تطبيقات نُظم الذكاء الاصطناعي في تحليل المحتوي وعمليات التكشيف
دراسة تطبيقية لنظم معالجة اللغة الطبيعية(1)

د. إهداء صلاح ناجي

مدرس المكتبات والمعلومات، كلية الآداب – جامعة القاهرة

Ehdaa.salah@gmail.com

تاريخ الاستلام: 4 إبريل 2022 **تاريخ القبول: 23 مايو 2022**

المستخلص:

مع تطور تقنيات الذكاء الاصطناعي ومن أهمها: معالجة اللغة الطبيعية التي تم إدراك دورها الكبير في تحليل المحتوى، وإدخال العديد من التحسينات إليها، جاءت هذه الدراسة لتهدف إلى التعرف على دور تطبيقات نظم الذكاء الاصطناعي في تحليل المحتوي وعمليات التكشيف، واستخدامات معالجة اللغة الطبيعية بصفة عامة وفي المكتبات بصفة خاصة، ورصد منصات ومكتبات الذكاء الاصطناعي التي يمكن الاستفادة منها في مجال تحليل المحتوى، مع العمل على تطبيق أحد نظم معالجة اللغة الطبيعية واستخدامها في تحليل المحتوى لمصادر المعلومات؛ لتحقيق الإفادة منها، وذلك اعتمادًا على المنهج الوصفي التحليلي.

ومن أهم النتائج التي خرجت بها الدراسة: إمكانية استخدام معالجة اللغة الطبيعية في المكتبات في الحصول على المصادر، ومعالجة المحتوى، وإحاطة المستفيدين بالمصادر والإجابة على الاستفسارات، وبتطبيق منصة IBM Watson Knowledge Studio لإنشاء نموذج لتعلم الآلة واختباره باستخدام أداة Watson Discovery وكانت كفاءته عالية في البحث والاسترجاع للمصادر باللغة الإنجليزية، أما بالنسبة للمصادر باللغة العربية فإن الأداة تحتاج لمزيد من التحسينات.

الكلمات المفتاحية: تطبيقات الذكاء الاصطناعي (1)؛ نظم معالجة اللغة الطبيعية (2)؛ تحليل المحتوى(3)؛ التعلم الآلي (4).

0/ المقدمة:

تم استخدام تقنيات الذكاء الاصطناعي في العديد من المجالات، مثل: الأعمال، والإدارة، والطب، والمهام العسكرية، وما إلى ذلك، وبدأت أفكار استخدام النظم الذكية بدلاً من النظم التقليدية في المكتبات بداية من عام 1990م؛ حيث تم استخدامها لتوفير خدمات قائمة على المعرفة للمستفيدين من المكتبات، والعاملين، والمتخصصين في المجال.

---

(1) **تم عرضه في الندوة العلمية الأولي عن تطبيقات الذكاء الاصطناعي فى مجالات العلوم الاجتماعية والإنسانيات :الخصائص . البرمجيات، آليات التنفيذ، كلية الآداب، جامعة القاهرة، في الفترة من 29 – 30 ديسمبر 2021.**

وفي ظل سعي المكتبات نحو تقديم خدمات سريعة وفعالة وذات كفاءة، والحرص على تحسين خدماتها ومزاياها التنافسية اعتمدت المكتبات الأكاديمية على تقنيات كثيرة في الماضي، ولكن كانت تقنيات الذكاء الاصطناعي هي الأحدث بين التقنيات التي يتم تقديمها حاليًا في المكتبات، وهي تشمل نظما متخصصة للخدمات المرجعية، وروبوتات القراءة على الرفوف، والواقع الافتراضي للتعلم، ومعالجة اللغة الطبيعية؛ لاستخراج المعرفة ومعالجة المحتوى.

وعملت بعض المكتبات الأكاديمية بالدول المتقدمة في السنوات الأخيرة على إنشاء معامل للذكاء الاصطناعي، من أجل زيادة فرص تحسين الخدمات الحالية بها، على سبيل المثال: قامت مكتبة جامعة رود آيلاند University of Rhode Island (URI) بإنشاء معملًا للذكاء الاصطناعي في عام 2018م؛ من أجل تشجيع التعاون بين التخصصات المختلفة، ومساعدة الطلاب وأعضاء هيئة التدريس في التعرف على جميع القضايا المتعلقة بالذكاء الاصطناعي، وتطوير مشروعات لخوارزميات ذكية في حقل الذكاء الاصطناعي، وكذلك من أجل تقديم خدمات أفضل داخل المكتبة وجعلها أكثر دقة وأسرع (Enis, 2018).

فالذكاء الاصطناعي هو التكنولوجيا الحالية التي تطورت مع آفاق ضخمة وتطبيقات واعدة في المكتبات، وإذا كان للمكتبات أن تزدهر في عصر اقتصاد المعرفة فإنها بحاجة إلى استكشاف هذه التقنية وتطبيقها، من أجل تعظيم فوائدها الغنية بشكل مناسب لتقديم خدمات مبتكرة ومثالية في المكتبات مثل: تطوير نظم الذكاء الاصطناعي للمساعدة في العمليات الفنية والخدمات المرجعية، وإدارة المصادر وتحليل محتواها وبحثها واسترجاعها.

وتعد معالجة اللغة الطبيعية من أهم التقنيات التي تم إدراك دورها الكبير في تحليل المحتوى، وإدخال العديد من التحسينات إليها مثل تقنيات الويب الدلالي، وقد جاءت هذه الدراسة لتهدف إلى التعرف على تطبيقات نظم الذكاء الاصطناعي في تحليل المحتوي وعمليات التكشيف، والتعرف على استخدامات معالجة اللغة الطبيعية بصفة عامة وفي المكتبات بصفة خاصة، ورصد منصات والمكتبات البرمجية للذكاء الاصطناعي التي يمكن الاستفادة منها في مجال تحليل المحتوى، مع العمل على تطبيق أحد نظم معالجة اللغة الطبيعية، واستخدامها في تحليل المحتوى لمصادر المعلومات لتحقيق الإفادة منها.

أولًا: الإطار المنهجي للدراسة:

1/1مشكلة الدراسة وأهميتها:

تسعى المكتبات بكل الطرق إلى إضافة قيمة لمصادر المعلومات بها، من خلال اختيار مصادرها الأكثر فائدة للمستفيدين، مع تقديم البيانات الوصفية لها بجودة عالية، بالإضافة إلى القيام بعمليات التحليل الموضوعي للمصادر، واستخراج المعلومات منها؛ لتحسين عمليات البحث والتصفح لها، ولكن في ظل الزخم المعلوماتي الهائل والتطور التكنولوجي أصبح من الصعب القيام بعمليات التحليل الموضوعي لكافة المصادر، وأصبحت هناك حاجة ماسة إلى نظم تساعد في ذلك؛ من أجل تسهيل عمليات الوصول للمصادر وإتاحتها بما يتوافق مع احتياجات المستفيدين.

ومع تطور تقنيات الذكاء الاصطناعي ومن أهمها: معالجة اللغة الطبيعية التي تم إدراك دورها الكبير في تحليل المحتوى، وإدخال العديد من التحسينات إليها، وجاءت الحاجة إلى التعرف على تطبيقات نظم الذكاء الاصطناعي في تحليل المحتوي وعمليات التكشيف، والتعرف على استخدامات معالجة اللغة الطبيعية، ورصد منصات ومكتبات الذكاء الاصطناعي التي يمكن الاستفادة منها في مجال تحليل المحتوى.

ومن ثم تستمد هذه الدراسة أهميتها من الحاجة إلى معرفة الأدوات التي يمكن استخدامها وتطبيقها من أجل القيام بتحليل المحتوى باستخدام تقنيات الذكاء الاصطناعي، ونظم معالجة اللغة الطبيعية من أجل تحسين كفاءة عمليات البحث والاسترجاع.

2/1 أهداف الدراسة:

تسعى هذه الدراسة إلى تطبيق نظم معالجة اللغة الطبيعية في عمليات تحليل المحتوى لمصادر المعلومات، وذلك من خلال تحقيق مجموعة من الأهداف الفرعية المتمثلة في:

1. بيان تطبيقات الذكاء الاصطناعي وتعلم الآلة داخل المكتبات.
2. تحليل استخدامات نظم معالجة اللغة الطبيعية بصفة عامة، وفي المكتبات بصفة خاصة.
3. دراسة دور نظم معالجة اللغة الطبيعية في تحليل المحتوى، وآليات عملها.
4. رصد المنصات والمكتبات البرمجية للذكاء الاصطناعي المستخدمة في مجال تحليل المحتوى.
5. تطبيق أحد نظم معالجة اللغة الطبيعية في عمليات تحليل المحتوى، وإنشاء نموذج لتعلم الآلة.
6. اختبار نموذج تعلم الآلة الذي تم إنشاؤه، وبحث إمكانات تطويره.

3/1 تساؤلات الدراسة:

في ضوء الأهداف التي تسعى الدراسة لتحقيقها، فإن التساؤلات التي حاولت الدراسة الإجابة عليها هي:

1. ما تطبيقات الذكاء الاصطناعي وتعلم الآلة داخل المكتبات؟
2. ما استخدامات نظم معالجة اللغة الطبيعية بصفة عامة، وفي المكتبات بصفة خاصة؟
3. ما دور نظم معالجة اللغة الطبيعية في تحليل المحتوى؟ وما آليات عملها؟
4. ما أبرز المنصات والمكتبات البرمجية للذكاء الاصطناعي المستخدمة في مجال تحليل المحتوى؟
5. ما فاعلية نموذج تعلم الآلة المقترح لتحليل المحتوى؟ وكيف يمكن تطويره؟

4/1 مجال الدراسة وحدودها:

تتناول الدراسة تطبيقات نظم الذكاء الاصطناعي وتعلم الآلة ودورها في تحليل المحتوي، في ضوء الحدود التالية:

* الحدود الموضوعية: دراسة تطبيقات نظم الذكاء الاصطناعي وتعلم الآلة، وبيان دورها في تحليل المحتوى مع تطبيق أحد نظم معالجة اللغة الطبيعية وإنشاء نموذج لتعلم الآلة.

* الحدود اللغوية: تغطي الدراسة استخدام نظم معالجة اللغة الطبيعية لتحليل المحتوى لمجموعة من مصادر المعلومات باللغتين العربية، والإنجليزية.

* الحدود النوعية: تشمل الدراسة تحليل محتوى مجموعة من المقالات العلمية في موضوع "إدارة المعرفة".

5/1 منهج الدراسة وأدوات جمع البيانات:

اعتمدت الدراسة على المنهج الوصفي التحليلي لدراسة تطبيقات نظم الذكاء الاصطناعي وتعلم الآلة، وبيان دورها في تحليل المحتوى، ورصد أبرز المنصات والمكتبات البرمجية لنظم معالجة اللغة الطبيعية، بالإضافة إلى استخدام طريقة تحليل المحتوى؛ لإجراء تحليل للمصادر التي تم إدراجها في الدراسة التطبيقية؛ من أجل تنفيذ نموذج تعلم الآلة ML Model باستخدام أداة Watson Knowledge Studio.

وتم الاعتماد على الملاحظة المباشرة لمواقع الويب لرصد أبرز المنصات والمكتبات البرمجية لنظم معالجة اللغة الطبيعية، بالإضافة إلى مسح الإنتاج الفكري في موضوع الدراسة.

6/1 مصطلحات الدراسة:

* **الذكاء الاصطناعي**: المجال المعني بتطوير التقنيات التي تسمح لأجهزة الحاسب بالتصرف بطريقة تبدو وكأنها كائن حي ذكي، مثل الإنسان؛ وذلك بهدف محاولة تطوير كيان واعٍ وذكي قائم على الحاسب الآلي(William, 1999) .

* **تعلم الآلة**: مجال في علوم الحاسب يعمل على بناء نماذج لديها القدرة على التعلم من البيانات، ثم تقديم التنبؤات، ويمكن تقسيم التعلم الآلي إلى ثلاث فئات: التعلم تحت الإشراف، والتعلم غير الخاضع للإشراف، والتعلم المعزز (Guo, Farhang-Razi, & Algra, 2019)

* **معالجة اللغة الطبيعية**: تقنية تجعل النظم قادرة على تحليل اللغة البشرية، وفهمها، وتوليدها بشكل طبيعي؛ مما يتيح التواصل مع الحاسب الآلي كما لو كان إنسانًا، وتشمل التطورات الحديثة في هذا المجال برامج التعرف على الصوت، وترجمة اللغة البشرية، واسترجاع المعلومات، والذكاء الاصطناعي (Guo et al., 2019)

* **معالجة المحتوى**: العملية الآلية لتحليل وفرز بيانات النص غير المهيكل لاكتساب رؤى قيمة، عن طريق استخدام معالجة اللغة الطبيعية (NLP) والتعلم الآلي؛ من أجل فهم اللغة البشرية تلقائيًا، وإجراء تحليل للمحتوى، وإنتاج معرفات مناسبة له، ويمكن استخدامها فيما بعد لتدريب خوارزميات التعلم الآلي(تعريف إجرائي).

7/1 صياغة الاستشهادات المرجعية:

اعتمدت الدراسة على معيار الجمعية الأمريكية لعلم النفس American Psychological Association – APA لصياغة الاستشهادات المرجعية الواردة في الدراسة، وتم الاستعانة ببرنامج ENDNOTE X7 لصياغة تلك الاستشهادات.

8/1 الدراسات السابقة:

تم البحث في عدد كبير من الأدوات العربية والأجنبية للتعرف على الدراسات والجهود التي ترتبط بشكل مباشر أو غير مباشر بموضوع الدراسة، ومن أهمها قواعد البيانات المتاحة من خلال بنك المعرفة المصري مثل: دار المنظومة، والعبيكان، EBSCO،Jstor، ومجموعة من المستودعات الرقمية في المجال، وأهمها E-lis Repository، وكذلك محركات البحث العلمية، وأهمها: Google Scholar، وأسفرت هذه الخطوة عن مجموعة من الدراسات السابقة ذات الصلة بموضوع الدراسة، وهي على النحو التالي:

1/8/1 أولًا: الدراسات العربية:

❋ تطبيقات الذكاء الاصطناعي وتعلم الآلة في المكتبات:

هدفت دراسة (حسين، 2022) إلى تقديم نموذج مقترح لاستخراج البيانات الببليوجرافية من المصادر العربية مع توضيح دور الناشرين في عملية إنشاء التسجيلات الببليوجرافية، والاستفادة من إمكانيات التعلم الآلي في استخراج البيانات الببليوجرافية من مصادر المعلومات النصية، اعتمادًا على المنهج الوصفي التحليلي، وخرجت الدراسة بمجموعة من النتائج أهمها: عدم توفر مجموعات بيانات Datasets لكافة المواد النصية الصادرة عن دور النشر العربية لاسيما المصرية، ووجود مشكلة في تنوع مخططات الكتب أثناء تدريب نموذج التعلم الآلي، وأوصت الدراسة بضرورة توافر نماذج الذكاء الاصطناعي التي تسهل من العمل وتوفر الوقت والجهد.

وانطلقت دراسة (صدقي، 2021) حول الاستفادة من تقنيات الذكاء الاصطناعي في تطوير أساليب بحث المعلومات واسترجاعها، وتقدم الدراسة بالاعتماد على منهج تحليل المحتوى عرضًا لمفهوم الذكاء الاصطناعي وتطبيقاته، وعرض خدمات الذكاء الاصطناعي المقدمة من شركة IBM وعرض خدمة Watson Discovery المختصة بالبحث والتحليل لمحتوى الملفات، وقد خلصت الدراسة إلى: يجب توجيه المؤسسات نحو الاستفادة من خدمات الذكاء الاصطناعي، وتوجيه المكتبات نحو الاستفادة من تلك التقنيات نظرًا لما توفره من وقت وجهد للعاملين والمستفيدين.

وسعت دراسة (عامر، 2021) إلى إبراز أهمية الذكاء الاصطناعي، وتقنياته، وتطبيقاته في مجال المكتبات والمعلومات؛ حيث تعنى الدراسة برصد أشهر برامج الذكاء الاصطناعي وخدماته المتاحة في مجال أنشطة المكتبات؛ وذلك اعتمادًا على المنهج الوصفي التحليلي، وخلصت الدراسة إلى ضرورة الاهتمام بتقنيات الذكاء الاصطناعي في المكتبات، مع وضع تصور تخطيطي لكيفية تطبيقها في الخدمات المتعددة التي تؤديها المكتبات المصرية ومنها: البحث في الفهارس، وخدمات الإعارة، والخدمات المرجعية، وخدمات البحث عن المعلومات، وتطبيقات الروبوتات.

وسلطت دراسة (السلمى، 2017) الضوء على استفادة موقع جوجل من تطبيقات الذكاء الاصطناعي في تعزيز خدمات استرجاع المعلومات وتطويرها، اعتمادًا على المنهج الوصفي التحليلي من خلال الرجوع لموقع جوجل، واستخدام بعض التطبيقات المتوفرة، ومن أهم نتائج الدراسة: اتساع دعم الذكاء الاصطناعي للعديد من تطبيقات جوجل لتقديم الدعم للمستخدمين وكانت أبرز التوصيات هي: تسليط الضوء على الخدمات الجديدة التي أتاحها الذكاء الاصطناعي في مجال استرجاع المعلومات.

وهدفت دراسة (إبراهيم، 2010) إلى التعرف على تطبيقات الذكاء الاصطناعي ومجالات استخدامها في المكتبات، والتعرف على النظم الخبيرة وكيفية تصميمها في الخدمة المرجعية بالمكتبات، بالإضافة إلى التطبيقات الأخرى الممكنة للذكاء الاصطناعي، وذلك باستخدام المنهج الوصفي التحليلي، وتوصلت الدراسة إلى عدة نتائج أهمها: عدم استخدام أيًا من نظم الذكاء الاصطناعي أو النظم الخبيرة في المكتبات المدروسة، وعدم إلمام المكتبيين بتقنيات الذكاء الاصطناعي، وعدم معرفة المكتبيين بكيفية استخلاص المعرفة وتمثيلها في قواعد المعرفة في النظم الخبيرة في المكتبات.

وناقشت دراسة (يوسف، 2009) الذكاء الاصطناعي والنظم الخبيرة في المكتبات وتم الوقوف على تطبيقات النظم الخبيرة في المكتبات ومراكز المعلومات، وإعداد نموذج خبير تجريبي في التزويد في المكتبات اعتمادًا على المنهج الوصفي التحليلي، ومن أهم نتائج الدراسة: تكنولوجيا النظم الخبيرة في السودان مازلت في مراحلها الأولى، وعدم وجود الاهتمام الكامل في أقسام المكتبات والمعلومات

في الجامعات السودانية بتكنولوجيا النظم الخبيرة، وأوصت الدراسة بضرورة الاهتمام بتكنولوجيا الذكاء الاصطناعي، وإدخال مقرراتها في أقسام المكتبات والمعلومات.

*** تطبيق نظم معالجة اللغة الطبيعية:**

هدفت دراسة (العتيبى؛ العمرى؛ الغامدى، 2019) إلى التعرف على نظام ديبيتر Debater التابع لشركة IBM ووصفه وشرح خصائصه ومميزاته اعتمادًا على المنهج الوصفي، وتوصلت الدراسة إلى بالرغم من عدم اكتمال العمل على النظام إلا أنه حقق مهارات واضحة وتطورات ملحوظة في مجال الذكاء الاصطناعي، وعند اكتمال المشروع فإنه سيحقق منافع عدة للباحثين والطلاب وحتى الحكومات سواء في إجراء البحوث أو اتخاذ القرارات.

وقدمت دراسة (غازي، 2019) نموذجًا للبيئة اللغوية والتطويرية لنظام نوج NOOJ لكونه نظامًا آليا لمعالجة اللغة الطبيعية ومعجمًا إلكترونيًا يتضمن أدوات مصممة منهجيًا ووظائف توفر للمترجم فرصًا لمعالجة اللغة، ومن أهم ما خلصت إليه الدراسة: نظام نوج يوفر بيئة واعدة لترجمة اللغات آليًا لأنه يسمح للمستخدم ببناء الموارد اللغوية وباستخدام واجهات بسيطة وليس من الضروري استخدام سلسلة معقدة من المعالجة، وهذا يضمن التنسيق بين المعالجة اللغوية من جهة والبحث عن المعلومات وتوليدها بصفة تلقائية من السياق.

2/8/1 ثانيًا: الدراسات الأجنبية:

*** تطبيقات الذكاء الاصطناعي وتعلم الآلة في المكتبات:**

ركزت دراسة (ADEJO & MISAU, 2021) على كيفية تطبيق الذكاء الاصطناعي في المكتبات الأكاديمية في نيجيريا اعتمادًا على مراجعة الأدبيات في الموضوع، وخلصت الدراسة إلى إمكان تطبيق الذكاء الاصطناعي في خدمات المكتبات الأكاديمية في نيجيريا مثل: الأنظمة الخبيرة في الخدمات المرجعية والفهرسة والتزويد، وتطبيق نظم معالجة اللغة الطبيعية والتعرف على الأنماط والروبوتات في أنشطة المكتبة وعملياتها، وأوصت الدراسة بضرورة تدريب موظفي المكتبة على استخدام هذه التقنيات في تقديم خدمات المكتبة مع تضمين الإنفاق على تطبيقات الذكاء الاصطناعي في مخصصات الميزانية للمؤسسات التابعة لها المكتبات.

وتناولت دراسة (Das & Islam, 2021) أدبيات تطبيقات الذكاء الاصطناعي وتعلم الآلة في المكتبات ؛ وقد تم اختيار 32 ورقة علمية وتحليلها عن الموضوع، وخلصت الدراسة إلى تركيز الأبحاث الحالية بشكل أساسي على الموضوعات النظرية مع تركيز القليل منها على دراسات الحالة أو تنفيذ مشروعات الذكاء الاصطناعي لإدارة المجموعات في المكتبات، أو توليد البيانات الوصفية، أو اكتشاف المصادر، أو تحليل النصوص، أو الفهرسة، أو التكشيف، باستخدام التقنيات المتقدمة مثل: التعرف على الأنماط والتعلم العميق والشبكات العصبية.

وحاولت دراسة (Stafford, 2021) استكشاف دور تقنيات الذكاء الاصطناعي وتعلم الآلة في تحسين خدمات المكتبة مع التركيز على الروبوتات وتقنيات التنقيب عن النصوص وتوليد البيانات الوصفية، ومن أبرز النتائج التي خرجت بها الدراسة : إمكانية استخدام قدرات الذكاء الاصطناعي في أداء المهام الروتينية وقيام أخصائيي المكتبات بأداء المهام التي تتطلب مهارات أعلى؛ مما يعمل على توفير

وقت أخصائي المكتبات، كما يمكن لهذه الأدوات تقليل التكاليف وترشيد النفقات، وأوصت الدراسة بأهمية توسيع نطاق استخدام تقنيات الذكاء الاصطناعي الثلاث التي تم التركيز عليها في الدراسة من أجل تحسين خدمات المكتبات.

❋ تطبيقات نظم معالجة اللغة الطبيعية المختلفة:

عملت دراسة (Ba, J.Mb, & T, 2021) على تحليل تطبيقات معالجة اللغة الطبيعية في قطاعات مختلفة؛ و أوضحت أن معالجة اللغة الطبيعة تتعامل مع مفهوم الذكاء الاصطناعي الذي هو حقل فرعي من علوم الحاسب، مما يمكّن أجهزة الحاسب من فهم اللغة البشرية ومعالجتها، وتتضمن إنشاء خوارزميات تحول النص إلى كلمات ذات دلالات مع التطورات الناشئة في التعلم الآلي والتعلم العميق، وتلخص هذه الورقة الجوانب المختلفة من معالجة اللغة الطبيعية جنبًا إلى جنب مع دراسات الحالة المرتبطة بقطاع الأعمال والتعليم والتنقيب عن النصوص والتعلم العميق.

❋ تطبيقات نظم معالجة اللغة الطبيعية في عمليات التكشيف وتحليل المحتوى:

عرضت دراسة (Obeid, Zalmout, & Khalifa, 2020) سمات ومزايا أداة – Computational Approaches to Modeling CAMeL Language، وهي أداة مفتوحة المصدر لمعالجة اللغة العربية الطبيعية باستخدام لغة البرمجة بايثون، والتي تعمل على توفير أدوات مساعدة للمعالجة المسبقة للمصادر، والنمذجة المورفولوجية، وتحديد اللهجات، والتعرف على الكيانات المسماة وتحليل المشاعر، وتعد هذه الأداة في حالة تطور مستمر؛ حيث يعمل القائمون عليها على تحسين وظائفها بشكل دائم باستخدام معالجات اللغة الطبيعية.

وهدفت دراسة (Karoo, 2018) إلى تقديم وصف بالقضايا المتعلقة بمهمة إدارة المكتبات الرقمية، وتحديد المشاكل البحثية الجديدة المتعلقة بهذه القضايا، ومن أهم ما خلصت إليه الدراسة ضرورة العمل على إيجاد طرق جديدة لمعالجة المصادر في المكتبات الرقمية وكيفية استخدام تقنية معالجة اللغة الطبيعية (NLP) الحالية لبناء مهمة إدارة المصادر؛ حيث يمكن استخدامها في وصف المصادر، وتحليل المحتوى، والإحاطة الجارية، والاستفادة منها في أدوات تنظيم المعرفة مثل المكانز.

وأجرت دراسة (Hamandi, wahab, & Karim, 2016) عملية معالجة النصوص واستخراج نوع الكلمة من حيث أجزاء الكلام؛ حيث تم تطبيق عمليات معالجة اللغات الطبيعية على اللغة المصدر للحصول على ترجمة صحيحة لتبسيط عمليات معالجة اللغة الطبيعية اعتمادًا على لغة البايثون، وتناولت نظم معالجة اللغات الطبيعية التي تعمل على القيام بجميع عمليات المعالجة شائعة الاستخدام، والتي تضم مكتبات ونماذج توفر سهولة استخدام وتطبيق عمليات NLP، ومن أهم ما خرجت به الدراسة: هذه الأدوات سجلت نسبة دقة 70٪ لحالات اختبار تم اختيارها عشوائيًا.

وتناولت دراسة (Esposito, Ferilli, Basile, & Mauro, 2008) اقتراح إطار عمل يدعى DOMINUS وهو عبارة عن نموذج للتعلم الآلي يعمل على استثمار التقنيات الذكية لدعم المهام المختلفة للمعالجة التلقائية للمصادر بداية من الحصول على المصادر إلى فهمها وفهرستها وتكشيفها؛ وذلك من أجل تحسين فعالية الاسترجاع، كما يمكن تضمين النظام المقترح كمحرك بحث لإدارة المصادر داخل العديد من نظم إدارة المكتبات الرقمية، والمستودعات الرقمية، ونظم إدارة المؤتمرات العلمية.

وقامت دراسة (Jr., 2007) باستخدام أداة Stanford's JavaNLP tool من أجل إنشاء نموذج ثلاثي لمعالجة اللغة الطبيعية يتم استخدامه من قبل جميع تطبيقات الويب الدلالي؛ حيث يعمل النموذج المقترح كوسيلة لحفظ المصادر، بالإضافة إلى استخدامه لمعالجة

أية استفسارات صادرة عن المستخدم للنظام؛ مما يسمح للنظام بمعالجة أفضل لجميع العلاقات والكيانات الموجودة في النص، كما يسمح المنتج النهائي بالبحث والتكشيف بناءً على هذا المفهوم.

3/8/1 التعليق على الدراسات السابقة:

1- جاءت بدايات الاهتمام بتقنيات الذكاء الاصطناعي في المكتبات بالإنتاج الفكري الأجنبي مبكرًا بينما تأخر ذلك في الإنتاج الفكري العربي.

2- قلة الدراسات العربية التي تناولت بناء نماذج للذكاء الاصطناعي وتطبيقها في العمليات المكتبية.

3-عدم وجود دراسات على مستوى الإنتاج الفكري العربي عملت على الاهتمام باستخدام معالجة اللغة الطبيعية في عمليات تحليل المحتوى لمصادر المعلومات.

4- اختلفت الدراسة الحالية عن الدراسات السابقة في عملها على تطبيق أحد نظم معالجة اللغة الطبيعية في عمليات تحليل المحتوى لمصادر المعلومات؛ مما يسهل عمليات بحثها واسترجاعها.

ثانياً: الإطار النظري للدراسة:

1/2 تقنيات الذكاء الاصطناعي وتعلم الآلة في المكتبات:

1/1/2 تقنيات الذكاء الاصطناعي:

يقوم الذكاء الاصطناعي ككل على مبدأين أساسيين هما (عامر، 2021):

* تمثيل البيانات: يقصد بها طريقة تمثيل المشكلة للحاسب؛ بحيث يتمكن من فهمها، وتقديم حل مناسب لها، ومن اللغات الخاصة بتمثيل البيانات لغة RDF ولغة OWL المستخدمتين في الويب الدلالي.

* البحث: حيث يقوم الحاسب بالبحث في الاختيارات المتاحة له، وتقييمها وفقًا لمعايير موضوعة له مسبقًا، أو ما قام الحاسب باستنباطه بنفسه لاختيار الحل الأنسب.

ويمكن تقسيم موجات ظهور تقنيات الذكاء الاصطناعي إلى ثلاث موجات حسب التسلسل الزمني لتطورها، كالتالي (Boucher, 2020):

* **الموجة الأولى**: تصف تقنيات الذكاء الاصطناعي المبكرة، وهي تشير إلى تطوير آلات ذكية من خلال ترميز معرفة وخبرة الخبراء في مجموعات من القواعد التي يمكن للحاسب تنفيذها، وتضم: النظم الخبيرة والمنطق الضبابي fuzzy logic، ففي النظم الخبيرة يقوم خبير بشري في مجال التطبيق لقواعد دقيقة تسمى خوارزميات يمكن للحاسب اتباعها، خطوة بخطوة، لتحديد كيفية الاستجابة بذكاء لموقف معين، أما المنطق الضبابي Fuzzy Logic: هو نهج آخر للنظم الخبيرة التي تسمح للمتغيرات أن يكون لها "قيمة حقيقة" تتراوح بين 0 و 1، والتي تلتقط مدى تناسب الفئة.

* **الموجة الثانية**: تصف الموجة الثانية الأساليب الأكثر حداثة "المستندة إلى البيانات" والتي تطورت بسرعة على مدار العقدين الماضيين وهي تضم التعلم الآلي والذكاء الاصطناعي القائم على البيانات؛ ويشير التعلم الآلي (ML) إلى مجموعة واسعة من التقنيات

التي تعمل على ميكنة عملية تعلم الخوارزميات، وهذا يختلف عن نهج الموجة الأولى حيث يتم تحسين الأداء بالخوارزمية من خلال تدريب نفسها على البيانات.

* **الموجة الثالثة**: الموجات المستقبلية المحتملة للذكاء الاصطناعي "نحو الذكاء الخارق superintelligence " وهو يشير إلى الخوارزميات التي يمكن أن تظهر الذكاء في مجموعة واسعة من السياقات ومساحات المشكلات؛ ونظرًا لأنه غير موجود بعد، فهو ينتمي إلى عالم الذكاء الاصطناعي التأملي، ومن المفترض أنه سيأتي بمستويات أعلى من الذكاء العام للبشر العاديين.

2/1/2 الذكاء الاصطناعي ومجالات استخدامه في المكتبات:

يعد الذكاء الاصطناعي هو التقنية الحديثة التي تستخدم لإدارة المكتبة الرقمية، ويتمثل المنتج النهائي لها في تطوير نظم أو آلات تفكر وتتصرف مثل البشر، وتنافس الذكاء البشري، وتعتمد النظم الآلية للمكتبات الذكية على تقنيات الذكاء الاصطناعي؛ لتقديم خدمات قائمة على المعرفة لمستخدمي المكتبات وموظفيها، وهي تعمل على تقليل التدخل البشري في أداء المهام الروتينية بالمكتبة وتوفير وقت أخصائي المعلومات للتواصل مع المستفيدين(Sridevi & Shanmugam, 2017) .

ويمكن تحديد مجالات استخدام الذكاء الاصطناعي في المكتبات في المجالات التالية (Omame & Alex-Nmecha, 2020):

* **النظم الخبيرة Expert system**: هي التي تحاكي جانب صنع القرار للدماغ البشري، ويمكن إنشاء نظام خبير في العمليات المكتبية مثل: التزويد، والفهرسة، والخدمة المرجعية.

* **معالجة اللغة الطبيعية Natural Language Processing**: يمكن استخدامها في معالجة المحتوى؛ مما يؤدي إلى تحسين عمليات البحث والاسترجاع.

* **التعرف على الأنماط Pattern Recognition**: يركز على التفاعلات بين الإنسان والتطبيقات الصديقة للمستخدم وسهلة الاستخدام والرسومية.

* **الروبوتات Robotics**: وهي تتعامل مع المهام الحركية والإدراكية، ويمكن استثمارها في المكتبات لأداء العديد من الأنشطة مثل: عمليات التنظيم والاسترجاع للمصادر، وعمليات الجرد.

ويتم التركيز في الدراسة الحالية على نظم معالجة اللغة الطبيعية – أحد تقنيات الذكاء الاصطناعي – ودورها في تحليل المحتوى لمصادر المعلومات، وآليات عملها، واستخداماتها داخل المكتبات.

3/1/2العلاقة بين الذكاء الاصطناعي وتعلم الآلة ومعالجة اللغة الطبيعية :

يتمثل دور التعلم الآلي والذكاء الاصطناعي في معالجة اللغة الطبيعية وتحليلات النصوص في تحسين وتسريع وميكنة وظائف تحليلات النص الأساسية مع الاستفادة من مزايا معالجة اللغة الطبيعية التي تحول هذا النص غير المنظم إلى بيانات ورؤى قابلة للاستخدام.

وتعتمد خوارزميات معالجة اللغة الطبيعية–عادةً–على خوارزميات التعلم الآلي، فبدلا من ترميز مجموعات كبيرة من القواعد يدويًا، فإن معالجة اللغة الطبيعية يمكن أن تعتمد على التعلم الآلي لتعلم هذه القواعد تلقائيًا عن طريق تحليل مجموعة من المصادر مثل: كتب أو مقالات، وعمل استنتاج إحصائي، وإنشاء نموذج للتعلم الآلي، ونموذج التعلم الآلي هو مجموع التعلم الذي تم الحصول عليه

من بيانات التدريب الخاصة به، ويتغير هذا النموذج مع اكتساب المزيد من التعلم وبشكل عام، كلما زادت البيانات التي يتم تحليلها، كلما كان النموذج أكثر دقة ("Introduction natural language processing nlp," 2016).

وعلى عكس خوارزميات البرمجة، فإن نموذج التعلم الآلي قادر على التعميم والتعامل مع الحالات الجديدة، إذا كانت الحالة تشبه شيئًا ما شاهده النموذج من قبل، فيمكن للنموذج استخدام "التعلم" السابق لتقييم الحالة؛ فالهدف هنا هو إنشاء نموذج يمكن تحسينه باستمرار للقيام بمهمته، ويتضمن التعلم الآلي لـ NLP والتحليلات النصية مجموعة من التقنيات الإحصائية لتحديد أجزاء الكلام، والكيانات، والمشاعر، والجوانب الأخرى للنص، ويمكن التعبير عن التقنيات كنموذج يتم تطبيقه بعد ذلك على نص آخر، ويُعرف أيضًا باسم التعلم الآلي الخاضع للإشراف، ويمكن أن تكون أيضًا مجموعة من الخوارزميات التي تعمل عبر مجموعات كبيرة من البيانات لاستخراج المعنى، وهو ما يُعرف باسم التعلم الآلي غير الخاضع للإشراف.(Barba, 2020)

ففي التعلم الآلي الخاضع للإشراف، يتم وضع التيجان لمجموعة من المستندات النصية أو وضع تعليقات توضيحية عليها بأمثلة لما يجب أن تبحث عنه الآلة وكيف يجب أن تفسر هذا الجانب، تُستخدم هذه الوثائق "التدريب" نموذجا إحصائيا، أما التعلم الآلي غير الخاضع للإشراف فيتضمن تدريب نموذج دون وضع علامات أو تعليقات توضيحية مسبقًا، وهو يعتمد على تجميع المصادر المتشابهة معًا في مجموعات أو فئات، ثم يتم فرز هذه المجموعات على أساس الأهمية والملاءمة (Emms & Luz, 2007).

2/2 معالجة اللغة الطبيعية ودورها في تحليل المحتوى:

1/2/2 تقنيات معالجة اللغة الطبيعية:

تقع معالجة اللغة الطبيعية بين عدد من التخصصات، منها: علوم الحاسب والمعلومات، واللغويات، والرياضيات، والهندسة، والذكاء الاصطناعي والروبوتات، وعلم النفس، وما إلى ذلك، وتشمل تطبيقاتها عددًا من المجالات، مثل: الترجمة الآلية، ومعالجة النصوص باللغة الطبيعية وتلخيصها، وواجهات المستخدم، واسترجاع المعلومات بلغات متعددة، والتعرف على الكلام، والذكاء الاصطناعي والنظم الخبيرة، وما إلى ذلك(Chowdhury, 2003) .

ولقد تم تطوير دراسات معالجة اللغة الطبيعية منذ السنوات الخمسين الماضية بداية من تطبيقات للترجمة الآلية واسترجاع المعلومات حتى الآن، وقد اتسع نطاق هذين المجالين من البحث وانتشرا على نطاق واسع، وفي إطار عملية حل القضايا المتعلقة بفهم اللغة الطبيعية، بالنسبة لكل من الترجمة والاسترجاع ظهرت العديد من المجالات الفرعية مثل: التحليل الصرفي، والجناس، وتحليل الجمل، والمعاني الدلالية، وتوليد اللغة الطبيعية والتعرف على الكيانات (Karoo, 2018).

وتنبع أهمية معالجة اللغة الطبيعية بأنها تساعد أجهزة الحاسب على التواصل مع البشر بلغاتهم الخاصة وتوسيع نطاق أنواع المهام الأخرى فيما يتعلق باللغة، فعلى سبيل المثال، تعمل معالجة اللغة الطبيعية على تمكين جميع أجهزة الكمبيوتر من قراءة النصوص، وسماع الكلمات، وتفسيرها، وقياس المشاعر، وكذلك تحديد الأجزاء المهمة التي يجب الاهتمام بها في النص (Redondo & Sandoval, 2016).

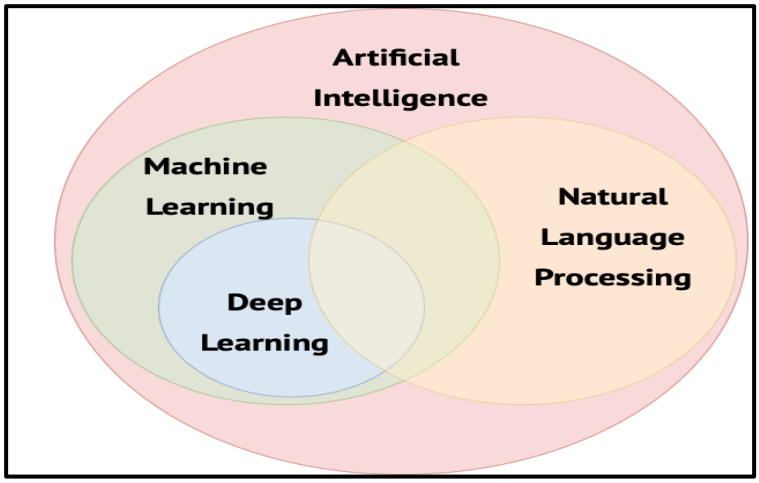
وإذا كانت اللغات البشرية تعد معقدة للغاية ومتنوعة في طبيعتها للدرجة التي تجعلنا نعبر عن أنفسنا بعدة أنواع من الطرق سواء شفهيًا أو كتابيًا؛ لذا تعد معالجة اللغة الطبيعية مهمة جدًا في الذكاء الاصطناعي؛ لأنها تساعد في حل الغموض في اللغة، وإضافة بنية مفيدة للبيانات داخل العديد من أنواع التطبيقات مثل: التعرف على الكلام أو تحليلات النصوص (Orobor, 2016).

وتعد معالجة اللغة الطبيعية هي إحدى الطرق المستخدمة في التنقيب عن النصوص؛ من أجل فحص عدد كبير من المصادر النصية لإنتاج معرفة جديدة، وأدوات تنظيم المعلومات كالمكانز، والمعاجم، والأنطولوجيات كلها تتطلب استخدام نظم معالجة لغة طبيعية كي تؤدي مهامها ووظيفتها (Fisher, Garnsey, Goel, & Tam, 2010)، يتضمن تحليل المحتوى اندماج عدد من التكنولوجيات وهي (Tyagi, 2021):

* **الذكاء الاصطناعي**: قدرة نظام الحاسب على تنفيذ الأنشطة التي تتطلب عادةً ذكاءً بشريًا، وتتضمن هذه المهام التعرف على الصوت واتخاذ القرار، ويستخدم في معالجة كميات كبيرة من النصوص وتصنيفها تلقائيًا.

* **تعلم الآلة**: أحد مكونات الذكاء الاصطناعي الذي يركز على قدرة خوارزميات الحاسب على التعلم الفعال من التجربة، والتكيف تلقائيًا لتحسين الأداء دون برمجة بشرية، وتستخدم تحليلات النصوص Text Analytics التعلم الآلي لتحديد كيفية تصنيف الأجزاء النصية الجديدة بناءً على النص الذي تمت معالجته مسبقًا، ولتقييم ما إذا كان يجب تحسين الفئات المستخدمة لتصنيف هذه الأجزاء النصية وفقًا للأنماط المحددة.

* **التعلم العميق**: مجموعة فرعية متخصصة ومضبوطة من الذكاء الاصطناعي تشمل قدرة نظام الحاسب على تحليل البيانات وإصدار الأحكام على البيانات الأخرى في تحليلات النص، ويمكن استخدام التعلم العميق لفهم السياق بشكل أفضل في التعليقات غير المنظمة ولتعزيز دقة التحليل التلقائي للنص.



شكل (1) الاندماج بين الذكاء الاصطناعي وتعلم الآلة ومعالجة اللغة الطبيعية والتعلم العميق

وهناك نوعان أساسيان من تقنيات معالجة اللغة الطبيعية، وهما: التحليل النحوي والتحليل الدلالي (Tyagi, 2021):

أولًا: **التحليل النحوي**: يفحص النص باستخدام المبادئ النحوية الأساسية لاكتشاف بنية الجملة وترتيب الكلمات والارتباط بينها، ويشمل بعض المهام الفرعية مثل:

1. **الترميز**: يتضمن تقسيم النص إلى أجزاء أصغر تسمى الرموز (والتي قد تكون عبارات أو كلمات) لتبسيط التعامل مع المواد.
2. **أجزاء من علامات الكلام Part of the speech**: مثل فعل، ظرف، صفة، مادة، إلخ، وهذه العلامات تساعد في تحديد معنى الكلمات.

3. الاشتقاق Lemmatizaton: يتضمن اختزال العبارات إلى شكلها الأساسي لتسهيل التحليل.

4. كلمات التوقف Stop-word: تتضمن حذف الكلمات التي لا تضيف قيمة للنص.

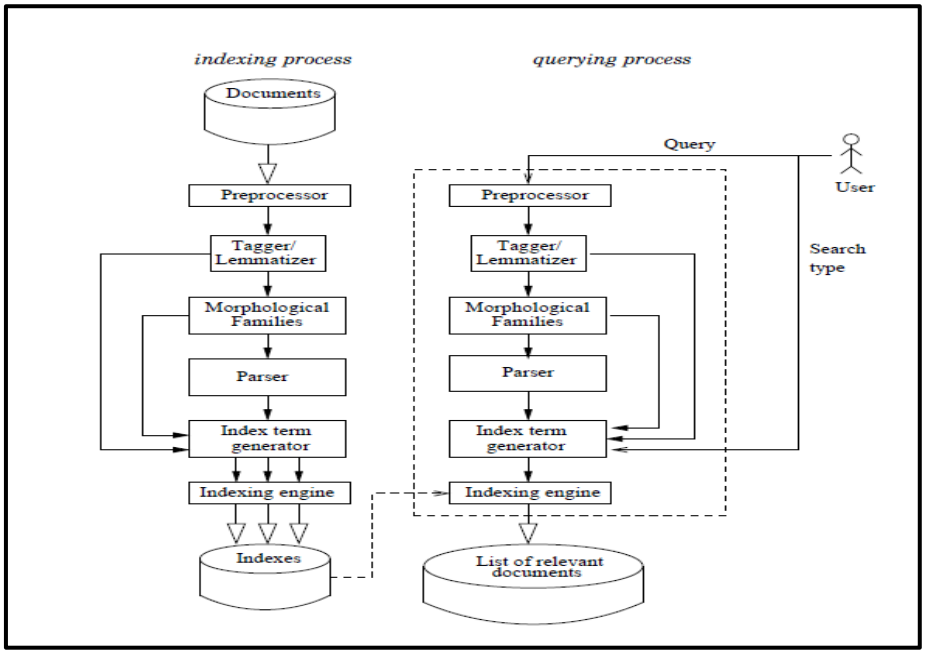
ثانيًا: التحليل الدلالي: يعتمد على التقاط النص ثم فحص دلالة كل مصطلح (الدلالات المعجمية)، لكي يتم كشف الغموض عن المصطلحات بتحديد المعنى الذي تُستخدم فيه الكلمة في سياق معين.

وتعد معالجة النصوص من أجل استخلاص المعرفة لأغراض التكشيف الآلي وتحليل المحتوى حقلًا مهمًا في مجال بحوث معالجة اللغة الطبيعية؛ حيث تم إدراك دورها مبكرًا في عمليات تحليل المحتوى مع إدخال الكثير من التحسينات إليها مثل: تقنيات الويب الدلالي؛ وذلك نظرًا للقيمة الاقتصادية المتزايدة للمعلومات الرقمية، ويتم تصنيف هذا على نطاق واسع بأنه مجال معالجة نصوص اللغة الطبيعية التي تسمح بهيكلة مجموعات كبيرة من المعلومات النصية؛ بهدف استرجاع معلومات معينة أو لاشتقاق هياكل المعرفة التي يمكن استخدامها لغرض معين(Karoo, 2018).

وتسير معالجة اللغة الطبيعية جنبًا إلى جنب مع تحليلات النصوص التي تعمل على حساب الكلمات وتجميعها وتصنيفها لاستخراج المعلومات من المحتوى، ويتم استخدام تحليل النصوص لاستكشاف المحتوى النصي، واشتقاق نوع جديد من المتغيرات من البيانات غير المهيكلة التي يمكن عرضها وفلترتها واستخدامها بشكل أكبر في شكل مدخلات للتنبؤ بالنماذج وأنواع أخرى من الأساليب الإحصائية (Tyagi, 2021).

2/2/2 آلية عمل نظم معالجة اللغة الطبيعية في عمليات تحليل المحتوى:

قد يبدأ نظام معالجة نصوص اللغة الطبيعية بالتحليلات الصرفية morphological analyses، ويتم اشتقاق المصطلحات Stemming of terms، في كل من الاستعلامات والوثائق، تتضمن المعالجة المعجمية والنحوية استخدام المعاجم لتحديد خصائص الكلمات، والتعرف على أجزاء كلامها، وتحديد الكلمات والعبارات، وتحليل الجمل (Chowdhury, 2003)، ويوضح الشكل التالي آلية عمل نظم معالجة اللغة الطبيعية في عمليات التكشيف وتحليل المحتوى:



شكل (2) آلية عمل نظم معالجة اللغة الطبيعية في عمليات التكشيف وتحليل المحتوى (Vilares et al., 2002)

حيث يتم إدخال مجموعة الوثائق إلى نظام معالجة اللغة الطبيعية ثم يتم إجراء مجموعة من العمليات عليها، هي (Vilares et al., 2002):

1- المعالجة الأولية The Preprocessor:

وتتضمن العمليات التالية:

* الفلترة: يتم تحويل النصوص من تنسيق المصدر (مثل HTML أو XML) إلى نص بسيط plain text، وإزالة الفراغات من النص.

* الترميز: يتم تقسيم الجمل إلى كلمات يفصل بينها مجموعة من الرموز المحددة مع مراعاة الاختصارات والاستهلالات والأرقام.

* تجزئة الجملة: يتم الفصل بين الجمل من خلال محددات معينة، مثل: النقطة المتبوعة بحرف كبير.

* المعالجة الصرفية: يتم تحليل الكلمة لمعرفة جذرها، ووزنها الصرفي، وما اتصل بها من سوابق، وما لحقها من لواحق.

2- محدد العلامات The Tagger:

يتم استخدام نموذج ماركوف المخفي Hidden Markov Model (HMM)، من أجل التعرف على الأنماط مثل: أجزاء الكلام والكتابة اليدوية، وبمجرد وضع العلامات على النص يتم استخراج كلمات المحتوى: (الأسماء، الأفعال، الصفات) ليتم تكشيفها.

3- العائلات الصرفية Morphological Families:

يتم تحديد مجموعة من الكلمات التي تم الحصول عليها من نفس الجذر الصرفي من خلال آليات الاشتقاق؛ و من المتوقع أن تظل العلاقة الدلالية الأساسية بين كلمات عائلة معينة، بالإضافة إلى ربط هذه المصطلحات ببعض بالكشاف.

4- المحلل اللغوي The Parser:

يتم تحديد البنية النحوية للنص عن طريق تحليل الكلمات المكونة له بناءً على القواعد النحوية الأساسية.

5- إنشاء مصطلحات الكشاف Index:

عندما يتم إجراء استعلام من قبل المستخدم يتم البحث داخل النظام وإجراء الخطوات السابقة على استعلام المستخدم ليتم استرجاع الوثائق ذات الصلة بموضوع البحث.

3/2/2 استخدامات معالجة اللغة الطبيعية :

* التعرف على الكلام: Speech recognition

يتم استخدام هذه التقنية من أجل إدخال أي كلام صوتي مستمر، وإخراج المكافئ النصي؛ و أصبح من الممكن إعطاء الأوامر الصوتية أو طرح بعض الأسئلة بشكل صوتي دون الحاجة إلى الكتابة (Sawhney, 2021).

* تصنيف النصوص: Text Classification

باستخدام معالجة اللغات الطبيعية، يمكن لتصنيف النص تحليل النص تلقائيًا ثم تعيين مجموعة من التيجان أو الفئات المحددة مسبقًا بناءً على سياقها، وهناك ثلاثة طرق لتصنيف النصوص وهي: النظم القائمة على القواعد، ونظم الآلة، والنظم الهجينة (Neupane, 2020).

* استخلاص المعلومات: Information Extraction

يتم استخدام مجموعة متنوعة من التقنيات مثل : الترميز، وتحديد أجزاء من الكلام؛ لتحليل البيانات غير المهيكلة لاستخراج المعلومات الأساسية من المصدر، ويمكن استخدام هذه المعلومات بعد ذلك لإعداد ملخصات للنصوص، واستكمال الميتاداتا لقواعد البيانات، وتحديد الكلمات الرئيسية، وما إلى ذلك(Kurama, 2021).

* تحليل النصوص Text analytics:

على الرغم من استخدامها بالتبادل في بعض الأحيان مع التنقيب عن النصوص، إلا أن هناك تمييزًا بينهما حيث يوفر تحليل النص مزيدًا من المعلومات الكمية المتعمقة لاتخاذ القرارات، فإذا ما تم تطبيقه على مدخلات غير مهيكلة، فإن تحليل النص يقدم نظرة ثاقبة للاتجاهات والأنماط والمشاعر لاكتشاف الأساليب وتحديد أولوياتها (Lee, 2002).

* الترجمة الآلية: Machine Translation

تقدمت تقنية الترجمة الآلية بشكل كبير؛ حيث تسمح هذه التقنيات بالتفاعل بلغات مختلفة، وتعزيز الاتصالات العالمية دون وجود عوائق اللغة(Fisher et al., 2010).

* تلخيص النص الآلي: Automatic Text Summarization

وذلك بهدف إنشاء ملخصات للنصوص؛ حيث يتم في هذه النظم تحديد العبارات المهمة في النص واستغلالها للكشف عن المعلومات ذات الصلة لإضافتها في الملخص بناءً على بعض المعايير اللغوية و / أو الإحصائية، وتقدم هذه التقنية عددًا من القضايا ، بما في ذلك تحديد النص والتفسير وتوليد الملخص ، بالإضافة إلى تحليل الملخص الناتج، (Mishra, 2022).

* تحليلات الشبكات الاجتماعية: Social Media Analytics:

تتبع الوعي بما في ذلك المشاعر المتعلقة بأنواع معينة من الموضوعات وتحديد الأشخاص المؤثرين الرئيسيين (Lee, 2002).

4/2/2 استخدامات معالجة اللغة الطبيعية في المكتبات:

يمكن تقسيم استخدامات تطبيقات معالجة اللغة الطبيعية في المكتبات إلى أربع قطاعات، كالتالي (Karoo, 2018):

* الحصول على المصادر:

يمكن استخدام تقنيات NLP في اختيار المصادر بشكل آلي من خلال تعريف معايير اختيارها للنظام والسياسات التي تم تحديدها من قبل المكتبة ليتم تطبيقها في عملية التزويد، بالإضافة إلى المعالجة الوصفية للمصادر واستخراج عناصر الميتاداتا.

* معالجة المحتوى:

تعد معالجة المحتوى جزءً رئيسيًا في معالجة اللغة الطبيعية، وهي تتمثل في إنتاج بيانات وصفية محسنة، من أجل تسهيل استرجاع المصادر من قبل المستخدمين، بالإضافة إلى إمكانات الاسترجاع التي يوفرها البحث عن النص الكامل، وتتضمن عملية معالجة المحتوى إجراء تحليل للمحتويات اللغوية و / أو المفاهيمية، وإنتاج معرفات مناسبة لـهذه المصادر مثل: مصطلحات الفهرسة، والملخصات: ورموز التصنيف،)، بالإضافة إلى تلك المهام التقليدية التي تغطيها معالجة اللغة الطبيعية؛ فهي تعمل على تكشيف المحتوى؛ وتقدم وصفا للمصادر مع قائمة مختصرة من المصطلحات أو الكلمات الرئيسية التي تمثل الموضوعات الرئيسية التي تمت مناقشتها في مصدر المعلومات، كما تقدم تلخيصا للمحتوى الموجود.

* إحاطة المستفيدين بالمصادر والإجابة على الاستفسارات:

يمكن لنظم معالجة اللغة الطبيعية المساعدة في تقديم خدمة المراجع من خلال غرف الدردشة وخدمات (اسأل أمين المكتبة) واستخدام الـشات بوت Chatbot.

* أدوات تنظيم المعرفة:

يمكن تحقيق الاستفادة من أدوات تنظيم المعرفة مثل: المكانز في مهام تحليل المحتوى مثل: التكشيف، والتصنيف، والتلخيص؛ من أجل الكشف عن العلاقات الدلالية بين المصطلحات، ومن ثم تحسين مستوى المعالجة.

6/2 المنصات والمكتبات البرمجية للذكاء الاصطناعي التي يمكن الاستفادة منها في مجال تحليل المحتوى:

يمكن استخدام أدوات معالجة اللغة الطبيعية من خلال أدوات SaaS (البرنامج كخدمة software as a service) أو من خلال المكتبات البرمجية، وبعض هذه الأدوات مفتوح المصدر والبعض الآخر تجاري وقليل منها ما هو متاح للباحثين داخل مراكز البحوث فقط؛ فأدوات SaaS هي حلول جاهزة للاستخدام قائمة على السحابة يمكن تنفيذها بعدد قليل من الأكواد، وهي غالبًا ما تقدم نماذج مدربة مسبقًا في معالجة اللغة الطبيعية والتي يمكن استخدامها دون أكواد، وهي موجهة أكثر نحو أولئك الذين يريدون خيارًا أكثر مرونة وأقل كودًا، أما المكتبات البرمجية فهي تسمح بالتخصيص الكامل لأدوات معالجة اللغة الطبيعية، إلا أنها تستهدف المطورين، لأنها تحتاج إلى خبرة في التعلم الآلي لبناء أدوات معالجة اللغة الطبيعية (The App Solutions Inc., 2021)

1/6/2 منصات معالجة اللغة الطبيعية:

هناك مجموعة من منصات معالجة اللغة الطبيعية المتاحة للتطبيق والتي يمكن استخدامها في تحليل المحتوى، ويعرض الجدول التالي مقارنة بين أشهر هذه المنصات، كالتالي:

جدول (1) منصات معالجة اللغة الطبيعية المستخدمة في تحليل المحتوى

اسم المنصة	اللغة البرمجية	الإمكانيات والوظائف	اللغات التي تدعمها	الإتاحة	نوع الترخيص
MonkeyLearn (MonkeyLearn Inc., 2020)	Python , Ruby , PHP , Java	تستخدم لأداء مهام تحليل النصوص مثل تحليل المشاعر أو تصنيف الموضوعات أو استخراج الكلمات الأساسية. يمكن للمستخدم إنشاء نموذج تعلم آلي مخصص بواسطتها.	الإنجليزية، الإسبانية، الفرنسية، الألمانية، العربية، اليابانية، الصينية، الروسية، الكورية، السويدية، الإيطالية، الرومانية.	تجارية مع توفر ديمو تجريبي.	GDPR
Aylien (Aylien Ltd., 2020)	Python, Ruby	تستخدم التعلم العميق ومعالجة اللغة الطبيعية لتحليل كميات كبيرة من البيانات المستندة إلى النصوص، مثل المحتوى الأكاديمي والمحتوى في الوقت الفعلي من المنافذ الإخبارية وبيانات الوسائط الاجتماعية. تستخدم في تلخيص واستخراج معلومات المقالات واستخراج الكيانات وتحليل المشاعر.	الإنجليزية – الإسبانية – الفرنسية – الألمانية – العربية – اليابانية – الصينية – الروسية – الفارسية – السويدية – الإيطالية – التركية.	تجارية مع توفر ديمو تجريبي	--
IBM Watson ("About IBM Watson," n.d)	Python, R , OpenCL	يتميز بقدرته على فهم اللغة الطبيعية، والتي تسمح بتحديد واستخراج الكلمات الرئيسة والفئات والمشاعر والكيانات وغيرها.		تجارية مع توفر خطة مجانية بسمات محددة.	--
Google Cloud Natural Language ("Cloud Natural Language documentation")	Go , Java, Node.JS, PHP, and Python (UpBuild)	يستخدم لتحليل المشاعر وتصنيف المحتوى واستخراج الكيانات. يوفر نماذج للتعلم الآلي AutoML Natural Language، والتي تسمح ببناء نماذج مخصصة لتعلم الآلة.	الإنجليزية – الإسبانية – اليابانية – الصينية – الفرنسية – الألمانية – لإيطالية – الكورية – البرتغالية – الروسية	تجارية مع توفر خطة مجانية بسمات محددة	

2/6/2 المكتبات البرمجية لمعالجة اللغة الطبيعية:

هناك مجموعة من المكتبات البرمجية لمعالجة اللغة الطبيعية المتاحة للتطبيق، والتي يمكن استخدامها ضمن إطار عمل Platform في تحليل المحتوى، ويعرض الجدول التالي مقارنة بين أشهر هذه المكتبات، كالتالي:

جدول (2) المكتبات البرمجية لمعالجة اللغة الطبيعية

اسم المكتبة	اللغة البرمجية	الإمكانيات والوظائف	اللغات التي تدعمها	الإتاحة	نوع الترخيص
Stanford CoreNLP (Stanford NLP Group, 2020b)	Java	تُمكّن المستخدمين من اشتقاق التعليقات التوضيحية اللغوية للنص، بما في ذلك الترميز وأجزاء الجملة، وأجزاء الكلام، والكيانات المسماة، والقيم الرقمية والوقت، وتحليل العلاقات والمشاعر، وخصائص الاقتباس.	العربية، والصينية، والإنجليزية، والفرنسية، والألمانية، والمجرية، والإيطالية	مفتوحة المصدر	GNU General Public License Stanford NLP) (Group, 2020a
Natural Language Toolkit Documentation - ") Natural Language (Toolkit," n.d	Python	تحتوي على الكثير من الخوارزميات التي تتعامل مع النصوص. توفر واجهات سهلة الاستخدام لأكثر من 50 مجموعة بيانات ومعجم مثل WordNet، جنبًا إلى جنب مع مجموعة من مكتبات معالجة النصوص للتصنيف والترميز والاشتقاق والعلامات والتحليل والتفكير الدلالي	الدنماركية، الهولندية، الإنجليزية، الفنلندية، الفرنسية، الألمانية، المجرية، الإيطالية، النرويجية، البرتغالية، الرومانية، الروسية، الإسبانية، السويدية (Devopedia, 2019)	مفتوحة المصدر	Apache 2.0 License
Apache OpenNLP Apache OpenNLP ") Developer (Documentation," n.d	JAVA	تدعم مهام NLP الأكثر شيوعًا، مثل الترميز، وتجزئة الجملة، وعلامات جزء من الكلام، واستخراج الكيانات المسماة، والتحليل، عادة ما تكون هذه المهام مطلوبة لإنشاء خدمات معالجة نصية أكثر تقدمًا. من أهم مزاياها: التعرف على الكيانات المسماة، والتلخيص حيث يمكن تلخيص فقرات أو مقالات، والبحث داخل نص معين الترجمة إلى لغات أخرى. توليد اللغة الطبيعية من قواعد البيانات (OpenNLP- Overview," n.d")	لا تدعم لغة بعينها	مفتوحة المصدر	Apache License, Version 2.0
SPACY ("spacy - Industrial-Strength Natural Language Processing ", n.d)	Python	تعمل على التحليل النحوي، وهو مفيد لتحليل المشاعر وتحسين واجهة مستخدم المحادثة، والتعرف على الكيانات المسماة، وتصنيف النصوص. تستخدم في التحليل العميق للنصوص وتلخيص النصوص. تحتوي على العديد من نماذج التدريب	تدعم أكثر من 64 لغة	مفتوحة المصدر	MIT License
Linguakit (Otero, 2019)	Perl	تمتاز بإمكانية التعرف على الكيانات المسماة، وتصنيف الكيانات، وتحليل المشاعر، واستخراج الكلمات الرئيسية، والتعرف على العلاقات، وتجزئة الجملة، والتلخيص، تدقيق اللغة من حيث التهجئة والأخطاء النحوية، التعرف الدلالي.	البرتغالية والإنجليزية والإسبانية البرتغالية	مفتوحة المصدر	GNU General Public License v3.0

3/6/2 التعليق على أبرز المنصات والمكتبات البرمجية لمعالجة اللغة الطبيعية:

هناك العديد من أدوات معالجة اللغة الطبيعية المتاحة والمتوفرة على الويب، والتي تنوعت بين منصات سهلة الاستخدام ومكتبات برمجية لدمجها ضمن منصات لاستخدامها، وتسمح تلك الأدوات بتحليل المحتوى والتعرف على الكيانات والعلاقات بينها، واعتمدت كثير من تلك الأدوات على لغة البرمجة بايثون باعتبارها اللغة الشائعة في عالم الذكاء الاصطناعي، وبعضها اعتمد على لغة الجافا وغيرها من اللغات البرمجية.

ثالثاً: الإطار التطبيقي للدراسة:

تطبيق أحد نظم معالجة اللغة الطبيعية لتحليل محتوى مصادر المعلومات:

1/3 الأداة المستخدمة في التطبيق:

تم اختيار منصة IBM Watson – الخطة المجانية منها – والتي تعد من أشهر منصات الذكاء الاصطناعي التي تدعم معالجة اللغة الطبيعية للعمل على تحليل المحتوى نظرًا لما تتميز به من قدرتها على فهم اللغة الطبيعية، والتي تسمح بتحديد واستخراج الكلمات الرئيسية والفئات والمشاعر والكيانات وغيرها، واعتمادها على لغة PYTHON أشهر اللغات البرمجية في نظم الذكاء الاصطناعي، وكذلك لسهولة استخدامها والمرونة التي تتميز بها، وتوفر الأدلة الإرشادية للتطبيق.

وتم العمل على تطبيق أداتين من ضمن الأدوات التي تقدمها منصة IBM Watson وهما :

1- أداة IBM Watson Knowledge Studio : تستخدم لتحديد الكيانات والعلاقات فيما بينها داخل النصوص، وتمكن من إنشاء نماذج التعلم الآلي ، وتم استخدامها وتطبيقها من أجل تحليل محتوى مجموعة من المقالات باللغة العربية والإنجليزية في موضوع "إدارة المعرفة" Knowledge Management، وإنشاء نموذج للتعلم الآلي لتحديد الكيانات والعلاقات فيما بينها.

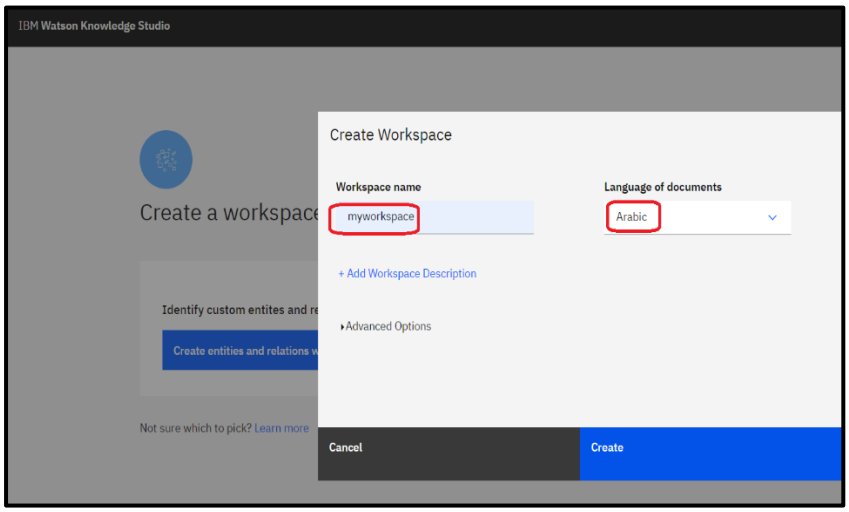
2- أداة IBM Watson Discovery : تستخدم في بحث وتحليل المحتوى واسترجاع النتائج، وتم من خلالها تطبيق نموذج تعلم الآلة – الذي تم إنشاؤه بواسطة أداة IBM Watson Knowledge Studio – على مجموعة من المقالات الأخرى في نفس المجال؛ من أجل تحسين عمليات تحليل المحتوى، وبحثه، واسترجاعه.

2/3 خطوات العمل:

تم تقسيم العمل على المنصة إلى مرحلتين، هما :

1/2/3 المرحلة الأولى: تطبيق أداة Waston Knowledge Studio

1. إنشاء مساحة العمل التي سيتم داخلها تهيئة أدوات العمل وتحليل محتوى المقالات.

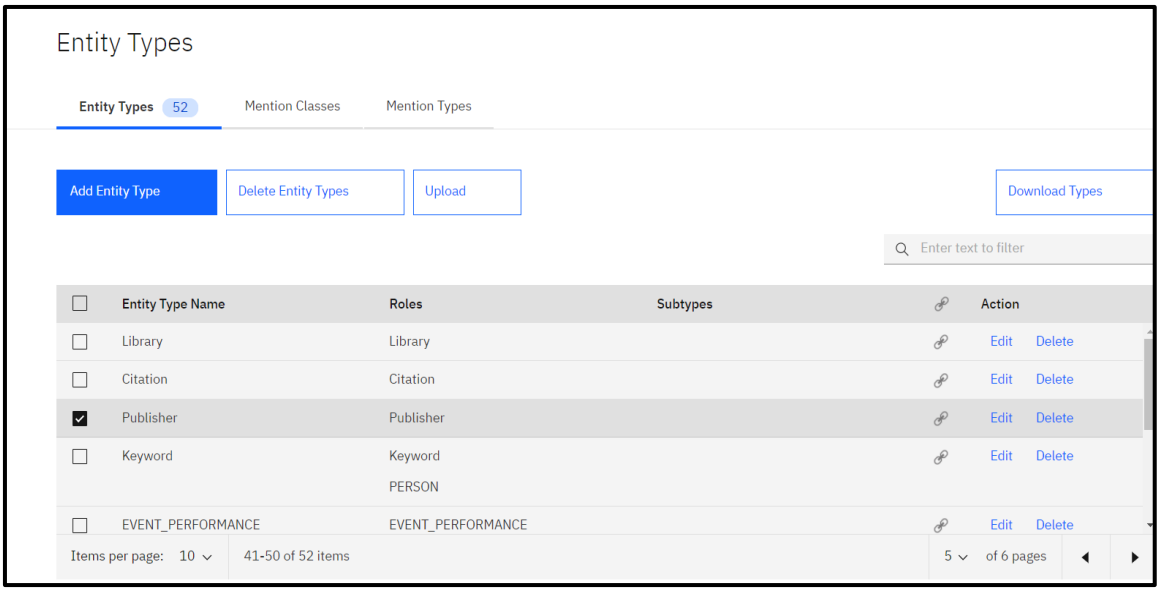


شكل (3) إنشاء مساحة العمل Workspace داخل المنصة.

2. إنشاء ملف typesytem وهو الملف الذي يعرف الكيانات والعلاقات بصيغة json وتحميله على النظام، وقد تم تعريف 15 مفردة داخل الملف بالإضافة إلى المفردات المتاحة على النظام، وهي:

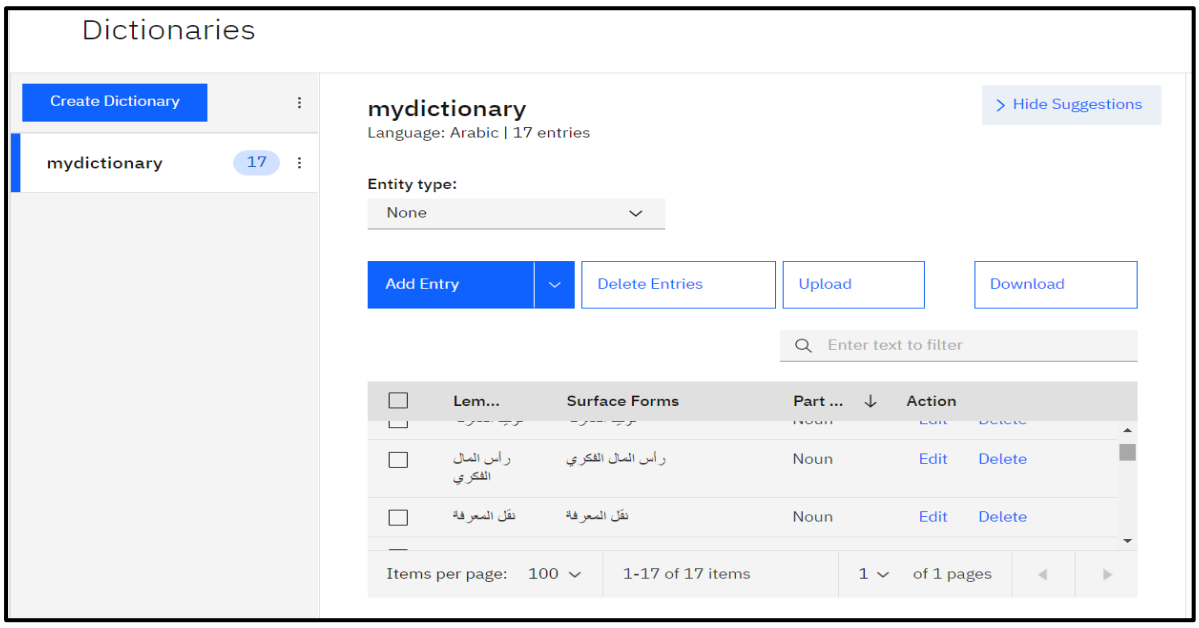
جدول (3) ملف تعريف الكيانات بصيغة JSon.

Author المؤلف	Keywords الكلمات المفتاحية
Title عنوان العمل	E-mail البريد الإلكتروني
Place مكان	Citation الاستشهاد
Publisher ناشر	Persons أشخاص
Organization مؤسسة	Libraryمكتبة
System برنامج/ نظام	Date تاريخ
Information Resource مصادر معلومات	Time_Period فترة زمنية
	Database قاعدة بيانات



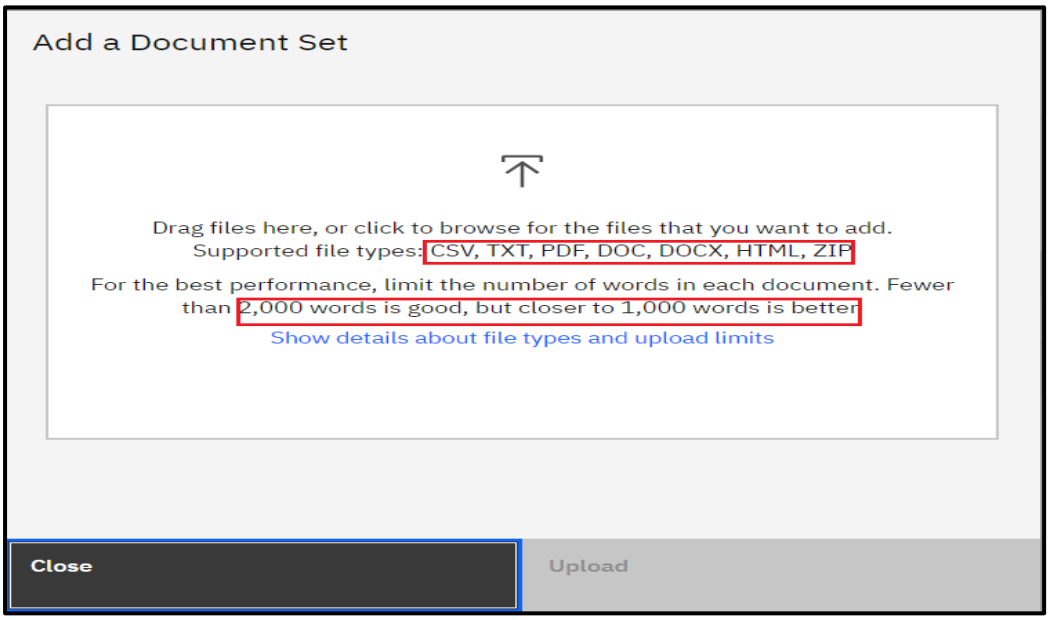
شكل (4) ملف أنواع الكيانات التي تم تحديدها

3. إنشاء القاموس Dictionary لمساعدة محلل المحتوى human annotator على البدء في مهام تحليل المحتوى، و تم إنشاء قاموس مبسط مزود بمجموعة من المصطلحات في مجال إدارة المعرفة.



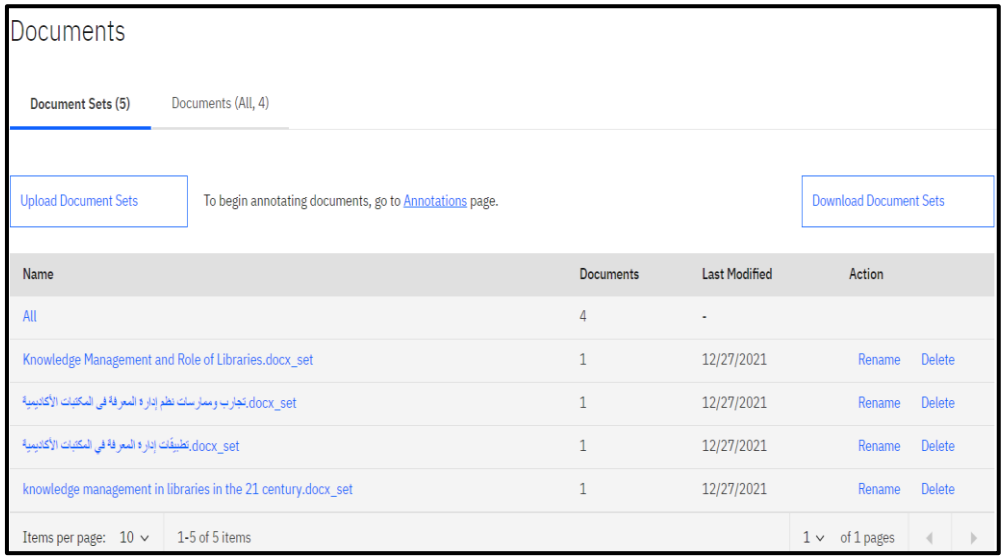
شكل (5) إنشاء قاموس المصطلحات في مجال إدارة المعرفة

4. إضافة مجموعات المقالات من أجل فحصها وتحليل محتواها، فمن أجل تدريب نموذج التعلم الآلي ML Model، يجب إضافة المقالات التي تمثل محتوى المجال الذي تم اختياره، وهو (إدارة المعرفة) إلى مساحة العمل؛ و تم تحميل المقالات التي تم اختيارها في شكل.word ، مع ملاحظة قبول النظام الملفات من نوع CSV, TXT, PDF, DOC, DOCX, HTML, ZIP ويفضل ألا يزيد حجم الكلمات بالملف الواحد عن 2000 كلمة.



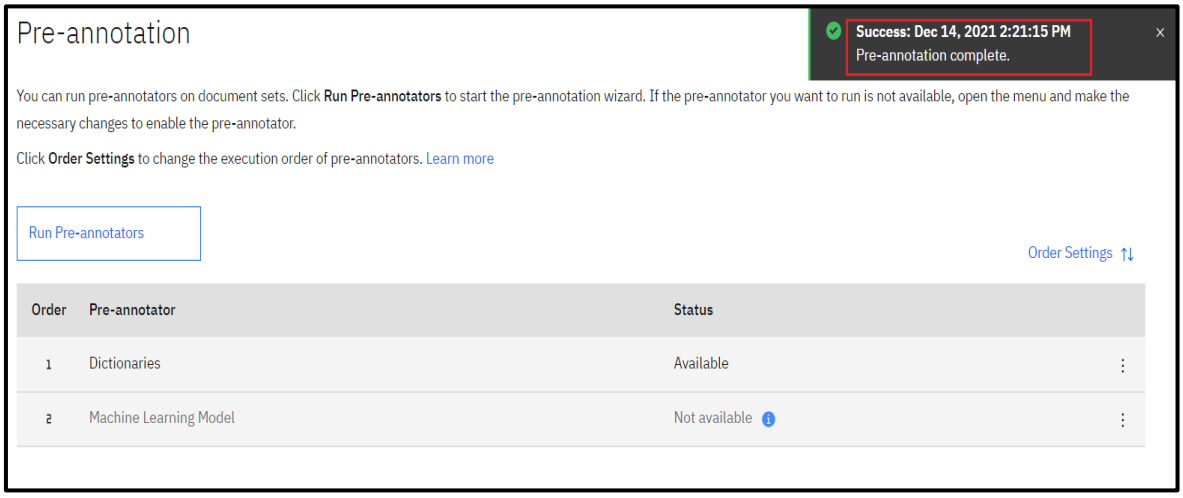
شكل (6) تحميل وإضافة المقالات للمنصة

ومن أجل بناء نموذج التعلم الآلي تم رفع 4 مقالات من نوعdocx. (مقالان باللغة العربية وآخران باللغة الإنجليزية).



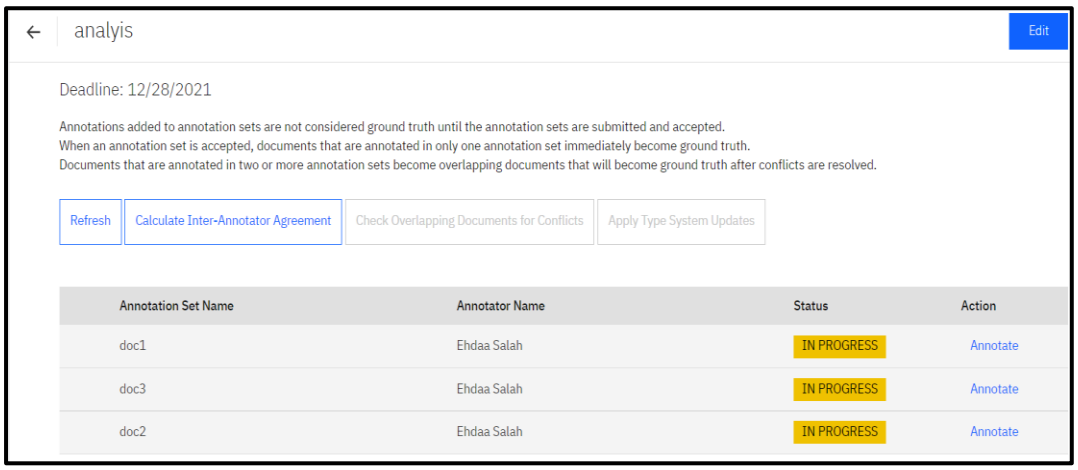
شكل (7) مجموع المقالات التي تم تحميلها على المنصة.

5. عمل تحليل مسبق preannotating لجميع المقالات باستخدام القاموس الذي تم إنشاؤه من أجل تسهيل عملية تحليل المحتوى.



شكل (8) التحليل المسبق للمقالات باستخدام قاموس المصطلحات.

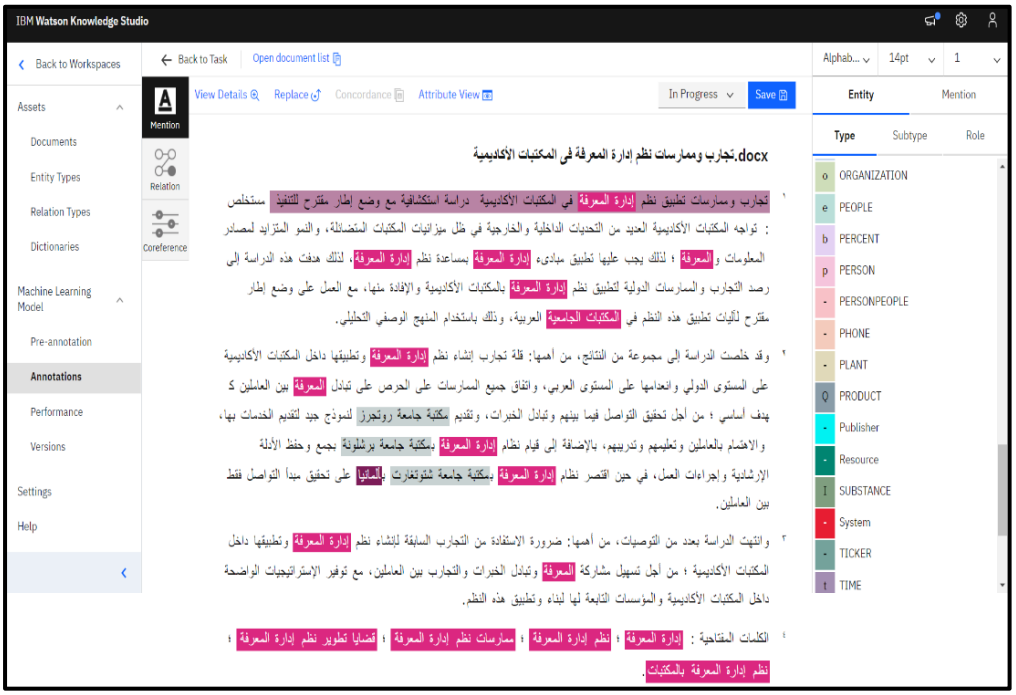
6. القيام بمهمة التحليل Creating an annotation task؛ و تم توزيع العمل على الملفات الموجودة كلًا على حدة؛ و تم تقسيم كل ملف (مقال) إلى مهمة قائمة بذاتها للتحليل.



شكل (9) توزيع مهمة تحليل المحتوى والمقالات.

7. يبدأ عمل المحلل annotator باختيار المقال الأول ويتم فتحه من خلال النظام، ويلاحظ تظليل المصطلحات الواردة بالقاموس؛ و تتمثل مهمته في تسهيل عمل المحلل ليبدأ عمله البشري في تحليل المحتوى، واستخراج الكيانات والعلاقات بينها.

8. يتم تحليل المحتوى وتحديد الكيانات الرئيسية؛ و يتم استخراج الكيانات الرئيسية مثل: أسماء المكتبات، والمواقع، والتواريخ المهمة، والاستشهادات، والكلمات المفتاحية.



شكل (10) القيام بمهمة تحليل المحتوى وتحديد الكيانات الرئيسية والعلاقات بينها

الخاصة بالمؤسسة، بالإضافة إلى مشاركة المعرفة فيما بينهم ؛ مما يؤدي إلى توفير حلول أكثر فعالية في حل المشكلات وتوليد أفكار جديدة في العمل.

وتتناول هذه الدراسة استعراض بعض التجارب والممارسات الدولية لتطبيق نظم إدارة المعرفة بالمكتبات الأكاديمية، والإفادة منها، والعمل على وضع إطار مقترح لآليات تطبيق تلك النظم في المكتبات الأكاديمية العربية، بالإضافة إلى إنشاء نموذج لتنفيذ عمليات

تصبح ذات صلة بمؤسساتها الأم، في ظل البيئة التنافسية الحالية ، وخاصة في الدول التي تعاني من عجز في الإمكانيات والموارد (Poonkothai, 2016)، كما يمكنها أن تصبح عنصرًا فعالًا داخل المجتمع العلمي، وأكثر مشاركة في جميع مراحل و سياقات خلق المعرفة ونشرها واستخدامها (Lougee, 2002).

ويمكن للمكتبات أن تتبوأ مكانة بارزة في المجتمع العلمي، من خلال مجموعة واسعة من المصادر مثل: وجهات نظر ثقافية و اجتماعية واقتصادية وتكنولوجية متعددة، واعتماد مصطلحات مثل: مجتمع المعلومات و مجتمع المعرفة التي تعتمد على مفهوم توليد وتطبيق المعلومات والمعرفة في مجموعة متنوعة من الأشكال ((EldhoJoy, 2012.

وعملًا بمبادىء إدارة المعرفة التي تتمثل في مشاركة المعرفة وتبادلها وإتاحتها، ظهرت نظم إدارة المعرفة التي تعمل على دعم حفظ الخبرات والمعارف وإتاحتها، وتقوم فكرتها على تمكين العاملين في المؤسسات من الوصول إلى الحقائق ومصادر المعلومات الخاصة بالمؤسسة، بالإضافة إلى مشاركة المعرفة فيما بينهم ؛ مما يؤدي إلى توفير حلول أكثر فعالية في حل المشكلات وتوليد أفكار جديدة في العمل.

وتتناول هذه الدراسة استعراض بعض التجارب والممارسات الدولية لتطبيق نظم إدارة المعرفة بالمكتبات الأكاديمية، والإفادة منها، والعمل على وضع إطار مقترح لآليات تطبيق تلك النظم في المكتبات الأكاديمية العربية، بالإضافة إلى إنشاء نموذج لتنفيذ عمليات

تصبح ذات صلة بمؤسساتها الأم، في ظل البيئة التنافسية الحالية ، وخاصة في الدول التي تعاني من عجز في الإمكانيات والموارد

شكل (11) القيام بمهمة تحليل المحتوى وتحديد الكيانات الرئيسية والعلاقات بينها

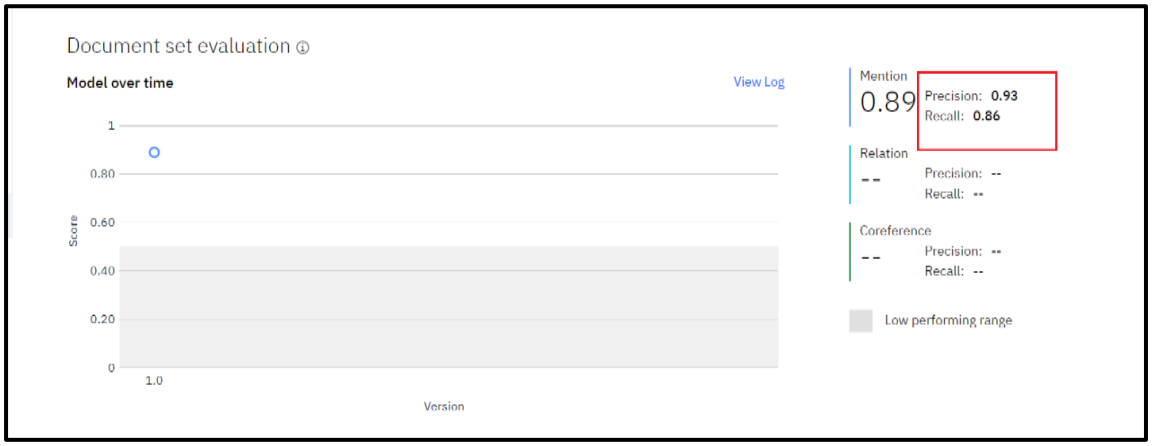
9. بعد ذلك تبدأ محاولة إيجاد العلاقات بين هذه الكيانات مثل: المكتبات الجامعية، وأماكن وجودها، والكلمات المفتاحية وعلاقاتها بالنص، و تتعدد العلاقات المتاحة داخل النظام فشملت:

جدول (4) أنواع العلاقات بين الكيانات داخل المنصة.

Relative	partOf	partOfMany	ownerOf	memberOf	locatedAt
educatedAt	capitalOf	Before	basedIn	authorOf	agentOf
playsRoleOf	participantIn	spokespersonFor	overlaps	affiliatedWith	AffectedBy

كما يمكن إضافة مزيد من العلاقات حسب الحاجة إليها.

10. بعد الانتهاء من عملية تحليل المحتوى للمقالات الموجودة، يتم عمل تدريب للنموذج وتقييمه train and Evaluate model من خلال النظام، وبعد أن تم تدريب النموذج بناءً على مجموعات البيانات الموجودة تم عمل تقييم له، وظهرت النتيجة كالتالي:



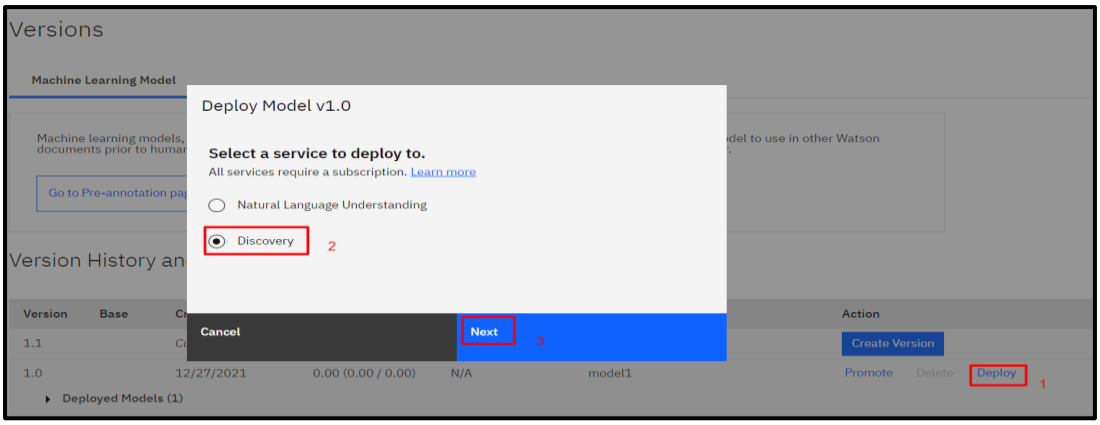
شكل (12) تدريب نموذج التعلم الآلي وتقييمه داخل المنصة.

11. بعد الانتهاء من تقييم النموذج والرضا عن النتائج، يمكن الانتقال للمرحلة التالية من التطبيق وهي: استخدام نموذج التعلم الآلي الذي تم إنشاؤه من أجل تحليل المحتوى لمقالات في نفس الموضوع، واستخراج الكيانات والعلاقات فيما بينها.

2/2/3 المرحلة الثانية: استخدام نموذج التعلم الآلي في عمليات تحليل المحتوى:

يتم في هذه المرحلة استخدام نموذج التعلم الآلي الذي تم إنشاؤه سابقًا لإثراء عمليات استخراج الكيانات والعلاقات فيما بينها من خلال تحقيق الاتصال والتكامل بين أداتي the Watson Knowledge Studio machine learning &Watson Discovery model، من خلال الخطوات التالية:

1. تم تحميل ملفات مجموعة من المقالات الجديدة في مجال (إدارة المعرفة) إلى أداة Watson Discovery، وبلغ عدد هذه المقالات 4 مقالات باللغتين العربية والإنجليزية، وتم عمل نشر لنموذج التعلم الآلي الذي تم إنشاؤه بواسطة knowledge studio



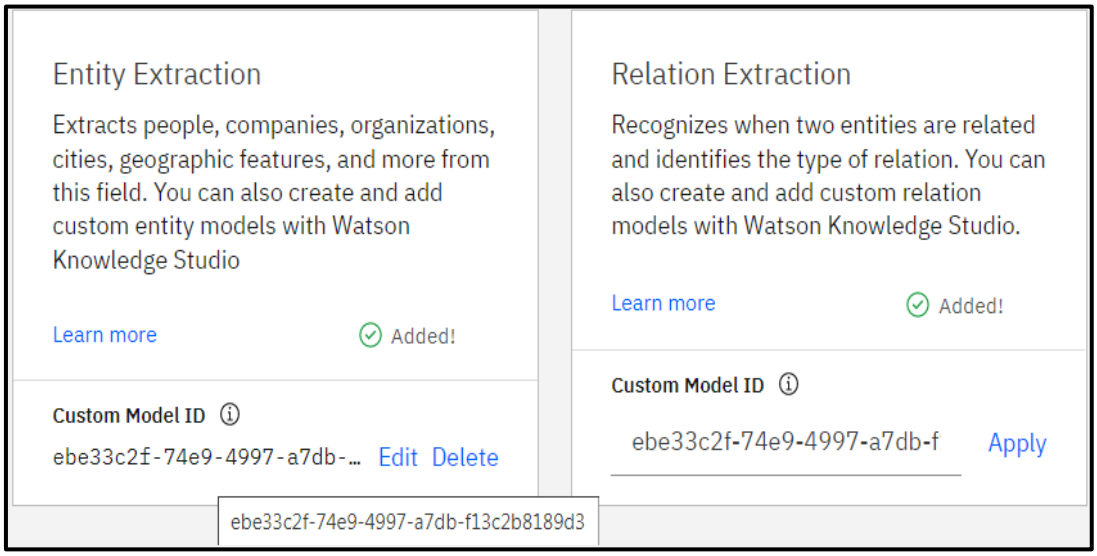
شكل (13) نشر نموذج التعلم الآلي المنشئ واستخدامه داخل خدمة Watson Discovery

2. ثم إضافة تحسينات النموذج إلى ثلاثة حقول رئيسية، هي:

* Keyword Extraction

❋ Entity Extraction

❋ Relation Extraction



شكل (14) إضافة التحسينات إلى نموذج التعلم الآلي

3. بتحليل محتوى المقالات الجديدة تم التعرف على الكيانات بها نتيجة التحديد المسبق لها في خدمة knowledge studio، فبالنسبة لكيان الكلمات المفتاحية تم تحديد الكلمات المفتاحية الواردة داخل النص ومدى علاقتها بالنص وتكرار ورودها.

enriched_text.entities.count	1
enriched_text.entities.sentiment.score	0.803161
enriched_text.entities.sentiment.label	positive
enriched_text.entities.text	knowledge extraction
enriched_text.entities.confidence	0.946298
enriched_text.entities.type	Keyword
enriched_text.entities.disambiguation.subtype.0	NONE

شكل (15) نتيجة استخدام نموذج التعلم الآلي على مجموعات المقالات الجديدة

enriched_text.keywords.text	information technology
enriched_text.keywords.relevance	0.557481
enriched_text.keywords.count	3
enriched_text.keywords.text	Knowledge resources management
enriched_text.keywords.relevance	0.556645
enriched_text.keywords.count	1
enriched_text.keywords.text	information management
enriched_text.keywords.relevance	0.546673
enriched_text.keywords.count	3

شكل (16) نتيجة استخدام نموذج التعلم الآلي على مجموعات المقالات الجديدة

4. تم اختبار عملية البحث على مجموعة من الوثائق الجديدة اعتمادًا على تحسينات نموذج التعلم الآلي؛ و تم عمل استعلام عن أبرز الكيانات الواردة في المقالات بالصيغة التالية:

Write an aggregation query using the Discovery Query Language

```
nested(enriched_text.entities).term(enriched_text.entities.type,count:5).term(enriched_text.entities.text)
```

شكل (17) صيغة الاستعلام عن الكيانات الموجودة

5. فكانت النتيجة؛ التعرف على الكيانات المحددة مسبقًا في نموذج التعلم الآلي، كما يتضح من الشكل التالي:

Aggregations

term(enriched_text.entities.type) **Keyword** (162)

- term(enriched_text.entities.text) **knowledge** (107)
- term(enriched_text.entities.text) **knowledge management** (11)
- term(enriched_text.entities.text) **Knowledge** (7)
- term(enriched_text.entities.text) **Knowledge management** (4)
- term(enriched_text.entities.text) **Libraries** (3)
- term(enriched_text.entities.text) **knowledge management systems** (3)

شكل (18) التعرف على الكيانات المعرفة مسبقًا في نموذج التعلم الآلي

3/3 نتائج عملية التطبيق على منصة IBM :

بعد تطبيق منصة IBM Watson Knowledge Studio لإنشاء نموذج لتعلم الآلة وتطبيقه باستخدام أداة Watson Discovery كانت كفاءته عالية في البحث والاسترجاع للمصادر باللغة الإنجليزية إلا أنه على الرغم من أن منصة IBM Watson Discovery تدعم

اللغة العربية إلا أنها لم تنجح في استخراج الكيانات، والعلاقات بينها، والفئات التي تم تحديدها في المرحلة الأولى من التطبيق باستخدام أداة knowledge studio، إلا أنها نجحت في بحث النص الكامل؛ فعند عمل الاستعلام والبحث بأي مصطلح موجود في أحد الوثائق التي تم تحليلها كان يتم استخرج النتائج، لذلك فإن هذه الأداة تحتاج لمزيد من التحسينات بالنسبة للمصادر باللغة العربية.

ويمكن تحسين النموذج المقترح لتعلم الآلة من خلال إثرائه بمزيد من الدراسات في المجال المحدد "إدارة المعرفة" وإثراء القاموس بمزيد من المصطلحات الجديدة في الموضوع.

رابعاً: النتائج والتوصيات:

1/4 النتائج :

* إمكانية استخدام تقنيات الذكاء الاصطناعي داخل المكتبات في العديد من الأعمال مثل: بناء النظم الخبيرة، وتقديم خدمة المراجع، ومعالجة اللغة الطبيعية، والتعرف على الأنماط، والروبوتات.

* تعد معالجة النصوص من أجل استخلاص المعرفة لأغراض التكشيف الآلي وتحليل المحتوى حقلًا مهمًا في مجال بحوث معالجة اللغة الطبيعية؛ فقد تم إدراك دورها مبكرًا في عمليات تحليل المحتوى مع إدخال الكثير من التحسينات إليها مثل: تقنيات الويب الدلالي.

* تستخدم معالجة اللغة الطبيعية في: التعرف على الكلام، وتصنيف النصوص، واستخلاص المعلومات، وتحليل النصوص، والترجمة الآلية، وتلخيص النص الآلي، وتحليلات الشبكات الاجتماعية.

* يمكن استخدام معالجة اللغة الطبيعية في المكتبات في الحصول على المصادر، ومعالجة المحتوى، وإحاطة المستفيدين بالمصادر والإجابة على الاستفسارات، وتحسين المعالجة باستخدام أدوات تنظيم المعرفة مثل المكانز.

* هناك العديد من أدوات معالجة اللغة الطبيعية المتاحة والمتوفرة على الويب والتي تنوعت بين منصات سهلة الاستخدام، ومكتبات برمجية لدمجها ضمن منصات لاستخدامها، وتسمح تلك الأدوات بتحليل المحتوى والتعرف على الكيانات والعلاقات بينها، واعتمدت كثير من تلك الأدوات على لغة البرمجة بايثون باعتبارها اللغة الشائعة في عالم الذكاء الاصطناعي، وبعضها اعتمد على لغة الجافا وغيرها من اللغات البرمجية.

* بعد تطبيق منصة IBM Watson Knowledge Studio لإنشاء نموذج لتعلم الآلة وتطبيقه باستخدام أداة Watson Discovery كانت كفاءته عالية في البحث والاسترجاع للمصادر باللغة الإنجليزية، أما بالنسبة للمصادر باللغة العربية فإن الأداة تحتاج المزيد من التحسينات.

2/4 التوصيات:

* العمل على تطبيق تقنيات الذكاء الاصطناعي والتعلم الآلي لتقديم خدمات جديدة وتحسين الوظائف الموجودة بالمكتبات والاستفادة منها في تحقيق النفع للمكتبات نفسها أو المستفيدين منها.

* بحث إمكانية تخصيص جزء من ميزانية المكتبات لإدخال تقنيات الذكاء الاصطناعي بها، وتدريب أخصائي المعلومات على استخدامها.

* توجيه المزيد من الاهتمام نحو الدراسات التي تتناول تطبيقات معالجة اللغة الطبيعية في تحليل المحتوى.

* تطوير نماذج لتعلم الآلة باستخدام المنصات والمكتبات البرمجية المتوفرة لمعالجة المحتوى والعمل على تحسينها باستمرار.

* الاهتمام بتطوير نظم معالجة اللغة الطبيعية باللغة العربية.

خامساً: **قائمة المصادر**:

1/5 المصادر العربية :

إبراهيم، عفاف محمد الحسن، (2010)، استخدام تطبيقات الذكاء الاصطناعي في المكتبات الجامعية: تصميم نموذج لنظام خبير في المراجع لمكتبة جامعة الخرطوم،جامعة الخرطوم، كلية الآداب، قسم علوم المكتبات والمعلومات (رسالة دكتوراه).

حسين، محمد، (2022)، التعلم الآلي واستخراج البيانات الببليوجرافية من مصادر المعلومات النصية: نموذج مقترح لمصادر المعلومات النصية المكتوبة باللغة العربية، بحث مقدم إلى: المؤتمر العلمي الثاني عشر لقسم المكتبات والوثائق وتقنية المعلومات، كلية الآداب، جامعة القاهرة في الفترة من 30-31 مارس 2022م.

السلمى، عفاف سفر، (2017)، تطبيقات الذكاء الاصطناعي لاسترجاع المعلومات في جوجل، مجلة دراسات المعلومات، ع 19 (يوليو 2017).

صدقي، كريمان بكام، (2021)، الاستفادة من تقنيات الذكاء الاصطناعي في تطوير أساليب البحث واسترجاع المعلومات بحث مقدم إلى: المؤتمر السنوي الثاني والثلاثون للاتحاد العربي للمكتبات والمعلومات، القاهرة 15-16 ديسمبر.

عامر،ياسمين أحمد، (2021)، تطبيقات الذكاء الاصطناعى فى المكتبات المصرية:دراسة تخطيطية، أطروحة ماجستير، جامعة القاهرة، كلية الآداب، قسم المكتبات والوثائق وتقنية المعلومات.

العتيبي، شروق زايد، والعمرى، أمل حسن، والغامدى، إيمان سعد. (2019)، الذكاء الاصطناعى واستخدامه فى المجالات البحثية والمعلوماتية: دراسة استشرافية حول نظام **Debater**، بحث مقدم إلى: مؤتمر الابتكار واتجاهات التجديد في المكتبات، المدينة المنورة، السعودية.

غازي، عز الدين، (2019)، المعالجة الآلية للغات الطبيعية والترجمة الآلية: تطبيقات نظام نوج**Nooj** المجلة الدولية للآداب والعلوم الإنسانية والاجتماعية,ع 31.

يوسف، الفاتح يوسف حماد، (2009)، تطبيقات الذكاء الاصطناعي والنظم الخبيرة: نموذج خبير في التزويد، جامعة النيلين، كلية الدراسات العليا، قسم المكتبات والمعلومات (رسالة دكتوراه).

2/5 المصادر الأجنبية:

ADEJO, A. A., & MISAU, A. Y. (2021). APPLICATION OF ARTIFICIAL INTELLIGENCE IN ACADEMIC LIBRARIES IN NIGERIA. Library Philosophy and Practice (e-journal). 6639, From: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=12669&context=libphilprac.

About IBM Watson. (n.d). Retrieved Dec 24, 2021, from: https://www.ibm.com/watson/about

Apache OpenNLP Developer Documentation. (n.d). Retrieved Dec 21, 2021, from: https://opennlp.apache.org/docs/1.9.0/manual/opennlp.html#opennlp

Aylien Ltd. (2020). Aylien: AI-Powered News Intelligence. Retrieved Dec 20,, 2021, from https://aylien.com/

Ba, P., J.Mb, N., & T, G. (2021). An Analysis of the Applications of Natural Language Processing in Various Sectors. Smart Intelligent Computing and Communication Technology Conference .

Barba, P. (2020). Machine Learning (ML) for Natural Language Processing (NLP). Retrieved Jan 24, 2022, from https://www.lexalytics.com/lexablog/machine-learning-natural-language-processing

Boucher, P. (2020). Artificial intelligence :How does it work, why does it matter, and what can we do about it? Panel for the Future of Science and Technology Retrieved from https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf

Chowdhury, G. G. (2003). Natural Language Processing. Annual Review of Information Science and Technology, 37. pp. 51-89 .

Cloud Natural Language documentation. Retrieved Jan 20, 2022, from https://cloud.google.com/natural-language/docs.

Das, R. K., & Islam, M. S. (2021). Application of Artificial Intelligence and Machine Learning in Libraries: A Systematic Review. arXiv, Doi: https://doi.org/10.48550/arXiv.2112.04573.

Devopedia. (2019). "Natural Language Toolkit ".Version 12. Retrieved Dec 21, 2021, from https://devopedia.org/natural-language-toolkit

Documentation - Natural Language Toolkit. (n.d). Retrieved Dec 20, 2021, from:

https://www.nltk.org/

Emms, M., & Luz, S. (2007). Machine Learning for Natural Language Processing. European Summer School of Logic, Language and Information ESSLLI 2007 ,Dublin, Ireland Course .

Enis, M. (2018). University of Rhode Island Opens AI Lab in Library. Retrieved Dec 20, 2021, from https://www.libraryjournal.com/story/180926URIlibraryAIlab

Esposito, F., Ferilli, S., Basile, T. M. A., & Mauro, N. D. (2008). Machine Learning for Digital Document Processing: From Layout Analysis To Metadata Extraction

Fisher, I. E., Garnsey ,M. R., Goel, S., & Tam, K. (2010). The Role of Text Analytics and Information Retrieval in the Accounting Domain Journal of Emerging Technologies in Accounting (2010) 7 (1): 1–24 .

Guo, J., Farhang-Razi, V., & Algra, P. (Eds.). (2019). Springer Nature Switzerland AG.

Hamandi, A. M., wahab, H. B. A., & Karim, A. (2016). Natural Language Processing Using Natural Language Toolkit. Iraqi Journal of Information Technology, vol. 7, No. (2).

Introduction natural language processing nlp. (2016). Retrieved Jan 24, 2022 ,from https://algorithmia.com/blog/introduction-natural-language-processing-nlp

Jr., D. C. H. (2007). EXTRACTION AND INDEXING OF TRIPLET-BASED KNOWLEDGE USING NATURAL LANGUAGE PROCESSING. (Master Degree) .

Karoo, K. (2018). Natural Language Processing and Digital Library Management System. International Journal of Science and Research (IJSR), Volume 7 Issue 11, November 2018 .

Kurama, V. (2021). Information Extraction. Retrieved Feb 20, 2022, from https://nanonets.com/blog/information-extraction/

Lee, L. (2002). A non-programming introduction to computer science via NLP,IR,and AI. Paper presented at the Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, Pennsylvania, USA .

Mishra, U. (2022). What Is Text Summarization in NLP? Retrieved Feb 2022, 2022, from https://www.analyticssteps.com/blogs/what-text-summarization-nlp

MonkeyLearn Inc. (2020). MonkeyLearn. Retrieved Dec 20, 2021, from : https://monkeylearn.com/

Neupane, P. (202 ،(0Understanding text classification in NLP with Movie Review Example. Retrieved Jan 7, 2022, from https://www.analyticsvidhya.com/blog/2020/12/understanding-text-classification-in-nlp-with-movie-review-example-example/

Obeid, O., Zalmout, N., & Khalifa, S. (2020). CAMeL Tools: An Open Source Python Toolkit for Arabic Natural Language Processing. Proceedings of the 12th Language Resources and Evaluation Conference,Marseille, France .

Omame, I. M., & Alex-Nmecha, J. C. (2020). Artificial Intelligence in Libraries. In book: MANAGING AND ADAPTING LIBRARY INFORMATION SERVICES FOR FUTURE USERS, IGI Global .

OpenNLP - Overview. (n.d). Retrieved Dec 21, 2021, from: https://www.tutorialspoint.com/opennlp/opennlp_overview.htm

Orobor, I. A. (2016). Integration and Analysis of Unstructured Data for Decision Making: Text Analytics Approach. International Journal of Open Information Technologies, Vol 4, No 10 .(2016)

Otero, P. G. (2019). LinguaKit. Retrieved Jan 2, 2022, from https://citius.usc.es/transferencia/software/linguakit

Redondo, T., & Sandoval, A. M. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE, vol. 3, Special Issue on Big Data and AI .

Sawhney, P. (2021). Introduction to Automatic Speech Recognition and Natural Language Processing. Retrieved Jan 5, 2022, from : https://www.analyticsvidhya.com/blog/2021/01/introduction-to-automatic-speech-recognition-and-natural-language-processing/spacy -Industrial-Strength Natural Language Processing (n.d). Retrieved Dec 29, 2022, from https://spacy.io/

Sridevi, P. C., & Shanmugam, A. P. (2017). Artificial Intelligence and its applications in Libraries. In book: E- Resources Management, Department of Library, Thiruvalluvar University College of Arts &Science Villupuram .

Stafford, C. (2021). Artificial intelligence and machine learning: improving the efficiency of library services. HUMANITIES COMMONS Repository, Doi: https://doi.org/10.17613/9h7p-xn21

Stanford NLP Group. (2020a). About CoreNLP. Retrieved Dec 18, 2021, from: https://stanfordnlp.github.io/CoreNLP/index.html

Stanford NLP Group. (2020b). CoreNLP. Retrieved Dec 18, 2021, from: https://stanfordnlp.github.io/CoreNLP/

The App Solutions Inc. (2021). NATURAL LANGUAGE PROCESSING TOOLS AND LIBRARIES. from: https://theappsolutions.com/blog/development/nlp-tools/

Tyagi, A. (2021). A Review Study of Natural Language Processing Techniques for Text Mining. INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), Volume 10, Issue 09 (September 2021) .

UpBuild, L. An Introduction to Google's Cloud Natural Language API. Retrieved Dec 18,, 2021, from https://www.upbuild.io/blog/google-cloud-natural-language-api/

Vilares, J., Barcala, F .M., Alonso, M. A., Gra~na, J., & Vilares, M. (2002). Practical NLP-Based Text Indexing. Lecture Notes in Computer Science .

William, J. R. (Ed.) (1999). London: Glenlake Publishing Company, Ltd.

●●●

Applications of Artificial Intelligence Systems in Content Analysis and Indexing Processes:

An Applied Study of Natural Language Processing systems

Dr. Ehdaa Salah Nagy

Lecturer of Libraries & Information Technology, Faculty of Arts, Cairo University
Ehdaa.salah@gmail.com

Based on the development of Artificial Intelligence techniques, the most important of which is natural language processing which has a significant role in content analysis, this study aims at identifying the role of applying Artificial Intelligence systems in content analysis and indexing processes as well as identifying the uses of natural language processing in general and in libraries in particular. Additionally, the study attempts to monitor Artificial Intelligence platforms and libraries that can be used in the field of content analysis, along with the application of one of the natural language processing systems and using it in content analysis for information resources to achieve benefit from them, by using the descriptive-analytical method.

Among the most important results of the study is the possibility of using natural language processing in libraries for obtaining resources, content processing, informing users of resources and answering inquiries. After using the IBM Watson Knowledge Studio platform to create a machine learning model and testing it via the Watson Discovery tool, its high efficiency was proven in searching for and retrieving English language resources. However, it needs further improvements for Arabic language resources.

***Keywords*: *Natural Language Processing Systems* (1); *Content Analysis* (2); *Artificial Intelligence* (3); *Machine Learning* (4).**