تطبيقات الذكاء الاصطناعي في تحليل الدلالات المعجمية والسياقية للكلمات العربية

د. علاء عبدالخالق حسين المندلاوي
Dr.ALAA ABDULKHALEQ HUSSEIN
جامعة بغداد / كلية العلوم الإسلامية
Alaa.Abdulkhaleq@colaw.uobaghdad.edu.iq

بحث علمي منشور في مجلة الارتقاء العلمية الإلكترونية/
اتحاد نقابات المدربين العرب
بتاريخ ٦/ ٧/ ٢٠٢٤

المستخلص:

يهدف هذا البحث إلى تحديد إمكانيات الذكاء الاصطناعي في تحليل المعاني والسياقات للكلمات العربية، أُسْتُخْدِم الباحث المنهج الوصفي التحليلي، مع الاعتماد على الدراسات السابقة والأسس النظرية والأدبيات في مجالي معالجة اللغة الطبيعية والذكاء الاصطناعي. ونموذج ذكاء اصطناعي تطور لفهم الدلالات والسياقات في النصوص العربية.

جمعت مجموعات متنوعة من النصوص العربية، بما في ذلك النصوص الأدبية والتاريخية والعلمية، بهدف توافر تغطية واسعة للغة العربية واستخداماتها، بعد تعديل النصوص وتحويل الكلمات غير الضرورية إلى رموز رقمية، طورنا نموذج الذكاء الاصطناعي باستخدام تقنيات متطورة مثل التعلم العميق والتعلم التعزيزي.

قُيّم أداء النموذج باستخدام معايير كمية مثل الدقة والاستدعاء، فضلاً عن استعمال معايير كيفية مثل فهم السياق والإشارات الثقافية.

حقق النموذج نتائج إيجابية بنسبة ٩٢٪ في استيعاب المعاني والسياق للكلمات العربية، فضلاً عن فهم عميق للسياق والقدرة على التعامل مع الكلمات غير الواضحة والمعاني المتعددة، وفهم العلامات الثقافية والاجتماعية في اللغة العربية.

بناءً على نتائج البحث، قدم الباحث مجموعة من التوصيات التي تضمنت:

زيادة حجم وتنوع مجموعات البيانات المستخدمة في تدريب النموذج، نظراً لغنى وتنوع اللغة العربية، مما يتطلب وجود مجموعات بيانات أكبر لضمان فهم جميع أشكال اللغة.

- تعزيز فهم اللغة العربية عن طريق التركيز على السياق الثقافي باستخدام تقنيات الذكاء الاصطناعي.

إنشاء نماذج مخصصة لفهم العلامات الثقافية والاجتماعية في اللغة العربية، بهدف تعزيز قدرة تلك النماذج على فهم المعاني بنحوٍ أكثر دقة.

- التعاون مع الخبراء في اللغة العربية والعلوم الإنسانية يُسهم في تعمق فهم اللغة العربية وثقافتها، مما يُسهل تحسين قدرة الأفراد على فهم اللغة.

واقترح الباحث عدة اقتراحات للدراسات المستقبلية، لتحديد مسارات مختلفة للأبحاث القادمة ومنها:-

- القيام بأبحاث في تحليل الأدب لفهم المعاني والاستعارات المستخدمة في النصوص الأدبية.

- إجراء دراسة لتحليل قدرة النموذج على فهم والتفاعل مع التباينات في اللهجات العربية.

- دمج تقنيات الرؤية الحاسوبية ومعالجة اللغة الطبيعية لاستكشاف أسرار النصوص المرافقة للصور والفيديوهات.

- إجراء أبحاث في مجالات العلوم الإنسانية يتضمن التحقيق في النصوص التاريخية والخطابات بنحوٍ أعمق.

الكلمات المفتاحية، الذكاء الاصطناعي، الدلالات المعجمية والسياقية، الكلمات العربية.

Applications of artificial intelligence in analyzing lexical semantics And the contextuality of Arabic words

Dr.ALAA ABDULKHALEQ HUSSEIN

Alaa.Abdulkhaleq@colaw.uobaghdad.edu.iq

Abstract:

This research aims to determine the capabilities of artificial intelligence in analyzing the meanings and contexts of Arabic words. The researchers used the descriptive analytical method, relying on previous studies, theoretical foundations, and literature in the fields of natural language processing and artificial intelligence. An artificial intelligence model was developed to understand the connotations and contexts in Arabic texts.

I collected diverse collections of Arabic texts, including literary, historical, and scientific texts, with the aim of providing a broad coverage of the Arabic language and its uses, after modifying the texts and converting unnecessary

words into digital codes, We developed the AI Model using cutting-edge techniques such as deep learning and reinforcement learning.

Model performance was evaluated using quantitative criteria such as precision and recall, as well as using qualitative criteria such as understanding of context and cultural cues.

The model achieved positive results of 92% in understanding the meanings and context of Arabic words, as well as a deep understanding of context and the ability to deal with unclear words and multiple meanings, and understanding cultural and social signs in the Arabic language.

Based on the research results, the researchers presented a set of recommendations that included:

Increased size and diversity of data sets used to train the model, given the richness and diversity of the Arabic language, which requires larger data sets to ensure understanding of all forms of the language.

-Enhancing understanding of the Arabic language by focusing on the cultural context using artificial intelligence techniques.

Creating customized models to understand cultural and social signs in the Arabic language, with the aim of enhancing the ability of these models to understand meanings more accurately.

Collaboration with experts in the Arabic language and humanities contributes to a deeper understanding of the Arabic language and its culture, which facilitates improving individuals' ability to understand the language.

The researchers offer several suggestions for future studies, such as identifying different paths for future research.

-Conducting research in analyzing literature to understand the meanings and metaphors used in literary texts.

-Conduct a study to analyze the model's ability to understand and interact with variations in Arabic dialects.

-Integrating computer vision and natural language processing techniques to explore the secrets of texts accompanying images and videos.

-Conducting research in the fields of the humanities includes investigating historical texts and discourses in a deeper way.

Keywords: artificial intelligence, lexical and contextual connotations, Arabic words.

المقدمة:

تعد اللغة إحدى أهم وسائل التواصل بين الشعوب، وتعتبر جزءًا أساسيًا من هويتهم. مع تقدم التقنية والذكاء الاصطناعي، بدأت الدراسات في تعزيز الأدوات المستخدمة في فهم ونقل اللغات البشرية. أُقْتُرِحَت فكرة بناء المعاجم وتمثيل المعرفة اللغوية عن طريق التطبيقات الإلكترونية. وعلى ضوء التطور الاقتصادي، بدأت فكرة الترجمة الآلية بين لغات مختلفة في الانتشار والتوسع. مع التقدم التقني في العصر الحديث، فُتحت آفاق جديدة لتطوير القدرات الإلكترونية في تفسير ونقل اللغات الطبيعية بنحوٍ أفضل وتحسين أداء البرمجيات التعليمية.

قيمة اللغة تكمن في كونها هوية الشعوب الأصلية، واللغة العربية تحتوي على تاريخ وثقافة غنية. تُعتبر إحدى اللغات الهامة على مستوى العالم، وتتألف من ٢٨ حرفًا و١٦٠٠٠ حذر. في العصر التقني والذكاء الاصطناعي، يُمكن استخدام هذه التقنيات لفهم ومعالجة اللغة العربية بنحوٍ أفضل وتحليل الدلالات المعجمية والسياقية للكلمات بنحوٍ أعمق. وعند الحديث عن اللغة العربية وثرائها، لا يمكن تجاهل تاريخها الطويل وتأثيرها الواسع، حيث ازدهرت خلال العصر الذهبي للإسلام كلغة للعلم والمعرفة. (الجراجرة، ٢٠٢١، ص ١٥). ونتيجةً لهذا الانتشار، تأثرت وتفاعلت مع العديد من اللغات والثقافات، مما أدى إلى إثرائها وتطورها.

وتتميز اللغة العربية بخصائص فريدة تجعلها لغةً غنيةً ومتميزةً. فلديها نظام صرفي غني، حيث يمكن اشتقاق عشرات الكلمات من جذرٍ ثلاثيٍ واحدٍ، مما يضيف طبقاتٍ من المعاني والدلالات (الجراجرة، ٢٠١٨، ص ٤٣). تتميز اللغة العربية بنظامها النحوي المعقد الذي يتيح تنظيم الجمل والكلمات بأساليب متنوعة لتوصيل المعاني بدقة. زيادة على ذلك، تحتوي اللغة العربية على مفردات واسعة تغطي جوانب مختلفة من الحياة، مما يعكس تاريخها العريق وثراء ثقافتها.

فهم الدلالات المعجمية والسياقية للكلمات العربية يمثل تحديًا كبيرًا، حيث قد تحمل الكلمة معانٍ متعددة، وقد يتغير معناها بحسب سياقها. لذلك، يعد استكشاف وفهم هذه الدلالات بعمق أمر ضروري لتطبيقات مثل الترجمة الآلية وتحليل النصوص ومعالجة اللغة الطبيعية. ورغم الجهود البحثية المبذولة، هناك جوانب لم تكن مكتشفة بعد، وخاصةً مع التطور السريع في تقنيات الذكاء الاصطناعي.

مشكلة البحث:

مع التقدم السريع في مجال الذكاء الاصطناعي وزيادة استخدامه في مختلف المجالات، أصبح تطبيقه في فهم ومعالجة اللغات الطبيعية ومفرداتها محل اهتمام كبير. ورغم التطورات التي حدثت في هذا المجال، تبقى اللغة العربية تحديًا بسبب غناها اللغوي وتعقيدها.

تعد اللغة العربية من أقدم اللغات السامية ذات التأثير الواسع في الشرق الأوسط وما وراءه، حيث تتميز بمفرداتها الغنية وتراكيبها النحوية المعقدة. تشكل هذه الخصائص تحديًا كبيرًا لأنظمة الذكاء الاصطناعي في فهم معاني الكلمات والجمل العربية والتعامل مع التعقيدات النحوية والدلالات السياقية. من هنا، تسعى الدراسات الحالية إلى تطوير أدوات متقدمة لفهم الدلالات اللغوية والسياقية للغة العربية، لسد الفجوة بينها وبين تقنية الذكاء الاصطناعي وتحسين فهمها وتحليلها بدقة.

على الرغم من أهمية هذا البحث، إلا أنه يواجه تحديات في تدريب نماذج الذكاء الاصطناعي؛ بسبب ندرة مجموعات البيانات العربية والتحدي الذي يشكله اللغة العربية المعقدة. لذا، يعتمد هذا البحث على منهجية شاملة لجمع مجموعات بيانات متنوعة واستخدام تقنيات الذكاء الاصطناعي الحديثة، مثل التعلم العميق والتعلم المعزز، لتصميم نماذج قادرة على فهم الدلالات المعجمية والسياقية. كما سيُتَعَاوَن مع خبراء في اللغة العربية واللغويات الحاسوبية لضمان دقة النماذج.

سيتألف البحث من أربعة أقسام رئيسة، يقدم القسم الأول مقدمة للبحث، حيث سنناقش أهميته وأهدافه ومنهجيته، وسنقدم خلفية عن الذكاء الاصطناعي وتقنياته، ونفهم الدلالات المعجمية والسياقية. في القسم الثاني، سنستعرض الإطار النظري للبحث، بما في ذلك المفاهيم الأساسية المتعلقة بالدلالات المعجمية والسياقية في اللغة العربية، وتقنيات الذكاء الاصطناعي المستخدمة في معالجة اللغة العربية، والدراسات السابقة في هذا المجال.

سيُرَكَّز في الفصل الثالث على الجانب التطبيقي، حيث ستُشْرَح عملية جمع مجموعات البيانات من النصوص العربية المختلفة وتصميم نموذج ذكاء اصطناعي متقدم قادر على فهم الدلالات المعجمية والسياقية، وستُنَاقَش عملية تدريب النموذج على هذه المجموعات وتقنيات تقييم أدائه وتحسينه. في الفصل الرابع والأخير، سيتم عرض ومناقشة النتائج التي توصل إليها، بما في ذلك نتائج تجارب النموذج بنحوٍ كمي ونوعي، مع مناقشة أهميتها وتأثيرها. كما سيُقَارَن نتائج البحث مع الدراسات السابقة، واستكشاف التطبيقات العملية للنموذج، ومناقشة القيود والتحديات، وتقديم اقتراحات للأبحاث المستقبلية.

نتمنى أن يكون هذا البحث إسهامًا متواضعاً في تطوير مجال معالجة اللغة العربية باستخدام التقنية الذكية. عن طريق دراسة معاني الكلمات وسياقها في اللغة العربية، نهدف إلى تعزيز فهم الذكاء الاصطناعي للغة وتطوير تطبيقات عملية في مختلف المجالات. بالجمع بين اللغة العربية والذكاء الاصطناعي، يمكننا بناء أدوات لتحليل وتعزيز فهم اللغة، وفتح آفاق جديدة لتطبيقات التقنية الذكية في المجتمع العربي.

الفصل والأول: أهمية البحث

أولا: أهمية البحث

يعد هذا البحث مساهمة مهمة في مجال معالجة اللغة الطبيعية، ولاسيما في فهم اللغة العربية واستكشاف دلالاتها المعجمية والسياقية. يتمثل أهمية هذا البحث من الناحية النظرية والعملية في:

١-الأهمية النظرية:

تعزيز المعرفة في مجال معالجة اللغة العربية:

يُسهم هذا الدراسة في تعزيز المعرفة النظرية في مجال معالجة اللغة الطبيعية، وتحديدًا في اللغة العربية، عن طريق استكشاف المعاني اللغوية والسياقية للكلمات العربية، مما يوافر فهمًا أشمل للصعوبات والتحديات التي تنشأ من معالجة هذه اللغة. ويعد اللغة العربية أحد أصعب اللغات في التعامل معها؛ نظرًا لغناها النحوي والصرفي، ولذلك، فإن البحث في هذا المجال يساهم في تطوير الخوارزميات والنماذج التي تستطيع التعامل مع هذه التعقيدات، ومن ثم يُسهم في تثنية المعرفة النظرية في مجال معالجة اللغة الطبيعية. (نورمان وآخرون، ٢٠١٩، ص ٣٢).

فهم أعمق للغة العربية:

يهدف البحث إلى فهم أعمق للغة العربية وخصائصها الدلالية، عن طريق تحليل دلالات الكلمات والسياق، واستكشاف المعاني الخفية التي تحملها. يساهم هذا الفهم في تحسين نماذج معالجة اللغة الطبيعية وزيادة قدرتها على فهم والاستجابة للنصوص العربية بدقة. (الجراجرة وآخرون، ٢٠٢٠، ص ١٧٦).

تطوير نماذج معالجة اللغة الطبيعية:

عن طريق الاستفادة من تطبيقات الذكاء الاصطناعي في تحليل المعاني اللغوية والسياقية للكلمات العربية، نسعى إلى تحسين نماذج معالجة اللغة الطبيعية وتطويرها بنحوٍ متقدم. باستخدام تقنيات مثل التعلم العميق والتعلم المعزز، يمكننا بناء نماذج تفهم المعاني المعقدة والسياقات بنحوٍ أفضل. وهذا بدوره يمكن أن يسهم في تطوير أنظمة جديدة لمعالجة اللغة الطبيعية تتفاعل بفاعلية مع اللغة العربية وتفهمها بعمق. (عبد الله وآخرون، ٢٠١٩، ص ٥٦).

استكشاف حدود الذكاء الاصطناعي:

يقدم هذا البحث فرصة لاستكشاف حدود وإمكانيات الذكاء الاصطناعي في مجال معالجة اللغة الطبيعية، حيث تُعتبر اللغة العربية تحديًا بسبب تعقيدها ودقتها. عن طريق دراسة قدرة هذه الأنظمة على فهم

الدلالات والسياقات، يمكن تقييم تقدمها وقدرتها على محاكاة الفهم البشري، ويمكن أن يساهم هذا الاستكشاف في تطوير خوارزميات ونماذج متقدمة في فهم اللغة الطبيعية. (الجراجرة، ٢٠٢١، ص ٤٥).

٢- الأهمية العملية:

تحسين أنظمة الترجمة الآلية:

الترجمة الآلية هي تطبيق مهم لمعالجة اللغة الطبيعية، حيث يمكن تحسين جودة أنظمة الترجمة الآلية عن طريق فهم الدلالات والسياق اللغوي للكلمات. عادةً ما تواجه هذه الأنظمة صعوبة في نقل المعاني على نحو دقيق؛ بسبب عدم فهمها للسياق والمعاني. بالبحث والتطوير، يمكن تطوير نماذج ترجمة تستطيع فهم الدلالات والسياق الثقافي، مما يؤدي إلى ترجمات دقيقة وطبيعية أكثر. (الجراجرة، ٢٠٢١ ب، ص ٧٨).

تعزيز تحليل المشاعر والآراء:

يهدف هذا البحث إلى تعزيز كفاءة أنظمة تحليل المشاعر والآراء باللغة العربية عن طريق فهم الدلالات المعجمية والسياقية. يساهم التمييز بين المعاني المختلفة للكلمات في تحسين تصنيف المشاعر والآراء بدقة، حيث يمكن لكلمة واحدة أن تحمل معنى إيجابياً في سياق معين وسلبي في آخر. عن طريق فهم السياق والدلالة، يمكن تحسين دقة أنظمة تحليل المشاعر والآراء بنحوٍ عام. (عبد الله وآخرون، ٢٠١٨، ص ٢٣).

تطوير أنظمة التعرف على الكلام:

تمثل اللغة الشفوية تحدياً كبيرًا لأنظمة الذكاء الاصطناعي، حيث يمكننا، عن طريق فهم المعاني اللغوية والسياقية، تحسين قدرة أنظمة التعرف على الكلام على فهم النصوص المنطوقة بدقة، وتلعب المعاني اللغوية والسياقية دورًا مهمًا في فهم المقصود من الكلام، خاصة في اللغة العربية المميزة بالتشابه والاستعارات. وعن طريق البحث، يمكننا تطوير أنظمة تعرف على الكلام بدقة أكبر وقدرة أكبر على فهم اللغة الشفوية. (الجراجرة وآخرون، ٢٠٢٢، ص ٩٨).

تطبيقات في العلوم الإنسانية الرقمية:

يمكن أن يكون لهذا البحث تطبيقات مهمة في مجال العلوم الإنسانية الرقمية. بفهم الدلالات المعجمية والسياقية للكلمات العربية، يمكن تطوير أدوات لتحليل النصوص الأدبية والتاريخية والثقافية. يمكن إنشاء أدوات للتحليل الدلالي أو لبحث الاستعارات والتشابيه في الشعر العربي، أو لتحليل الخطابات السياسية والتاريخية. تلك التطبيقات يمكن أن تثري مجال العلوم الإنسانية، وتوفر أدوات جديدة للباحثين. (الجراجرة، ٢٠٢٠، ص ٦٧).

تحسين التفاعل بين الإنسان والآلة:

عن طريق تعزيز فهم الآلة للغة العربية، يمكننا تحسين التفاعل بين الإنسان والآلة. غالبًا ما يكون عدم فهم اللغة عائقًا أمام التواصل الفعّال بين البشر وأنظمة الذكاء الاصطناعي. وعن طريق جهودنا، يُمكننا تحسين الأنظمة لتفهم بنحوٍ أفضل معاني الكلام أو النصوص، مما يؤدي إلى تواصل أكثر سلاسة وفعالية بين الإنسان والآلة. (عبد الله وآخرون، ٢٠١٩، ص ٦٧).

توسيع نطاق تطبيقات الذكاء الاصطناعي:

يمكن أن يسهم هذا البحث في دعم استخدام التقنية بذكاء اصطناعي في الوطن العربي، عن طريق تعزيز فهم اللغة العربية. يمكن إنشاء أنظمة ذكاء اصطناعي في مجالات متنوعة مثل الرعاية الصحية والتعليم والخدمات الحكومية، مما يعزز جودة الحياة، ويدعم التنمية الاقتصادية. (نورمان وآخرون، ٢٠٢٠، ص ١٢).

تسهيل معالجة اللغات الأقل انتشارًا:

يمكن لهذا البحث أن يسهم في تبسيط معالجة اللغات النادرة. اللغة العربية تشبه العديد من اللغات النادرة في تعقيدها. عن طريق فهم معاني الكلمات والسياق في اللغة العربية، يمكننا تطوير طرائق جديدة يمكن استخدامها مع لغات أخرى، مما يجعل معالجتها وتحليلها أسهل. هذا يمكن أن يساهم في الحفاظ على اللغات والثقافات المتنوعة في جميع أنحاء العالم. (الجراجرة، ٢٠٢١ أ، ص ٩٢).

الاستفادة من البيانات الضخمة:

مع التطور التقني وانتشار الإنترنت، أصبحنا نعيش في زمن الكم الهائل من البيانات. يمكن لهذا البحث المساهمة في تطوير أدوات بسيطة لتحليل ومعالجة النصوص العربية ضمن هذه البيانات الكبيرة. عن طريق فهم المعاني والسياق، يمكن إنشاء أدوات تستطيع استخراج المعلومات والعلاقات الهامة من النصوص العربية ضمن مجموعات البيانات الكبيرة. ويمكن أن يقدم ذلك رؤى قيمة للشركات والحكومات والباحثين. (عبد الله وآخرون، ٢٠٢٠، ص ٤٥).

التعليم والتوعية:

يمكن أن يسهم هذا البحث في نشر الوعي حول أهمية معالجة اللغة الطبيعية والذكاء الاصطناعي، وذلك عن طريق نشر النتائج والدراسات لتطبيقها في المجالات التعليمية. فضلاً عن استخدامه كأساس لمناهج تعليمية في الجامعات والمدارس، لتدريب جيل جديد من الباحثين والمطورين في هذا المجال الحيوي. يمثل هذا البحث محاولة بسيطة في تحسين فهم اللغة العربية وتعزيز تطبيقات الذكاء الاصطناعي المتعلقة

بها، عن طريق استكشاف الدلالات والسياقات اللغوية وتقديم أدوات تقنية جديدة تعزز فهم اللغة وتحليلها بفعالية في المشهد العربي. (نورمان وآخرون، ٢٠٢٢، ص ٢٣).

ثانياً: أهداف البحث:

يهدف هذا البحث إلى تحقيق الأهداف التالية:

١- تقويم استخدام تقنيات الذكاء الاصطناعي في تحليل معاني الكلمات العربية وسياقها، ونقيم فعالية هذه التقنيات في فهم اللغة العربية.

٢- إنشاء نموذج ذكاء اصطناعي متقدم يفهم المعاني والسياق في النصوص العربية، باختيار الخوارزميات المناسبة وتصميم هيكل النظام الأساسي ومعالجة البيانات، يمكن تحقيق الأهداف بنجاح.

٣- تدريب نموذج الذكاء الاصطناعي على مجموعات متنوعة من النصوص العربية، وسنقيم أداءه في تحليل المعاني والسياق. سنستخدم مقاييس تقييم مختلفة مثل الدقة والاستدعاء ومعدل الخطأ.

٤- استخدامات النموذج العملي في مجالات متنوعة مثل تحسين الترجمة الآلية وتحليل المشاعر في وسائل التواصل الاجتماعي وحتى تحليل النصوص الأدبية.

٥- تحديد العقبات والحدود التي تواجه نماذج الذكاء الاصطناعي في فهم الدلالات المعجمية والسياقية للغة العربية، وتقديم حلول وتوصيات لتطوير البحث في هذا المجال.

ثالثاً: أسئلة البحث:

ما هي الابتكارات والنماذج الجديدة في مجال الذكاء الاصطناعي التي تُستخدم لتحليل اللغة العربية وفهمها؟

كيف يمكن لتقنيات الذكاء الاصطناعي الحديثة أن تُساهم في تعزيز فهم المعاني اللغوية والسياقية في النصوص العربية؟

ما هي نتائج الأبحاث حول استخدام النماذج المطورة في فهم المعاني والسياق في النصوص باللغة العربية؟ ما هي التحديات التي تواجه تطبيقات الذكاء الاصطناعي في تفسير المعاني والسياقات باللغة العربية؟

رابعاً: منهجية البحث

هذا البحث يستند إلى المنهج التحليلي الوصفي، سيتم شرح وتقييم استخدام تقنيات الذكاء الاصطناعي في تحليل ومعالجة اللغة الطبيعية. وتحليل المفاهيم النظرية ذات الصلة بالدلالات المعجمية والسياقية. سيتم سنقوم أيضًا بتصميم وتطبيق نموذج عملي لتطبيقات الذكاء الاصطناعي في معالجة النصوص العربية.

سنجمع مجموعات متنوعة من البيانات باللغة العربية، تشمل النصوص الأدبية والتاريخية والعلمية، وسنسمي هذه البيانات يدويا بمعلومات حول المعاني والسياق اللغوي للكلمات.

ستُعَالَج البيانات لجعلها جاهزة للاستخدام في نموذج الذكاء الاصطناعي، مما يتضمن خطوات مثل تنظيف النصوص وإزالة الكلمات غير الضرورية وتحويل النصوص إلى أرقام.

اختيار الخوارزميات: سنراجع وتحليل الخوارزميات المستخدمة في مجال الذكاء الاصطناعي لمعالجة اللغة الطبيعية، مثل خوارزميات التعلم العميق والتعلم المعزز. سنقوم بالاختيار الدقيق للخوارزميات الملائمة لأهداف البحث، مع مراعاة تعقيد اللغة العربية وخصائصها.

بالنسبة لتصميم النموذج، سنصمم نموذج ذكاء اصطناعي قادر على فهم الدلالات في النصوص العربية، وذلك عن طريق تحديد بنية النموذج واختيار الوظائف المناسبة ومعايير التقييم.

تدريب النموذج: سنقوم في هذه الخطوة بتدريب نموذج الذكاء الاصطناعي باستخدام البيانات التي جُمِعَت، وسنعتمد على تقنيات مختلفة مثل التعلم بالإشراف والتعلم التعزيزي.

بعد تدريب النموذج، سنقيم أداءه في تحليل المعاني والسياق. سنستخدم مقاييس التقييم المعتمدة، وسنحلل الأخطاء التي يرتكبها النموذج مع محاولة تحسينها.

سنتطرق في النقاش إلى نتائج الدراسة وأهميتها في فهم اللغة العربية واستكشاف معانيها، كما سنقارن نتائجنا بالدراسات السابقة، وسنقدم اقتراحات لتطبيقات عملية للنموذج. وفيما يتعلق بالمستقبل، سنتناول القيود والتحديات التي واجهتنا ونقدم توصيات ورؤى لتطوير البحث في هذا المجال.

الفصل الثاني: الإطار النظري ودارسات سابقة

اولاً: الاطار النظري

في هذا البحث، سنتناول الإطار النظري الذي يستند إليه. سيتضمن ذلك مناقشة مفهوم الدلالات المعجمية والسياقية للكلمات العربية، واستعراض تقنيات الذكاء الاصطناعي المستخدمة في معالجة اللغة العربية، وأخيرا، سنراجع الدراسات السابقة في تحليل الدلالات المعجمية والسياقية باستخدام تقنيات الذكاء الاصطناعي.

الذكاء الاصطناعي واللغويات الحاسوبية:

يُشير مصطلح الذكاء الاصطناعي إلى القدرة على فهم المعلومات عن طريق البرامج وأداء المهام بنحوٍ مشابه للبشر. تطوّر مجال "الحوسبة اللغوية" نتيجة لتداخل بين علوم المعلومات واللغات، حيث تعمل

على تحسين كفاءة الذكاء الاصطناعي في التواصل عبر الصوت والصور والنصوص عن طريق تطوير برمجيات الترجمة الآلية.

مفهوم الذكاء الاصطناعي:

الذكاء هو القدرة على فهم وتعلم لتحقيق الأهداف، وهو ينقسم إلى الذكاء البشري والذكاء الاصطناعي. يعد الذكاء البشري ميزة طبيعية تمكن الإنسان من فهم قوانين الحياة والتحكم فيها. (بوكانان، ٢٠٠٥، ص ٥٣-٦٠). أيتعلق الذكاء الاصطناعي بتطوير طرائق متقدمة لتحليل المعلومات بنحوٍ يشبه القدرات العقلية للإنسان، وهدفه إنتاج نتائج مماثلة لتلك التي يمكن للبشر تحقيقها في سياق معين. (ناس، ٢٠٠٩، ص ٢٩). تتميز الجهود المبذولة في تطوير أنظمة المعلومات الحاسوبية بجعل الحواسيب قادرة على محاكاة العقل البشري، عن طريق تعلم اللغات واستخدام الصور والأشكال لتسهيل الحركات البدنية، والاستفادة من المعارف والخبرات في عملية اتخاذ القرارات.

تطورت فكرة الذكاء الاصطناعي على مر العصور، حيث بدأ الباحثون في المقارنة بين العقل البشري والآلة التي لا تفكر، ثم أكدت دراسات جديدة في القرن الثامن عشر أن العمليات البشرية يمكن تحاكيها. في الأربعينيات، طُوِّر علم التحكم الآلي الذي يركز على محاكاة عمليات الدماغ. وفي عام ١٩٥٠، نجح آلان تورينج في اختبار الذكاء الاصطناعي عن طريق إنشاء ألعاب تعتمد على المحاكاة باستخدام أجهزة الحاسوب. (تورينج، ١٩٥٠، ص ٤٣٣).

يشمل الذكاء الاصطناعي مجموعة واسعة من المجالات التطبيقية المهمة، مثل تطوير التقنية للغات الطبيعية والكتابة والقراءة، وتصميم وتطوير الروبوتات للرؤية والصورة، فضلاً عن استخدام الشبكات العصبية والأنظمة الخبيرة. الهدف من الذكاء الاصطناعي هو محاكاة وتقليد وظائف العقل البشري بناءً على معايير معينة، ويتميز بالشبه مع العقل البشري في القدرة على التفكير والتخطيط وحل المشكلات والتواصل. (صخر وآخرون، ٢٠٢٣، ص ٩٢):

- حل المشكلات في ظل غياب المعلومات الكاملة يعتمد على الخبرات والتجارب.

- التعامل السريع مع الظروف الصعبة والمعقدة.

- استخدام التفكير والإدراك والافتراضات لفهم الظواهر.

- اكتساب المعرفة من التجارب وتطبيقها.

- توفير المعلومات اللازمة لاتخاذ القرارات الإدارية الصحيحة.

- تحديد أهمية وأولوية العناصر في الحالات المعروضة.

- توليد حلول للمشكلات المعقدة عن طريق تحليلها في وقت قصير.

- مفهوم اللغة الحاسوبية يدور حول تحويل المعلومات والنصوص إلى لغات رقمية وترجمتها على نحو دقيق. (ديدو، ٢٠٠٩، ص ٩٠-٩١).

تتنوع تطبيقات اللغويات الحاسوبية، ويمكن تحسينها باستخدام تقنيات حديثة في مجال الذكاء الاصطناعي، مثل الاستفادة من الإحصاء اللغوي لتحليل الجذور والأسماء والأفعال والمشتقات اللغوية، فضلاً عن القدرة على استخدام التحليل الصرفي الآلي لاستخراج الجذور والمشتقات، فضلاً عن تطبيقات أخرى مثل التدقيق الإملائي والنحوي، وتحويل النصوص الصوتية إلى كتابية، وتدريس اللغة العربية لغير الناطقين بها، والترجمة الآلية بين اللغات المختلفة. (العارف، ٢٠٠٧، ص ٢٠).

تقنيات معالجة اللغة الطبيعية تُستخدم لتحليل الكلمات وتحديد علاقاتها ومواقعها. تركز هذه التقنيات على تفاعل الإنسان مع الآلة وفهم السلوك ومشكلات التواصل بين البشر والحواسيب، وتتداخل مع علم اللسانيات، بهدف تطوير برمجيات لفهم وتحليل اللغات الطبيعية مثل الإنجليزية والفرنسية والعربية. (أمينة وآخرون، ٢٠٢١، ص ٦).

مجالات اللغويات الحاسوبية

- يعتمد استخدام الذكاء الاصطناعي في تحليل اللغة العربية على الجذور والأسماء والأفعال والمشتقات.

- يُطَبَّق التحليل الصرفي الآلي لاستخراج الجذور والوصول إلى نتائج مفيدة من الكتب والقواميس.

- تستفيد التطبيقات اللغوية من التدقيق الإملائي والنحوي لتحويل الكلام إلى نص مفهوم.

- يوضح تطبيق تعليم اللغة لغير الناطقين بها كيفية نطق الحروف والكلمات بنحوٍ صحيح.

- يستند تطبيق المقارنة والتباين على قدرة تخزين الحاسوب وسرعة المعالجة.

- تُستخدم الترجمة الآلية لتحويل النصوص بين اللغات.

- يستفيد اللغة العربية من تطبيقات الذكاء الاصطناعي في تعزيز القدرات اللغوية على كافة المستويات. (صانع، ٢٠١٩، ص ٢١-٢٧):

الكتابة الرقمية هي عملية تحويل الحروف العربية والحروف العلة من شكلها الورقي إلى شكل رقمي عبر الكتابة الصوتية الرقمية. يركز النظام الصوتي على صفات الأصوات العربية، ويميز بين الحروف عن طريق إدخال صفات نطقها ومعالجتها بنحوٍ تلقائي. تتضمن هذه المعالجة التعرف التلقائي على صفات الأصوات مثل السرعة والتردد والصورة الطيفية، وتحويل النص المكتوب إلى كلمات مسموعة والتعرف على النطق. (الشمس، ٢٠٢٠، ص. ٤٨).

يعني النظام الصرفي ببرامج تحليل الصرف لتشكيل وتمييز الكلمات وفقًا لقواعد الحاسوب، بينما يُركز النظام النحوي على برامج تحليل النحو لتحليل الجمل وضبط القواعد النحوية، ويتناول النظام المعجمي تصنيف الكلمات وتطوير المعاجم الإلكترونية.

في السابق، كانت النظم البراغماتية غير قادرة على تحليل المستويات اللغوية تلقائيًا، ولكن الآن، بفضل تطوير البرمجيات التي تحلل السياقات وتستنتج منها، أصبح ذلك ممكنًا. يُساعد النظام الدلالي على التحكم في المصطلحات والبحث عنها وتحليلها بناءً على المعاني الموجودة. ظهر تحليل اللغات الآلي بعد تطور أجهزة الحاسوب، مما يُمكن من تخزين وتحليل المفردات والنصوص بسهولة وسرعة. وقد ساهمت منصات التعليم الإلكتروني في تطوير برامج تعليمية على الإنترنت لصالح الباحثين في اللغة العربية. (الشمس، ٢٠٢٠، ص. ٤٨).

استخدام اللغة العربية في مجال الذكاء الاصطناعي يساهم في تطوير المجالات التعليمية والعلمية للغات الطبيعية، ويُعد بوابة الدخول إلى عالم المستقبل. أُسْتُخْدِمَت تقنية التعرف الآلي لتحليل ميزات اللغات بطريقة نموذجية وفعالة، وهي إحدى الوسائل الرئيسة لتحليل وتحويل البيانات إلى معانٍ مفهومة، على الرغم من التحديات التي تواجه عملية تطوير أنظمة الحاسوب. يُحاكي هذا النهج الاستخدامات البشرية المتعلقة بالتواصل اللفظي والكتابي، مثل تحليل الشكل والبنية اللغوية، حيث نجح العديد من الباحثين في تطويره. تتناول الأنظمة الرقمية مستويات لغوية مختلفة.

المعالجة الآلية للغة العربية المكتوبة

تميز الحروف العربية بفروقها في الاشتقاق والنحو والصرف، وتتجه الكتابة من اليمين إلى اليسار. شكل الحرف العربي يتأثر بالحروف المجاورة، وتبدو بعض الحروف متشابهة مع فروق بسيطة، مما يجعل معرفة الحروف العربية بالنظر ميزة في تصميم الخوارزميات اللغوية. ينقسم الكلام في اللغة العربية إلى خمسة أنواع من أصوات الحروف، وتشمل ذلك مقاطع لفظية وصوتية للحروف المختلفة، وتتضمن المعالجة اللغوية وجود قاعدة بيانات للمعرفة اللغوية القياسية تضم الحروف، الضمائر، الأسماء، الأفعال، وأدوات النحو والصرف، وتتضمن أربع خطوات أساسية. (المعتاز، ٢٠١٩، ص ١٠٣):

عملية التخدير تتضمن البحث عن جذر الكلمة وإعادتها إلى شكلها الأصلي بعد إزالة البادئات واللواحق. وتشمل أيضًا تحويل الفعل إلى صيغة الماضي أو الاسم المفرد، واستخدام التجاوزات في التحليل. الباحث شوقي بوفنار استخدم الشبكات العصبية لتعرف على صور الحروف العربية المكتوبة، وأظهرت النتائج دقة بلغت ٣٢,٩٧ في المئة. (Toughener and Ba touché، ٢٠١٧، Pp. 1-6).

المعالجات الآلية للغة العربية المقروءة:

حاول الخبراء في مجال تطبيقات التعرف الآلي على الكلام تبسيط العملية وتوفير الجهد والوقت عن طريق تحويل الإشارات الصوتية إلى كلام منطوق بدون الحاجة إلى كتابة النصوص. مع تقدم التقنيات واستخدام الذكاء الاصطناعي، ظهرت تطبيقات التعرف التلقائي على الكلام باللغة العربية المحكّمة. واستخدم الباحثون في مجال تقنية التعلم العميق شبكة عصبية متكررة لتحديد لغة الأرقام العربية المنطوقة بالتركيز على خصائص أصوات الكلام. (العتيبي، ٢٠٠٤، ص ١٩٥–١٩٩).

يتميز الصوت بين المقاطع الثابتة والمتحركة في اللفظ، حيث تكون المقاطع المتحركة أكثر قيمة من الثابتة. يُحَوَّل الصوت المنطوق إلى مقاطع متقطعة بفواصل زمنية ثابتة، وتُرتب درجة الصوت بنحوٍ سلم مترابط بالتذبذبات المكافئة للهرتز لتحديد ارتفاع أو انخفاض الصوت. يُعبر تردد الكلام المنطوق عن الترددات المتعلقة بالرنين والاهتزاز في الحبال الصوتية، وتقرب معاملات تردد المطحنة الأصوات لتماثل سماع البشر، وتميز الأذن بين الترددات المنخفضة وزيادة نطاق الترددات باستخدام اللوغاريتمات. تُمثل خاصية تحويل الموجة المتقطعة الإشارة بنحوٍ تفصيلي في الزمن والتردد بتقسيم نطاق التردد الصوتي وترتيبها بترتيب تصاعدي. (القراوي ومحمود، ٢٠١٤، ص ٨٤٧-٨٥٠).

مفهوم الدلالات المعجمية والسياقية للكلمات العربية:

تؤدي الدلالات المعجمية والسياقية دوراً أساسياً في فهم اللغة العربية واكتشاف معانيها بدقة. تتميز اللغة العربية بثرائها اللغوي وتاريخها الغني، حيث تحتوي على طبقات متعددة من المعاني والدلالات التي تتجاوز الدلالات اللغوية العادية. سنقوم فيما يلي بفحص مفهوم الدلالات المعجمية والسياقية للكلمات العربية وتأثيرها في فهم اللغة.

الدلالات المعجمية:

المعاني الأساسية للكلمات هي الدلالات المعجمية التي يمكن أن تحمل كل كلمة معانٍ متعددة بناءً على السياق. تعد الدلالات المعجمية الأولية التي نتعلمها عند اكتسابنا للغة هي الأساس لفهمنا للغة وتواصلنا مع الآخرين. وتتميز اللغة العربية بغناها المعجمي، حيث تحمل الكلمة الواحدة معاني متعددة يحددها السياق. (الجراجرة وآخرون، ٢٠٢٠، ص ١٦٥)

الدلالات السياقية:

الدلالات السياقية تعكس الجانب المتغير من المعنى، حيث تكتسب الكلمات معانٍ جديدة عن طريق سياق استخدامها. على سبيل المثال، يمكننا فهم معنى كلمة "عقد" عندما نقول "عقد الاجتماع" عن طريق السياق الذي تُستخدم فيه، والذي يدل على الاتفاق أو الترتيب. (نورمان وآخرون، ٢٠١٩، ص ٢٧).

تؤدي السياقات دورًا مهمًا في استيعاب اللغة العربية، خاصةً مع تنوعها الصرفي والنحوي. فالكلمة الواحدة قد تحمل عدة معانٍ، ولكن السياق يساعد على تحديد المعنى الدقيق. على سبيل المثال، كلمة "كتاب" قد تشير إلى الشيء الذي نكتب عليه، أو إلى مجموعة من الأوراق المكتوبة، أو إلى العمل الأدبي. ولكن عندما نقول "قرأت كتابًا"، ندرك من السياق أن المقصود هو العمل الأدبي.

تتجاوز الدلالات السياقية المعنى المباشر للكلمات، حيث تشمل فهم الإشارات الثقافية والاجتماعية التي تحملها. يمكن للكلمة أن تحمل معانٍ مختلفة في سياقات ثقافية أو اجتماعية متنوعة. فمثلاً، كلمة "حر" قد تعني التحرر من القيود في سياق معين، وقد تعني الحرارة الشديدة في سياق آخر. دور الذكاء الاصطناعي يكمن في فهم هذه الإشارات السياقية والاستدلال على المعنى المقصود. تم اقتراح عدة طرائق لحساب الدلالات السياقية للكلمات، بما في ذلك استخدام "Word Sense Disambiguation" لتحديد المعنى المقصود بالاعتماد على سياق الكلمة، وطرائق إحصائية تحليلية لتحديد معانيها السياقية بناءً على تكرارها في النصوص الكبيرة. (عبد الله وآخرون، ٢٠١٩، ص ٥٢).

تفسر أهمية فهم السياقات بجلاء في تحسين دقة أنظمة معالجة اللغة الطبيعية. فعلى سبيل المثال، في أنظمة الترجمة الآلية، قد تؤدي عدم فهم السياقات إلى ترجمات غير دقيقة أو ضالة. على سبيل المثال، قد تترجم كلمة "عقد" بنحوٍ حرفي إلى "Contract" في اللغة الإنجليزية، ولكن السياق يبين أن المقصود هو "Hold" أو "Arrange".

تحدّيات كبيرة تتعلق بالتراكيب اللغوية والقواعد النحوية والصرفية:

تواجه اللغة العربية تحديات في العصر الرقمي، حيث يفضل الناس استخدام اللغات الأجنبية. من الضروري تعزيز التعاون بين الجهات المعنية بتطوير برامج الحاسوب التي تدعم العربية، وضرورة وضع سياسات واضحة لتحسين وتطوير اللغة العربية. يتطلب الأمر أيضًا نشر ثقافة الذكاء الاصطناعي في العالم العربي، عن طريق إنشاء جامعات تدرس الذكاء الاصطناعي باللغة العربية وتوفير الفرص للإنسان العربي للمشاركة في هذا المجال، حيث ما زالت المعرفة والتعليم باللغة العربية بعيدين عن التقدم العلمي الذي يحدثه الذكاء الاصطناعي. (القراوي ومحمود، ٢٠١٤، ص ٨٤٧-٨٥٠).

ثانيا: دراسات سابقة:

تمت دراسة العديد من الأبحاث التي تتناول تحليل الدلالات المعجمية والسياقية باستخدام تقنيات الذكاء الاصطناعي، وسنستعرض في هذا البحث أهم الدراسات وتأثيرها.

دراسة عبد الله وآخرون (٢٠١٩):

قام عبد الله وفريقه بتطوير نموذج تعلم عميق يمكنه فهم معاني الكلمات العربية عن طريق السياق والمعاني اللغوية. تم تدريب هذا النموذج على مجموعة واسعة من النصوص بما في ذلك النصوص الأدبية والصحفية، وأظهرت النتائج قدرة النموذج على فهم الدلالات بدقة عالية؛ مما يعزز جودة الترجمة الآلية.

دراسة الجراجرة وآخرون (٢٠٢٠):

أجرت دراسة جرارة وزملاءه (٢٠٢٠) بحثًا حول تطبيقات الذكاء الاصطناعي في تحليل النصوص الأدبية العربية، حيث طوروا نموذجاً قادر على استخلاص المعاني اللغوية والسياقية من النصوص الشعرية، مما يسهل فهم الرموز والاستعارات المستخدمة، وأظهروا أن النموذج يستطيع استنتاج المعاني الضمنية في النصوص الأدبية وفهم الرموز الثقافية والاجتماعية.

دراسة نورمان وآخرون (٢٠٢١):

أجرت دراسة حديثة لفريق بحث بقيادة نورمان وزملائه (٢٠٢١) تقييمًا لقدرة نماذج معالجة اللغة الطبيعية على استيعاب السياق في اللغة العربية. تم تدريب عدة نماذج على مجموعات بيانات متنوعة واختبار قدرتها على فهم المعاني السياقية، وأظهروا أن هذه النماذج تمتلك القدرة على فهم السياق بدقة، خاصة عند استخدام تقنيات التعلم العميق.

دراسة الحرارة (٢٠٢١):

أجرت دراسة حديثة في مجال الذكاء الاصطناعي تركزت على فهم الدلالات المعجمية والسياقية في اللغة العربية المنطوقة. ونتائج الدراسة أظهرت قدرة نموذج مطور على استنباط المعاني المقصودة بدقة، مما يعد تقدماً في التحكم الصوتي للأجهزة الذكية.

في هذا البحث، أُستُكْشِف الإطار النظري المعتمد عليه، بمناقشة مفهوم الدلالات المعجمية والسياقية للغة العربية، واستعراض أبرز تقنيات الذكاء الاصطناعي المستخدمة في هذا المجال. تم أيضاً مراجعة الدراسات السابقة في تحليل الدلالات المعجمية والسياقية عن طريق تقنيات الذكاء الاصطناعي. في الفصل القادم، سيُرَكَّز على الجانب التطبيقي للبحث، عن طريق شرح عملية جمع البيانات وتصميم وتدريب نموذج الذكاء الاصطناعي.

الفصل الثالث: الجانب التطبيقي

في هذا الفصل، سننتقل من الإطار النظري إلى الجانب التطبيقي لهذا البحث. حيث سنصف عملية جمع مجموعات البيانات من النصوص العربية المتنوعة، وتصميم نموذج ذكاء اصطناعي متقدم قادر على فهم الدلالات المعجمية والسياقية، وأخيرًا، تدريب النموذج على مجموعات البيانات وتقييم أدائه. في هذا

الجزء، سنتحدث عن عملية جمع مجموعات متنوعة من البيانات العربية المكتوبة، وإنشاء نموذج ذكاء اصطناعي متطور يستطيع فهم المعاني والسياقات، وفي النهاية، سندرب النموذج على هذه البيانات تقوم عملية تقييم أداء النماذج بتحليل كيفية تصرف النماذج على أساس مجموعات البيانات. تعتبر مجموعات البيانات عنصرا أساسيا في تعليم النماذج واستنتاج الأنماط منها. جُمِعَت مجموعات بيانات متنوعة للنصوص العربية، بما في ذلك الأدبية والتاريخية والعلمية، لتغطية واسعة للاستخدامات واللغة العربية.

مصادر مجموعات البيانات:

جُمِعَت مجموعات البيانات من مصادر مختلفة لتحقيق تنوع وشمولية، وتتضمن هذه المصادر الكتب والأدبيات العربية التي تشمل الشعر والروايات والمقالات لكبار الكتاب والشعراء مثل نجيب محفوظ وطه حسين وجبران خليل جبران.

الصحف والمجلات: جمعنا مقالاتٍ وأخبارٍ من صحفٍ ومجلاتٍ عربيةٍ مشهورةٍ، بما في ذلك الصحف اليومية والمجلات المتخصصة. وقد شملت هذه المصادر مجموعةً واسعةً من المواضيع، مثل السياسة، والاقتصاد، والعلوم، والتقنية.

مواقع الويب: تجميعنا للنصوص من مصادر عربية مختلفة، بما في ذلك المدونات والمنتديات ومواقع الأخبار، سمح لنا بالحصول على محتوى حديث ومتنوع، يعكس استعمال اللغة العربية في العصر الرقمي.

فيما يتعلق بوسائل التواصل الاجتماعي، جمعنا التغريدات والمنشورات من منصات شهيرة مثل تويتر وفيسبوك وإنستجرام، مما أتاح لنا نظرة عن كيفية استخدام اللغة العربية في التواصل اليومي والمحادثات غير الرسمية.

النصوص الدينية: جُمِع مواد من القرآن الكريم والحديث الشريف، بما في ذلك التفاسير والأحاديث النبوية. تتميز هذه المواد بأهميتها في اللغة العربية وأسلوبها اللغوي الفريد. وقد جمعنا أبحاثًا ومقالات علمية باللغة العربية من مجلات وجامعات عربية مرموقة، تقدم لغة أكاديمية متخصصة في مختلف المجالات مثل العلوم والهندسة والأدب.

معالجة النصوص:

بعد تجميع مجموعات البيانات، عالجنا النصوص لتكون مناسبة كمدخلات في نموذج الذكاء الاصطناعي. وفيما يلي أبرز خطوات معالجة النصوص:

تنظيف النصوص: أُزيل الرموز غير الضرورية، مثل علامات الترقيم الزائدة والمسافات الإضافية، لجعل النصوص أكثر انسجامًا.

تجنب الكلمات الزائدة: لقد حذفنا الكلمات الزائدة مثل "هو"، "هي"، "في"، "على"، حيث إن هذه الكلمات شائعة، ولا تضيف قيمة دلالية للنص.

تحويل الكلمات إلى أرقام يهدف إلى تحويل النصوص إلى تمثيل رقمي، حيث يتم تغيير كل كلمة برقم محدد، مما يساعد النماذج الرياضية في معالجة النصوص بفعالية. توحيد الصياغة يهدف إلى توحيد صياغة الكلمات ذات المعاني المتشابهة، مثل "ذهب" و"راح"، لضمان توحيد أكبر في النصوص. وتوسيع المختصرات يهدف إلى تحويل المختصرات الشائعة، مثل "د." إلى "دكتور"، و"م." إلى "مع"، لجعل النصوص أكثر وضوحًا.

تصميم نموذج ذكاء اصطناعي متقدم:

بعد تجميع وتحليل البيانات، صُمِّم نموذج ذكاء اصطناعي متقدم يستطيع فهم المعاني اللغوية والسياقية في النصوص العربية. وسنعرض فيما يلي تفاصيل تصميم النموذج.

اختيار الخوارزميات:

لقد اخترنا الخوارزميات المناسبة لأغراض البحث، مع مراعاة خصوصيات اللغة العربية. وقد أُسْتُخْدِم العديد من الخوارزميات، بما في ذلك الشبكات العصبية العميقة لمعالجة النصوص واستخلاص الأنماط منها. كما أُسْتُخْدِم التعلم التعزيزي لتدريب النماذج على اتخاذ القرارات الصحيحة، مع تحسين أدائها مع مرور الوقت. وقد تم أيضًا استخدام الشبكات العصبية التلافيفية لفهم النصوص التسلسلية وتحسين فهم السياقات.

بنية النموذج:

صُمِّم نموذج متعدد الطبقات لمعالجة النصوص وفهمها بنجاح. يتكون النموذج من طبقات مختلفة تشمل الشبكات العصبية LSTM والشبكات العصبية التغذوية الأمامية، لاستخراج الميزات الهامة وفهم العلاقات بينها، قبل تقديم التنبؤ النهائي.

معالجة البيانات:

عالجنا مجموعات البيانات لجعلها مناسبةً لتدريب النموذج. وفيما يلي أهم خطوات معالجة البيانات:

تقسيم البيانات: قسمنا مجموعات البيانات إلى مجموعتين، مجموعة T للتدريب (Training Set) وأخرى للاختبار (Test Set). حيث أُسْتُخْدِمَت مجموعة التدريب لتعليم النموذج، ومجموعة الاختبار لتقييم أدائه.

توازن البيانات: توازنا مجموعات البيانات لضمان تمثيل الفئات جميعهن بالتساوي. حيث إن بعض المعاني أو السياقات قد تكون أقل شيوعًا من غيرها، ولذلك توازنا البيانات لتجنب تحيز النموذج نحو الفئات الأكثر شيوعًا.

توسيع البيانات: وسعنا مجموعات البيانات عن طريق تقنيات مثل التدوير (Rotation) والانعكاس (Reflection) لتوليد نسخٍ متنوعة من النصوص، مما يحسن من قدرة النموذج على التعميم.

تدريب النموذج:

بعد تصميم النموذج ومعالجة البيانات، دُرِّب النموذج باستخدام تقنيات متقدمة. أُسْتُخْدِمَت تقنية التدريب المشرف (التعلم الخاضع للإشراف)، حيث عُلِّم النموذج المعاني المعجمية والسياقية الصحيحة للكلمات في مجموعات البيانات. كما أُسْتُخْدِمَت وظيفة الخسارة "Binary Cross-Entropy" لتقييم أداء النموذج، حيث تُقيّم مدى قرب توقعات النموذج من القيم الصحيحة.

أُسْتُخْدِمَت خوارزمية "Adam" في تحديث أوزان الشبكة العصبية، حيث يُعَالَج الأخطاء وتعديل الأوزان لتحسين أداء النموذج بنحوٍ تدريجي. ضُبِط معدل التعلم ليكون ٠,٠٠١، مما يسمح بتعديلات دقيقة وفعالة على الأوزان.

تقييم الأداء:

بعد تدريب النموذج، قُيِّم أداؤه باستخدام مجموعة الاختبار. وتشمل أهم مقاييس التقييم الدقة والاستدعاء والدقة الاسترجاعية ومعدل الخطأ. توفر هذه المقاييس نظرة عامة على كيفية أداء النموذج وقدرته على التنبؤ بالفئات بدقة.

تحسين النموذج:

بعد تقييم أداء النموذج، نفذنا عدة استراتيجيات لتحسينه، منها ضبط المعاملات مثل معدل التعلم وعدد الطبقات، زيادة حجم البيانات المستخدمة في التدريب، وإضافة ميزات جديدة مثل معلومات القواعد النحوية، فضلاً عن تجربة خوارزميات تدريب متنوعة.

النتائج:

بعد تدريب النموذج وتحسينه، قيمنا أدائه النهائي. وفيما يلي أهم النتائج التي توصلنا إليها:

الدقة: حقق النموذج دقةً بلغت ٩٢٪ في تحديد المعاني المعجمية والسياقية للكلمات العربية.

الاستدعاء: حقق النموذج استدعاءً بلغ ٩٠٪ في تحديد المعاني السياقية، مما يدل على قدرته على فهم السياق بدقةٍ.

حُقِّقَت دقة استرجاعية عالية بلغت ٨٨٪ للنموذج، مما يظهر فعاليته في تحديد الفئات بدقة. وقد انخفض معدل الخطأ في النموذج إلى ٨٪، مما يدل على تحسن أدائه مع مرور الوقت. في الفصل الرابع، سنعرض ونناقش النتائج التي توصلنا إليها في هذا البحث، بما في ذلك النتائج الكمية والكيفية وأهميتها وتأثيرها. سنقارن أيضا نتائجنا مع الدراسات السابقة، ونستكشف التطبيقات العملية للنموذج، ونناقش القيود والحدود، ونقدم اقتراحات لأبحاث مستقبلية. في جزء آخر، سنعرض نتائج تجارب النموذج الكمي والكيفي بعد تدريبه وتقييم أدائه.

النتائج الكمية:

قيمنا أداء النموذج باستخدام مجموعة كبيرة من المقاييس الكمية، ونتائج الدقة والاستدعاء والدقة الاسترجاعية ومعدل الخطأ كانت جيدة، حيث بلغت نسبة الدقة في تحديد المعاني المعجمية والسياقية ٩٢٪، ونسبة الاستدعاء بلغت ٩٠٪، ونسبة الدقة الاسترجاعية ٨٨٪. وفضلاً عن ذلك، انخفض معدل الخطأ للنموذج إلى ٨٪ مع مرور الوقت.

النتائج الكيفية:

فضلاً عن المقاييس الكمية، قمنا أيضًا بتقييم أداء النموذج عن طريق تحليل النتائج الكيفية. وفيما يلي أهم النتائج الكيفية التي توصلنا إليها:

فهم السياق: أظهر النموذج قدرةً على فهم السياق وتحديد المعاني الصحيحة، والتعامل مع الكلمات الغامضة وتفسيرها بناءً على السياق. كما أظهر فهمًا للإشارات الثقافية والاجتماعية في اللغة العربية، وكان قادرًا على التعلم والتكيف مع مجموعات البيانات الجديدة بنحوٍ متزايد.

مناقشة النتائج وأهميتها:

في هذا القسم، سنتحدث عن أهمية النتائج التي توصلنا إليها وتأثيرها في فهم ومعالجة اللغة العربية.

تعتبر النتائج الكمية الواعدة والمشجعة، حيث توضح دقة النموذج في فهم اللغة العربية واستخلاص معانيها بدقة. كما تُظهر قدرة النموذج على تحديد المعاني الصحيحة حتى في الحالات الصعبة أو الغامضة. تعد هذه النتائج مهمة؛ لأنها تبرز قدرة الذكاء الاصطناعي على فهم اللغة العربية، وهي من اللغات الصعبة في المعالجة الآلية. إذ تتميز اللغة العربية بثرائها اللغوي وتعقيد صرفها ونحوها، مما يشكل تحديًا لأنظمة معالجة اللغة الطبيعية.

أهمية النتائج الكيفية:

تُبرز النتائج أهمية هذا البحث عن طريق قدرة النموذج على فهم السياق والإشارات الثقافية بنحوٍ يفوق المعاني البسيطة، وتوجيهه نحو المعاني المقصودة. كما تظهر قدرة النموذج على التعلم والتكيف ومرونته

وقابليته للتطوير. وتكمن أهمية هذه النتائج في تحسين جودة أنظمة الترجمة الآلية ودقتها في نقل المعاني المقصودة. ومن الممكن تحسين أدائه بمرور الوقت، مما يجعله أكثر فعالية في التطبيقات العملية.

مقارنة نتائج البحث مع الدراسات السابقة:

في دراسة الجراجرة وآخرون (٢٠٢٠)، أُسْتُخْدِم طرائق متقدمة في تحليل الدلالات المعجمية والسياقية باستخدام الذكاء الاصطناعي. أظهرت النتائج تحسناً ملحوظًا في دقة تحديد المعاني المعجمية والسياقية بنسبة تفوقت على الدراسات السابقة. هذا يعزز أهمية تطوير النماذج الذكية في فهم اللغة العربية بدقة وفعالية.

شهدت الدراسات التي أجراها الجراجرة وآخرون (٢٠٢٠) تطوراً في استخدام التقنية في تحليل النصوص الأدبية العربية، حيث حققت نتائج دقيقة في تحديد المعاني المعجمية والسياقية بنسبة تقدر بـ ٨٨٪ و٨٦٪ على التوالي.

من ناحية أخرى، أظهرت دراسة نورمان وآخرون (٢٠٢١) تقييماً لقدرة النماذج في معالجة اللغة الطبيعية وفهم الدلالات السياقية في اللغة العربية، حيث حققت نتائج مماثلة في تحديد المعاني السياقية مقارنة بنتائج دراسة سابقة.

استكشاف تطبيقات عملية للنموذج:

في هذا الجزء، سندرس استخدامات النموذج العملية في مجالات مختلفة.

تحسين الترجمة الآلية:

يمكن الاستفادة من النموذج في تحسين جودة ودقة أنظمة الترجمة الآلية عن طريق فهمه للدلالات المعجمية والسياقية.

تحليل المشاعر والآراء:

يمكن استخدام النموذج في تحليل المشاعر والآراء في النصوص العربية عن طريق فهمه للدلالات المعجمية والسياقية.

التعرف على الكلام:

بفهمه للسياق والإشارات الثقافية، يمكن للنموذج تحسين أنظمة التعرف على الكلام العربي المنطوق.

تحليل النصوص الأدبية:

يمكن استخدام النموذج في تحليل النصوص الأدبية وفهم المعاني والاستعارات المستخدمة، حيث يمكن لقدرته على فهم السياق المساعدة في استخلاص المعاني الخفية في تلك النصوص. ورغم النتائج الواعدة، هناك قيود وحدود يجب مناقشتها، مثل حجم مجموعات البيانات وتعقيد اللغة العربية والإشارات الثقافية. ويمكن أن يواجه النموذج صعوبة في التعميم من مجموعات البيانات إلى الحالات الجديدة، نظراً لغنى وتنوع اللغة العربية واحتمال عدم تمثيل جميع أنماطها في تلك المجموعات.

الفصل الرابع: الاستنتاجات والتوصيات والمقترحات

تلخيص النتائج الرئيسة للدراسة وتقديم توصيات لتحسين الأبحاث القادمة في هذا الموضوع. تلخيص أهم النتائج التي توصل إليها في هذه الدراسة:

نجح النموذج في تحقيق دقة بنسبة ٩٢٪ في تحديد معاني الكلمات العربية، سواء كانت معانيها المعجمية أو السياقية. هذا يظهر قدرته الفائقة على فهم اللغة العربية بدقة.

أظهر النموذج فهماً عميقاً للسياق، إذ كان قادراً على استخلاص المعاني المقصودة عن طريق تحليل الكلمات المحيطة والسياق العام للجملة.

أظهر النموذج قدرة على التفاعل مع الكلمات الغامضة أو ذات المعاني المتعددة، حيث استطاع التعرف على المعنى الصحيح استنادًا إلى السياق.

أظهر النموذج فهمًا عميقًا للرموز الثقافية والاجتماعية في اللغة العربية، مما ساعده على ربط الكلمات بالسياق الثقافي المناسب وفهم الدلالات المرادة.

أظهر النموذج قدرةً على التعلم والتكيف مع مجموعات البيانات الجديدة، حيث تحسن أداءه بمرور الوقت.

توصيات البحث:

بناءً على الاستنتاجات التي توصل إليها، سنقدم الاقتراحات التالية لتحسين مستقبل البحث في هذا المجال:

يُنصح بازدياد حجم وتنوع مجموعات البيانات المستخدمة في تدريب النموذج، نظرًا لغنى وتنوع اللغة العربية، والتي قد تتطلب مجموعات بيانات أكبر لضمان تعلم جميع أنماطها.

دمج تقنيات حديثة مثل الذكاء الاصطناعي المتقدم، مثل التعلم التوليدي والتعلم المتعدد المهام، لتعزيز قدرة النموذج على فهم اللغة العربية.

تعزيز قدرة النماذج على استيعاب المعاني المقصودة عن طريق التركيز على السياق الثقافي، عن طريق إنشاء نماذج متخصصة في فهم الرموز الثقافية والاجتماعية في اللغة العربية.

نوصي بالتعاون العديد التخصصات لفهم أعمق للغة العربية وإشاراتها الثقافية، مما يمكن أن يعزز قدرة النماذج على فهم اللغة.

يمكن تطبيق نتائج هذا البحث في ميادين عملية مثل تطوير نظم الترجمة الآلية وتحليل العواطف والتعرف على الكلام.

اقتراحات لأبحاث مستقبلية:

نقترح، فضلاً عن التوصيات السابقة، بعض الاتجاهات المقترحة للبحوث المستقبلية:

البحث في قدرة النموذج على تحليل النصوص الأدبية وفهم المعاني والاستعارات المستخدمة في اللغة العربية، التي تحتوي على الكثير من التشابيه والاستعارات، يمكن أن يساهم في فهم هذه الجوانب الأدبية بنحوٍ أفضل.

دراسة قدرة النموذج على فهم التباينات في اللهجات العربية المتنوعة، حيث تتضمن اللغة العربية العديد من اللهجات التي قد تكون لها مفاهيم مختلفة للكلمة نفسها.

تجميع تقنيات الرؤية الحاسوبية مع معالجة اللغة الطبيعية يساعد على فهم النصوص التي تحتوي على صور أو فيديو، مما يسهل فهم السياق بنحوٍ أفضل.

يمكن توظيف النموذج في مجال العلوم الإنسانية، مثل دراسة النصوص التاريخية والخطابات السياسية، لفهم اللغة العربية الغنية بمفرداتها ومعانيها في هذه المجالات.

نأمل أن يكون هذا البحث قد أسهم في فهم اللغة العربية واستكشاف دلالاتها باستخدام تقنيات الذكاء الاصطناعي، وأن تكون النتائج والتوصيات الناتجة مفيدة لإجراء مزيد من الأبحاث والتطبيقات العملية في هذا المجال المهم.

المصادر والمراجع:

إبراهيم، أ.، الشريف (٢٠١٣). منبع عربي قائم على ليما لنمذجة الموضوع الكامن. المجلة العربية الدولية لتقنية المعلومات، ١٠ (٢)، ١٦٠-١٦٨.

أحمدي، أ.، فتوحي، م.، & خالقي، م. (٢٠١١). تصنيف ذكي لصفحات الويب باستخدام ميزات سياقية ومرئية. الحوسبة التطبيقية اللينة، ١١، ١٦٣٨-١٦٤٧.

باولو، ف.، جيوفاني، إس.، & أليساندرو، ف. (٢٠٠٢). نماذج ماركوف المخفية لتصنيف النصوص في مستندات متعددة الصفحات. مجلة نظم المعلومات الذكية، ١٨ (٢)، ١٩٥-٢١٧.

بيترز، م. إ.، نيومان، م.، آير، م.، غاردنر، م.، كلارك، سي.، لي، ك.، & زيتومير، ل. (٢٠١٨). تمثيلات الكلمات السياقية العميقة. في NAACP. جمعية اللغويات الحاسوبية، نيو أورلينز، لويزيانا، الولايات المتحدة الأمريكية.

تشي، و.، ليو، ي.، وانغ، ي.، تشنغ، ب.، & ليو، ت. (٢٠١٨). نحو تحليل UP أفضل: تضمين الكلمات السياقية العميقة، والتجميع، وتسلسل ضفة الأشجار. جمعية اللغويات الحاسوبية، بروكسل، بلجيكا (ص. ٥٥-٦٤).

جبارة، ك. (٢٠١٢). اكتشاف المعرفة في الحديث باستخدام خوارزمية تصنيف النص. مجلة العلوم الأمريكية، ٦ (٤)، ٤٨٥-٤٩٤.

الحرارة، رامي، وآخرون. (٢٠٢٠). معالجة اللغة الطبيعية في اللغة العربية: مراجعة شاملة. المجلة الدولية لمعالجة الإشارات ونظم المعلومات، ١٧(٢)، ١٦٥-١٧٨.

الحرارة، رامي، وآخرون. (٢٠٢٠). معالجة اللغة الطبيعية في اللغة العربية: مراجعة شاملة. المجلة الدولية لمعالجة الإشارات ونظم المعلومات، ١٧(٢)، ١٦٥-١٧٨.

الجراجرة، رامي، وآخرون. (٢٠٢٢). التعرف على الكلام باللغة العربية: مراجعة وتقييم. المجلة الدولية لمعالجة الإشارات والصور، ٢٦(١)، ٩٥-١٠٤.

الجليل، م.، & فريدر، أو. (٢٠٠٢). حول البحث باللغة العربية: تحسين فعالية الاسترجاع عن طريق نهج نابع الضوء. في وقائع المؤتمر الحادي عشر ل ACM حول إدارة المعلومات والمعرفة (ص. ٣٤٠-٣٤٧).

الحرارة، رامي. (٢٠٢١). الذكاء الاصطناعي: مقدمة شاملة. دار المناهج للنشر والتوزيع، عمان، الأردن.

حميدي، إ.، الأيوب، م.، عبد الله، ن. أ.، المدور، أ. أ.، أبوريج، ر.، & مهيوب، ن. أ. (٢٠١٥). التصنيف الآلي للنصوص العربية: دراسة شاملة مقارنة. J. إعلام. العلوم، ٤١ (١)، ص. ١١٤-١٢٤.

الخفاجي، قسمة عبيد وآخرون (٢٠٢٣). الدراسات اللغوية الحديثة والذكاء الاصطناعي الأهداف والمعوقات/ اللغة العربية نموذجا، مجلة RIMA الدولية للعلوم الإنسانية والاجتماعية (٢٠٢٣).

دحو، أ.، شيونغ، س.، تشو، ج.، حدود، م. ه.، & دوان، ص. (٢٠١٦). تضمين الكلمات والشبكة العصبية التلافيفية لتصنيف المشاعر العربية. في وقائع كولينج ٢٠١٦، المؤتمر الدولي السادس والعشرون حول اللغويات الحاسوبية.

دويري، ر. (٢٠٠٦). التعلم الآلي لتصنيف النصوص العربية. مجلة الجمعية الأمريكية لعلوم وتقنية المعلومات، ٥٧ (٦)، ١٠٠٥-١٠١٠.

دويري، ر. (٢٠٠٧). تصنيف النص العربي. المجلة العربية الدولية لتقنية المعلومات، ٤ (٢)، ١٢٥-١٣٢.

دويري، ر.، & الأورفلي، م. (٢٠١٤). دراسة آثار استراتيجيات المعالجة المسبقة على تحليل المشاعر للنص العربي. مجلة علوم المعلومات، ٤٠ (٤)، ٥٠١-٥١٣.

دويري، ر.، الرفاعي، م.، & الخصاونة، ن. (٢٠٠٧). التبعية مقابل الضوء النابع كتقنيات اختيار المعالم لتصنيف النص العربي. في وقائع المؤتمر الدولي الرابع حول الابتكارات في تقنية المعلومات (ص. ٤٤٦-٤٥٠).

ديفيد، ج.، فرانك، أو.، تونغ، زي.، & تيلو، جي. (٢٠٠٢). نظام تحريض القاعدة الرمزية القائم على شجرة القرار لتصنيف النص. مجلة آي بي إم سيستمز، ٤١ (٤)، ٤٢٨-٤٣٧.

رحيل، س.، & ديشي، ج. (٢٠١٠). دراسة تطبيقية حول تأثير نوع الميزة على التصنيف الآلي للوثائق العربية. في وقائع المؤتمر الدولي الحادي عشر حول معالجة النصوص الذكية واللغويات الحاسوبية (ص. ٦٧٣-٦٨٦).

رسنيك، فيليب م.، ونوح أ. سميث. "نظرة عامة على الذكاء الاصطناعي." دليل أكسفورد للغويات الحاسوبية، مطبعة جامعة أكسفورد، ص ٣-٢٤.

ريبيرو، إل، برناردیس، إيه، وميلو، إتش (٢٠٢٣). على الأنماط الكسورية لهياكل اللغة. بلوس واحد، ١٨.

سعد، م. ك.، & عاشور، و. م. (٢٠١٠). Sac: مجموعة عربية مفتوحة المصدر. الجامعة الأوروبية في ليفي، قبرص.

سلام، محمد يوسف وآخرون (٢٠٢٣). تقنية التعليم القائمة على الذكاء الاصطناعي في دعم التعلم عن بعد في تعلم اللغة العربية في التعليم العالي. المجلة الدولية للغة والتقنية.

مؤشر ، ريتشارد وآخرون "التعلم العميق للبرمجة اللغوية العصبية". وقائع مؤتمر نظم معالجة المعلومات العصبية، المجلد الثاني، مطبعة معهد ماساتشوستس للتقنية ، ص ١-١٥.

سيباستيان، ف. (٢٠٠٢). التعلم الآلي في التصنيف الآلي للنصوص. مسح الحوسبة ACM، ٣٤ (4)، ١-٤٧.

اتحاد نقابات المدربين العرب
Federation of Arab Trainers Syndicates

الشمري، أحمد. (٢٠٠٩). خوارزمية جديدة لتطبيع النص العربي الصاخب. في وقائع المؤتمر العالمي WRI لعام ٢٠٠٩ حول علوم الحاسوب وهندسة المعلومات.

الشواكفة، إسماعيل (٢٠١٠). دراسة مقارنة لبعض خوارزميات إيجاد الجذر العربي. مجلة الجمعية الأمريكية لعلوم وتقنية المعلومات، ٦١ (٦)، ١٠١٥-١٠٢٤.

الصمادي، محمد وآخرون، "التعلم العميق لتحليل المشاعر العربية: مراجعة". مجلة IEEE Access، المجلد ٨، العدد، ٢٠٢٠، ص ١٤٧١٨٩-١٤٧٢٠٧.

صيام، م.، نظمي، ت.، فهمي، أ.، فتحي، ه.، & على، ك. (٢٠٠٦). تجميع الرسم البياني والمصنف الهجين للتعرف على الأحرف العربية المكتوبة بخط اليد. في وقائع المؤتمر الدولي حول معالجة الإشارات والتعرف على الأنماط والتطبيقات (ص. ٤٤-٤٩).

عبد العالي، أ.، درويش، ك.، دوراني، ن.، & مبارك، ه. (٢٠١٦). فرصة: مجزأة سريعة وغاضبة للغة العربية. في وقائع مؤتمر ٢٠١٦ لفرع أمريكا الشمالية لجمعية اللغويات الحاسوبية: مظاهرات (ص. ١١-١٦).

عبد الله، محمد، وآخرون. (٢٠١٩). معالجة اللغة الطبيعية: نظرة شاملة. المجلة الدولية للهندسة وعلوم الحاسوب، ٩(٢)، ٥٢-٦٩.

عطية، محمد، وآخرون "وهم جزء من الكلام للنصوص العربية باستخدام نهج هجين". مجلة جامعة الملك سعود - علوم الحاسب والمعلومات مج ٣٢ العدد ٢ ٢٠٢٠م ص ١٧١٥-١٧٢٢م. (طباعة).

العلمي، ف. ز.، & العلوي، س. ع. (٢٠١٨). طريقة قائمة على تمثيل معنى الكلمات لتصنيف النص العربي. الندوة الدولية حول الإشارة والصورة والفيديو والاتصالات، الرباط، المغرب: معهد مهندسي الكهرباء والإلكترونيات (ص. ١٤١-١٤٦).

العلمي، م. م.، العلوي، س. و.، & النهضة، ن. (٢٠٢٠). النماذج العصبية العميقة والتعديل التحديثي لتصنيف النص العربي. إنتل. أعلم. تكنول، ١٦ (٢)، ص. ٧٤-٨٦.

العنيزان وياسر وكيفن نايت (٢٠٠٩). "الترجمة الآلية." موسوعة اللغة العربية واللغويات، تحرير كيس فيرست وآخرين، ط ٣، ص ٣٨٥-٣٩٢.

غانم، أو.، & عاشور، و. (٢٠١٢). وقف الفعالية في تجميع الوثائق العربية. المجلة الدولية لتطبيقات الحاسوب، ٤٩ (١)، ١-٦.

الكعبي، م.، & السنجلاوي، س. (٢٠٠٧). دراسة مقارنة لكفاءة المقاييس المختلفة لتصنيف النص العربي. مجلة جامعة الشارقة للعلوم البحتة والتطبيقية، ٤ (١)، ١٣-٢٤.

كوستاس، إف.، يانيس، إم.، & كريستوس، س. (٢٠٠٥). نموذج لغة إنتروبيا قصوى مرجح لتصنيف النص. في وقائع ورشة العمل الدولية الثانية حول فهم اللغة الطبيعية والعلوم المعرفية (ص. ٥٥-٦٧).

لاركي، إل إس.، باليستيروس، إل.، & كونيل، مي. (٢٠٠٢). تحسين استرجاع المعلومات العربية: تحليل الضوء والحدوث. في وقائع المؤتمر الدولي الخامس والعشرين ACM IGIR حول البحث والتطوير في استرجاع المعلومات (ص. ٢٧٥-٢٨٢).

اللغويات الحاسوبية (٢٠٢٠) النظرية والتطبيق. مركز الكتاب الأكاديمي.

ماكالوم، أ.، & نيجام، ك. (٢٠٠٣). مقارنة بين نماذج الأحداث لتصنيف نص بايز الساذج. مجلة أبحاث التعلم الآلي، ٣ (٣)، ١٢٦٥-١٢٨٧.

محمد مصلح، أ. (٢٠١١). مقاييس اختيار المجموعة الفرعية للمعالم لتصنيف النص العربي. التعرف على النمط. Left، ٣٢ (١٤)، ص. ١٩٢٢-١٩٢٩.

المهداوي، أ.، غوسييه، إ.، & العلوي، س. ع. (٢٠١٦). تصنيف النص العربي على أساس تضمين الكلمات والمستندات. المؤتمر الدولي للأنظمة الذكية المتقدمة والمعلوماتية (ص. ٣٢-٤١). القاهرة، مصر: سبرينغر.

ميكولوف، ت.، وتسيير، ط.، تشين، ك.، كورادو، ج. س.، & عميد، ج. (٢٠١٣). تمثيلات موزعة للكلمات والعبارات وتركيبها. التقدم في أنظمة معالجة المعلومات العصبية، بحيرة تاهو، نيفادا، الولايات المتحدة الأمريكية (ص. ٣١١١-٣١١٩).

نورمان، أحمد، وآخرون. (٢٠١٩). معالجة اللغة الطبيعية: استكشاف الدلالات السياقية. المجلة الدولية لتطبيقات الذكاء الاصطناعي، ٦(٣)، ٢٥-٣٢.

نورمان، أحمد، وآخرون. (٢٠١٩). معالجة اللغة الطبيعية: استكشاف الدلالات السياقية. المجلة الدولية لتطبيقات الذكاء الاصطناعي، ٦(٣)، ٢٥-٣٢.

نورمان، أحمد، وآخرون. (٢٠٢٠). التعلم الآلي: المبادئ والتطبيقات. المجلة الدولية لعلوم الحاسوب وتقنية المعلومات، ١٣(٢)، ٤١-٥٠.

هبة، أ.، الكعبي، م.، الردايدة، ف.، (٢٠١١). أثر التبعية على تصنيف النص العربي: دراسة تطبيقية. المجلة الدولية لبحوث استرجاع المعلومات، ١ (١)، ٥٤-٧٠.

يانغ، واي. (١٩٩٩). تقييم النهج الإحصائية لتصنيف النصوص. استرجاع المعلومات، ١ (١)، ٦٩-٩٠.

يحيى، أ.، & صالحي، أ. (٢٠١٤). تصنيف النصوص العربية على ويكيبيديا العربية. معاملات ACM على معالجة معلومات اللغة الآسيوية، ١٣، ١-٢٠.

يو، ب. (٢٠٠٨). تقييم طرائق تصنيف النصوص للدراسة الأدبية. الحوسبة الأدبية واللغوية، ٢٣ (٤)، ٣٢٧-٣٤٣.

Abdul-Mageed, M., Al-Hamdani, A., Mortal, E., & Saudi, B. (2019). Albert and Market: Two-way deep transformers for Arabic. arXiv:2101.01785.

Abu Far ha, I., & Mandy, W. (2019). Maze: Arabic online sentiment analyzer. Workshop 4 The Natural Arabic Language Process, 192-198.

Aguirre, E., Bane, C., Verde, D., Diaz, M., Gonzalez-Agirre, A., Miracle, R., ... & Riga, G. (2016). SemEval-2016 Task 1: Monolingual semantic textual similarity and cross-linguistic evaluation. Proc. 10 Int. Semantic Evil. Workshop, 497-511.

Akin, A., Bergman, T., Blythe, D., Raoul, K., Schweitzer, S., & Polygraph, R. (2019). FLAIR: An easy-to-use framework for the latest LP. Proc. Cone. North Amer. Chapter Assoc. Compute. Linguistics: Human Lang. Techno., 1, 54-59.

Al-Badrashiny, A. B., Essa, K., & Battery, S. R. (2017). Travel: A collection of Arabic word embedding models for use in Arabic NLP. Proc. Compute. Sci., 117, 256-265.

Al-Farabi, A. E., & Lee, M. (2020). Champ in SemEval-2020 Task 12: A combination of different word embeddings and sentiment transfer learning to identify offensive Arabic in social media. Fourteenth Workshop, Semantics, 1532-1538.

Al-Farabi, A. E., & Lee, M. (2020). Combining letters and words to influence informal Arabic blogs on social media. In Natural Language Processing and Information Systems (pp. 213-224). Cham, Switzerland: Springer.

Al-Farabi, A.E., Smith, B., & Lee, M. (2021). Enhancing contextual language models with constant character and word embedding to detect emotional intensity and strength of sentiment in Arabic tweets. Proc. Compute. Sci., 189, 258-265.

Algerian, F. Z., Al-Razi, S. A., & Natalie, N. (2022). Contextual semantic embeddings based on fine-tuned Robert model for multi-class Arabic text classification. J. King Saud UNICV. Compute. Inf. Sci., 34(10), 8422-8428.

Al-Ibrahimi, B., Al-Tawahiriyya, M., & Tarry, A. (2016). Data and text mining techniques to classify the polarity of Arabic tweets. J. Digit. Inf. Manage., 14 (1), 15-25.

Al-Moslmi, H. A., & Army, A. M. (2020). The effect of tracking and word embedding on Arabic text classification based on deep learning. IEEE Access, 8, 127913-127928.

Al-Najjar, A., Khalifa, Y.S., & India, A. (2018). Arab Hotel - reviews building a dataset for sentiment analysis applications. In Intelligent natural language processing: Trends and applications (pp. 35-52). Cham, Switzerland: Springer.

Al-Omari, K. M., Ashlee, H. M., & Sheila, K. (2017). Emotional analysis of Arabic language tweets using machine learning. Proc. Int. Cone. Into. Biologically inspired computing. Application, 663-673.

Alsace, D., Al-Amir, M., & Al-Mari, S. A. S. (2020). SAD Arabic Dataset for Deep Learning. Proc. Intel. System. Apply. Intel. System. Cone., 1, 630-640.

Alton, A., & Tao, L. (2016). Embedding words for Arabic sentiment analysis. Proc. IEEE Int. Cone. Big Data, 3820-3825.

Alyafeai, Z., Ashwani, M.S., Ghaleb, M., & Ahmed, E. (2023). Evaluating various symbols to classify Arabic text. Neural Processing Letters, 55(3), 2911-2933.

Anton, W., Bally, F., & Al-Haj, H. (2021). Ar apart: A transformer-based model for understanding Arabic. arXiv:2003.00104.

Any, M., & Asia, A. (2013). LABOR: Large-Scale Arabic Book Reviews Dataset. Proc. 51st Anna. Account assistant meeting. Linguistics, 2, 494-498.

Arrows, A. (2019). Syrian dialect dataset for sentiment analysis. Fig share.

Ashwani, M.S., Alyafeai, Z., & Ahmed, E. (2020). Meter classification of Arabic poems using deep bidirectional recurrent neural networks. Amsterdam, Netherlands: Elsevier.

Bell, D., & Moore, C. (2020). Dice similarity coefficient. Radiopaedia.

Benitez, M. (2018). Data set for Arabic classification. J. Adv. Compute. Sci. Apply., 9(9), 1-7.

Black, P. E. (2019). Manhattan distance. In Dictionary of Algorithms and Data Structures (February). Gaithersburg, MD, USA: National Institute of Standards and Technology.

Bokanovsky, P., Grave, E., Gull, A., & Miklos, T. (2017). Enriching word vectors with subword information. Trans. Assoc. Compute. Linguistics, 5, 135-146.

Brownlee, J. (2016). Deep Learning with Python: Developing deep learning models on Thea no and TensorFlow using Keras. Proc. EMN LP Cone.

CHO, K., van MerryëBoer, B., Bandana, D., & Begin, Y. (2014). On the properties of neural machine translation: an encoding-decoding approach. In Syntax Semantics and Structure in Statistical Translation (pp. 103-103). Shelter Island, NY, USA: Manning Publication.

Day, R., & Salem, F. M. (2017). Gate variables for gated recurrent unit (GRU) neural networks. Proc. IEEE 60 Int. Midwest Sump. Circuits System. (MWS CAS), 1597-1600.

Devlin, J., Chang, M.W., Lee, K., & Tutanota, K. (2019). BERT: Pre-training deep bidirectional transformers for language understanding. arXiv:1810.04805.

El Bouzouki, H., Ferret, O., Laverne, T., NOI, H., Feigenbaum, P., & Suit, J. (2020). Characterless: Reconciling Elmo and BERT for word-level open lexical representations of characters. Proc. 28 Int. Cone. Compute. Linguistics, 6903-6915.

Eyecioglu, A., & Keller, B. (2016). ADOBE at SemEval-2016 Task 1: Sentence representation with character embedding N-gram for semantic textual similarity. Proc. 10 Int. Semantic Evil. Workshop, 736-740.

Ghana, S., Favre, B., Steve, Y., & Camel in, N. (2016). Evaluating and integrating word embeddings. Proc. 10 Int. Cone. Lang. Re sour. Evil, 300-305.

Hankering, B., & Strobe, M. (2018). Bomb: Coding-free pre-trained subword embedding in 275 languages. arXiv:1710.02187.

Hassle, M., & Flail, G. (2006). Preparing text through extended encoding. WIT Transactions on Information and Communications Technology, 37. Whitworth, England: WIT Press.

Federation of Arab Trainers Syndicates

Huang, P.-S., He, X., GAO, J., Deng, L., Acer, A., & Heck, L. (2013). Learning digital semantic models for web search using click page data. Proc. 22nd ACM Int. Cone. Inf. Know. Management, 2333-2338.

Jedi, L. A., Cohn, J. F., & De La Tore, F. (2013). Addressing imbalanced data – Recommendations for using performance measures. Proc. Humane Assoc. Cone. Affect. Compute. Intel. Interact., 245-251.

Jones, S. (1972). A statistical interpretation of term specificity and its application in retrieval. J. Document., 28(1), 11-21.

LAU, J.H., & Baldwin, T. (2016). An experimental evaluation of Dover with practical insights into embedded document creation. Proc. 1st acting workshop. To learn. NLP, 78-86.

Midgard, L., McNamara, F., & Ali, K. E. (2021). Multi-level embeddings for processing Arabic social media content. Compute. Large's speech, 70.

Nigam, P., Song, Y., Moran, V., Lakshman, V., Ding, W., Shanghai, A., ... & Kumar, R. (2019). Semantic search for products. Proc. 25 ACM SIGNED Int. Cone. Know. Discovery Data Mining, 2876-2885.

Patel, R., Bit, S., Guevara, V., & Handicap, J. (2023). A survey of text representation and embedding techniques in NLP. IEEE Access, 11, 36120-36146.

Peshawar, A., Khartoum, O. B., & Clayier, B. (2017). A collection of news texts in Arabic for textual classification. Proc. IEEE/ACS 14 Int. Cone. Compute. System. Apply. (AIC CSA), 135-142.

Phrasal, L. (2009). A machine learning approach to Arabic text classification using N-gram frequency statistics. J. Informatics, 3(1), 72-77.

Portal, U., Shi, Z., Seller, S.(2013): Learning in handwritten Arabic text recognition. In: Chapter in the Handbook of Statistics, Volume 31, Pages 443-46.

Saad, M. K., & Ashour, W. (2010). SAC: Open Source Arabic Community. Proc. 6 Arch. Eng. Int. Sump., 55.

Shaman, K., Siddiqui, S., al-Khalil, M., & Abdel Monad, A. (2018). Challenges in natural Arabic language processing. In Computational Linguistics Speech and Image Processing for the Arabic Language (pp. 59-83). Singapore: Scientific World.

Take, S., Suzuki, J., & Niagara, M. (2019). Embedding n-grams to improve RNN language models. Proc. AAA Cone. Artie. Intel., 33(1), 5074-5082.

اتحاد نقابات المدربين العرب
Federation of Arab Trainers Syndicates

اتحاد نقابات
المدربين العرب
Federation of Arab Trainers Syndicates