الذكاء الاصطناعي والمعالجة الآلية للغة العربية: بحث في أنظمة التعرف الآلي على الكلام

Artificial intelligence and automated processing of the Arabic language: research on automated speech recognition systems.

[dr.ahmedkobbi@gmail.com]

احمد قبي Ahmed Kobbi

طالب دكتوراه بجامعة مولاي اسماعيل بمكناس الكلية متعددة التخصصات بالراشيدية

إشرف الدكتور: علي بولعلام

ملخص

تهدف هذه الدراسة المعنونة ب "المعالجة الالية للغة العربية بحث في انظمة التعرف الآلي على الكلام" إلى تسليط الضوء على العديد من التحديات التي تواجهها المعالجة الآلية للغات الطبيعية واللغة العربية بشكل خاص، حيث تسعى المعالجة الآلية للغات الطبيعية لتحليل النصوص ومعالجتها في مستوياتها الصوتية والصرفية والنحوية باستخدام الذكاء الاصطناعي، والذي يزيد من هذه الصعوبات هو خصوصية اللغة العربية وتفردها عن باقي اللغات. ولتحديد اكثر تم التركيز على التعرف الآلي على الكلام والكلام العربي بشكل خاص حيث تدخل هذه الدراسات في مجال الذكاء الاصطناعي التي تتوخى تطوير الآلة بشكل عام لتحاكي الذكاء البشري في فهم وتحليل ومعالجة اللغة.

ويعد تحويل الكلام / المنطوق إلى نص/ Speech -to-text مكتوب تقنية لسانية حاسوبية تستخدم للتعرف الآلي على الكلام البشري أو التعرف على المتواليات الصوتية إلى نص، وعلى الرغم من أن طريقة تحويل المنطوق الى مكتوب تبدو متشابهة مع طريقة تحويل المكتوب إلى المنطوق Text -to- Speech ، ومن الملاحظ أن المواقع التي تم العمل عليها لتحويل الكلام المنطوق إلى نص مكتوب والنص المكتوب إلى كلام منطوق، تواجه تحديات جمة في معالجة اللغة العربية وخصوصا لغة الشعر القديم، فيما نجد صعوبات أقل في معالجة النثر العربي حيث توجد أخطاء لكنها أقل مقاربة بتلك الواردة في الشعر.

الكلمات المفتاحية

المعالجة الآلية للغات الطبيعية، المعالجة الآلية للغة العربية، الذكاء الاصطناعي، أنظمة التعرف على الكلام العربي، تحويل المنطوق إلى نص، تحويل النص إلى منطوق،

Abstract

Automated processing of natural languages is one of the vital fields that has received great attention by researchers and engineers because of its great impact on advanced artificial intelligence methods, especially machine learning and deep learning supported by an artificial neural network. In front of the huge increase of applications that convert spoken to written and written to spoken, in addition to applications of automatic translation, spelling and grammar, automatic formation of texts, morphological and grammatical analyzers and other various computer applications that have become the basics of linguistic engineering.

This study entitled "automated processing of the Arabic language research in automated speech recognition systems" aims to shed light on many of the challenges faced by automated processing of natural languages and the Arabic language in particular, as automated processing of natural languages seeks to analyze texts and process them at their phonetic, morphological and grammatical levels using artificial intelligence, which increases these difficulties is the specificity of the Arabic language and its uniqueness from other languages. To be more specific, the emphasis was placed on automated speech recognition and Arabic speech in particular, as these studies enter the field of artificial intelligence, which envisages the development of the machine in general to mimic human intelligence in understanding, analyzing and processing language.

Speech-to-text is a computer language technology used for automatic recognition of human speech or recognition of sound sequences into text, although the method of converting spoken to written sounds similar to the method of converting written to spoken Text-to- Speech, however, the conversion process itself differs technologically and in practice, which makes it Dictate the algorithmic paths it takes, as it changes according to the context of automated processing, so input / analysis alternates with output/ generation in each way, the best method followed by this conversion technology was the dictation dictation method adopted by Microsoft in the person of "Word Microsoft", which allows users to dictate or spell a word aloud instead of typing it in their text files, the artificial intelligence engine and machine learning algorithms process the Spoken Word and converted into accurate text and so used in translation applications and in existing conversation applications It can easily convert the language spoken by the user into text in a different language using speech to text, and it can also help people with motor disabilities to use the grid by creating text files and filling out forms on a digital platform without having to type them.

Key words

Automated processing of natural languages, automated processing of the Arabic language, artificial intelligence, Arabic speech recognition systems, speech-to-text, text-to-speech.

References

بشر كمال: دراسات في علم اللغة، دط، 1998، دار غريب للطباعة والنشر والتوزيع، القاهرة مصر.

لحلح محمد ، مدخل للذكاء الاصطناعي وتعلم الآلة، تحرير: جميل بيلوني،الطبعة الأولى، أكاديمية حسوب، 2020.

غازي عز الدين ، الذكاء الاصطناعي: هل هو تكنولوجية رمزية؟، مجلة فكر- العلوم الإنسانية والإجتماعية- المغرب، العدد 6، 2005.

إيه بودين مارجريت: الذكاء الاصطناعي، ترجمة إبراهيم سند أحمد،مراجعة هاني فتحي سليمان، دط،2022،مؤسسة هنداوي.

العريان يوسف سالم وأخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، تحرير يوسف سالم العريان، الطبعة الأولى، 2019، دار وجوه للنشر والتوزيع، الرياض

نص المقال

حظيت اللغة العربية بعناية الباحثين في معالجة اللغات الطبيعية في المشرق والمغرب؛حيث نالت حظا وافرا في حوسبة اللغة وتقنياتها، سواء في صورتها المنطوقة او المكتوبة، وسواء على مستوى محارفها أو مبانيها أو على مستوى تراكيبها ومعانيها . وساعدت بعض خصائص اللغة العربية على تطوير أدوات ناجعة لمعالجتها آليا عبر عدة مستويات.

وتجدر الإشارة إلى أن كل نسق لغوي يتميز بتضافر وانصهار خمس مستويات للدرس اللساني المكون لهذا النسق: الصوتي والصرفي والتركيبي والدلالي والتداولي، بحيث لا يتأتى لأي دارس لغوي أن يتعرف على مميزات هذه اللغة دون أن يلم بكل أو ببعض مستوياتها اللسانية.

غير أن أي تطور في مجال معالجة اللغة العربية وتقاناتها الحاسوبية ينبغي أن يظل مرتبطا بمعالجات تراعي خصائص اللغة العربية وسماتها من دون إغفال لما توصلت إليه معالجة اللغات الأخرى من نتائج،حتى لا تكون هناك فروق شاسعة بين التقانات اللغوية في اللغة العربية ومثيلاتها في اللغات العالمية التي حققت تقدما كبيرا في هذا المجال.

خصص لهذا المقال هدف عام من خلال طرح إشكالي لإعداد البيانات من أجل تطوير المعالجة الآلية للغات الطبيعية وبشكل خاص المعالجة الآلية للغة العربية، بالتركيز على التعرف الآلي على الكلام العربي بغرض حل بعض الصعوبات التي تعترض اللغة العربية ،من قبيل اللبس الدلالي بسبب غياب التشكيل والإعجام بالحركات ...، مع الأخذ بعين الاعتبار أن اللغة العربية لغة بصرية.

تم الانتقال من الهدف العام إلى الهدف الخاص بوضع فرضية للبحث التي تهدف تطوير أدوات وتطبيقات التعرف الآلي على الكلام العربي، وذلك عبر تخريج نصوص عربية واختبارها في مواقع تحويل الكلام إلى نص و النص إلى كلام وقد اعتمدت في البحث على منهج وصفي.

إن استخدام التقنيات الحديثة للذكاء الاصطناعي في مجال المعالجة الآلية للغة العربية سيسهم بشكل كبير في تحسين الأداء والدقة في تحليل وفهم النصوص والكلام العربي، مما يفتح آفاقًا جديدة لتطبيقات اللغة العربية في مختلف المجالات.

تعود البديات الاولى للذكاء الاصطناعي إلى " الربع الثاني من القرن التاسع عشر الذي شهد بداية ظهور بذور الذكاء الاصطناعي على الرغم من أن الحواسيب لم تك موجودة بعد،إلا أن الذكاء الاصطناعي –كمنطق رياضي- أوجد لنفسه المكان المناسب وشق طريقه في الوسط العلمي فعندما قدم لنا العالم جورج بول (George Boole) نظريته في المنطق الجبري والتي سميت لاحقا بالجبر البولياني (Boolean Algebra) وهي نظرية تعتمد في أساسها على تمثيل أي متغيرات في أي عميلة رياضية على قيمتين وهما 1و 0، وتشكل مجموعة متغيرات عبارات منطقية."[1]

يسعى الذكاء الاصطناعي إلى تمكين أجهزة الحاسوب من تنفيد المهام التي يستطيع العقل تنفيدها ."ويعتبر الذكاء الاصطناعي فرعا من المعلوميات يقوم بإعادة إنتاج بعض مظاهر الذكاء الإنساني من إدراك للعلامات اللغوية وفهمها واستيعابها وتخزينها ثم إنتاجها واستخدامها في ظروف جديدة."[2]

وثمة هدفان أساسيان للذكاء الاصطناعي. الهدف الأول <<تكنولوجي>>؛ استخدام أجهزة الحاسوب لإنجاز مهام مفيدة (وتوظف في بعض الأحيان طرقا غير التي يستخدمها العقل تماما). الهدف الثاني <<علمي>>؛ استخدام مفاهيم الذكاء الاصطناعي ونماذجه للمساعدة في الإجابة عن أسئلة تتعلق بالإنسان وغيره من الكائنات الحية. لا يركز معظم العاملين في الذكاء الاصطناعي إلا على هدف من هذين الهدفين، ولكن بعضهم يركز على كليهما.[3]

"يعرف الذكاء الاصطناعي على أنه علم يهتم بتصميم خوارزميات تستطيع أداء مهام محددة بنفس كفاءة البشر أو أفضل."[4] بناء عليه فإن أي خوارزمية تحاكي سلوكا يختص به الإنسان تدخل تحت مظلة الذكاء الاصطناعي. فعلى سبيل المثال، الإنسان يستطيع فهم الكلام، فأي خوارزمية تستطيع عمل هذه المهمة تعتبر خوارزمية ذكاء اصطناعي.

والحديث عن الذكاء الاصطناعي سيسلمنا لا محالة للحديث عن المعالجة الآلية للغات الطبيعية حيث توجد العديد من المصطلحات التي تترادف تقريبًا مع معالجة اللغة الطبيعية. يشير فهم اللغة الطبيعية (Natural Language Understanding)وتوليد اللغة الطبيعية (Natural Language Generation) إلى استخدام أجهزة الحاسوب لفهم وإنتاج اللغة البشرية على التوالي.

توليد اللغات الطبيعية هو عملية إنتاج نصوص مكتوبة تشبه اللغة الطبيعية بواسطة الحواسيب. يمكن استخدام توليد اللغات الطبيعية في مجالات مثل الكتابة الإعلامية، التحليل المالي، النشر الإلكتروني، التسويق، والترفيه.

تتضمن تقنيات توليد اللغات الطبيعية تطبيق الذكاء الاصطناعي وتحليل البيانات لتوليد نصوص مكتوبة تشبه اللغة الطبيعية. وتتطلب هذه العملية تحليل مجموعات كبيرة من النصوص واستخلاص الأنماط اللغوية والمعنوية المختلفة. ومن ثم يمكن تدريب الحواسيب على إنتاج النصوص الجديدة بناءً على هذه الأنماط.

يتم استخدام توليد اللغات الطبيعية بشكل واسع في مجالات مثل توليد الشرح التلقائي لمقاطع الفيديو، وإنشاء محتوى مقالات ومدونات وتقارير مختلفة، وتوليد التعليقات والردود الآلية على المواقع الإلكترونية، وتوليد التقارير والمخططات في المجالات المالية والاقتصادية. وبما أنه يمكن توليد النصوص الطبيعية بشكل سريع ودقيق، فإن هذه التقنيات تساعد على توفير الوقت والجهد في مجالات مختلفة وتحسين كفاءة العمل.

بدأ البحث عن (Natural Language Processing)بعد فترة وجيزة من اختراع أجهزة الحاسوب الرقمية في الخمسينيات، ويستند على كل من اللسانيات والذكاء الاصطناعي. ومع ذلك ، تم دعم الإنجازات الرئيسية في السنوات القليلة الماضية من خلال التعلم الآلي ، وهو فرع من الذكاء الاصطناعي الذي يطور الأنظمة التي تتعلم من البيانات .التعلم العميق هو نوع من التعلم الآلي الذي يمكنه تعلم أنماط معقدة للغاية من مجموعات البيانات الكبيرة، مما يعني أنه مناسب بشكل مثالي لتعلم تعقيدات اللغة الطبيعية من مجموعات البيانات التي يتم الحصول عليها من الويب[5].

والمعالجة الآلية اللغات الطبيعية تهتم بدراسة اللسان البشري بحيث يتميز كل نسق لغوي بتضافر وانصهار خمس مستويات للدرس اللساني المكون لهذا النسق: الصوتي والصرفي والتركيبي والدلالي والتداولي، بحيث لا يتأتى لأي دارس لغوي أن يتعرف على مميزات هذه اللغة دون أن يلم بكل أو ببعض مستوياتها اللسانية.

يعد فهم اللغات الطبيعية أحد أكثر التحديات في مجال الذكاء الاصطناعي، حيث تتعقد المسألة بسبب تعددية اللغات وتنوع الطرق التي يستخدمها البشر في التواصل، ولذلك يعتمد العلماء في هذا المجال على استخدام الأساليب الحديثة والتقنيات الحديثة مثل تقنيات التعلم العميق[6] وتقنيات تعلم الآلة[7] ، والشبكات العصبية لتطوير النماذج الحاسوبية التي تقوم بفهم اللغات الطبيعية. ويستخدم هذا المجال في مختلف المجالات مثل الرعاية الصحية والتجارة الإلكترونية والترجمة والتحليل اللغوي وغيرها .وتستخدم (Natural Language Understanding) لتمكين أجهزة الحاسوب من فهم هيكل ومعنى جميع اللغات البشرية، مما يسمح للمطورين والمستخدمين بالتفاعل مع أجهزة الحاسوب باستخدام الجمل الطبيعية والاتصالات. وعلم اللغة الحاسوبي (Computational Linguistics) هو المجال العلمي الذي يدرس الجوانب الحاسوبية للغة البشرية ، في حين أن (Natural Language Processing) هو الانضباط الهندسي المتعلق ببناء البيانات الاصطناعية التي تفهم اللغة البشرية أو تولدها أو تتلاعب بها.

تواجه المعالجة الآلية للغة تحديات عديدة نقف عند بعضها على سبيل المثال لا الحصر، حيث أن فهم الحاسوب للغات الطبيعية من المشكلات الكبرى التي تواجه نظام المعالجة الآلية للغات الطبيعية، لأن ذلك يتطلب معرفة عميقة بالعالم

الخارجي مع القدرة على استغلال هذه المعرفة بشكل ذكي، وفيما يلي نماذج المشكلات التي تعترض المعالجة الآلية للغات بواسطة الحاسوب ومن ضمنها اللغة العربية:

- تقطيع الكلام والأصوات والوحدات المعجمية
- فك الغموض أو اللبس
- المطابقة بين المنطوق والمكتوب
- الحقيقة والمجاز:

عموما فالمعالجةالآلية للغات الطبيعية علم فرعي للذكاء الاصطناعي،الذي بدوره فرعا من المعلوماتية، ويتداخل بشكل كبير مع اللسانيات الحاسوبية التي توفر الوصف اللغوي المطلوب لجهاز الحاسوب وعن طريق المعالجة الآلية يمكن صناعة برمجيات تساعد على تحليل ومحاكاة اللغات الطبيعية .

تشترك اللغات البشرية في كثير من الخصائص والسمات، ويمكن معالجتها آليا بنفس القواعد والآليات الحاسوبية من خلال البحث في تلك الخصائص المشتركة بين كل اللغات البشرية وعلى الرغم من ذلك نجدأن"لكل لغة سماتها ومميزاتها الخاصة بها، ويستوي في ذلك أن تكون هذه الخواص صوتية أو صرفية أو نحوية أو أسلوبية أو على مستوى الألفاظ ودلالتها"[8]، واللغة العربية لها خصائص تنفرد بها عن باقي اللغات، قد يشكل بعضها تحديا أمام المعالجة الآلية لها من قبيل:

- ❖ التشكيل في اللغة العربية أو الاعجام بالحركات
- ❖ توظيف الضمائر والسوابق واللواحق في اللغة العربية.

إن تطوير أنظمة التعرف الآلي على الكلام العربي المنطوق رهين باستخدام أنظمة الذكاء الاصطناعي وتطبيقه في المعالجة الآلية للغة العربية،من بين هذه التقنيات نجد الطريقة التقليدية لمعالجة الكلام،تعتمد على تقسيم الجملة الصوتية إلى مجموعة ثابته من الأطر،بينما في الطرق الحديثة يتم الاعتماد على المقاطع الصوتية والتي تعرف على أنها الجزء الأصغر في اللغة المكونة من مقاطع ساكنة وأخرى متحركة. ولا يتسنى ذلك إلا عبر الوقوف على خصائص المقاطع الصوتية.ولا يتأتى ذلك إلا عبر دراسة الأصوات الكلامية من ناحية سماتها وصفاتها ومخارجها، ودراسة أيضا القواعد التي تتحكم في تآلفها وتجاورها...إلا أن اللسانيات الحاسوبية اهتمت بالجانب الأكوستيكي والنطقي للصوت ؛ أي بالذبذبات الصوتية دون الاهتمام بمكوناتها وخصائصها الفيزيائية؛ وذلك ببرمجة الصوت في الحاسوب ثم محاولة التعرف على خصائصها بتحديد أجزائها ودراستها قطعة، قطعة، إلا أن المعلوماتي يواجه بعض المشاكل في هذا الجانب ، وترتبط بالخصوص بطريقة نطق الصوت كالظواهر اللسانية فوق المقطعية مثل : النبر، والتنغيم.

تحويل الكلام المنطوق من موجات صوتية إلى نص مكتوب يستخدم الآن في العديد من التطبيقات كالمساعدات الشخصية وتحويل الكلام المسجل إلى نصوص. وقد كان الاعتماد سابقا بشكل كبير على خوارزميات (Hidden Markov Model)، ولكن في الآونة الأخيرة تم تبني التعلم العميق بشكل أساسي، حيث تستخدمه الآن كبرى الشركات في منتجاتها للتعرف على الكلام .

عكس تحويل الكلام إلى نصوص، الهدف من توليد الكلام هو تحويل النص المكتوب إلى مقطع صوتي منطوق. كان في السابق يتم إلصاق الفونيمات (الوحدات الصوتية) لإنشاء الكلام، ولكن في السنوات الاخيرة تم ابتكار العديد من الخوارزميات باستخدام التعلم العميق تعطي نتائج مقاربة بشكل كبير للصوت البشري. من الخوارزميات المهمة لتوليد الكلام باستخدام التعلم العميق هي خوارزمية Wavenet من قوقل، وهي خوارزمية توليدية تتعلم التوزيعة المشروطة التالية:"[9]

$$p(\boldsymbol{x}) = \prod_t p(x_t | x_{<t}, \boldsymbol{\theta})$$

خلال البحث تم التركيز على التعرف الآلي على الكلام والكلام العربي بشكل خاص وعلى فهم وتحليل ومعالجة اللغة العربية، بتقييم واختبار أنظمة التعرف الآلي على الكلام وتقنيات تحويل الكلام إلى نص(https://www.speechtexter.com https://dictation.io/speech -https://speechlogger.appspot.com -)

والنص الى كلام(-https://www.naturalreaders.com/ -https://ttsreader.com/ -https://ttsmp3.com) المتاحة عالميا ، فكان الجانب التطبيقي فرصة لتقييم نصوص عربية شعرية (معلقات) ونثرية (لابن خلدون) وعرضها على مواقع لتحويل الكلام الى نص أخرى لتحويل النص الى كلام، وذلك بالكشف عن الأخطاء الممكنة والمتكررة في عملية التحويل التي تعتمد في هذه المواقع؛ بالاستعانة بمدقق إملائي لتصحيح هذه الأخطاء، التي تنتج عادة بسبب صعوبة التعرف على بعض المتواليات الصوتية وأنساق الكلمات التي يعثر عليها القارئ الآلي في هذه النصوص وخصوصا وأن النص الشعري /المعلقات يتميز بلغة شعرية عروضية وبلاغية جمالية تحتاج الى جهد علمي عميق في تطوير الأدوات والطرائق والمحللات الكفيلة بتحسين تقنية التعرف الآلي على الكلام الشعري و على خصوصياته الجمالية. ما دفعنا للتساؤل عن هذا المدقق ولماذا لم يتعرف على كلمات من صلب العربية الفصحى رغم قدرته النسبية في التعرف على الكلام المنثور، فهل هناك مشكل في قاعدة البيانات أم أخطاء في التوصيف أم معلومات غير كافية في البرمجة أم غير ذلك ؟.

لعل الأدوات العقلانية و التقنيات و الخوارزميات التي تبنتها أساليب الذكاء الاصطناعي والتحليل الرياضي الاحصائي والاحتمالي وطرق التعلم العصبي العميق، قد اثبتت نجاعتها مع نصوص كثيرة ومختلفة في تحويل الكلام المنطوق إلى نص مكتوب والنص المكتوب إلى كلام منطوق ، لكنها مع ذلك تتعثر في العديد من التطبيقات مثل الإدخال الصوتي للنصوص والترجمة الصوتية والتعامل مع الأوامر الصوتية لأجهزة الإلكترونيات الذكية، كما تختلف أنظمة التعرف على الكلام من حيث الدقة والكفاءة والاستجابة بين الأنظمة المختلفة.

إن محاولة الكشف عن العثرات الممكنة في العديد من الأدوات والتطبيقات الداعمة للنصوص العربية كانت فرصة للبحث في طرق تحسين وتجويد التعرف الآلي على الكلام العربي، بالاستفادة من الأبحاث المنجزة في هذا المجال الحيوي وفي مجالات رديفة قد تسعف هذه الدقة بنسبة عالية جدا مثل الأبحاث الفونولوجية والفونيتيكية وخاصة منها التي عالجت الظواهر التطريزية من مماثلة ومخالفة وتنغيم ونبر واختزال وقلب مكاني وغيره من القضايا إضافة الى تقنيات الإيقاع والموسيقى ،

ومن جملة الصعوبات التي اعترضت البحث عملية اختيار المدونة العربية في الشابكة (موقع صحح لي / dictation) ومدى جودة الصوت ومؤثراته الجانبية وانعكاس الجودة على طبيعة النص وعلى نجاعة الموقع الذي يتوفر على

معدات عالية الدقة مستخلصا التحديات التي تواجهها المعالجة الآلية للغة العربية والتعرف الآلي على الكلام العربي خاصة النص الشعري لخصوصياته الجمالية والايقاعية ، والى طبيعة اللغة العربية الاشتقاقية والانصهارية والى غياب نجاعة المشكل الآلي الذي يودي الى مشاكل في الغموض الدلالي خلال التعرف الآلي على الكلام والنصوص العربية. فتحويل النص إلى كلام ينطوي على صعوبات شأنه شأن تحويل الكلام إلى نص ، وإن صح القول فإن الأخطاء التي تم جردها من تحويل الكلام إلى نص هي تقريبا نفس الأخطاء لكن في نطق الكلمات ، أضف إلى ذلك أن نطق بعض الكلمات حتى في سياق معين يجعلها تبدو بلا معنى وكأنها نطق لمتوالية من الحروف فقط وهذا بارز في المقاطع الصوتية سواء الخاصة بالشعر أو النثر، مما يستدعي العمل تطوير مواقع الناطق الآلي والعمل على قواعد بيانات من أجل الوصول لنطق صحيح وجعل النطق مشابه البشري وما يميزه من نبر وتنغيم ووقف ووصل تعمل على إيصال بعض المعلني وفهمها من خلال السياق، وقد يرجع ذلك (أي نطق الحروف فقط دون معنى)إلى طبيعة الناطق الآلي في الربط بين الحروف ونطقها، وفي أحيان أخرى يجمع بين كلمتين في النطق أو يقسم الكلمة إلى قسمين، وتتفاوت جودة النطق من موقع لآخر.

المراجع

---

[1] لحلح محمد ، مدخل للذكاء الاصطناعي وتعلم الآلة، تحرير: جميل بيلوني،الطبعة الأولى، أكاديمية حسوب، 2020، ص 41

[2] غازي عز الدين ، الذكاء الاصطناعي: هل هو تكنولوجية رمزية؟، مجلة فكر– العلوم الإنسانية والإجتماعية– المغرب، العدد 6، 2005، ص 44

[3] إيه بودين مارجريت: الذكاء الاصطناعي، ترجمة إبراهيم سند أحمد،مراجعة هاني فتحي سليمان، دط،2022،مؤسسة هنداوي، ، ص 12

[4] العريان يوسف سالم وأخرون: تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، تحرير يوسف سالم العريان، الطبعة الأولى، 2019، دار وجوه للنشر والتوزيع، الرياض ، ص 146

[5] https://www.oracle.com/bh-ar/artificial-intelligence/what-is-natural-language-processing/

[6] يعتمد في أساسه على طريقة تعلم مؤلفة من عدة طبقات من التمثيلات المقابلة لبنية هرمية من السمات، ويتم تعريف السمات والمفاهيم عالية المستوى نزولا إلى المفاهيم ذات المستوى الأدنى.

[7] أسلوب يهدف الى تدريب خوارزمية تعلم الآلة الخاصة ببيانات التعلم، اي بيانات نظام خبير معين. ومن أساسيات الذكاء الصناعي ايضا التنبؤ بالبيانات باستخدام القواعد المعرفية.

[8] بشر كمال: دراسات في علم اللغة، دط، 1998، دار غريب للطباعة والنشر والتوزيع، القاهرة مصر، ، ص 193

[9] الصانع وليد بن عبد الله و آخرون: خوارزميات الذكاء الاصطناعي في تحليل النص العربي، ص65 .