الذَّكَاءُ الاصْطِنَاعِيُّ وَمَهَامُّ مُعَالَجَةِ اللُّغَاتِ الطَّبِيعِيَّةِ

تحت إشراف :

الأستاذ الدكتور محسن عبد الرازق رشوان الخبير بالمجمع

بالاشتراك مع:
لجنة اللغة العربية والذكاء الاصطناعي بالمجمع

الذكاء الاصطناعي ومهام معالجة اللغات الطبيعية

ملخص البحث

يَتَعامـلُ الذَّكَـــــاءُ الاصْــطِنَاعِيُّ مــع لُغـاتِ البشـرِ، فَيقـدِّمُ لهـا حُلـولًا عمليـة تعتمـد علـى الخوارزميـات الرياضـية والمنطـق وتجريـد الأفكـــار وترميزهـــا، وَكُـــلُّ مَهـــام الـــذكاءِ الاصْــطِناعيِّ تَنْطَلِـقُ مِـنْ تَـدْرِيبِ الآلَـةِ أولًا لِتُحَاكِيَ العقلَ الْبَشَرِيَّ، وَقَدْ نَجَحَتْ فِي كَثِيرٍ مِن الْمَهَامِّ التي اخْتَرْنَا مِنْهَا أَرْبَعَ عَشْرَةَ مهمَّةً، أَبْرَزُهَا التَّرْجَمَةُ الآلِيَّةُ وَتَعرُّفُ الصَّوْتِ وَالصُّورَةِ والنص... وغير ذلك من الْمَهَامِّ التي حَاوَلْنَا أَنْ نُعَرِّفَ بها بِشَكْلٍ مُوجَزٍ وَشَرْحٍ مُبَسَّطٍ لِكَيْفِيَّة اقتِنائِهَا وَالاسْتِفَادَةِ مِنْهَا.

كلمات مِفتَاحِية:

اللغات الطبيعية – الذكاء الاصطناعي – الخوارزمية – الشبكات العصبية

الْمَبْحَثُ الْأَوَّلُ: مُعَالَجَةُ اللُّغَاتِ الطَّبِيعِيَّةِ آليًّا..الْوَاقِع وَالتَّحَدِّيَات

مدخل

(١.١) تَعْرِيفُ اللُّغَةِ

التَّعْرِيفَاتُ الْمُخْتَلِفَةُ للُّغةِ تتخذ منطلقات عدة، ويصل غالبها إلى نتيجة واحدة، هي أن اللغة أداة للتواصل الإنساني، فمثلًا ابن جني يرى أن اللغة أصوات يعبر بها كل قوم عن أغراضهم (ابن جني، ١/٣٣)، وابن خلدون يراها فعلًا لسانيًّا بقصد إفادة الكلام (ابن خلدون، ١٩٥٢، ص٣٣)، ودي سوسير يرى أنها تنظيم من الإشارات المفارقة (دي سوسير،١٩٨٥، ١١١)، أو هي نظام من الرموز الاعتباطية يتم بواسطتها التعارف بين أفراد المجتمع، وتخضع هذه الأصوات للوصف من حيث المخارجُ والحركاتُ التي يقوم بها جهاز النطق، من حيث الصفات والظواهر الصوتية المصاحبة لهذا النطق (عمايرة: ١٩٨٧: ١٥).

(٢.١) إِنْتَاجُ اللُّغَةِ

وإذا كانت اللغة أصواتًا أو إشارات متفارقة أو رموزًا اعتباطية، فهذا يعني أنها تعمل وَفْقَ نظام يقوم بوظيفة اتصالية بين أفراد المجتمع، وهذه الرموز يخزنها العقل البشري؛ لينتج أصواتًا أو كلامًا له معنى ودلالة وفق آلية معقدة تناولها الباحثون على مستويين: **الأول** يتمثل في إنتاج الكلام من المخ، **والثاني** فسيولوجية النطق، فأما إنتاج الكلام من الدماغ فيكون بتفعيل ترميزات لغوية للتحدث والفهم والقراءة والكتابة، فالمتحدث يفكر في معنى ما يريد أن يتحدث فيه، ثم يختار المفردات المناسبة لهذا المعنى، وهذه المفردات هي فونيمات ومورفيمات[١]، ثم بعد ذلك يُرَكِّبُ الجملة في رأسه لينطقها في نهاية المطاف، وهذه العملية تترجم إلى حركة الفم والفكين واللسان والحنجرة، وتتم في الإنسان الطبيعي بشكل سريع. (نعيم عبد الغني، ٢٠١٩: ١٧٩).

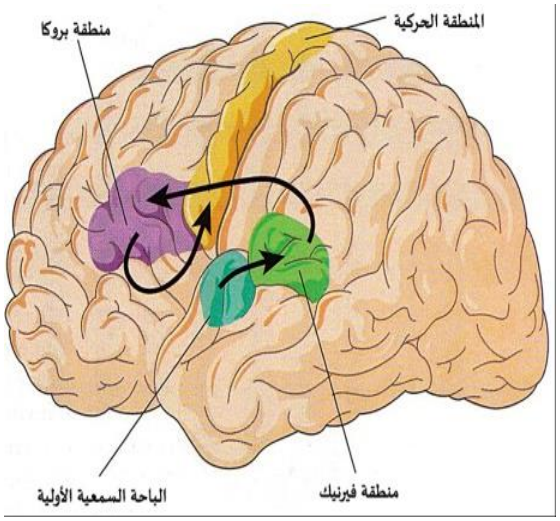
إن آلية إنتاج الكلام واستقباله عملية معقدة يقف أمامها علماء اللغة والطب كثيرًا، فعالِم النفس فيجوتسكي ذكر أن اللغة تتم من خلال: **الاتصال الخارجي مع البشر، وترتيب الأفكار في العقل**، فهناك نظامان يعملان:

(١) الفونيم أصغر وحدة صوتية، والمورفيم أصغر وحدة صرفية.

نظام داخلي، وآخر خارجي، وكلا النظامين يستخدمان شفرة لغوية واحدة (جوديت جرين، ١٩٩٢: ١٣ -١٥).

أما تشومسكي فإنه يرى أن إنتاج اللغة يمر من خلال الكفاءة والأداء، فكل إنسان يحتوي دماغه على معرفة لغوية يتم تغذيتها من محيط مجتمعه، وبعد إضافة المدخلات اللغوية يصبح الإنسان قادرًا على التكلم. (Chomsky. -2000- : ٧٧-٧٩).

وعلم اللغة العصبي (Neuro Linguistique) في دراسته لهذه القضية يرى أن هناك مناطقَ في الدماغ مسؤولة عن اللغة هي منطقة بروكا واللحاء السمعي واللحاء الحركي الرئيسي ومنطقة فيرينك وعند نطق أية كلمة مقروءة يجب أن تصل المعلومات أولًا إلى القشرة البصرية التي ترسل بدورها المعلومات إلى منطقة الكلام الخلفية، بما فيها منطقة فيرينك، ثم تنتقل المعلومات إلى منطقة بروكا، ثم إلى اللحاء الرئيسي؛ فهذه هي المناطق المسؤولة عن السمع والنطق، وهي أدوات الاتصال المباشر، وأي خلل فيها يُحدِثُ صممًا أو فقدًا للنطق أو على الأقل اضطرابات فيهما (شهداء، ٢٠١٥، ٨٣ – ٨٩).



(شكل: ١) مناطق إنتاج اللغة في المخ

(١.٣) الذكاء الاصطناعي وإنتاج اللغة

تبين لنا بشكل مقتضب أن إنتاج اللغة في الدماغ البشري عملية معقدة جدًّا، وهذا التعقيد لم يكن حائلًا دون الاجتهاد في بحوث الذكاء الاصطناعي؛ لمعالجة اللغات الطبيعية، وكان المنطقُ الرموزَ التي تستخدِمُها اللغة، حيث إننا يمكن أن ندرب الحاسوب على هذه الرموز ونحاكي الدماغ البشرية؛ فالعلاقة بين اللغة والحاسوب تبادلية والالتقاء بينهما أمر حتمي؛ لأن الإنسان هو محور النشاط اللغوي، وهو الذي منحه الله القدرة على تصميم الحاسوب وتشغيله، وهذا يُمَكِّنُه من صناعة لغات البرمجة التي تخدم لغته التي يتكلم بها (نبيل علي، ١٩٨٨: ١١٣، ١١٥).

(١.٤) محاولات علمية

في النصف الثاني من أربعينيات القرن العشرين والنصف الأول من خمسينياته بدأ عدد متزايد من علماء الرياضيات والهندسة الكهربية والميكانيكية والاقتصاد وعلم النفس والعلوم السياسية في نقاشات مكثفة لمحاولة إنتاج دماغ آلية، وبعدها جاء "مارفين مينسكي" أحد الآباء المؤسسين لعلم الذكاء الاصطناعي فحاول بناء شبكة عصبية اصطناعية كهربائية؛ لإظهار قدرات المخ الطبيعي، واستطاعت هذه الشبكة بعد عمل طويل في تطويرها إظهار بعض التصرفات التي تكافئ العمليات المنطقية الثنائية البسيطة (عطية: ٢٠١٩ ص٣١).

وجاءت بعد ذلك مؤتمرات ومحاولات جادة في الذكاء الاصطناعي بشكل عام ومحاكاة الدماغ البشرية، وكانت معالجة اللغات الطبيعية (Natural languages) في قلب هذه المحاولات والدراسات، حيث بدأت مهام الذكاء الاصطناعي في معالجة اللغات الطبيعية بدراسة المستوى الدلالي (Semantic) والأنطولوجي (Ontological).

فلما كانت نظرية الاتصال المعلوماتي (Information theory) التي تأسست واستقرت قبل تلك الحقبة بعقدين، وحقّقَت نجاحاتٍ هائلةً في هندسة الاتصالات وعلوم الحاسب قد عالجت مسائل الترميز على مستوى متقدم وطوَّرت في ذلك السياق آلياتٍ رياضيةً رفيعةً فإن الباحثين نظروا إلى المستويات الشكلية من اللغة الطبيعية مثل "النسق الكتابي Orthography" و"النسق الصرفي Morphology" على أنها مسائل ترميزية تسهل معالجتها بنفس آليات نظرية الاتصال المعلوماتي. (عطية: ٢٠١٩، ص٣٧).

(1.5) الذكاء الاصطناعي ومهام معالجة اللغات الطبيعية (Natural language Processing) - (NLP)

انصبت أبحاث الذكاء الاصطناعي في مهام معالجة اللغات الطبيعية على المستوى الدلالي، حيث تجمع الوحدات الترميزية المتمثلة في مفرداتها تحت مفاهيم "Concepts"، ثم تدرس وتوصف العلاقات بين هذه المفاهيم بروابط دلالية فيكون حاصلها ما استقرت تسميته بالشبكة الدلالية "Semantic network" وهذه الشبكة ابتكرها ريتشارد ريتشنز عام ١٩٥٦م من جامعة كامبريدج البريطانية لتكون جسرًا بين اللغات المختلفة في الترجمة الآلية.

أما الدرس الأنطولوجي فيُعْنَى بوصف العالم الواقعي حاسوبيًّا وربط وحدات اللغة بهذا النموذج الحاسوبي للعالم. (عطية: ٢٠١٩، ٣٨).

(6.1) تحديات الذكاء الاصطناعي في معالجة اللغات الطبيعية (Natural language Processing)

وإذا كان الترميز لغة مشتركة بين العقل البشري والآلي فإن ذلك لا يعني أن الأمر سهل، بل هناك عقبات تقف أمام الباحثين نحاول إجمالها في النقاط التالية:

أولًا: اللغة وعاء الفكر والمعرفة، ومعارف الناس متنوعة ولا يمكن لأي نظام رياضي أن يحيط بها؛ ومن ثم فتنميطها في قواعد بيانات جامعة مانعة أمر بعيد المنال.

ثانيًا: لأن الخيال جزء من التفكير البشري فإن التفكير فيه والتعبير عنه ينعكس كذلك في اللغات الطبيعية؛ والتحدي لا يقتصر فقط على بناء أنطولوجيا للعالم الحقيقي بل التحدي الأكثر إعجازًا هو بناء أنطولوجيا للخيال، وللجزء المستور من تفكير البشر.

ثالثًا: لا يُضمِّن منتج اللغة –المتحدث أو الكاتب– عند استخدامه اللغة أداة اتصالية أو قناة لنقل محتوى معرفي على كل المعلومات التي يريد توصيلها للمتلقي السامع أو القارئ في رسالته، فلو فعل لصارت العملية ثقيلة العبء لدرجة تقارب الاستحالة، ولكنه يفترض أن السامع يشترك معه في معرفة سياق رسالته؛ ولذلك فإنه يضمن في رسالته فقط المعلومات المضافة على سياقها، فتبقى المادة اللغوية خلوًّا من السياق الذي تجري فيه الرسالة، وتكون بذلك المعلومات المتاحة للآلة التي تحاول معالجة اللغةِ الطبيعيةِ ناقصةً بشكل كبير.

رابعًا: يفترض منتج اللغة أن يكون المتلقي فاهمًا مجملَ خطابه وما فيه من حذف يظهره السياق الخارجي للكلام، وهذا شائع جدًّا في اللغات، مثل اللغة العربية التي يقرر فيها ابن مالك هذه القاعدة الاتصالية، فيقول:

وَحَذْفُ ما يُعْلَمُ جائزٌ كما

تقولُ: زَيْدٌ بَعْدَ مَنْ عِنْدَكُما؟

وفي جَوَابِ كَيْفَ زَيْدٌ؟ قُلْ: دَنِفْ

فَزَيْدٌ اسْتُغنِيَ عَنْهُ إذ عُرِفْ

فهذان البيتان يمثلان قاعدة اتصالية تعتمد على السياق، إذ الحوار بين المنتج والمتلقي فيه حذفٌ مفهومٌ من السياق اللغوي

والخارجي، وهذا تحدٍّ كبيرٌ جدًّا للآلة؛ إذ نحتاج إلى جمع أنماط الكلام ومحذوفاته لتكون مدخلات نستطيع من خلالها أن نحصل على نتيجة مشابهة للعقل البشري.

خامسًا: يمثل المجاز جزءًا أساسيًّا من كلام الناس، وهذا المجاز ليس مقصورًا على النصوص الأدبية من شعر ونثر، بل هو شائع ذائع في كلام الناس، في كل اللغات تقريبًا، والنمذجة الآلية للغة الطبيعية تقوم على تزويد الحاسوب بمعانٍ حقيقية لا تحتمل التأويل، ومن هنا تأتي مشكلة اللبس الدلالي، وتدريب الحاسوب على هذا اللبس يحتاج إلى قاعدة بيانات ضخمة جدًّا.

سادسًا: ولأن اللغة هي وعاء ثقافة المجتمع الذي يتغير دائمًا، فإننا نجد التطور في مستويات اللغة – وخاصة الدلالي– يجعل أنماط الكلام لا يمكن حصرها وإحصاؤها، ومن ثم فالإحاطة الآلية بكلام البشر على اختلاف أزمانهم وتطور لغاتهم يعد أمرًا بعيد المنال.

سابعًا: إن التحدي الأكبر الذي يواجه الذكاء الاصطناعي في معالجة اللغات الطبيعية يتمثل في إقامة حوار مع الآلة دون واسطة، ودون اضطرار لتعلم إحدى لغات البرمجة، فكل كلمة لها من الصفات الدلالية ما يجعلها تطلب متلازمة لها، فلو قلنا مثلًا: محمد، فعلينا أن نعلم الحاسوب، أنه أولا إنسان، وأنه مذكر، وأنه اسم عربي، وأن هذا الاسم يتطلب أن نجد له جملة تناسبه فنقول: محمد يمشي، وكلمة يمشي تستدعي أن نكتب بجوارها على الأرض، وهذا مستوى من مستويات اللغة، أما إن قلنا مثلا: (محمد يمشي على لحم بطنه) فإن معنى الجملة أنه جائع، وهي جملة تقال في ثقافة مجتمع دون آخر، ومثل هذه التعبيرات المجازية نحتاج أن نعلمها للآلة.

ثامنًا: بناء الجملة ودلالتها من التحديات الكبرى التي تواجهنا في معالجة اللغات الطبيعية المنطقية في الكلام، وهي ملاحظة نجد لها نصوصًا قديمة، ففي اللغة العربية نجد مثلًا سيبويه يفرق في الكلام بين المستقيم الكذب والمستقيم المحال، فيقول: "والكلام منه: مستقيم حسن، ومحال، ومستقيم كذب، ومستقيم محال، ومستقيم قبيح، وما هو محال كذب، فأما المستقيم الحسن فقولك: أتيتك أمس وسآتيك غدًا، وأما المحال فأن تنقض أول كلامك بآخره فتقول: أتيتك غدًا وسآتيك أمس، وأما المستقيم الكذب فقولك حملت

الجبل وشربت ماء البحر ونحوه، وأما المستقيم القبيح فأن تضع اللفظ في غير موضعه نحو قولك: قد زيد رأيت ، وكي زيد يأتيك، وأشباه هذا، وأما المحال الكذب فأن تقول: سوف أشرب ماء البحر أمس (سيبويه ١٩٧٨، ٢٥/١-٢٦)

وما أشار إليه سيبويه قبل مئات السنين يعد عقبة كبرى في المعالجة الآلية للغات الطبيعية، فنحن أمام تحدٍّ جديد لمعرفة نظم الكلمة في داخل الجملة بحيث تكون مفهومة بين منتج اللغة ومتلقيها، وهذا التحدي أشار إليه أيضًا عبد القاهر قبل مئات السنين بقوله: "وإنا إن بقينا الدهر نجهد أفكارنا حتى نعلم للكلمة المفردة سلكًا ينظمها وجامعًا يجمع شملها ويؤلفها، ويجعل بعضها بسبب من بعض غير توخي معاني النحو وأحكامه فيها طلبنا ما كل محال دونه) (عبد القاهر، ٣٦٠).

وإذا كان العقل البشري اللغوي وهو في قمة نضجه يواجه التحديات في معرفة سلوك الكلمة، ثم يتخذ النحو وسيلة لهذه الغاية، فإن الأمر سيكون أكثر صعوبة على الآلة؛ حيث سنحتاج لملايين التوسيمات للمفردات والجمل ورصدا للمتلازمات من مدونات ضخمة تجمع بين القديم والمعاصر، وهذه تأخذ وقتًا طويلًا وجهدًا كبيرًا، لكن كلما كانت قاعدة البيانات كبيرة كانت المعالجة الآلية للغات الطبيعية أكثر دقة وأقرب إلى اللغات البشرية، وستكون البحوث في هذه الناحية متجهة نحو التنميط وصياغة القواعد المنطقية، وهي عملية تستوعب كثيرًا من الظواهر بشكل تدريجي، فالمدونة عندما تستخرج كل قواعد اللغة التي بها ويتدرب عليها الحاسوب، ثم تزود المدونة لاحقًا بنصوص أخرى تظهر معها قواعد جديدة تعطي قوة للنظام ودقة في التعامل مع سلوك الكلمة داخل الجملة.

والنتيجة التي نخلص إليها بعد هذه التحديات أننا لا يمكن أن نجد نموذجًا رياضيًّا يمكنه استيعاب أنماط كلام البشر الذي تملؤه المشاعر والأحاسيس والحقائق والمجاز، ولا مفر من معالجة اللغات الطبيعية بشكل جزئي مرحلي.

(7.1) أقسام المعالجة الآلية للغات الطبيعية

رغم التعقيدات التي عرضناها في المعالجة الآلية للغات الطبيعية، فإننا بشكل عام نستطيع أن نصنف هذه المعالجات في إطارين هما: **الأول:** التعرف (Recognitions)، وفي هذا الإطار نجري

مجموعة من العمليات على النص المدخل؛ لإيجاد نموذج صوري يستوعبه الحاسوب ويتعامل معه بسهولة.

والإطار الثاني: التوليد (Generation)، ونعني به إنتاج النص من الحاسب اعتمادًا على النموذج الذي قمنا بتخزينه من عملية التعرف، وهذا التوليد ينقسم إلى شقين:

أولًا: التوليد العميق (Deep generation) ويقصد به: ماذا نريد أن نقول؟ وما هو مضمون النص المراد توليده؟

ثانيًا: التوليد السطحي (GENERATION DE SURFACE)، ونعني به: كيف نعبر عما نريد أن نقوله؟

(عيجولي حسين، ٢٠٢١، ٥٠٥)

وخلاصة القول فإن المعالجة الآلية للغات الطبيعية تحتاج إلى معارف كثيرة حول قواعد اللغة وهيكلتها وماهية الكلمات وكيفية تركيبها في بناء الجملة، ثم ترتيب الجمل داخل النص والحصول على جمل مقبولة المعنى.

وبناء على ما سبق من بناء نموذج ثم إنتاج نص نجد التطبيقات المختلفة في معالجة اللغات الطبيعية، وهذا ما سنحاول عرضه في المبحث الثاني.

المبحث الثاني معالجة اللغات الطبيعية.. المهام والتطبيقات

أولًا: الترجمة الآليةMachineTranslation

تمثل الترجمة جسرًا لنقل الثقافات بين اللغات، وهي من أهم مظاهر التواصل الحضاري؛ ولذا فهي تحظى بالاهتمام الواسع في كل زمان، ومن هذا الاهتمام ما حظيت به في الذكاء الاصطناعي من جهود تدل بلا مبالغة على أن الترجمة قطب الرحى في مهام معالجة اللغات الطبيعية على الإطلاق.

مفهوم الترجمة البشرية

الترجمة نقل معاني نص من لغة إلى أخرى مع مراعاة الدقة والأسلوب، ويتطلب ذلك فهم النص والتعبير عن المحتوى والأسلوب بلغة أخرى؛ فالمترجم يجب أن يتقن اللغتين: المترجم منها وإليها.

وبشريًّا عندنا طريقتان في الترجمة: الأولى الترجمة الحرفية، والأخرى الترجمة المعتمدة على فهم فحوى النص والتعبير بأسلوب المترجم نفسه.

والمترجم البشري يرجع إلى الوحدات المعجمية المتمثلة في الكلمات والتعبيرات الاصطلاحية ويفهم معانيها في سياقاتها اللغوية والاجتماعية المختلفة، وتظهر أمامه مشكلات على مستوى المفردات والتعابير الاصطلاحية، مثل المشتركات اللفظية التي تتشابه كتابة ونطقًا وتختلف في المعنى، مثل كلمة عين التي تعني (عين الإنسان وعين الماء) (إسماعيل صالح، ٢٠١٦).

وإذا كان العنصر البشري يجد مشكلات دلالية أثناء الترجمة العادية، فإن هذه المشكلة ستواجهها الترجمة الآلية؛ من أجل ذلك فإن غالبية التطبيقات المتاحة على الإنترنت لا تحظى بدقة عالية في الترجمة.

وإذا اتفقنا على أن الترجمة الآلية الحالية ليست بالدقة المطلوبة، فإننا ينبغي أن نذكر أن هذه الترجمة في سبيل تطويرها والسعي نحو غايتها المنشودة فإنها تمر بمراحل ثلاث: التحليل اللغوي، والمقابلة، والتوليد للغة المقصودة.

من تاريخ الترجمة الآلية

ليس هناك تاريخ محدد يمكن أن ننطلق منه في التأصيل للترجمة الآلية، غير أن بعض الباحثين يرى أن إرهاصات الفكرة جاءت في مطلع القرن السابع عشر الذي شهد اقتراح معاجم إلكترونية للقضاء على الحواجز اللغوية (Zughoul، ٢٠٠٥، P 1024.).

ولسنا ممن يرجعون إلى الوراء كثيرًا، فالترجمة الآلية هي نتاج تراكم معرفي في الرياضيات والمنطق، وما يهمنا متى بدأت بشكل فعلي بالمفهوم الذي نحن بصدده، وهو مفهوم الذكاء الاصطناعي.

وإذا ارتضينا هذه المنهجية في البحث فإن الوثائق التي بين أيدينا تشير إلى أن عام ١٩٥١م شهد معهد ماساتشوسش للتكنولوجيا في الولايات المتحدة مشروع الترجمة الآلية، وبعد ذلك بعام عقد أول مؤتمر في المعهد نفسه عرضت فيه تجارب في الترجمة الآلية حضره ثمانية عشر عالمًا (سلوى حمادة، ٢٠٠٧)، وفي عام ١٩٥٤م، عقدت شركة IBM أول عرض توضيحي عام للترجمة الآلية، وكان النظام يحتوي على مفردات صغيرة جدًّا من ٢٥٠ كلمة فقط، ويمكن أن يترجم فقط ٤٩ جملة روسية منتقاة بعناية إلى الإنجليزية.(Hutchins, J (2004.

تطور الترجمة الآلية:

ومنذ ذلك التاريخ والتطور جارٍ بقوة في الترجمة الآلية، هذا التطور الذي شهد طفرة في التسعينيات من القرن الماضي، حيث ظهرت برامج تترجم اللغة المنطوقة، مثل التكنولوجيا التي طورتها شركة (NEC شركة مختصة في كل ما يتعلق بالحاسوب والاتصال) وهي عبارة عن نظام يستطيع الترجمة شفويًّا بين لغتين عاميتين.

كما ظهر نظام (بولا) Paula وهو مترجم إلكتروني من الإنجليزية إلى لغة الإشارة الأمريكية، وهناك أيضًا برامج avatars التي تحول الإنجليزية المنطوقة إلى نص مكتوب على شاشة الكمبيوتر، و كذا تطبيقات أخرى على صلة وثيقة بالترجمة الآلية مثل التعرف الآلي على الكلام. (Zughoul، ٢٠٠٥، P1027-1031.).

وما زالت الجهود تتوالى في الترجمة الآلية على مستوى الأفراد والمؤسسات، حيث تكثر التطبيقات المتاحة للجميع بشكل يجعل من استخدامها تطويرًا في عملية النمذجة الآلية وتدريب الآلة على رفع مستوى دقتها في الترجمة.

أنواع الترجمة الآلية

الترجمة الآلية لها ثلاثة أنواع هي:

١. **الترجمة الآلية القائمة على القواعد (RBMT)** حيث تستخدم مجموعات القواعد لنقل البنية النحوية لمحتوى المصدر إلى اللغة المراد الترجمة

إليها، وفي هذا النوع يستخدم قاموس للغة المصدر لتحديد الكلمات والقواعد النحوية المناسبة في اللغة الهدف، ويتم تكييف الكلمات المأخوذة من القاموس بشكل مناسب (الجنس، العدد النحوي، ترتيب الكلمات، إلخ).

٢. **الترجمة الآلية الإحصائية (SMT)** **وفيها** تحدد الكلمات الأكثر شيوعًا من خلال التعلم والمقارنة بوثائق التدريب، ويتم ترجمة النص الأصلي المطلوب بناءً على احتمالية حدوثه في لغة الهدف.

٣. **الترجمة الآلية العصبية (NMT)**؛ وفيها يتم استخدم نظام التعلم العميق والتحسين المستمر؛ حيث يعمل NMT بشكل مشابه للدماغ البشري، وذلك باستخدام نماذج الشبكة العصبية لإنشاء نماذج ترجمة إحصائية.

مكونات نظم الترجمة الآلية

برامج الترجمة الآلية تتكون من:

أولا: قواعد لغوية فيها ترتيب المفردات والتراكيب النحوية والصرفية والتحليل الدقيق لعناصر بناء الجملة، وكلما كان التحليل دقيقًا كانت الترجمة عاكسة لهذه الدقة.

ثانيًا: قاموس لغوي خاص بنظام للترجمة الآلية، فكل كلمة لا بد أن توسم بنيويًّا ودلاليًّا في اللغة المصدر واللغة المترجم إليها، وهذه المعلومات يتم توليدها من محرك البحث، فإذا حدث التوسيم الدقيق لمدخلات الكلمات في النظام فإن هذا الترجمة الآلية ستتجاوز كثيرًا من مشكلات اللبس الدلالي وتعدد المعاني؛ إذ سيتدرب النظام على جمل حقيقية ونصوص مستخدمة، وستنمو قاعدة من المتلازمات والسياقات، حينها سيكون الإحصاء عاملا مرجحًا في اختيار المعنى المراد.

ثالثًا: الواجهة التي من خلالها نحافظ على شكل النص وتحديث ترجمته وتصويبها بشكل مستمر[١].

أنظمة الترجمة الآلية

تعتمد الترجمة الآلية على نظامين: إما ترجمة بشكل مباشر أو غير مباشر، ففي النظام المباشر نجد تفكيك النص إلى جزئيات ثم كلمات مستقلة، ثم تحديد المعنى الأكثر شيوعًا لكل مفردة، فهو نظام ينظر إلى جزئيات النص ولا ينظر إلى النص باعتباره كتلة واحدة، ومن ثم يعتري هذا النظام الكثير

(1) -www.softissimo.com4

من القصور، أما النظام غير المباشر فيتم عبر لغة وسيطة، حيث تتم عملية فصل تحليل النص المراد ترجمته عن عملية توليد النص المترجم، واستخدام لغة وسيطة هي عبارة عن أرقام مجردة، فتكون لدينا ثلاث خطوات:

1- تحليل النص الأصلي.

٢- ترميز التحليل.

٣- تحويل الرموز إلى الترجمة في اللغة المقابلة. (محمود إسماعيل: ١٩٩٩، ص ٩).

شكل ٢

نماذج من أنظمة الترجمة الآلية

١. نظام متيو-تاوم (الترجمة الآلية لجامعة مونتريال): بدأ هذا المشروع عام ١٩٦٥، وهو خاص بترجمة الأحوال الجوية، ويترجم من الإنجليزية إلى الفرنسية، ويعتمد على مفردات جديدة تقارب ألفي كلمة، وتعبير بمعدل ثلاثين ألف كلمة في اليوم، وهو كما نرى شديد التخصص، فتعبيراته ومصطلحاته المحدودة تجعل الترجمة الآلية غاية في الدقة[1].

٢. نظام ريفارسو (شركة سوفتيسيمو): يترجم مختلف أنواع النصوص (وثائق، مواقع الويب، رسائل إلكترونية...) على الإنترنت من الفرنسية إلى اللغات الأوربية الأخرى، وبالعكس (الإنجليزية، والألمانية، والإسبانية، والإيطالية).

٣. نظام اوروترا: تسميته هي اختزال لعبارة الترجمة وأوربا، يترجم من وإلى اللغات الأوربية كالفرنسية والإنجليزية والإيطالية والإسبانية...إلخ. وميزة البرنامج أنه بني انطلاقًا من نظرية متميزة تعتمد على [تمثيل] تصوير وسطي ودلالي مجرد بين اللّغة الأصل ولغة الهدف ونمط التحويل، بعكس البرامج المطورة حول محور التي تتطلب لغة ثالثة شديدة الهيكلة. (آمنة طالبي، ٢٠٠٧، ٥٤).

(1) - Anne-Marie LOFFLER-LAURIAN .La traduction automatique, Les divers systèmes. Presses Universitaires du Septentrion.1996, p21-22

ثانيًا: تعرف الكيان المحدد " Named entity recognition[1]

تعرف الكيان المحدد من المهمات الفرعية لاستخراج المعلومات المهيكلة تلقائيًّا من الوثائق غير المهيكلة المقروءة آليًّا، فتسعى NER إلى تحديد الكيانات المحددة في نص غير منظم، وتصنيفها في فئات محددة مسبقًا مثل: أسماء الأشخاص، والمؤسسات، والمواقع، والرموز الطبية، وتعبيرات الوقت، والكميات، والقيم النقدية، والنسب المئوية.....إلخ.

وهذه التقنية التي تعد من مهام اللغات الطبيعية تعتمد عليها كثير من الوظائف بشكل أساسي، فمثلًا في مجال الموارد البشرية، تستطيع هذه التقنية تلخيص السِّيرَ الذاتية للمتقدمين، كما أنها تستخدم في خدمة العملاء؛ حيث تصنف طلبات المستخدمين والشكاوى والأسئلة وتنظمها حسب الكلمات الرئيسية والأولويات المدخلة للنظام، فهذه التقنية لها قدرة على تصنيف المحتوى وتحديد موضوعاته، والحقول الدلالية التي ينتمي إليها، وذلك من خلال خطوتين:

الأولى: تحديد الكيان المحدد: تتضمن الخطوة الأولي اكتشاف كلمة أو سلسلة من الكلمات وتمثل كل كلمة رمزًا مميزًا.

الثانية: تصنيف الكيان: من خلال إنشاء فئات الكيانات وفيما يلي سنعرض بعض الكيانات والفئات الشائعة. (**S. Sekine," 1998.)**

الفئات	أمثلة
شخص	أحمد، سارة، شيماء، محمد، إبراهيم .... إلخ
مؤسسات	جامعة القاهرة، البنك الأهلي، جوجل ... إلخ
وقت	٦ am، ٢٠١١
موقع	الإسكندرية، ميدان الجيزة ... إلخ

مثال:

اشترى أحمد كتابًا من معرض الكتاب بمدينة نصر عام ٢٠١٩م.

تقوم أبحاث الـ NER بأخذ كتلة نصية كالمثال السابق وإنتاج جزء توضيحي من النص يبرز أسماء الكيانات التي تم ذكرها مسبقًا فيصبح المثال كالآتي:

(1) -S. Sekine, "NYU: Description of the Japanese NE system used for MET-2," 1998

اشترى أحمد [شخص] كتاب من معرض الكتاب [مؤسسة] بمدينة نصر [موقع] عام ٢٠١٩[وقت].

ولكي تجري التقنية هذه المهمة فإن الباحث في الذكاء الاصطناعي يبني نموذجًا تجرى فيه التوسيمات الدقيقة؛ حتى يميز بين الأرقام مثلًا إن كانت دالة على أعوام أو دالة على شيء آخر، وأن يوسم الأسماء لنعلم إن كانت أسماء لأشخاص أو لحيوانات أو لمؤسسات .. إلخ.

وفي المثال السابق استطاعت التقنية أن تحدد أن أحمد "اسم لشخص"، وأن معرض الكتاب "مؤسسة" وأن مدينة نصر "موقع"، وكل هذه مدخلات تخزينية سابقة في النظام، والسؤال: كيف تتم؟

عملية التخزين وبناء قواعد البيانات؟

في السابق كان الاعتماد في الغالب على قواعد مصنوعة يدويًّا، وحاليًّا هناك دراسات تعتمد على التعلم الآلي الخاضع للإشراف (Supervised machine Learning) كطريقة للبحث التلقائي للأنظمة القائمة على القواعد أو على خوارزميات، ووضع العلامات التسلسلية باستخدام مجموعة من أمثلة التدريب.

وتتمثل فكرة التعلم الخاضع للإشراف(SL) في دراسة ميزات الأمثلة الإيجابية والسلبية لـ NE على مجموعة كبيرة من المستندات المشروحة وقواعد التصميم التي تلتقط حالات من نوع معين، مثل نماذج ماركوف المستترة (HMM)[1]، وأشجار القرار Decision Trees، ونماذج الإنتروبيا القصوي (ME)[2] وآلات المتجهات الداعمة (SVM)، والحقول العشوائية المشروطة (CRF).

والعيب الرئيسي في (SL) الحاجة إلى مجموعة كبيرة مشروحة؛ حيث يؤدي عدم توافر هذه الموارد والتكلفة الباهظة لإنشائها إلى طريقتين تعليميتين بديلتين: التعلم شبه الخاضع للإشراف (Semi Supervised ML) والتعلم غير الخاضع للإشراف (Unsupervised ML).

(1) - A. Borthwick, J. Sterling, E. Agichtein, and R. Grishman, "Description of the MENE Named Entity System as used in MUC-7," 1998.
(2) - M. Asahara and Y. Matsumoto, "Japanese named entity extraction with redundant morphological analysis," in *Proceedings of the 2003 human language technology conference of the North American chapter of the association for computational linguistics*, 2003, pp. 8–15.

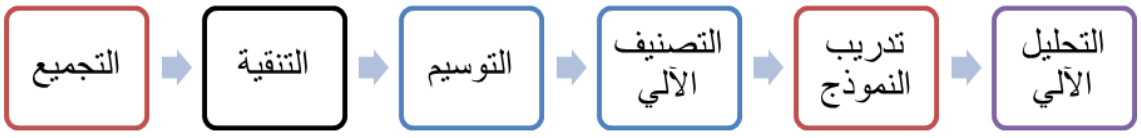
ثالثًا: تحليل المشاعر Sentiment analysis

من مهام معالجة اللغات الطبيعية تحليل المشاعر (sentiment analysis)، ويطلق عليه أيضًا الحوسبة العاطفية، حيث يتم تحليل النص لمعرفة مشاعر المتلقي تجاه هذا النص إن كانت سلبية أو إيجابية، وغالبا ما يستخدم ذلك في التسويق التجاري أو السياسي أو الأمني، فتستخدم التقنية لمراقبة وسائل التواصل الاجتماعي، وقياس الرأي العام تجاه مشكلة من المشكلات، كما تستخدم لتحليل دعم العملاء وآرائهم والبحث عن أماكن للبيع والشراء، أو البحث في توجهات الناخبين تجاه مرشح معين.

ويستخدم الذكاء الاصطناعي النهج المعجمي لبناء قواعد بياناته، فينشئ معجمًا للآراء سواء أكانت سلبية أم إيجابية أم محايدة، مع بناء خوارزمية لأكثر الكلمات والتراكيب التي تعبر عن رأي معين.

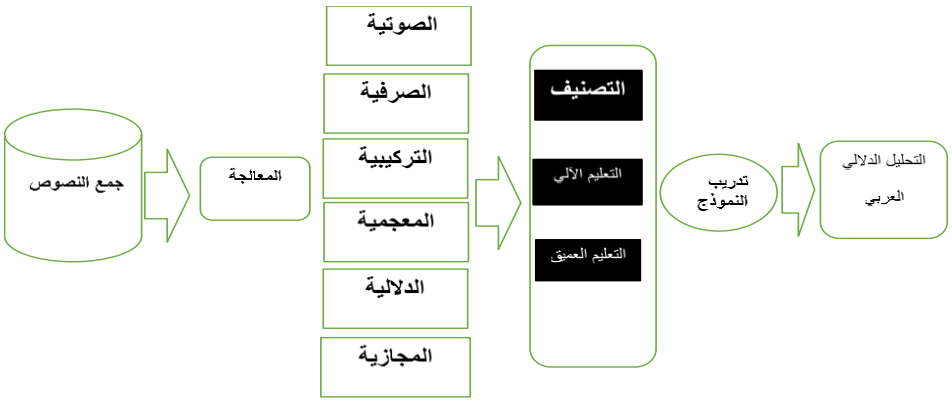
أما عن كيفية بناء نظام ذكاء اصطناعي لهذه المهمة في معالجة اللغات الطبيعية فإننا نقوم بالخطوات التالية:

١- تجميع النصوص.
٢- تنقية النص من الشوائب.
٣- المعالجة التوسيمية على المستوى الصوتي، والصرفي، والنحوي، والمعجمي، والدلالي.
٤- ثم التصنيف من خلال التعليم الآلي.
٥- ثم تدريب النموذج على معرفة نوع المشاعر.
٦- ثم التحليل الدلالي.



"الشكل:٣"

وهذا ما قام به نوح صبري وأحمد السياط وسعد العنزي، حيث قدموا ورقة بحثية بعنوان: "دراسة شاملة لتحليل المشاعر البلاغية في اللغة العربية، وقدموا النموذج التالي في آلية بناء نظام ذكي لتحليل المشاعر.



الشكل: ٤

والنموذج السابق الذي قدمه الباحثون كان على النصوص المكتوبة والمقروءة؛ حيث بنَوْا مكنزًا يحتوي على ألفي كلمة من نصوص قرآنية وأحاديث نبوية وتغريدات مقتبسة من موقع تويتر، مع الاستعانة بلغويين قاموا بما يأتي:

1- معالجة النصوص واستخراج الكلمات المكررة مثل (ما – كلا – لا...إلخ).
2- استخلاص الخصائص من النصوص مثل أسلوب صيغة المبالغة والاستعارة والتشبيه والكناية وتصنيفها إلى مشاعر إيجابية وسلبية.
3- تدريب البيانات باستخدام خوارزميات تعلم الآلة.
4- تقييم النتائج.

وقد خلصوا بعد تطبيقهم على البلاغة العربية إلى ما يأتي:

1- هناك اهتمام واسع بهذه التقنية في اللغات الإنجليزية والهندوأوربية مع انحسار في الاهتمام باللغة العربية.
2- هناك تفوق لبعض خوارزميات تعلم الآلة تتجلّى في معدلات الدقة العالية التي تصف المشاعر الكامنة وراء النصوص. **(نوح صبري وآخرون، ٢٠١٩)**.

وبشكل أكثر تفصيلًا لبناء نموذج لتحليل المشاعر يعرض علي فهمي هذه التقنية، وسنحاول استخلاص أهم ما عرضه على النحو التالي:

1- لكي نكون قادرين على تحديد المشاعر والآراء ينبغي أن نتعامل على مستوى عبارة داخل الجملة؛ لنعرف الكائن داخل الجملة هل هو شخص أو جهة أو منتج ... إلخ، كما نعرف خصائص هذا الكائن من حيث

مثلًا درجةُ الحرارة أو تصميم السيارة أو سعة الذاكرة ... إلخ، كما نتعامل على مستوى الجملة للوصول إلى الرأي.

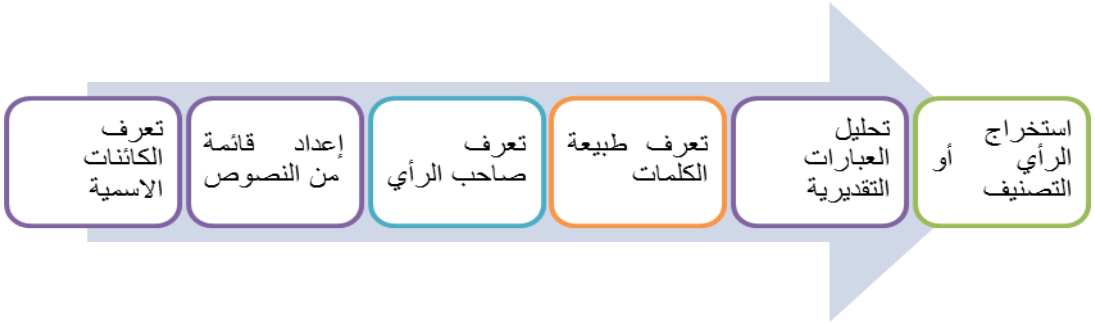
2- تتعامل تقنيات الذكاء الاصطناعي على مستوى الوثيقة للوصول إلى تصنيفات الرأي المستخلصة من الجمل.

3- أحيانًا تحتوي الجملة على أكثر من رأي أو مقارنة بين رأي وآخر، مثل محمد يحب كرة القدم ولكنْ عادل لا يكترث (علي فهمي، ٢٠١٩، ١٦٢-١٦٤)

ومما سبق فإن الذكاء الاصطناعي يتعامل مع ثلاثة عناصر: صاحب الرأي، وموضوع الرأي، والرأي ذاته، ولكي تتمكن الآلة من التعامل مع هذه العناصر بدقة فإن هناك خطواتٍ تتمثل في:

- **تعرف الكائنات الاسمية** بواسطة القواعد النحوية المبنية يدويًّا من متخصصين في اللغة،
- **إعداد قائمة كبيرة من النصوص** مع تمييز الكائنات الاسمية فيها يدويًّا، ويبقى دور البرمجيات في تعليم الآلة لاستخلاص نماذج التعرف على الكائنات الاسمية.
- **تعرف صاحب الرأي** من خلال عدة تقنيات إما نماذج ماركوف المستترة، أو تقنيات الحقول الشرطية العشوائية، أو المنهج القائم على قواعد المعرفة وتقنيات جمع البيانات ووجود معجم أو معلم، كما أن هناك تقنيات التعلم الذاتي أو التعلم المختلط ، بالإضافة إلى التقنيات الدلالية التي نحتاجها في فهرسة الدلالات الكامنة والترجيح بين المعاملات، والباحث يختار من هذه التقنيات ما يناسبه سواء أكانت مفردة أم مهجنة مع تقنية أو أكثر.
- **تعرف طبيعة الكلمات** والعبارات اللغوية، من حيث الموضوعية أو التقديرية.
- **تحليل العبارات التقديرية** ومعرفة الكلمات الإيجابية والسلبية والحيادية، وهذه مهمة سهلة على العقل البشري الذي يقرأ الكلمة في سياقها، لكنها بالغة الصعوبة عند تعليم الآلة التمييز بين المشاعر، خاصة إذا كانت المفردة أو العبارة أو الجملة المستخدمة في نطاق المجاز.

- وأخيرًا **استخراج الرأي** أو التصنيف (علي فهمي، ٢٠١٩، ١٦٤ - ١٦٩)



الشكل: ٥

رابعًا: المدقق الإملائي spell checker

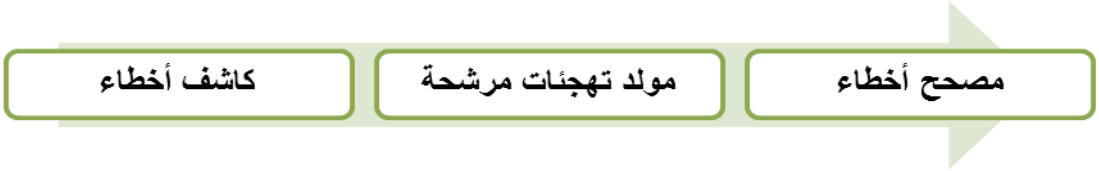
التدقيق الإملائي يكون على مستويين: **الأول** بنية الكلمة وتحليل ما بها من أخطاء، **والثاني** تصحيح تركيبي لموقع الكلمة في الجملة، والمدقق الإملائي يستطيع أن يقوم بالمهمة الأولى بشكل أكثر نجاحًا من المهمة الثانية؛ لأن تحديد شكل الكلمة في بناء الجملة يتطلب من المدقق معرفة إعرابها، والإعراب مقترن بالسياق الدلالي للكلمة.

ويتكون المدقق الإملائي من ثلاثة مكونات:

١- **كاشف الأخطاء** الذي يكتشف الكلمات التي بها أخطاء إملائية.

٢- **مولد تهجئات مرشحة** يوفر اقتراحات إملائية للكلمة المكتشف بها أخطاء إملائية.

٣- **مصحح الأخطاء** الذي يختار أفضل تصحيح من قائمة التهجئات المرشحة.



شكل: ٦

وتستخدم جميع المدققات الإملائية قاموسًا كقاعدة بيانات؛ حيث البحث عن كل كلمة من النص المكتوب في القاموس، وعند عدم العثور على كلمة في القاموس، يتم تحديدها على أنها خطأ؛ ثم يقوم المدقق الإملائي بالبحث في القاموس عن الكلمات التي تشبه أكثر الكلمات الخاطئة، ثم يقترح هذه الكلمات على المستخدم لاختيار أفضل كلمة كانت متوقعة.

اكتشاف الأخطاء

ويعتمد المدققون الإملائيون في الغالب على البحث في القاموس وتقنيات (النحو العددي) n-gram، وهي طريقة لاكتشاف الكلمات المكتوبة بشكل غير صحيح في كتلة نصية، بدلاً من مقارنة كل كلمة كاملة في النص بقاموس، وتتم مقارنة n–grams فقط بالقاموس؛ لأن مقارنة كل كلمة مفردة بالقاموس هي عملية تستغرق وقتًا طويلاً. ويتم إجراء الفحص باستخدام مصفوفة ذات أبعاد n حيث يتم تخزين ترددات n gram الحقيقية.

وإذا كان اكتشاف n-gram نادرًا أو غير موجود ، يتم وضع علامة على الكلمة على أنها خطأ أو بها خطأ إملائي.

إن n-gram عبارة عن مجموعة من الأحرف المتتالية مأخوذة من سلسلة طولها n. إذا تم تعيين n على واحد، فسيتم تسميته unigram، وإذا كان n هو اثنين، فسيكون Bigram، وبالمثل إذا كان n هو ثلاثة، فإن المصطلح هو trigram.

ويتم تقسيم كل سلسلة تشارك في عملية المقارنة إلى مجموعات متجاورة من n-grams. تتمتع خوارزميات n-grams بميزة رئيسية، هي أنها لا تتطلب معرفة باللغة التي يتم استخدامها بها؛ ولذا يطلق عليها غالبًا خوارزمية مستقلة عن اللغة.

تصحيح الأخطاء

عندما يتم اكتشاف خطأ في النص المكتوب، يتم تطبيق تقنيات التصحيح الإملائي على موضع الخطأ لتصحيح الكلمة أو تقديم اقتراحات صحيحة لتلك الكلمة. تسمى هذه العملية التصحيح الإملائي. (William B. Cavnar and John M. Trenkle:)

ويتصل بالتدقيق الإملائي –كما أشرنا– المصحح النحوي (Grammatical Error Correction)؛ حيث يأخذ نظام GEC جملة قد تكون خاطئة كمُدخل ومن المتوقع أن يحولها إلى نسختها المصححة.

خامسًا: تلخيص النص Text Summarization

(١.١) مفهوم التلخيص والحاجة إليه:

تلخيص النصوص عملية ذهنية معقدة تتطلب كثيرًا من المهارات، فأي قارئ عادي كي يعطي خلاصة ما قرأه فإنه لا بد أن يفهم النص أولًا، ويحدد أهم أفكاره ثانيًا، ثم ينتج نصًّا جديدًا موجزًا يتوافق مع النص الطويل الذي استطاع تلخيصه، وهذه مهمة نحتاجها

في حياتنا العامة كثيرًا، نحتاجها في الإعلام لتقديم موجز الأخبار، ونحتاجها في القضاء لتقديم موجز عن الأحكام، ونحتاجها في التعليم، وفي كل عمل تغلب عليه كثرة التدوين والتقارير.

إن هناك ملايين الوثائق التي تحتاج إلى تلخيص في حياتنا العامة، وإذا لخصت بشكل بشري فإن ذلك يأخذ من الوقت والجهد الكثير والكثير، من هنا كانت الحاجة إلى الذكاء الاصطناعي لحل هذه الإشكالية من خلال التلخيص الآلي الذي نحاول أن ندرب من خلاله الآلة على فهم أهم ما بالنص من معانٍ وأفكار.

(١.٢) أنواع التلخيص

التلخيص البشري يكون بطريقتين: إما بإسقاط الكلمات والجمل الفرعية والهامشية في النص، وإما بإعادة صياغة النص مختصرًا بعد فهمه.

والتلخيص الآلي يأخذ المفهوم نفسه حيث **إن هناك تلخيصًا يقوم على الاستخراج**، من خلال مجموعة فرعية من الكلمات التي تمثل أهم النقاط، يتم سحبها من جزء من النص ودمجها لعمل ملخص.

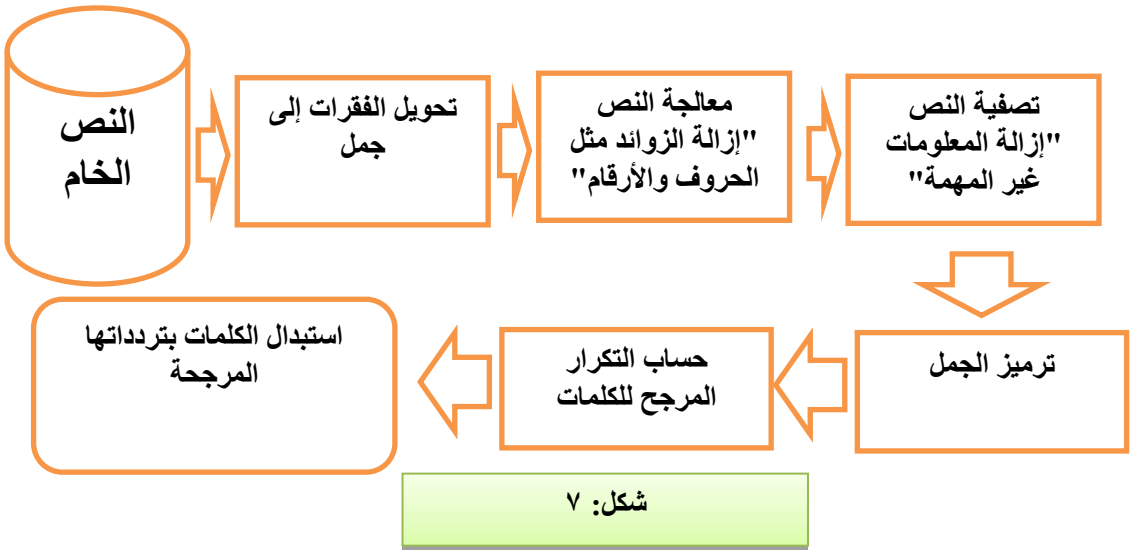
وفي التعلم الآلي، عادةً ما يتضمن التلخيص الاستخراجي موازنة الأقسام الأساسية للجمل، واستخدام النتائج لتوليد الملخصات، ويمكن استخدام أنواع مختلفة من الخوارزميات والأساليب لقياس أوزان الجمل، ثم ترتيبها وفقًا لمدى ملاءمتها وتشابهها بعضها مع بعض – ودمجها بشكل أكبر لإنشاء ملخص.

وهذا النوع هو الوسيلة المثلى لإنتاج نص موجز، لكن يحتاج إلى جهد كبير، وتقنيات معقدة ونمذجة متطورة؛ لأنه يحتاج إلى توليد نص جديد وتمثيل دلالي وحذف واستبدال وهي عمليات تحتاج خوارزميات معقدة.

(١.٣) كيفية تنفيذ التلخيص النصي آليًّا

- تحويل الفقرة إلى جمل.
- معالجة النص(Text processing) وإزالة كلمات التوقف (الكلمات الشائعة للغاية مع القليل من المعاني مثل "و" و "أو")، والأرقام، وعلامات الترقيم، والأحرف الخاصة الأخرى من الجمل.
- يساعد إجراء التصفية في إزالة المعلومات الزائدة عن الحاجة والتي قد لا توفر أي قيمة مضافة لمعنى النص.

- **الترميـــز (Tokenization)**: يـتم ترميز الجمل للحصول على جميع الكلمات الموجودة في الجمل.
- **حساب التكرار المرجح للكلمات**: بعد ذلك نحسب التكرار المرجح لجميع الكلمات، حيث نقسم تكرار حدوث كل كلمة على تكرار الكلمة الأكثر تكرارًا في الفقرة.
- **أن تُسْـتَبْدَلَ بالكلمـات ترددا تُهـا المرجحـة**: تُسْـتَبْدَلُ بكـل كلمـة مـن الكلمات الموجودة في الجمل الأصلية تردداتُها المرجحة. بعد ذلك، يحسب مجموعهـا، ونظـرًا لأن التـرددات الموزونة للكلمات غير المهمة، مثل كلمات التوقف والأحرف الخاصة، التـي تمـت إزالتهـا أثنـاء مرحلـة المعالجة، تساوي صفرًا، فليس من الضروري إضافتها.



شكل: ٧

سادسًـــا: اســـترجاع المعلومـــات Information Retrieval

(١.١) المفهوم والوظيفة:

يمكـن تعريـف اسـترداد المعلومـات (IR) علـى أنـه برنـامج برمجـي يتعامـل مـع تنظـيم المعلومـات، وتخزينهـا، واسـترجاعها، وتقييمهـا مـــن مستودعـــات المستنـدات، وخاصـة المعلومات النصية.

ويساعد النظـام المستخدمين فـي العثـور على المعلومات التي يحتاجون إليها، لكنه لا يعرض بشكل صريح إجابات الأسئلة.

يُعلـم النظـام بوجـود المسـتندات التـي قـد تتكـون مـن المعلومـات المطلوبـة، ويبـين

موقعها، والمستندات التي تفي بمتطلبات المستخدم تسمى المستندات ذات الصلة.

(١.٢) آلية العمل

نظام IR المثالي يسترجع المستندات ذات الصلة فقط، والمستخدم الذي يحتاج إلى معلومات سيتعين عليه صياغة طلب في شكل استعلام بلغة طبيعية، ثم سيستجيب نظام IR عن طريق استرداد المخرجات ذات الصلة، في شكل مستندات حول المعلومات المطلوبة.

(١.٣) مشكلة نظام استرجاع المعلومات (IR)

في الاسترداد المخصص، يجب على المستخدم إدخال استعلام بلغة طبيعية يصف المعلومات المطلوبة، ثم سيعيد نظام IR المستندات المطلوبة المتعلقة بالمعلومات المطلوبة، فعلى سبيل المثال، لنفترض أننا نبحث عن شيء ما على الإنترنت، ويعطي بعض الصفحات الدقيقة ذات الصلة وفقًا لمتطلباتنا، ولكن يمكن أن تظهر لدينا صفحات لا تتصل بما نبحث عنه، وهذه من مشاكل استرجاع المعلومات المخصص.

(١.٤) جوانب الاسترجاع المخصص Ad-hoc Retrieval

تتمثل جوانب الاسترجاع المخصص في عدة جوانب هي:

١. مساعدة المستخدمين في تحسين الصياغة الأصلية للاستعلام.
٢. كيفية دمج النتائج من قواعد بيانات نصية مختلفة في مجموعة واحدة.
٣. كيف يتم التعامل مع البيانات التالفة جزئيًّا؟

(١.٥) تحديد النماذج المناسبة؛ حيث إن هناك **نموذج IR الكلاسيكي الذي** يعتمد على المعرفة الرياضية وهناك **نموذج IR غير الكلاسيكي الذي** يعتمد على مبادئ أخرى غير التشابه والاحتمال والعمليات المنطقية، كما أن هناك **نموذج IR البديل، وهو** تحسين لنموذج IR الكلاسيكي باستخدام تقنيات محددة من بعض المجالات الأخرى. نموذج الكتلة والنموذج الضبابي ونماذج الفهرسة الدلالية الكامنة (LSI) هي مثال لنموذج الأشعة تحت الحمراء البديل.

(١.٦) مكونات نموذج استرجاع المعلومات (IR)

يتكون نموذج استرجاع المعلومات من:

١. نموذج للوثائق A model for documents.

٢. نموذج للاستعلامات A model for queries

٣. نموذج وظيفة مطابقة تقارن الاستعلامات بالمستندات A matching function that compares queries to documents.

سابعًا: التعرف الآلي على الكلام المنطوق Automated speech recognition

(١.١) **المفهوم:** من مهام اللغات الطبيعية التعرف الآلي على الكلام المنطوق؛ فالذكاء الاصطناعي الآن صار قادرا على استخدام أصوات البشر وبناء نماذج منها لتوليد كلام جديد، وهذا الكلام تطور من المفردة إلى الجملة إلى الحوار الذي يصاحبه أداء تنغيمي عال.

(١.٢) **تدريب التقنية:** وهذه التقنية تعتمد بشكل أساسي على تدريب النظام من خلال قاعدة بيانات صوتية وعنونتها بناء على معطيات التحليل الصوتي الحاسوبي، وغالبا ما يتم الاعتماد على عدد كبير من المتحدثين للغة أو اللهجة، بحيث تختلف أعمارهم وثقافاتهم، وتسجل لهم مفردات معزولة عن السياق أو نصا بأكمله، ثم تحلل هذه التسجيلات إلى فونيمات ومقاطع يتدرب عليها النظام الذي أدخل إليه نموذج صحيح يمكنه القياس عليه والمقارنة بينه وبين المدخلات التي يتدرب عليها، فمثلا إذا دربنا الحاسوب على أن التاء تكون مرققة ورمزنا لها بالرمز (t) في مقابل الطاء أنها مفخمة ورمزنا لها بالرمز (T) فإن التاء ستدخل للحاسوب من كلمات كثيرة وهي مجاورة

لكل حروف الهجاء، كما أن النموذج الصحيح الذي يحدد خصائص صوت التاء من حيث الترقيق والهمس، والتفرقة بين التاء الساكنة والمتحركة وغير ذلك من الخصائص اللغوية التي تكون قواعد للنظام يقيس عليها كل صحيح.

وتفيد التسجيلات التي تجمع حسب النوع والعمر والثقافة بالأخطاء المتنوعة للفونيمات، وهذه الأخطاء غير محصورة، بل هي كثيرة جدا، فلا يقتصر الأمر على حذف الحرف أو تبديله، بل هناك أخطاء نطقية غير متوقعة وغير متخيلة، والآلة تتدرب على اكتشاف هذه الأخطاء، بل وتستطيع تصنيفها.

(١.٣) **التطبيقات:** وهذه التقنية تفيد في تطبيقات كثيرة، بل صارت مستخدمة الآن في حياة الناس، حيث نجد الناس يخاطبون الآلات ويعطونها الأوامر، فمثلا في الهواتف المحمولة نجد أن المستخدم يأمر الهاتف بأن يفتح تطبيقا ما أو يتصل باسم معين في هاتفه والهاتف يتعرف الصوت واللغة ويستجيب للأمر (Lori Lamel,2008: 8)

وتفيد التقنية أيضا ذوي الاحتياجات الخاصة من المقعدين والمكفوفين، فبإمكانهم الآن إعطاء أوامر صوتية للكراسي المتحركة، فتستجيب الآلة للأمر وتوجه حسب رغبة المستفيد.

وطبقت التقنية أيضا لقراءة الوثائق إلكترونيا؛ مما أسهم في ظهور آلاف من الكتب الناطقة، وهذا أسهم في خدمة المكفوفين الذين يعانون من عدم وجود من يقرأ لهم ويعلمهم.

كما تفيد التقنية في تعليم النطق وتصحيح أخطائه، وهذا عنصر جوهري ومحوري في كل اللغات الطبيعية، حيث تسعى كل ثقافة أن تعلم أبناءها اللغة التي تتكلم بها.

وتفيد التقنية أيضا في الفهرسة الصوتية، والبحث الصوتي الآلي؛ حيث يمكن للمستخدم الحصول على مواضيع معينة من ملفات صوتية مطولة (Lori Lamel, 2008:8)

ثامنًا: تقنية تحويل النص إلى صوت (Text to Speech)

(١.١) **مفهوم التقنية:** إن هذه التقنية قادرة على قراءة أي نص مكتوب، والنصوص المكتوبة نوعان: إما نص مكتوب بصيغة (Text) أو (word) أو أية صيغة يمكن التعديل فيها أو يكون النص المكتوب مخطوطا أو

مصورا وهنا نلجأ إلى تقنية التعرف على الصور (OCR) لتحويل هذه الصور إلى شكل نصي، ومن ثم تزويد الذكاء الاصطناعي به لقراءته آليًّا.

(١.٢) كيفية عملها: يبدأ الذكاء الاصطناعي عمله بهذه التقنية من خلال الخطوات التالية:

١. تحديد المجال الذي نريد منه تحويل النص إلى صوت؛ حتى نركز نصوص المعالجة على هذا المجال، فإذا كان المقصود قراءة آلية للوثائق السياسية نركز في جمع النص على كل النصوص السياسية والإخبارية مع التركيز على أنماطها وأنواعها ومفرداتها وتراكيبها.

٢. تحديد نسب توزع حروف اللغة وخصائصها الصوتية، وعندنا في العربية مثلا لا بد أن تكون العينة ممثلة لكل الحروف العربية بحركاتها وسكونها وما فيها من نبر وتنغيم وغيرها من الخصائص.

٣. تسجيل الصوت من متحدث، وجعله نموذجًا.

٤. المطابقة بين النص المكتوب والصوت المتوقع نطقه من خلال تطبيق قاعدة كل ما يُقرَأُ يكتب حتى ولو كان مخالفا لقواعد اللغة، وتحويل نصوص اللغة إلى رموز يسهل بعد ذلك تحويل هذه الرموز إلى قيم رياضية.

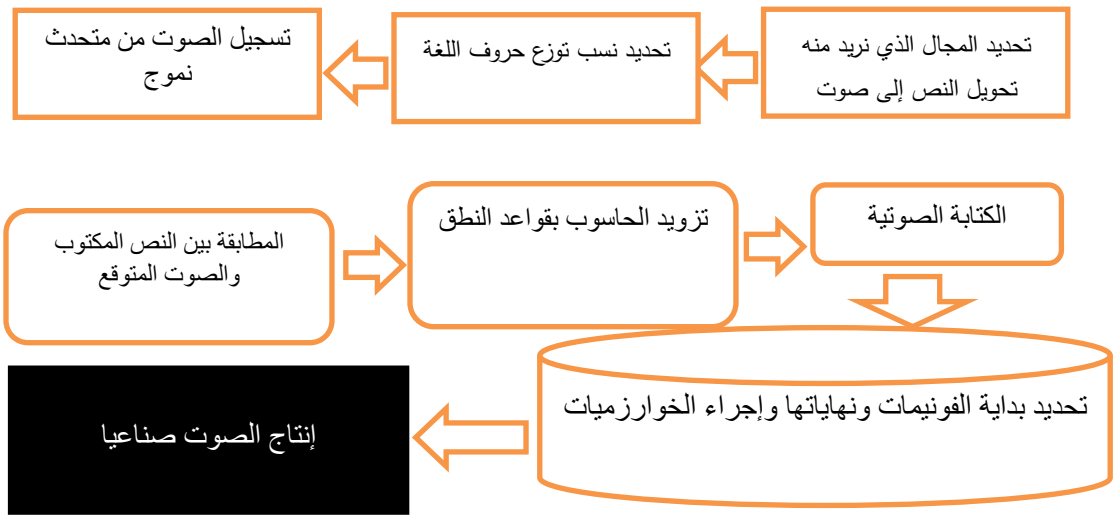
٥. استخلاص قواعد نطق عامة من خلال المادة المسجلة مثلا ندرب الحاسوب على أن واو (عمرو) لا تنطق وكذلك الواو في (قالوا) واللام الشمسية لا تنطق وغير ذلك من قواعد اللغة العربية، وقس على ذلك في كل اللغات لها القواعد التي تمثلها.

٦. تخليق الكتابة الصوتية بناء على تشكيل دقيق لكل ما سمع من المادة المسجلة.

تحديد بداية الفونيمات ونهايتها بشكل آلي عبر تقنية (HMMs) والهدف من ذلك معرفة الطول الزمني لكل فونيم، حتى تقوم الآلة بحساب المتوسطات لكل فونيم، ثم ينتج الصوت الصناعي بعد ذلك[١].

---

(١) تجدر الإشارة إلى أن الشركة الهندسية لتطوير نظم الحاسبات قامت بتطوير عدد مهم من التطبيقات اعتمادا على هذه التقنيات، ويمكن تجريبها عبر الرابط السابق المنتجات والخدمات – RDI (rdi-eg.ai)



شكل: ٨

تاسعًا: تقنية التعرف على الصور Optical Character Recognition (OCR)

(١.١) المفهوم والمنهج

تقنية التعرف على الصور هي إحدى التقنيات المهمة التي وظفها الذكاء الاصطناعي في تحويل النصوص المصورة إلى نصوص مكتوبة، وبنفس المفاهيم السابقة التي عرضناها فإن عملية التعرف الآلي على النص المكتوب يتم من خلال تدريب الآلة على نماذج صحيحة وخاطئة لتوليد النص المراد.

سنعرض هذه التقنية من خلال التعرف الآلي على الكتابة العربية وفق نماذج ماركوف المستترة والشبكات العصبية الاصطناعية، وقد أشرنا إلى هذين النموذجين في ثنايا هذا البحث، فنماذج ماركوف تسعى رغم اختلاف تنويعاتها إلى استنباط "تسلسل المدخلات" الأرجح احتمالا الذي يؤدي عبر الظاهرة العملية قيد الدراسة إلى تسلسل الملاحظات التي يناظرها ترميزات حروف الكلمات المطبوعة المراد التعرف عليها، أما تسلسل الملاحظات فهو تسلسل "متجهات" يصوغ كل منها رياضيا بعض خصائص المحتوى الرسومي داخل نافذة منزلقة على السطور الممسوحة ضوئيا.

(١.٢) آلية التطبيق

من أجل بناء نماذج رياضية لا بد من مادة لغوية تملأ وعاء واسعا من الأسئلة وهي الكتابة المرقمنة مطبوعة أو مخطوطة بالتوازي

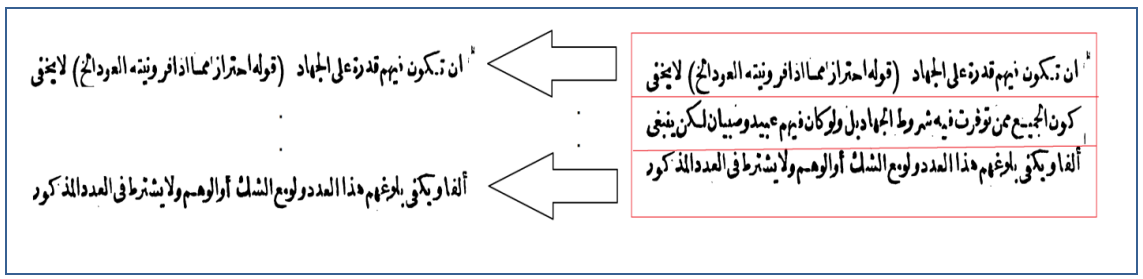
مع أجوبتها لتغذية أساليب التعلم الحاسوبي الموجه (supervised machine Learning)، ويطلق على عملية تشغيل هذه الأساليب الرياضية على المادة اللغوية من أجل بناء النماذج الاحتمالية الرياضية اسم التدريب (Training)، كما يطلق على الوعاء الممتلئ بتلك المادة اللغوية اسم المورد اللغوي (Language Resource)، وبطبيعة الحال فإن النماذج الاحتمالية سوف تحمل من الخصائص الإحصائية للمورد اللغوي الذي تكونت منه؛ ولذلك فإن بناء هذه الموارد اللغوية يشترط إضافة إلى الدقة العالية أن يجري التمثيل بصورة متوازنة لمختلف أبعاد التنوع في الكتابة مثل: أبناط الطباعة وأحجامها أو الخطوط اليدوية للكتاب ...إلخ.

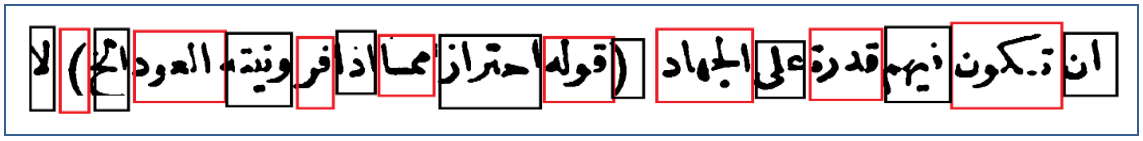
ويتشابه كثيرا بناء الموارد اللغوية لأنظمة التعرف على الكتابة العربية المخطوطة يدويا مع تلك المتعلقة بأنظمة التعرف على الكتابة العربية المطبوعة مع ملاحظة أن عمليات التقسيم الآلي للصفحات إلى كلمات وسطور ثم تصويب هذا التقسيم بل وتفسير النصوص في الصفحات المخطوطة يدويا تستغرق من اللغويين عملا ووقتا أكبر (عطية، ٢٠١٩: ٤٦-٥٣).

(١.٣) تطبيق عملي

في هذه النقطة سنعرض خطوات تحويل الصور إلى النص كما يأتي

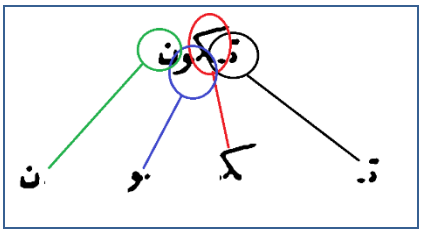
- تنقية الصور من الشوائب وتحويلها إلى صور نظيفة.
- تقسيم الصفحات والمستندات إلى سطور.





- تقسيم السطور إلى كلمات للتعرف على كلماتها.
- تقسيم الكلمات إلى حروف ومن ثم التعرف على تلك الحروف.

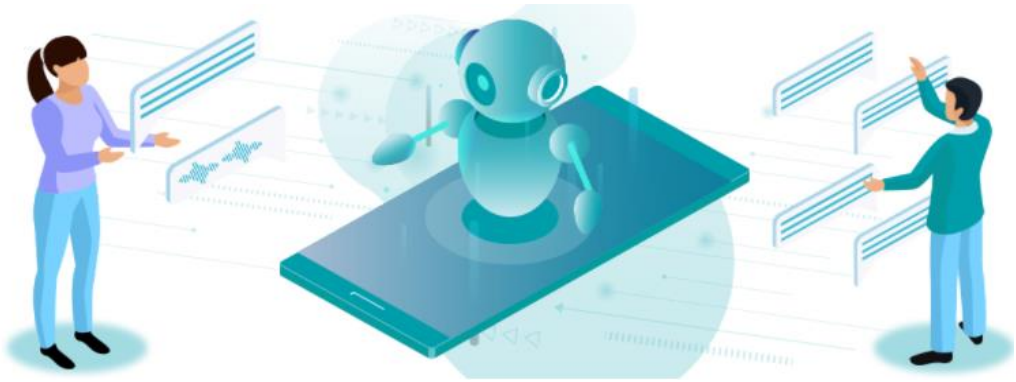
▶ تجميع الحروف المتعرف عليها لكل كلمة ومن ثم بقية كلمات السطر والصفحة.



عاشرًا: المساعد الافتراضى " Virtual Assistant"

المساعد الافتراضي (VA) هو برنامج كمبيوتر متقدم يستطيع الفهم والمعالجة والتعلم والاستجابة لمدخلات الصوت أو النص بطرق طبيعية.

ويزوَّد المساعد الافتراضي واجهات المحادثة بأزرار أو قوائم أو بدونها، ويمكن أن يفهم ويستجيب لألفاظ المستخدم المعقدة، ويجمع عادةً بين المعالجة المتقدمة للغة الطبيعية (NLP) وفهم اللغة الطبيعية (NLU) وتقنيات إدارة الحوار للتفاعل بطريقه أوتوماتيكيه دون التضحية بالدقة (Wolff, R. (2021, 1 8)).



رسم توضيحي ٩ المساعد الافتراضى

يستبدل المساعدون الافتراضيون واجهات المستخدم "UIs " المعقدة بلغة طبيعية؛ لذلك لا يوجد منحنى تعليمي ووقت ضائع، ويمكنهم التفاعل تقريبا مع أي برنامج يوفر واجهة برمجة تطبيقات "API"، وبالتالي فإن نطاق حالات الاستخدام لا حصر له.

ويستخدم المساعدون الافتراضيون مزيجًا من NLP والتعلم الآلي (ML) والتعرف على الكلام speech recognition وتقنيات أخرى لتفسير كلمات المستخدم ومطابقتها مع ما يريده وإنشاء استجابة مناسبة. على غرار التطبيقات، تشتمل المساعدين الظاهريين على

جملة تطبيقات application layer وقاعدة بيانات database و تطبيقات APIs وواجهة محادثة المستخدم conversational user interface (CUI) ، وهي متوافرة من خلال منصات مثل Facebook Messenger و Slack وغيرها.

الحادي عشر: توضيح معنى الكلمة "Word Sense Disambiguation"

من المعلوم أن الكلمات لها معانٍ مختلفة بناءً على سياق استخدامها في الجملة، وإذا تحدثنا عن اللغات البشرية، فهي غامضة أيضًا؛ لأن العديد من الكلمات يمكن تفسيرها بطرق متعددة اعتمادًا على سياق حدوثها، ويمكن تعريف توضيح معنى الكلمة، في معالجة اللغة الطبيعية (NLP)، على أنه القدرة على تحديد معنى الكلمة الذي يتم تنشيطه عن طريق استخدام الكلمة في سياق معين.

والغموض المعجمي أو النحوي أو الدلالي إحدى المشاكل الأولى التي يواجهها أي نظام معالجة لغوية (NLP). ويمكن لعلامات تمييز جزء من الكلام (POS) ذات المستوى العالي من الدقة حل الغموض النحوي syntactic ambiguity في الكلمة.

من ناحية أخرى يقوم توضيح معنى الكلمة بحل مشكلة الغموض الدلالي semantic ambiguity، وعادة ما يكون حل الغموض الدلالي أصعب من حل الغموض النحوي.

على سبيل المثال، ضع في اعتبارك المثالين للمعنى المميز الموجود لكلمة "bass" :-

I can hear bass sound - He likes to eat grilled bass.

يشير حدوث كلمة bass بوضوح إلى المعنى المميز، في الجملة الأولى ، تعني frequency وفي الثانية تعني fish. ومن ثم إذا تم توضيحها بواسطة WSD، فيمكن تعيين المعنى الصحيح للجمل أعلاه على النحو التالي:

I can hear bass/frequency sound - He likes to eat grilled bass/fish.

يتطلب تقييم WSD المدخلات التالية :-

١- **قاموس:** المدخل الأول لتقييم WSD هو القاموس، والذي يستخدم لتحديد الحواس التي يجب توضيحها.

٢- **مجموعة الاختبار:** الإدخال الآخر المطلوب من WSD هو مجموعة high-annotated test corpus التي تحتوي على الهدف أو الحواس

الصحيحة correct–senses. ويمكن أن تكون هيئة الاختبار من نوعين:

الأول: العينة المعجمية Lexical sample حيث يستخدم هذا النوع من المجموعات في النظام، عينة صغيرة من الكلمات.

والثاني: كل الكلمات All-words – إذ يتوقع في هذه المجموعة الكبيرة من الكلمات أن يزال الغموض عن جميع الكلمات في جزء من النص قيد التشغيل.

الثاني عشر: التقييم الآلي (Automatic Scoring)

(١.١) المفهوم والحاجة

مع كثرة عدد الطلاب في المدارس والجامعات والحاجة المستمرة إلى اختبارات وتقييمات يتزايد الطلب على الذكاء الاصطناعي لإنتاج تطبيق يقوم بإعداد الأسئلة وتقييمها آليا بشكل يحقق العدالة بين الطلاب، ويراعي تحقيق متطلبات الدراسة المنهجية.

وتتعامل أنظمة التقييم الآلي (AS) مع الطلاب من خلال ثلاث طرق: الأولى تصحيح الإجابات المكتوبة ومنها موضوعات التعبير والأسئلة المقالية وهذه تمثل تحديا كبيرا لأنها تقيس أسلوب الطالب وفهمه، والثاني تقييم الإجابات المنطوقة سواء أكان التعبير عن صورة أم نطق جملة مكتوبة، أما الطريقة الثالثة، فهي تصحيح المسائل الرياضية وهي تتطلب معادلات ومسائل وخطوات محددة.

فالذكاء الاصطناعي مطالب بحل هذه المشكلة التي سنعرض منها فقط الأسئلة المقالية التي قد تكون وصفية تتطلب سردا لنقاط أو أسئلة مناقشة تطلب من الطالب أن يوضح أو يفسر أو يستعرض قضية ما أو أسئلة تقييم وإبداء الرأي وعرض الدليل أو أسئلة مقارنة وغير ذلك من الأسئلة المقالية.

ومن الجدير بالذكر أن معظم أعمال تقييم المقال تتعامل مع اللغة الإنجليزية مع قليل من النظم التي صممت لدعم لغات أخرى مثل العبرية واليابانية ولغة الملايو "البهاسا" (علي فهمي: ١٢٤)

(١.٢) طرق التقييم الآلي

في هذا التقييم نحتاج من التقنية أن تضع درجات لكل إجابة، وتمر معظم أنظمة التقييم الآلي للموضوعات التعبيرية عبر نفس الخوارزميات الإشرافية (Supervised algorithms) التي تتطلب مرحلة التعليم،

فمراحل بناء نموذج التقييم الآلي (AES) هي على النحو التالي:

١- الإعداد وتشمل تجهيز عينة من المقالات التعبيرية التي تم تصحيحها سابقا بشكل يدوي
٢- تحديد السمات التي سيتدرب عليها النظام
٣- بناء نموذج برنامج التصحيح والتدريب وتحديد الخوارزم المناسب، مثل خوارزم بايز (Bayes Algorithm)
٤- بناء علاقة بين السمات المستخرجة يدويا ودرجة التقييم المستخرجة من النموذج
٥- الاستخدام الفعلي لبرنامج التصحيح.

الثالث عشر: الإجابة عن الأسئلة Question answering

(١٠١) المفهوم

الإجابة عن الأسئلة هي حقل فرعي من NLP والتعرف على الكلام يستخدم NLU لمساعدة أجهزة الكمبيوتر على فهم أسئلة اللغة الطبيعية تلقائيًا.

على سبيل المثال، إليك سؤالًا شائعًا قد تطرحه على Google Assistant: "كيف سيكون الطقس غدًا؟"

يمكن لأدوات NLP أن تقسم هذا السؤال إلى موضوع (الطقس) وتاريخ (غدًا)، وفهمه وجمع الإجابة الأكثر ملاءمة من مجموعات غير منظمة من " natural language documents": تقارير الأخبار عبر الإنترنت، وصفحات الويب المجمعة، والنصوص المرجعية ، إلخ.

بشكل افتراضي، يخبرك المساعدون الافتراضيون بحالة الطقس لموقعك الحالي، ما لم تحدد مدينة معينة، والهدف من الإجابة عن الأسئلة هو إعطاء المستخدم إجابة بلغته الطبيعية، بدلاً من قائمة الإجابات النصية.

إن نظام الإجابة تلقائيا عن الأسئلة – التي يقولها الإنسان بلغة طبيعية باستخدام قواعد البيانات المهيكلة مسبقا أو مجموعة من مستندات من اللغات الطبيعية– يُقَدِّمُ المعلومات المطلوبة بدلًا من البحث في الوثائق الكاملة للموضوع مثل محركات البحث وأشهرها جوجل.

تتضمن بعض الأمثلة على مجموعات مستندات اللغة الطبيعية المستخدمة لأنظمة الإجابة عن الأسئلة ما يلي:

- **مجموعة محلية من النصوص المرجعية**
- **وثائق التنظيم الداخلي وصفحات الويب**
- **تقارير إخبارية مجمعة.**
- **مجموعة من صفحات ويكيبيديا.**
- **مجموعة فرعية من صفحات الويب**

وفي هذا النظام إجابة عن مجموعة واسعة من أنواع الأسئلة بما في ذلك: الحقيقة، والقائمة، والتعريف، وكيف، ولماذا، والأسئلة الافتراضية، والمقيدة لغويًا، والأسئلة متعددة اللغات.

(١.٢) المجالات

ونظام (QA) يستخدم في عدة مجالات حيوية تخص التعامل مع الجمهور الذي تكون أسئلته مغلقة ومحددة مثل: البنوك أو الطب أو صيانة السيارات أو التغذية.

ويمكن أن يكون هذا النظام أيضا بأسئلة مفتوحة، لكن دقة إجابته ستكون منخفضة وأكثر عمومية ولا يمكن أن تقارن بالأسئلة المغلقة.

مصطلحات البحث[١]

أشجار القرار Decision Trees

شجرة القرار هي أداة لدعم القرار تستخدم نموذجًا شبيهاً للشجرة من القرارات وعواقبها المحتملة ، بما في ذلك نتائج الأحداث المصادفة وتكاليف الموارد والمنفعة.

التعلم الافتراضي propositional learning

الفهم الذي يتضمن الوظائف المعرفية العليا للتجريد والرموز.

التعلم العلائقي relational learning

يشير التعلم العلائقي إلى التعلم في سياق قد توجد فيه علاقات بين أمثلة التعلم ، أو حيث قد يكون لهذه الأمثلة بنية داخلية معقدة.

الترجمة الآلية Machine translation

نظام ذكاء اصطناعي له القدرة على نقل المعارف بين اللغات سواء كانت منطوقة أو مكتوبة.

(١) تم الاعتماد على كتاب معجم البيانات والذكاء الاصطناعي، مجمع الملك سلمان للغة العربية، ٢٠٢٢، ط: ١

التعـرف الآلـي علـى الكـلام (ASR) Automated speech recognition

هـو نظـام للـذكاء الاصـطناعي يـتم فيـه تـدريب الحاسـوب علـى أن يتعـرف آليـا علـى الكلام المنطوق

التعرف الضوئي على الأحرف Optical Character Recognition (OCR)

التحويل الإلكتروني أو الميكانيكي لصور نصوص أو كتابه بخط اليد، سواء من مستند ممسـوح ضـوئيًا أو صـورة لمسـتند أو صـورة مشـهد أو مـن نـص الترجمـة المتراكـب علـى صورة.

التعـرف علـى المشاعـر Emotion recognition

عمليـة التعـرف علـى المشـاعر المضـمنة في نص أو كلام أو صورة

تسمية الكيانات Entity Annotation

عمليـة تحديـد الكيانـات واسـتخراجها ووسـمها فـي الـنص لمسـاعدة الآلـة علـى فهـم الجمل غير المهكيلة.

حوسبـة عاطفيـة Affective computing

مجـال مـن مجـالات الـذكاء الاصـطناعي يهتم بتحليل المشاعر ومعالجتها

خوارزمية Algorithm:

مجموعـة مـن التعليمـات المحـددة لحـل مشكلة أو أداء مهمة معينة

الشبكات العصبية Neural networks

الشبكات العصبية (NNs) ، هـي أنظمـة حسابية مستوحاة قليلاً من الشبكات العصبية البيولوجية التي تشكل أدمغة الحيوانات

نايف بايز Naive Bayes

فـي الإحصـاء، تعتبـر مصـنفات بـايز السـاذجة مجموعـة مـن "المصـنفات الاحتماليـة" البسيطة القائمـة علـى تطبيـق نظريـة بـايز مـع افتراضـات اسـتقلالية قويـة (سـاذجة) بـين السمات

واجهه برمجه التطبيقات Application Programming Interface (API)

هي واجهة تحدد التفاعلات بين تطبيقات برمجيـة متعـددة أو وسـطاء بـرامج الأجهـزة المختلطة

واجهـة محادثـة المسـتخدم conversational user interface (CUI)

واجهـة مسـتخدم المحادثـة هـي واجهـة مستخدم لأجهزة الكمبيوتر التي تحاكي محادثة مع إنسان حقيقي.

المراجع العربية

1- آمنة طالبي، إشكالية حدود الترجمة الآلية، ترجمة نظام سيتران للمتلازمات اللفظية، ماجستير، جامعة قسطنطينة، الجزائر، ٢٠٠٧

2- ابن خلدون، المقدمة، تحقيق: علي عبد الواحد وافي، دار النهضة، ط:٣، عام ١٩٧٩

3- ديسوسير، علم اللغة، ترجمة مالك المطلب، بيت الموصل، ١٩٨٨.

4- علي، نبيل (١٩٨٨)، اللغة العربية والحاسوب، تعريب، القاهرة.

5- علي فهمي، التنقيب في النصوص، ضمن كتاب مشترك بعنوان: المعالجة الآلية للنصوص العربية، ٢٠١٩م، مركز الملك عبد الله لخدمة اللغة العربية.

6- عمايرة، خليل عمايرة، (١٩٨٥م)، في التحليل اللغوي، مكتبة المنارة، ط: ١

7- عطية (محمد عطية) (٢٠١٩م)، معالجة النص العربي المكتوب، ضمن كتاب تطبيقات أساسية في معالجة اللغة العربية، مركز الملك عبد الله لخدمة اللغة العربية، الرياض.

8- عيجولي حسين، ٢٠٢١، المعالجة الآلية للغة، مجلة دراسات وأبحاث المجلة العربية في العلوم الإنسانية والاجتماعية، مجلد ١٣، عدد ابريل ٢٠٢١

9- سلوى حمادة،"الترجمة الآلية كبنية أساسية في صرح التعريب" ،المجلة العربية لعلوم وهندسة الحاسوب ، المجلد الأول، العدد الأول،٢٠٠٧.

10- د. محمود إسماعيل صالح. الحاسوب في خدمة الترجمة و التعريب١٩٩٩م. عن موقع: int.who.emro.www . ص9

11- نوح صبري وأحمد السياط وسعد العنزي، ٢٠١٩، دراسة شاملة لتحليل المشاعر البلاغية في اللغة العربية، المجلة الدولية للتطبيقات الإسلامية في علم الحاسب والتقنية، المجلد: ٧، العدد: ٤، ديسمبر ٢٠١٩م.

المراجع الأجنبية

1- Zughoul, M.G, Miz'il Abu-Alshaar, A. (2005). English/Arabic/English Machine Translation: A Historical Perspective. Meta. L, 3. P 1024.

2- Hutchins, J (2004).The Georgetown-IBM experiment demonstrated in January 1954. Machine translation: from real users to research: 6th conference of the Association for Machine Translation in the Americas, AMTA 2004, Washington, DC, USA. Available from: www.hutchinsweb.me.uk P 01

3- D. M. Bikel, R. Schwartz, and R. M. Weischedel, "An Algorithm that Learns What's in a Name," *Machine Learning*, vol. 34, no. 1, pp. 211–231, Feb. 1999, doi: 10.1023/A:1007558221122

4- A. Borthwick, J. Sterling, E. Agichtein, and R. Grishman, "Description of the MENE Named Entity System as used in MUC-7," 1998.

5- M. Asahara and Y. Matsumoto, "Japanese named entity extraction with redundant morphological analysis," in *Proceedings of the 2003 human language technology conference of the North American chapter of the association for computational linguistics*, 2003, pp. 8–15

6- Hopcroft, J.E., R. Motwani, and J.D. Ullman. 2007. Introduction to Automata Theory, Languages, and

Computation, 3/E. Addison-Wesley: New York, NY.

7- Lori Lamel, Jean-Luc Gauvain, Speech Processing for Audio Indexing. International Conference Natural Language Processing, GoTAL 2008, Advances in Natural Language Processing,

8- Wolff, R. (2021, 1 8). *What Is Natural Language Understanding (NLU) & How Does It Work?* Retrieved from monkeylearn: https://monkeylearn.com/blog/natural-language-understanding/