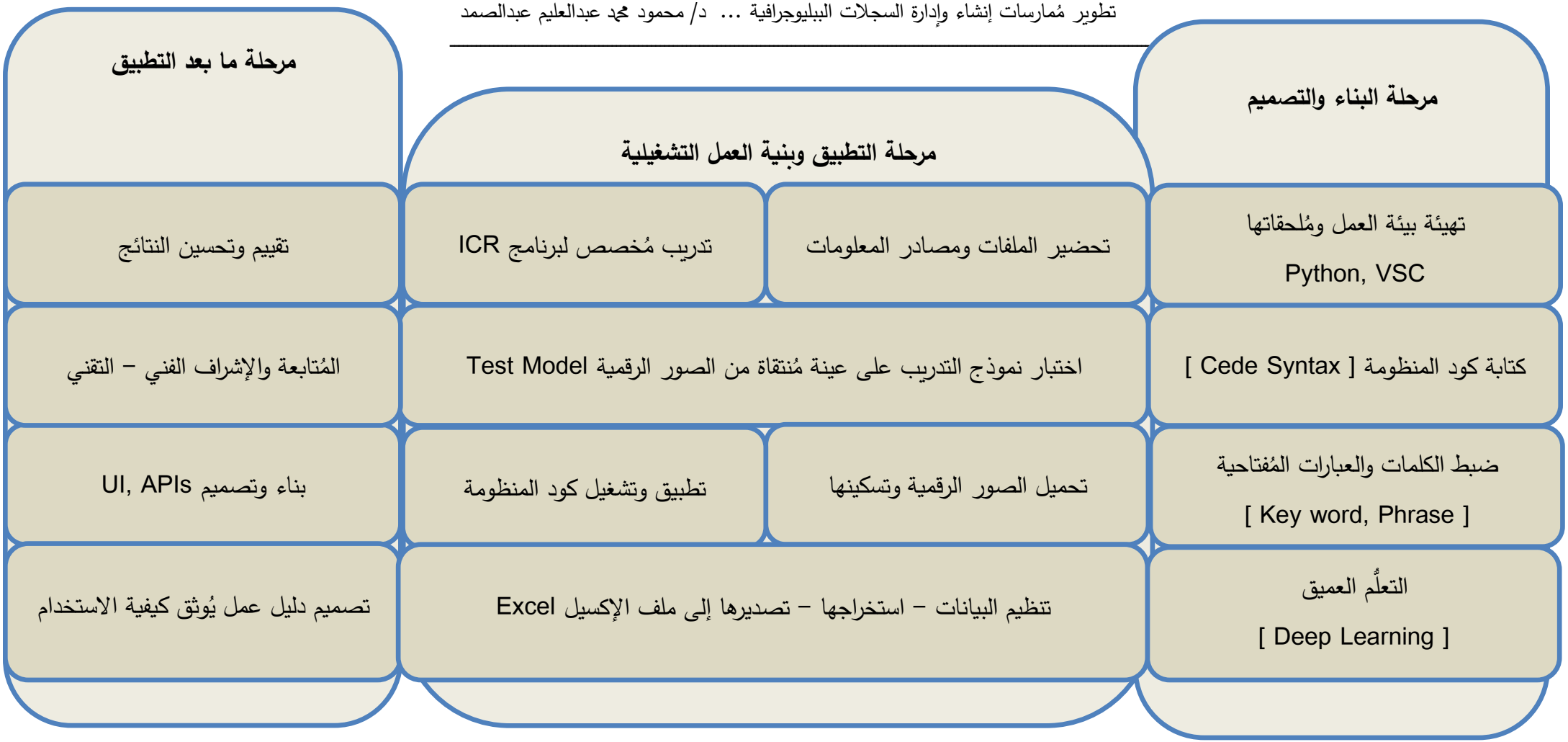
- **الأهداف العلمية والتطبيقية :**

1. تطبيق تقنيات الذكاء الاصطناعي في المكتبات لتحسين إدارة المعلومات الببليوجرافية.
2. نمو المفاهيم والتجارب العملية حول تأثير التقنيات الذكية على عمليات الفهرسة والتكشيف.
3. استخدام المنهج التجريبي التحليلي لتقييم فعالية المنظومة لبناء السجلات الببليوجرافية.
4. تطوير نموذج تطبيقي للذكاء الاصطناعي في المكتبات لتطوير نظم معلوماتية متقدمة.
5. تصميم مكتبة برمجية بلغة البايثون Python معنية بأسماء المؤلفين، وجهات النشر.

6. دمج المكتبات المُنشئة بالمنظومة لاستخدامها في استخراج البيانات الببليوجرافية.
7. توظيف تقنيات التعلم العميق Deep Learning في تحسين جودة عمل المنظومة.
8. تطوير نموذج مُتخصص للتعرف الضوئي على النص العربي يخدم أهداف المنظومة.
9. إتاحة الفرصة للفنيين، والباحثين للاستفادة من أحدث التطورات التقنية في هذا المجال.

13- منهجية تصميم المشروع – مراحل تطبيقه :

تعتمد المنظومة في منهجية عملها على استيعاب الصور الرقمية لمصادر المعلومات، ومُعالجتها، والتعرُّف الضوئي على محتواها، ثم تحليلها، باستخدام أدوات وتطبيقات الذكاء الاصطناعي وتقنيات التعلُّم الآلي ML، والتعلم العميق DL، بهدف استخراج البيانات الببليوجرافية بصورة ذكية، ثُم تصدير هذه البيانات في قوالب مُنظمة، لتسهيل إدارتها كسجل ببليوجرافي في أعمال إدارة المعلومات بالمكتبات ومؤسسات المعلومات، يُوضح الشكل التالي مراحل بناء المنظومة، والعمل عليها، كما يلي :



شكل 1 منهجية تصميم مشروع المنصة الذكية لإنشاء وإدارة السجلات الببليوجرافية ومراحل تطبيق المشروع
[الشكل من إعداد الباحث]

1.12 مرحلة البناء والتصميم :

تختص هذه المرحلة بتعريف المُتطلبات الأساسية لبناء المنظومة المُقترحة، (بالتعريف بماهية البرمجيات، والنُظم المُستخدمة، ولماذا تم اختيارها، وكيف سيتم التعامل معها)، كما يلي :

1.1.12 تهيئة بيئة العمل ومُلحقاتها [البنية التقنية للمشروع] :

استندت البنية التقنية للمنصة المُقترحة على عدة تقنيات أو برمجيات مُقترحة، وهي :

1- استخدم المشروع بيئة العمل فيجوال استوديو كود Visual Studio Code : VSCوذلك لأن بيئة العمل مُتاحة للاستخدام الكامل، والمجاني، ودعم العديد من اللغات البرمجية، مع إمكانية تطويع البرمجية لتتلاءم مع فكرة المشروع [12].

2- منصة ريبليت Replit – Build software faster وهي أحد أبرز المنصات السحابية لإنشاء وكتابة المشاريع البرمجية في بيئة مُتكاملة، تم استخدامها في المشروع لاختبار الكود والتحقق من صحته [13].

3- تقنية التعرُّف الضوئي على الحروف OCR، وقد تم التعامل مع عدة مُحركات مفتوحة المصدر، وتجارية، مثل Paddle OCR, OCR-

D, Tesseract OCR,, Easy OCR, Google Cloud Vision OCR لدعمهم للعديد من اللغات ومنها اللغة العربية، ومرونتهما الكافية للتخصيص، والتطوير طبقًا لخصائص المشروع، والمنصة المُقترحة، وطبيعة المحتوى والعمليات التي تتم بداخلها [14].

4- لغة البرمجة بايثون Python لكتابة كود المنظومة [15]، واستخدام مكتباتها لمعالجة وإدارة البيانات مثل : OpenCV, pandas, Farasa, Pytesseract، مع استخدام لغة البايثون لتصميم مكتبات مُتخصصة، وتطوير نموذج مُخصص للتعرف الضوئي على النصوص العربية.

5- قاعدة بيانات محلية استخدمها الباحث على حاسوب إلكتروني مُخصص للمشروع.

- **ضبط مسارات التعرُّف على بيئة العمل :**

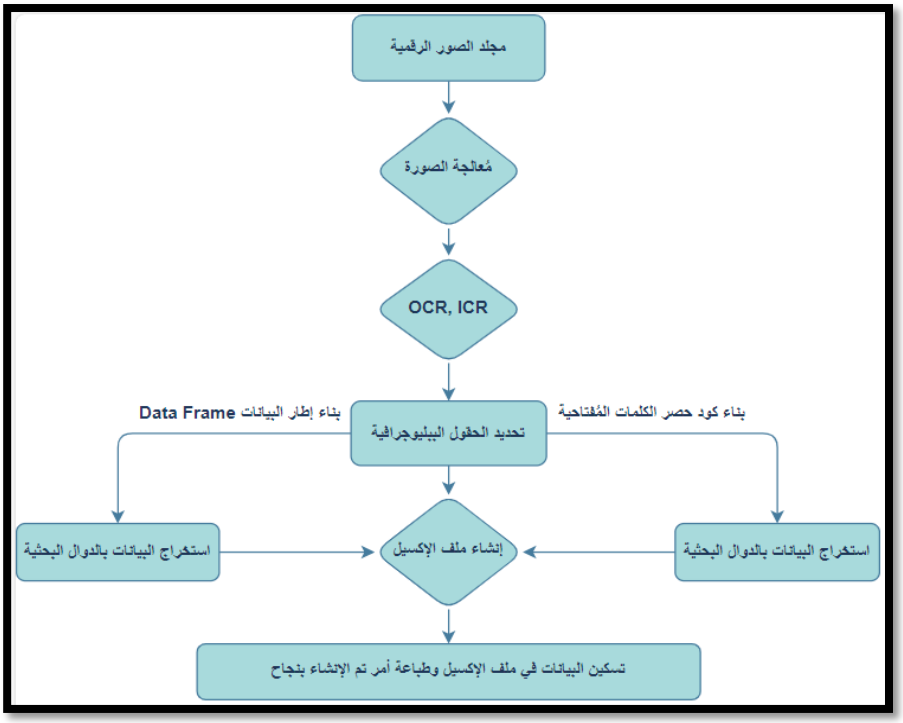
وهذه الخُطوة تتم طبقًا لمُحرك التعرُّف الذكي الذي سيتم تطبيقه في مرحلة التنفيذ، ومعرفة نظام التشغيل المُستخدم بالجهاز، فبعد تثبيت برمجيات Python, Tesseract على نظام Windows، يلزم ضبط متغيرات البيئة في النظام Environment Variables لاستخدام الأدوات من أي مكان في النظام، كما يلي [16]:

- التحقق من عملية تثبيت البرامج من موجه الأوامر Command Prompt بكتابة `cmd`
- كتابة الأوامر التالية : `pip --` ,`python, Tesseract – version` والتأكد من أنهما يعرضان الإصدار الصحيح.
- الوصول إلى مُتغيرات البيئة على النظام "Environment Variables".
- في "System variables" يتم الوصول إلى Path لإضافة مسار تثبيت البرامج.

إذا كانت الأوامر تعمل بشكل صحيح وتعرض الإصدارات المُناسبة، فهذا يعني أن إعداد متغيرات البيئة قد تم بنجاح.

2.1.12 كتابة كود المنظومة [Cede Syntax] :

لتصميم كود برمجي لمشروع تسبقه مرحلة من التخطيط، والتفكير المُسبق للتنفيذ، تبدأ ببناء مخطط رسومي Diagram لأهم الخطوات المرحلية لمسار المشروع المُستهدف، ليتم بعد ذلك تحويل هذا الشكل الرسومي إلى خطوات إجرائية أكثر تفصيلا في شكل رسومي يُعبر عن الخطوات بلغة تقترب من لغة الآلة Machine language، ولغة البرمجة programming language، تمهيدًا لقيام المُبرمج بتصميم كود المشروع، وضبط معامله، وتطويره، ويُوضح الشكل التالي مخططًا توضيحيًا Diagram لمراحل عمل المشروع موضوع الدراسة [17]:



شكل 2 مخططًا توضيحيًا Diagram لمراحل عمل المنظومة الذكية لاستخراج وإدارة المعلومات الببليوجرافية

[الشكل من إعداد الباحث]

تُمثل هذه الخطوة المُنطلق لتمثيل الخطوات التقنية المُستخدمة في بناء الكود، وتصميمه، ويُمكن بياناها من خلال الخطوات التالية :

- التأكد من تنصيب المُلحقات البرمجية وتحديد مساراتها بالنظام.
- تحديد المكتبات المطلوب استيرادها [تحديد اسم المكتبة – تحديد كود الاستيراد].

- تصميم المكتبات المطلوبة : العناوين – الناشرين – الشخصيات – تصميم الكود.
- بناء كود المشروع :
  - كود استيراد المكتبات.
  - كود ضبط مسارات المُلحقات البرمجية.
  - كود إعداد إطار البيانات طبقًا للحقول الببليوجرافية لمصادر المعلومات.
  - كود حصر الكلمات المُفتاحية طبقًا للحقول الببليوجرافية لمصادر المعلومات.
  - كود ضبط مسار الصور الرقمية لمصادر المعلومات.
  - كود معالجة الصور باستخدام أداة OpenCV لمُعالجة الصور وتحسين جودتها.
  - كود قراءة الصور وتحويلها من صورة إلى نص مقروء بواسطة الحاسب الآلي.
  - كود الدوال البحثية، والوظيفية، واستخراج البيانات الببليوجرافية :
    - دالة البحث عن العنوان واستخراجه.
    - دالة البحث عن المؤلف واستخراجه.
    - دالة البحث عن الناشر واستخراجه.
    - دالة البحث عن مكان النشر واستخراجه.
    - دالة البحث عن تاريخ النشر واستخراجه.

- دالة البحث عن رقم الطبعة واستخراجها.
- دالة البحث عن الرقم الدولي للكُتب أو الدوريات واستخراجه.

- كود تنفيذ الدوال الوظيفية، وتسكين البيانات المُستخرجة في قاموس البيانات.
- كود تحويل البيانات من إطار البيانات إلى حقول البيانات المُنظمة.
- كود تصدير البيانات إلى الصيغة المطلوبة [الإكسيل] طبقًا للحقول الببليوجرافية.

3.1.12 ضبط الكلمات والعبارات المُفتاحية [Key word, Phrase] :

تُعد قوائم الكلمات المُفتاحية في المشروع أداة حاسمة لتحسين دقة وكفاءة عملية استخراج البيانات وفقًا لطبيعة الحقول الببليوجرافية المُصممة، حيثُ تشمل هذه القوائم كلمات مفتاحية قد تسبق أو تُلحق بالمحتوى المُراد استخراجه، وبالتالي تُساعد على التعرُّف عليه وتحديده بدقة، تُعزز هذه القوائم من فعالية المنظومة في التعرُّف على البيانات، وتمييزها، وتنظيمها، مما يُساهم في تحويل البيانات المُستخرجة إلى معلومات مُنظمة، وقابلة للاستخدام في جداول البيانات مثل Excel أو أي مُنتج آخر، تقوم القوائم بدورًا محوريًا في ضمان دقة وسرعة مُعالجة المعلومات الببليوجرافية، وفي إطار المشروع موضوع الدراسة تختص قوائم الكلمات المُفتاحية بمُعالجة واستخراج البيانات الببليوجرافية المُتعلقة بالحقول التالية [مؤسسات النشر – المؤلفين – الطبعة] [18]، وتم تصميم بنيتها التنظيمية طبقًا لعدة تقسيمات كما بالشكل التالي :

شكل 3 البنية التنظيمية للكلمات المُفتاحية Key word, Phrase

[الشكل من إعداد الباحث]

4.1.12 التعلُّم العميق [Deep Learning] :

التعلُّم العميق أحد الفروع المُتقدمة من تقنيات الذكاء الاصطناعي، وهي تقنية أكثر تعقيدًا من تقنيات التعلُّم الآلي، ويستند إلى الشبكات العصبية العميقة DNN التي تُمكِّن الحواسيب من "التعلُّم" من البيانات دون الحاجة إلى تدخل بشري مُباشر، وهي تصور تطبيقي تنعكس أهميته في نتائجه في تحسين جودة أداء عمل المنظومة، بتعزيز دقة وجودة التعرُّف الضوئي والذكي على الحروف والخطوط، وتحليل المحتوى، والتي يتم تحسينها باستخدام نماذج التعلّم العميق لتحديد النصوص المُعقدة بشكل أدق، خاصة في الحالات التي تكون فيها جودة الصور مُنخفضة أو النصوص مكتوبة بخطوط أو لغات مختلفة، وتحسين جودة التمييز بين المعلومات الببليوجرافية المُختلفة [مثل العنوان، المؤلف، سنة النشر] من خلال تحليل بنية النصوص المكتوبة، والتعلُّم من التجارب السابقة، وبالتالي تحسين الأداء مع مرور الوقت عند التعامل مع أنواع مُتُعددة من الكتب أو المصادر، وتتمثل آلية تطبيق تقنية التعلُّم العميق فيما يلي :

- تحديد الأداة المُناسبة لاستخدامها ودمجها في المنظومة : PyTorch, TensorFlow [19].
- دمج تكنولوجيا التعلّم العميق لتحسين التعرُّف على النصوص غير الواضحة أو المكتوبة بخطوط اليد غير المألوفة، وبالحقول الببليوجرافية لتمييز البيانات الببليوجرافية والتعرُّف عليها.

- تدريب نماذج التعلُّم العميق على مجموعة مُتنوعة من الصور التي تمثل مصادر معلومات ببليوجرافية، مما يجعل النظام أكثر دقة في التعرُّف التلقائي لها مُستقبلًا، بالتعلُّم من الأنماط المُتكررة في البيانات، مثل كيف يمكن ترتيب البيانات الببليوجرافية في الكتب المختلفة، وتحسين استخراج المعلومات تلقائيًا.

2.12 مرحلة التطبيق ومنهجية العمل التنفيذية :

1.2.12 تحضير الملفات ومصادر المعلومات :

تتضمن هذه الخطوة جمع وتجهيز مجموعة مُعبرة عن كُل خصائص الخط وتنوعاته الشكلية من الصور ومصادر المعلومات المُتنوعة التي ستُستخدم كمُدخلات للنظام، وهو ما يُمثل المرحلة الأولية لضمان توفر البيانات الضرورية لاستكمال العمليات التالية في المشروع، تتطلب هذه المرحلة فحصًا دقيقًا للصور المُجمعة لضمان جودتها، ووضوح النصوص المطبوعة عليها، حيث يؤثر مستوى وضوح الصور بشكل مُباشر على دقة عملية التعرُّف الضوئي والذكي على الحروف OCR, ICR واستخراجها، يتضمن التحضير أيضًا تصنيف الصور، وتنظيمها بطريقة تُسهل الوصول إليها ومُعالجتها، مما يُساهم في تسريع مراحل التحليل والاستخراج.

2.2.12 تدريب مُخصص لتقنية OCR, ICR وتطبيقاتها المُستخدمة :

تعامل الباحث مع عدد من التطبيقات المفتوحة والمُغلقة المصدر لتقنيات OCR, ICR، مثل Paddle OCR, Tesseract OCR,, Easy OCR, Zinki OCR, Google Cloud Vision OCR لدعمهم العديد من اللغات ومنها اللغة العربية، ومرونتهما الكافية للتخصيص، والتطوير طبقًا لخصائص المشروع، والمنصة المُقترحة، وطبيعة المحتوى والعمليات التي تتم بداخلها، وذلك بهدف اختيار أفضلها، من أجل تحويل الصور الرقمية لصفحات العناوين المُصورة رقميًا إلى نصوص يسهُل استخراج محتواها، وطبقًا لعمليات الاختبار والتحقق التي قام بها الباحث على مُحركات التعرف الضوئي المذكورة لقياس قُدرتها على التعرُّف على طبيعة المحتوى العربي بالصور ذات النصوص العربية بالمنصة، للتأكد من جاهزيتها للاستخدام المُباشر، أم في حاجة لعملية التدريب والتخصيص للتعرف على النص العربي لتحسين دقة وجودة عملية التعرُّف، توصل الباحث إلى أن أفضل المُحركات التي يُمكن الاستفادة منها لتحقيق أهداف المشروع هي بالترتيب التالي :

1. مُحرك التعرف على النصوص الذكي زنكي Zinki OCR Engine.
2. مُحرك تيسيراكت Tesseract OCR Engine.
3. مُحرك جوجل للتعرُّف الضوئي على الحروف Google Cloud Vision OCR.

3.2.12 اختبار نموذج التدريب Test Model :

يُقصد بها اختبار النموذج بعد التدريب باستخدام صورة أو مجموعة صور جديدة بالمشروع، وتقييم النتائج، وإعادة التدريب إن لزم الأمر.

4.2.12 تحميل الصور الرقمية وتسكينها :

يُقصد بها عملية إدخال صور رقمية ممسوحة ضوئيًا أو مُصورة رقميًا لصفحات العناوين الخاصة بالكتب ومصادر المعلومات إلى مُجلد مُخصص ضمن مشروع البرمجة، وفي هذه الحالة، يُسمى المجلد "images" أو بالتسمية المرغوبة، تُمثل هذه الخطوة عملية تنظيمية هامة في إدارة بيانات المشروع، حيث يتم وضع الصور في مسار محدد ضمن بيئة التطوير المتكاملة، مثل فيجوال استوديو كود Visual Studio Code، بحيث تكون الصور جاهزة ويستطيع النظام الوصول إليها، ومُعالجتها، هذا يشمل التأكد من أن المسارات إلى الصور صحيحة ومطابقة للتكوين المطلوب في الشيفرة البرمجية، مما يُسهم في تسهيل عمليات قراءة وتحليل الصور ضمن المشروع. ذلك يضمن أن النظام يمكنه الوصول إلى الصور بسهولة وتنفيذ المهام المطلوبة.

5.2.12 تطبيق وتشغيل كود المنظومة :

في خطوة "تطبيق وتشغيل كود المنظومة"، يتم تنفيذ الشفرة البرمجية التي تم تطويرها في مشروع استخراج البيانات الببليوجرافية، تتضمن هذه الخطوة تشغيل الكود الذي يتولى مُعالجة الصور الرقمية لصفحات العناوين، للتعرُّف على أجزائها، ثم استخراج البيانات النصية منها، وتنظيمها في ملف Excel أو أي تنسيق آخر مطلوب طبقًا للمُحددات أو المعايير المطلوبة، وتُعتبر خطوة تطبيق وتشغيل الكود هامة لأنها تُحول الفكرة البرمجية إلى تطبيق عملي، وتضمن أن جميع أجزاء النظام تعمل كما هو متوقع لتحقيق نتائج دقيقة وموثوقة.

1.2.11 تنظيم البيانات – استخراجها – تصديرها إلى ملف الإكسيل Excel :

في خطوة "تنظيم البيانات وتصديرها إلى ملف Excel"، يقوم النظام بتحويل البيانات الببليوجرافية، والمحتوى الذي تم استخراجه من الصور الرقمية إلى ملف Excel مُنظم، ومن ثم إلى أي صيغة معيارية أُخرى مثل Marc21، وذلك طبقًا للحقول الببليوجرافية المعيارية المطلوبة والمُحددة طبقا لحاجة مؤسسة المعلومات، أو طبقًا لهيكل المعيار المُحدد، ففي هذه الخطوة بتم تحويل البيانات إلى صيغة مُنسقة وسهلة القراءة، مما يسهم في إدارتها وتحليلها بفعالية، وكذلك يُمكن مشاركة البيانات مع الآخرين طبقًا للحاجة، أو استخدامها في تحليلات إضافية.

3.12 مرحلة ما بعد التطبيق :

1.3.12. تقييم وتحسين النتائج :

يتم فحص النتائج المُستخلصة من عملية استخراج البيانات وتقييم جودتها وفعالية العملية، من خلال مراجعة البيانات المُستخرجة من صور الرقمية لصفحات عناوين الكتب للتحقق من دقتها واكتمالها، يتضمن ذلك مقارنة النتائج بالبيانات الأصلية والتأكد من صحة المعلومات المُستخرجة، وبناءً على عملية التقييم، يتم إجراء تحسينات على النظام مثل تُعديل خوارزميات التعرُّف الضوئي على الحروف OCR, ICR، تحسين معايير استخراج البيانات، أو تُعديل الكود البرمجي لتحسين دقة النتائج وجودتها.

2.3.12. المُتابعة والإشراف الفني والتقني :

يتم مراقبة وإدارة تقدم المشروع لضمان تنفيذ جميع جوانب النظام بشكل صحيح وفقًا للمعايير المُحددة، مع مراقبة أداء المنظومة البرمجية والتقنية المستخدمة، مثل التأكد من فعالية خوارزميات التعرُّف الضوئي على الحروف OCR, ICR بعد عمليات التقييم، وحل أي مشاكل تقنية قد تطرأ، تضمن هذه الخطوة تنفيذ المشروع بسلاسة واستمرارية، مع تحقيق أهدافه بكفاءة وفعالية.

3.3.12. تصميم وبناء واجهة المُستخدم UI وواجهة برمجة التطبيقات APIs :

تتضمن هذه الخطوة عدة مراحل، وهي :

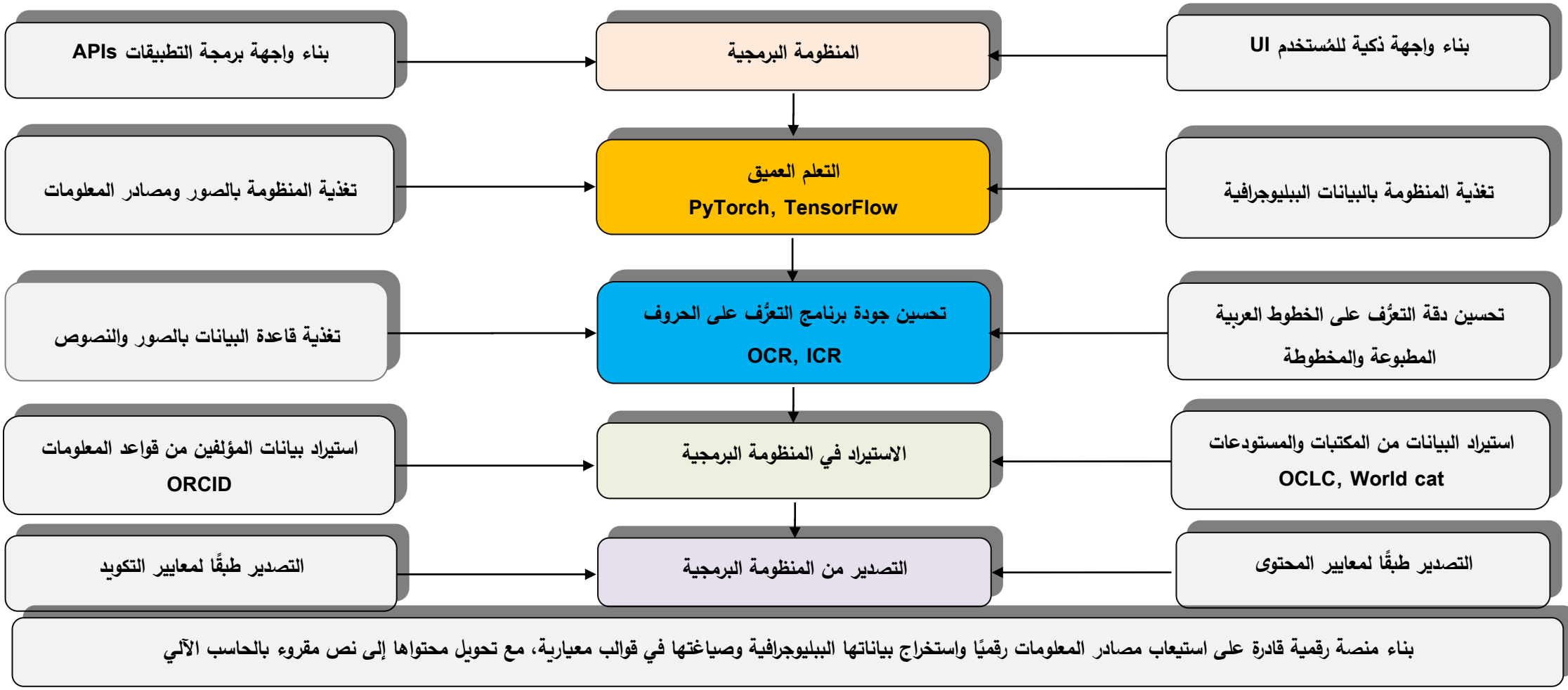
- بناء وتصميم واجهة المُستخدم User Interface والتي يستطيع المُستخدم من خلالها وبسهولة التفاعل مع النظام والتعامل معه
- بناء وتصميم واجهة برمجة التطبيقات Application Programming Interface والتي تُمكن المنظومة البرمجية في التعامل مع الأنظمة المُختلفة وتُيسر من عمليات تصدير البيانات والتكامل مع الأنظمة الأُخرى لإدارة المعلومات.
- دمج واجهة المُستخدم UI مع واجهة برمجة التطبيقات APIs لتيسير طُرق التعامل مع وظائف النظام بشكل كامل من خلال واجهة المُستخدم.

4.3.12. تصميم دليل العمل وتوثيق المنظومة :

في خطوة "تصميم دليل عمل يوضح كيفية الاستخدام"، يتم إنشاء مستند شامل يهدف إلى توجيه المستخدمين حول كيفية استخدام النظام أو التطبيق بشكل فعّال، يتضمن هذا الدليل تعليمات وإرشادات تساعد المستخدمين على فهم واستفادة كاملة من الوظائف والميزات المقدمة، تُعد خطوة تصميم دليل العمل أساسية لضمان أن جميع المستخدمين يمكنهم استخدام النظام أو التطبيق بكفاءة واستفادة كاملة من ميزاته، مما يساهم في تحسين تجربة المستخدم.

14- الاتجاهات والتطويرات الجارية والمُستقبلية بالمشروع :

يُحدد الشكل التالي التحضيرات التي يعمل عليها الباحث لتحسين المنظومة البرمجية بالعديد من الوظائف، والإضافات التي تُعزز من جودة أدائها، كما تستعرض التصورات المُستقبلية التي ستكون عليها المنظومة ضمن مُخطط التطوير المُستقبلي الذي يعمل عليه الباحث:



شكل 4 مُخطط التطورات الجارية والاتجاهات المُستقبلية التي يعمل عليها الباحث في بناء وتطوير المنظومة البرمجية

15- الخاتمة :

- **النتائج :**

قامت الباحث بالتوصل إلى النتائج الآتية :

1. تمكَّن الباحث من بناء منظومة برمجية ذكية باستخدام لغة بايثون Python قادرة على التعامل مع مصادر المعلومات العربية، واستخراج المعلومات الببليوجرافية منها.
2. انتقاء أنسب المكتبات البرمجية في لغة بايثون Python بما يتفق مع أهداف ومراحل المشروع ودمجها بكود المشروع لتعزيز فعالية عمل المنظومة، وتحسين جودة مُخرجاتها.
3. تصميم وبناء مكتبات استنادية بلغة البايثون لـ [عناوين مصادر المعلومات – أسماء الشخصيات – دور النشر]، قابلة للتحديث، مع دمجها ضمن كود المشروع، بما يُعزز قُدرة النظام على تحسين جودة محتوى السجل الببليوجرافي، مع واجهة استخدام سهلة ومُبسطة للتعامل مع المنصة بحيث يسهُل على المُستخدم أن يرى الصور الرقمية التي تم تحميلها جنبًا إلى جنب مع بياناتها الببليوجرافية المُستخرجة.

4. عالجت الدراسة تحديات اللغة العربية بتعاملها مع عدة مُحركات مفتوحة المصدر أو التجارية من مُحركات التعرُّف الذكي على النصوص بالمشروع، وأبرزها : Tesseract OCR, Zinki OCR, Google Cloud Vision , OCR لدعمهم للعديد من اللغات ومنها اللغة العربية، مع الأخذ بعين الاعتبار وضع منهجية لتقييم الأداء، والتطوير والتخصيص، طبقًا لخصائص المشروع، وطبيعة المحتوى، حيثُ تظهر فعالية المنظومة، وجودة مُخرجاتها في مدى التدريب والتخصيص الذي تم على مُحرك التعرُّف على النصوص العربية وهو ما ركزَّت عليه الدراسة في مشروعها.

5. تم بيان مُخطط لعمليات التطوير الجارية والمُستقبلية، بما في ذلك تصميم واجهة ذكية للمُستخدم (UI) وواجهة برمجة تطبيقات (API) لتحسين تجربة المُستخدم وتوسيع نطاق استخدام النظام، وتطوير عمليات التصدير بدمج المكتبات البرمجية اللازمة لذلك، لتشمل التصدير إلى كافة القوالب المعيارية في مجال علوم إدارة المعلومات MARC21, RDF.

- **التوصيات :**

1. الاستفادة من مشروع الدراسة وبنيته البرمجية والوظيفية لبناء منصة رقمية مُتاحة على الخط المُباشر قادرة على استيعاب مصادر المعلومات رقميًا، وتحويل محتواها إلى نص مقروء بالحاسب الآلي، وتكون قادرة على استخراج بياناتها الببليوجرافية وصياغتها في قوالب معيارية تدعم عمليات الفهرسة الوصفية والموضوعية، مع تمكين المكتبات ومؤسسات المعلومات من الاشتراك بالمنصة لبناء سجل ببليوجرافي طبقًا للحقول التي تُحددها كُل مكتبة أو مؤسسة معلومات، مع الحصول على المادة التي تم تحميلها لتكون قابلة للبحث في محتواها باستخدام تقنيات التعرُّف الذكي على المحتوى، وهي أحد المُخططات التي تسعى إليها الدراسة مُستقبلًا .
2. الاهتمام بالبحث الأكاديمي والتطبيقي في تقنيات الذكاء الاصطناعي وتطبيقاتها في التعرُّف الضوئي على الحروف، لبناء جيل قادرًا على فهم هذه التقنيات والتعامل معها، وبالتالي تعميق الفهم للتقنيات المُستخدمة في بناء أنظمة التعرُّف الضوئي على النصوص العربية، مع ضرورة دعم الأبحاث التي تستكشف خوارزميات جديدة وتقنيات للتعلُّم الآلي، والتعلُّم العميق، والتي تُساهم في تحسين دقة وجودة النصوص المُستخرجة.
3. إجراء مشاريع تعاونية ذات أهداف عليمة، وعقد ورش عمل، وجلسات نقاش وتحاور فيما بينهم، وتنمية وعي أخصائي المعلومات بالمهارات التقنية والبرمجية بتطبيقات الذكاء الاصطناعي، وأدواته البرمجية، بإجراء برامج تدريبية مُتخصصة تتدرج من التأسيس إلى التعمق لبناء وتكوين أخصائي معلومات قادر على فهم التقنيات الحديثة، وقادر على التعامل معها، وبناء، وتمثيل أفكاره بها.

16- الهوامش :

1. يتوجه الباحث بجزيل الشكر والتقدير لكُلًا من الأستاذ رياض طاهر رضوان رئيس قسم المُعالجة الفنية بمركز جُمعة الماجد للثقافة والتراث بدبي، لما قدمه لنا توجيهات ودعم ومُساندة لإتمام هذا العمل، والأستاذة منال فيصل محمد الأعرج، ووالأستاذ الدكتور هشام مكي لحُسن توجيهاتهم ونصائحهم ودعمهم، والمُستشار التقني الأستاذ مروان سلام الذي لم يبخل علينا بوقته وعلمه ومُساعداته التقنية فجزاهم الله عنا كُل خير.
2. أحمد، محمد حسين. "التعلم الآلي واستخراج البيانات الببليوجرافية من المواد النصية: نموذج مقترح للمواد النصية باللغة العربية." مجلة Cybrarians Journal. ع. 64 (2021). رابط الوصول : https://jstc.journals.ekb.eg/article_297924.html
3. الهادي، محمد محمد. "تأثير التكنولوجيا الذكية المتقدمة على تحول المكتبات التقليدية إلي مكتبات وخدمات معلومات ذكية." مجلة الجمعية المصرية لنظم المعلومات وتكنولوجيا الحاسبات. مج. 31. ع. 31 (2013). رابط الوصول : http://search.mandumah.com/Record/1219414
4. Tkaczyk, D., et al. "Machine Learning vs. Rules and Out-of-the-Box vs. Retrained: An Evaluation of Open - Source Bibliographic Reference and Citation Parsers." ,2018. Retrieved in : September 13, 2021, from : https://www.semanticscholar.org/paper/Machine-Learning-vs.-Rules-and-Out-of-the-Box

5. Vijayakumar, S., & Sheshadri, K. N. "Applications of Artificial Intelligence in Academic Libraries." International Journal of Computer Sciences and Engineering, 7. (16), p.p. 136 – 140, 2019, Retrieved in : 5 / 8 / 2024. From : https://www.ijcseonline.org

6. Singh, Mayank & Barua, , et al. "OCR++: A Robust Framework For Information Extraction from Scholarly Articles." Harvard university. Harvard, 2016.

7. Tkaczyk, Dominika, et al. "CERMINE: Automatic Extraction of Structured Metadata from Scientific Literature." International Journal on Document Analysis and Recognition, vol. 18, no. 4, Dec. 2015, pp. 317–335. Springer, from : https://link.springer.com/article/10.1007/s10032-015-0249-8

8. Cambridge University Press & Assessment. "Artificial Intelligence". Dictionary : Cambridge University Press & Assessment, 2024. Retrieved in : 13 / 6 / 2024. from : https://dictionary.cambridge.org/dictionary/english/artificial-intelligence

9. Amazon Web Services, Inc. ibid.

10. ياسين، مي أكرم. "المكتبات الرقمية : المفاهيم والتحديات." المجلة العربية للنشر العلمي. ع. 48. رابط الوصول : https://www.ajsp.net/research . تاريخ الوصول : 26 / 7 / 2024.

11. يوسف، محمد. "شريان الحياة للذكاء الاصطناعي : كيف تواجه شركات التقنية معضلة انتهاء البيانات؟." شبكة الجزيرة العالمية، تكنولوجيا. رابط الوصول : https://www.aljazeera.net/tech/2024/7/20. تاريخ الوصول : 26 / 7 / 2024.

12. Microsoft Corporation. "Visual Studio Code : VSC." United States of America : Microsoft Corporation, 2024. Retrieved from : https://code.visualstudio.com in : 18 / 8 / 2024.

13. Masa, Amjad. Faris Masad. Haya Odeh. " Replit – Build software faster." San Francisco : Repl.it, 2024. Retrieved from : https://replit.com in : 18 / 8 / 2024.

14. تم تطوير مُحرك Tesseract OCR بواسطة شركة Hewlett-Packard كبرنامج احتكاري في الثمانينيات، وتم إصداره كمصدر مفتوح في عام 2005 وتمت رعاية التطوير بواسطة Google في عام 2006، مُتاح عبر الرابط : https://github.com/tesseract-ocr/tesseract ,https://tesseract-ocr.github.io

15. Python Software Foundation. " Python." Python Software Foundation, 2024. Retrieved from : https://www.python.org in : 19 / 8 / 2024.

16. Scher, Nadav. "Change System variables on windows 11." Microsoft. Retrieved from : https://answers.microsoft.com/en-us/windows/forum/all/change-system-variables-on-windows-11

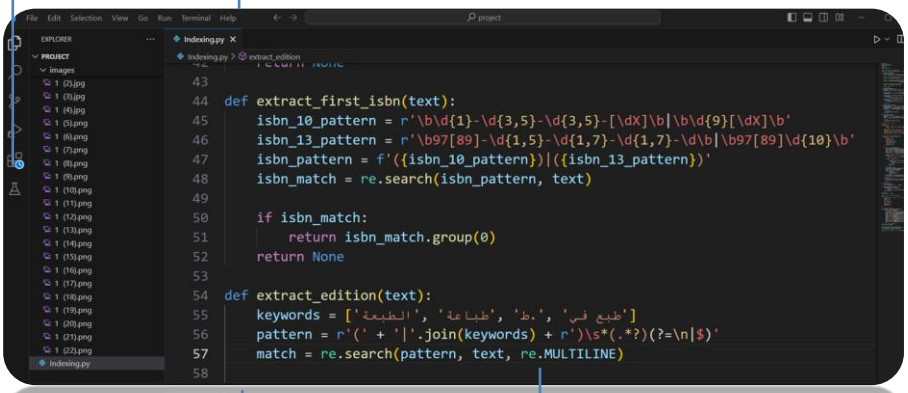
17. انظر ملحق كود المشروع ونماذجه التطبيقية".
18. أُنظر ملحق كود المشروع ونماذجه التطبيقية".
19. مكتبة تعلُّم الآلة مفتوحة المصدر مُستندة على تقنية Torch، وتُستعمل لتطبيقات الرؤية الحاسوبية، ومعالجة اللغات الطبيعية، وتدعم تطوير وتدريب النماذج العصبية العميقة لتحسين دقة استخراج البيانات من الصور الرقمية، مطورة أساسًا من مُختبر آل لأبحاث الفيسبوك، أول إصدار لها عام 2016م، وهي برمجية حرة مفتوحة المصدر مُرخصة تحت رخصة بي سي دي المُعدلة BSD License، ويُمكن الوصول إلى الأداة من خلال موقعها الرسمي : https://pytorch.org

17- الملاحق :

نماذج تطبيقية باستخدام المنظومة الذكية لإنشاء وإدارة المعلومات الببليوجرافية

قائمة الصور الرقمية لصفحات العناوين للكُتب بعد تحميلها في ملف النظام وظهورها في البيئة البرمجية Visual Studio

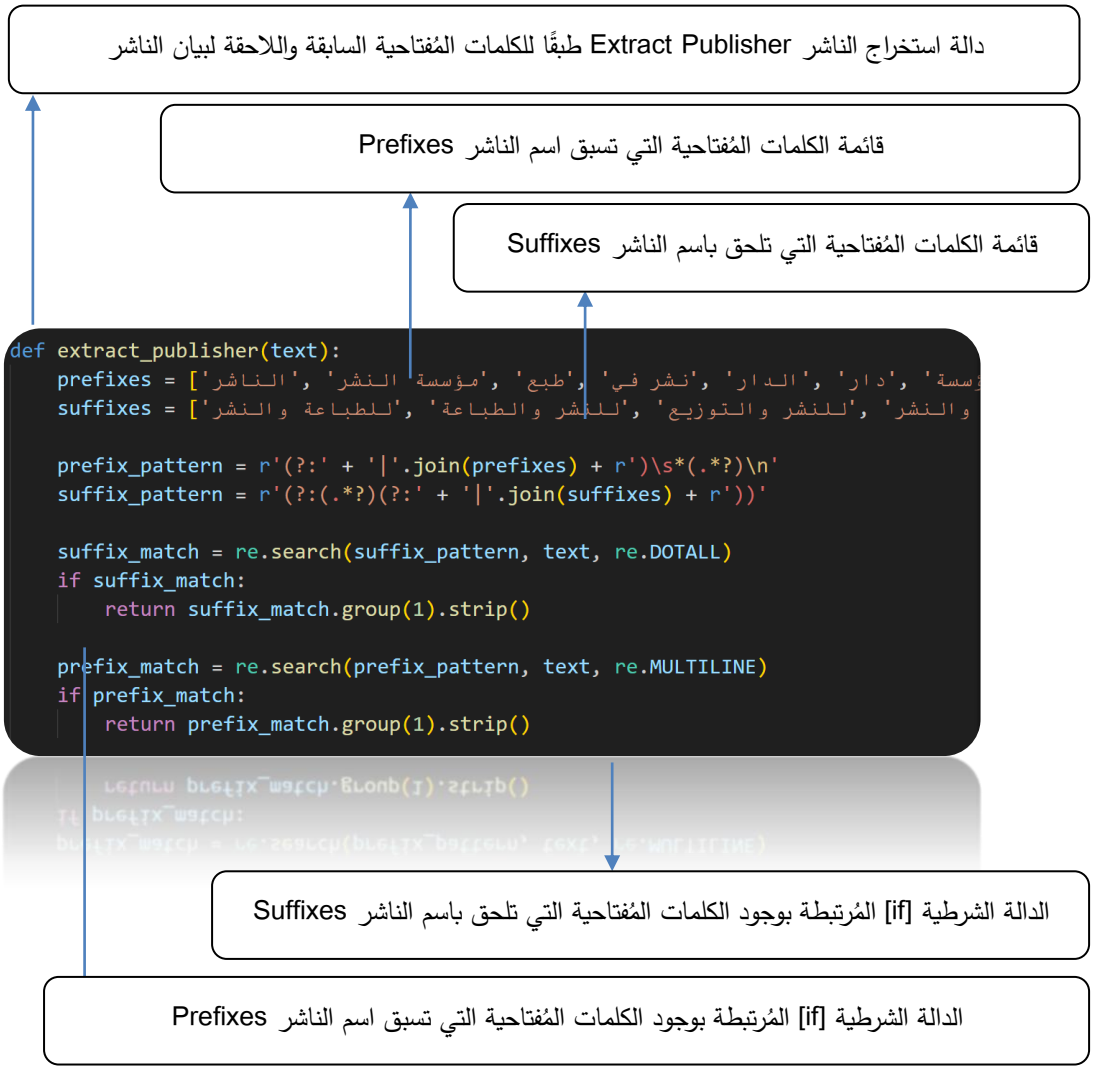
دالة استخراج الترقيم الدولي للكُتب ISBN طبقًا لنظام الترميز المُكون من 10 أو 13 رقم



أسطر الكود الخاصة بدالة استخراج رقم الطبعة Extract Edition

دالة استخراج رقم الطبعة Extract Edition

شكل 5 جُزء من الكود الخاص بالمنظومة الذكية لاستخراج وإدارة المعلومات الببليوجرافية



شكل 6 قائمة الكلمات المُفتاحية للناشرين طبقًا للتصنيف المُتفق عليه بالكود

شكل 7 نموذج مُدخلات مجموعة من الصور الرقمية لصفحات عناوين مجموعة من الكُتب

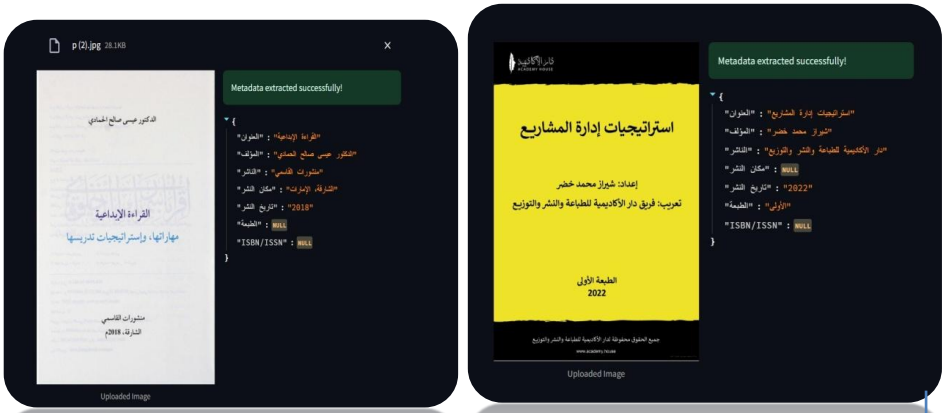
```
# تحويل البيانات إلى DataFrame
df = pd.DataFrame(data)

# وكتابة البيانات إليه Excel إنشاء ملف
excel_filename = os.path.join(image_folder, 'book_info.xlsx')
df.to_excel(excel_filename, index=False)

print(f"تم إنشاء ملف Excel بنجاح: {excel_filename}")
```

وظيفة تحويل البيانات إلى إطار البيانات Data Frame ثم تصديرها إلى ملف الإكسيل

شكل 8 وظيفة تحويل البيانات إلى إطار البيانات Data Frame ثم تصديرها إلى ملف الإكسيل



	A	B	C	D	E	F	G	H	I	J	K
	اسم الصورة	العنوان	المؤلف	الناشر	مكان النشر	تاريخ النشر	الطبعة	ISBN/ISSN	العدد	المجلد	المحتوى
1	(1).png	صرير القلم (مقالات في مناسبات)	الدكتور جاسم خليل ميرزا	مركز الكتاب للنشر		2021	الأولى	978-977-294-643-3			
1	(2).png	التصوير الإسلامي في الهند: تسليات البلاط وحياة الشعوب في التصوير المغولي	د. منى سيد علي حسن	دار النشر للجامعات	مصر	1424-2003	الأولى	977-316-103-X			
1	(3).png	مصباح معلق: شعر	قاسم سعودي	فضاءات	عمّان	2011	الأولى	978-995-30-271-9			
1	(4).png	إدارة الضغوط الحياتية من منظور اجتماعي معاصر	الأستاذ الدكتور نعيم عبد الوهاب شلبي	المكتبة العصرية							
1	(5).png	إدارة الموارد البشرية المعاصرة	حامد جودت أصرف	Brighter Horizon	الشارقة	2020	الأولى				

تحميل الصور على المنصة – الحقول الببليوجرافية يتم استخراجها تلقائيًا من صفحات العناوين بالمنصة

البيانات الببليوجرافية في ملف الإكسيل Excel الذي تم إنشاؤه تلقائيًا

شكل 9 المعلومات الببليوجرافية التي تم استخراجها وتصديرها إلى السجل الببليوجرافي Excel

Abstract :

This study aims to present a practical proposal for a smart technical system, based on artificial intelligence applications (AI), machine learning (ML), and deep learning (DL) techniques, which can be employed in libraries and information centers to perform one of the most important technical operations in the field of library and information sciences, which is managing and extracting bibliographic information from digital images of information sources, in order to build a standard bibliographic record for the huge information sources acquired by libraries.

The study deals in its theoretical framework with defining artificial intelligence (AI) techniques, and the applied framework comes to present a proposal for a smart system consisting of a set of functionally integrated technologies to achieve a set of objectives, namely: intelligent recognition (ICR) of digital images of information containers and sources, extracting bibliographic information from them, exporting bibliographic information in an Excel file according to the agreed standard fields, and the study uses the experimental applied approach for this purpose, to conclude the study by presenting a set of results and recommendations that would contribute to developing the processes and practices of dealing with information sources in libraries and information centers.

Keywords : Bibliographic data – Intelligent character recognition – Deep learning – Technical operations – Bibliographic records.