الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

"قياس أثر توفر موارد البيانات على جودة عمل تقنيات معالجة اللغات الطبيعية: دراسة تحليلية لموارد البيانات باستخدام خدمة Siri كنموذج"

إعداد الباحثتان:

د. هند بادي البادي[1]، رنا عبد الله العبيدي[2]

[1] قسم علم المعلومات، أستاذ مساعد، كلية الآداب والعلوم الإنسانية، جامعة الملك عبد العزيز، جده، المملكة العربية السعودية

[2] قسم علم المعلومات، ماجستير إدارة المعلومات، كلية الآداب والعلوم الإنسانية، جامعة الملك عبد العزيز، جده، المملكة العربية السعودية

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

مخلص:

تساهم تقنية معالجة اللغة الطبيعية في العديد من المجالات مثل الرعاية الصحية وخدمة العملاء، والإدارة المالية، والبحث الأكاديمي وغيرها الكثير. نظرًا للاختلافات في اللغات الموجودة في جميع أنحاء العالم، يمكن أن يختلف استخدام تطبيقات معالجة اللغة الطبيعية في الفعالية اعتمادًا على اللغة المستخدمة. هدفت الدراسة إلى الكشف عن الفروقات التي تنشأ عند استخدام تطبيقات معالجة اللغة الطبيعية بلغات مختلفة، وقياس أسباب هذه الاختلافات باستخدام أداة Siri كنموذج لأحدى تقنيات معالجة اللغات الطبيعية وذلك بالتركيز على ثلاث لغات؛ العربية والإنجليزية "الأمريكية" والإسبانية. اعتمدت الدراسة المنهج الوصفي التحليلي لتحليل عمل تطبيقات معالجة اللغات الطبيعية المتاحة لكل لغة من اللغات المختارة، وذلك من خلال أدوات الملاحظة والاختبار كوسيلة لرصد وتوثيق الملاحظات. بعد إجراء الاختبارات المحددة لـ Siri عن طريق 21 أمر مقسمة إلى ثلاث فئات (أوامر أساسية، أوامر مركبة، أوامر للترفيه) وكل أمر مقسم إلى 7 أسئلة، خلصت الدراسة إلى أن السبب الرئيسي للاختلافات أو الضعف في جودة النتيجة يعود إلى التفاوت بين توافر وجودة مصادر البيانات باللغة العربية مقارنة بالإنجليزية والإسبانية. في حين مقدرة اللغتين الإنجليزية والإسبانية في الوصول إلى العديد من مصادر البيانات بشكل أفضل. بغض النظر عن عدد الناطقين الهائل باللغة يعد توفر الموارد المصنفة والمقننة لغوياً وبلغة يمكن أن تفهمها الآلة أمراً آخراً من شأنه أن يؤثر على جودة الاستجابة. توصي الدراسة أنه يجب إيلاء الاهتمام إلى تعزيز مصطلحات اللغة العربية من قبل المتحدثين الأصليين واستخدامها لإنشاء مصادر بيانات من شأنها تحسين جودة تطبيقات معالجة اللغة الطبيعية لتحقيق أقصى استفادة من التقنيات المتاحة، والحصول على مخرجات عالية الجودة.

الكلمات الدالة: معالجة اللغات الطبيعية، الذكاء الاصطناعي، موارد البيانات، Siri، المساعدات الرقمية الصوتية

مقدمة البحث:

منذ بداية الثورة الصناعية الثالثة واستمراراً للثورة الصناعية الرابعة التي تسمى بالثورة الرقمية ظهرت العديد من الأدوات والتقنيات الناشئة في العديد من المجالات التي ساهمت في إيصال المليارات من الأشخاص إلى الويب والمساعدة في إدارة الأصول وتجديد البيئة الطبيعية وطريقة عملها، والرفع من كفاءة الأعمال بالمنظمات بشكل جذري، شملت التطورات العديد من التقنيات الناشئة وتعددت مجالاتها بما في ذلك إنترنت الأشياء، وتكنولوجيا النانو، والطباعة ثلاثية الأبعاد، والذكاء الاصطناعي، والروبوتات، والمركبات المستقلة والعديد غيرها. (Schwab, 2016). يعرف الذكاء الاصطناعي بأنه محاكاة للقدرات الذهنية البشرية وأنماط عملها وتتعدد مجالاته وأنواعه، ذكرت دراسة (الموجود وغريب، 2022) أن تطبيقات الذكاء الاصطناعي تتمثل في 7 أقسام أساسية وهي: تحليل النظم وتصميمها، والاختصاص في قواعد البيانات، والنظم الخبيرة، وهندسة الحاسوب، والعمل في مجال التعليم، ومعالجة اللغات الطبيعية، وأعمال الحاسوب والروبوتات. تعد تقنية معالجة اللغات الطبيعية أحد أهم فروع الذكاء الاصطناعي وأكبر مجالاته. فهو مجال يربط بين العديد من العلوم كعلم النفس والرياضيات واللسانيات وعلم الحاسوب وهو علم يجمع بين النظرية والتطبيقية ويمكن الاستفادة من تطبيقاته في دراسة ونمذجة أسلوب اكتساب البشر للمهارات.

تتعدد استخدامات وتقنيات معالجة اللغات من خلال تطبيقات الذكاء اصطناعي كالتعرف الآلي على الكلام، والترجمة الآلية الفورية، وChatbots (روبوت الدردشة) أو المساعدات الآلية مثل "اليكسا من أمازون" و"Siri من Apple" و"Cortana من مايكروسوفت". عند التعامل والتحدث مع هذه التطبيقات فإن تقنية معالجة اللغات الطبيعية NLP هي التي تسمح لهذه المُساعِدات بفهم الكلام الذي يقال من قِبل المستخدم، تعمل تقنية NLP مع الكلام المنطوق أو المكتوب ويمكن تطبيقها على جميع اللغات البشرية. تتعدد المهمات التي نقوم بها في حياتنا اليومية باستخدام تقنية معالجة اللغات الطبيعية دون الشعور بأننا نستخدم هذه التقنية، وتدعم تطبيقاتها العديد من المجالات مثل عملية تصفية البريد الإلكتروني العشوائي، الترجمة الفورية للكلام والنصوص، تحليل المشاعر، البحث على الإنترنت، التدقيق

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

الإملائي والنحوي، تلخيص المستندات. كما تقوم بعض برامج البريد الإلكتروني بالاستفادة من معالجة اللغات الطبيعية في قراءة الرسائل ومن ثم تحليلها وتقديم اقتراحات للرد المناسب للاستجابة على الرسالة. تناقش هذه الدراسة مفهوم تقنية معالجة اللغات الطبيعية واختلاف جودة نتائجها من خلال دراسة الموارد المتوفرة التي تستقي بياناتها منها وتعمل بواسطة استخدامها. (Oracle, 2023)

مشكلة الدراسة:

نبعت مشكلة الدراسة من خلال ملاحظة وجود اختلاف في جودة النتائج التي تولدها تطبيقات الذكاء الاصطناعي باختلاف أنواعها بحسب اللغة التي يتم استخدامها، ومع تكرار التجارب في عدة تطبيقات مثل تطبيقات الترجمة الفورية وأداة "Siri" و"ChatGPT" ظهرت فروقات واضحة في كمية وجودة النتائج. في تجربة قام بها اليافعي (2020) بالبحث في محرك Google باللغتين العربية والإنجليزية عن مصطلح معالجة اللغات الطبيعية NLP وهي أحد أكثر مجالات الذكاء الاصطناعي شيوعاً، ظهرت النتائج باللغة العربية بعدد 20,500 نتيجة و1,120,000 نتيجة باللغة الإنجليزية، وعند تكرار التجربة في شهر إبريل من عام 2023 أثناء إجراء هذه الدراسة ظهرت عدد النتائج باللغة العربية 611,000 نتيجة و34,100,000 نتيجة باللغة الإنجليزية، مما يعني وجود فارق بنسبة 56% في كمية النتائج بين اللغتين والتي كانت بنسبة 33% في عام 2020. مما يبرز لنا وجود مشكلتين أساسيتين وهي وجود فروقات واضحة في اختلاف عدد النتائج التي تظهر عند تغيير اللغة المستخدمة للبحث، بالإضافة إلى بروز مشكلة استمرار الفجوة الحاصلة وزيادتها مما يعني استمرار وجود أساس المشكلة. أدى هذا إلى تمحور مشكلة الدراسة في معرفة أسباب اختلاف جودة عمل تطبيقات معالجة اللغات الطبيعة عند اختلاف اللغة المستخدمة.

تساؤلات الدراسة:

1. من أين تستمد تطبيقات معالجة اللغات الطبيعية مواردها وماهي طبيعة عملها؟
2. ماهي العلاقة بين توفر موارد البيانات وجودة عمل تطبيقات معالجة اللغات الطبيعية؟
3. ماهي الفروقات التي تظهر في استجابة تطبيقات معالجة اللغات الطبيعية عند تغيير اللغة المستخدمة؟

أهداف الدراسة:

تسعى الدراسة الحالية إلى تحقيق الأهداف الآتية:

1. التعرف على طبيعة عمل تطبيقات معالجة اللغات الطبيعية والموارد التي تستمد منها البيانات.
2. التعرف على موارد البيانات باختلاف لغاتها ومدى توفرها.
3. قياس العلاقة بين جودة عمل تطبيقات معالجة اللغة الطبيعية وتوفر موارد البيانات.
4. تسليط الضوء على خدمة المساعد الذكي "Siri" من شركة Apple وقياس جودة النتائج باختلاف اللغة المستخدمة.

أهمية الدراسة:

تنبعث أهمية الدراسة الحالية من أهمية الاستفادة من جميع التقنيات المتاحة بجميع إمكانيتها الأساسية التي صُممت بها، وأحد هذه التقنيات هي الذكاء الاصطناعي الذي أصبح أحد أبرز الثورات التقنية في العالم الحديث والذي لا مفر من التعامل معه. لذا توجب علينا دراسة طبيعة عمله وفهم الأساليب والطرق التي يعمل الذكاء الاصطناعي من خلالها لمحاولة الوصول إلى أقصى استفادة منه. تعد تقنية معالجة اللغات الطبيعية أحد أكبر مجالات الذكاء الاصطناعي انتشاراً وتستخدم في العديد من الصناعات حيث تستخدمها العديد من المؤسسات في تحليل بياناتها بشكل دقيق وسريع لمساعدتها في اتخاذ القرارات الصحيحة. يمكن الاستفادة من استخدام تقنية معالجة اللغات الطبيعية في العديد من المجالات مثل مجال الرعاية الصحية، خدمة العملاء، الإدارة المالية، البحث الأكاديمي وغيرها من المجالات المهمة. (Oracle,2023) تعد موارد البيانات أو ما تسمى بتمثيل المعرفة الأساس التي تعتمد عليه تقنيات الذكاء الاصطناعي

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

لتقوم بتحليلها والتدرب عليها وإيجاد العلاقات بينها ومن ثم استخدامها وتخزين الإضافات المكتسبة من الإنسان عليها، حيث يعتمد نجاح البرنامج بشكل كبير عليها. (عبدالقادر، 2022)، تهدف هذه الدراسة إلى إلقاء الضوء على أهمية توفر موارد البيانات بشكل كافي ودورها في رفع مستوى جودة عمل تطبيقات الذكاء الاصطناعي، كما تفتح هذه الدراسة الآفاق للدراسات المستقبلية للمهتمين بمجال اللغات والحاسوب للبحث أكثر ودراسة موضوع توفر موارد البيانات على شبكة الإنترنت والاختلافات النسبية التي تظهر بين اللغات.

حدود الدراسة:

الحدود الموضوعية:

تنحصر حدود الدراسة الموضوعية في دراسة وجود علاقة بين توفر موارد كافية للبيانات مع جودة عمل تطبيقات معالجة اللغات الطبيعية، تقيس الدراسة موارد اللغات المتوفرة باللغة العربية والإنجليزية والإسبانية من خلال خدمة Siri كنموذج لمعالجة اللغات الطبيعية الذي يعد أول خدمة مساعد رقمي تم إطلاقها في شهر أكتوبر من عام 2011 ثم تلاه ظهور العديد من المساعدات الرقمية بعد ذلك، وبحسب دراسة (أحمد،2021) يعد Siri أحد أشهر ثلاث مساعدات رقمية شيوعاً برفقة "Google Now" و "Google Assistant"

الحدود الزمنية:

تختص الحدود الزمنية بموارد البيانات المتوفرة في وقت إجراء الدراسة من شهر إبريل إلى شهر يوليو من عام 2023.

مصطلحات الدراسة:

الذكاء الاصطناعي (AI):

عرفه (جبريل وآخرون، 2020) بأنه: "أحد فروع علم الحاسب الآلي التي تختص بتصميم البرمجيات التي تستطيع محاكاة القدرات العقلية للإنسان، وأنماط عملها هي القدرة على التعلم، وحل المشكلات المعقدة، والتخطيط، والاستنتاج، واتخاذ القرارات، والإدراك الحسي والتواصل، وتساعد هذه القدرات الآلات الذكية على القيام بمهام جديدة لم يبرمج عليها بدون تدخل العامل البشري".

كما يمكن تعريفه بأنه الأنظمة والأجهزة التي تُحاكي الذكاء البشري في أداء المهام والتي تقوم بالرفع من مستوى نفسها بالاعتماد على المعلومات التي تجمعها، وترتبط بالقدرة على تحليل البيانات والتفكير الهائل أكثر من قدرتها على القيام بوظيفة معينة بشكل معين. (الدلاهمة، 2019)

برنامج حاسوبي يستخدم المعرفة البشرية لتحقيق مستوى عالٍ من الإنجاز في مجال معين. (Waterman, 1986)

وعرفها قاموس كامبردج بأنها عملية دراسة كيفية إنتاج آلات لديها بعض الصفات التي يمتلكها العقل البشري، مثل القدرة على فهم اللغة والتعرف على الصور وحل المشكلات والتعلم. (Cambridge Dictionary, 2023)

وعليه تعرفه الدراسة إجرائياً بأنه "عملية تكرار منظمة تقوم بها الآلة بطريقة مميزة تُحاكي بها الذكاء البشري، وتملك القدرة على التطور والتعلم باستمرار وصقل نفسها بناءً على التجارب والظروف التي تمر بها دون الحاجة الى التدخل البشري"

اللغات الطبيعية:

أي لغة تطورت بشكل طبيعي من خلال الاستخدام، على عكس لغة الحاسوب. (الهيئة الاتحادية للموارد البشرية الحكومية، 2023)

ويمكن تعريفها: هي اللغات البشرية كالعربية والإنجليزية والإسبانية. (شرواني، 2021)

وعليه تعرفها الدراسة إجرائياً بأنها: اللغات الذي يستخدمها البشر بشكل طبيعي أثناء حواراتهم مثل اللغة العربية والإنجليزية وغيرها، وبالمقابل لها تأتي لغة الآلة أو لغة الحاسوب.

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

معالجة اللغات الطبيعية:

حقل فرعي من الذكاء الاصطناعي وعلم اللغويات وعلوم الكمبيوتر الذي يدرس تحليل ومعالجة بيانات اللغة الطبيعية. (الهيئة الاتحادية للموارد البشرية الحكومية، 2023)

هي فرع أو أحد مجالات الذكاء الاصطناعي، وتشير إلى مجال اللغات وعلوم الحاسوب، المعنية بالتفاعلات بين اللغات الطبيعية والحاسوب. (الموجود وغريب، 2022)

كما تعرف بأنها مجال في علوم الكمبيوتر والذكاء الاصطناعي واللغويات الحاسوبية المعنية بالتفاعلات بين اللغات البشرية وأجهزة الكمبيوتر. كما ترتبط بمجال التفاعل بين الإنسان والحاسوب. (Hamandi et al., 2016)

كما عرفها قاموس أكسفورد بأنها مجال علوم الكمبيوتر الذي يطور الأنظمة التي تنفذ فهم اللغة الطبيعية. إنه تخصص فرعي للذكاء الاصطناعي واللغويات الحاسوبية. (Oxford Reference,2023)

وعليه تعرفها الدراسة إجرائياً بأنها: عملية تحويل أو معالجة الكلام المكتوب أو المنطوق إلى بيانات يمكن فهمها من قبل الآلة.

موارد البيانات:

هي البيانات التي تتمثل من خلال جمع وقياس المعلومات. وتكون عادة على هيئة نص أو أرقام أو صور، ويمكن عرضها وتنظيمها بسهولة من خلال قواعد البيانات المنظمة (Structured) أو غير المنظمة (Unstructured) مثل الصور والنصوص ومقاطع الفيديو. (البرنامج الوطني للذكاء الاصطناعي،2020)

كما عرفتها (الهيئة الاتحادية للموارد البشرية الحكومية، 2023) بأنها: البيانات المستخدمة لتطوير تطبيقات الذكاء الاصطناعي، تستند عليها تطبيقات الذكاء الاصطناعي، والذي يتم تدريبه عليها، ويتعلم منها.

كما يمكن تعريفها بأنها: نوع خاص من البيانات اللغوية، وتتألف من آلاف الصفحات من النصوص أو النصوص من لغة محددة والتي يمكننا البحث عنها بواسطة الكمبيوتر، وهي مجموعة من النصوص الأصلية المقروءة والتي تم أخذ عينات منها لتكون ممثلة للغة أو مجموعة متنوعة من اللغات. (Brezina & Gablasova, 2018)

وعليه تعرفها الدراسة إجرائياً بأنها: البيانات التي يعتمد على الذكاء الاصطناعي في عملية التدريب وتوليد الحلول وتتوفر بعدة أشكال منها النصوص والصور والأرقام، والتي تعتمد جودة النتائج عليها.

المساعدات الرقمية الصوتية (DVAs):

عرفها (أحمد،2021) بأنها "برنامجاً متطوراً يُحاكي المحادثات التي تقوم بها المستخدمون، وتستخدم تقنيات الذكاء الاصطناعي المتقدم (AI)، ومعالجة وفهم اللغة الطبيعية، والتعلم الآلي وذلك لكي يتمكن من توفير بفعالية تجربة محادثة للمستخدم قائمة على التخصيص والشخصنة. ويمكن الإجابة على الأسئلة المعقدة وتقديم التوصيات ووضع التنبؤات وحتى بدء المحادثات من خلال جمع المعلومات من الطلبات، والأبحاث السابقة، والتفضيلات، وغيرها".

كما تعرف المساعدات الرقمية الصوتية بأنها تطبيقات برمجية تعتمد على الذكاء الاصطناعي (AI)، والتي تتواصل مع الناس من خلال اللغة الطبيعية المنطوقة، وقد تكون مدمجة في هاتف ذكي (على سبيل المثال DVA Siri من Apple أو مساعد Google). بالإضافة إلى ذلك، تتوفر DVAs أيضًا في شكل مكبرات صوت ذكية (مثل Amazon's Alexa). يمكن تنفيذ المهام، مثل إجراء المكالمات، أو إرسال الرسائل، تلقي التنبيهات أو فتح تطبيق محدد بواسطة DVA باستخدام التحكم الصوتي دون أي تدخل يدوي. (Ewers et al., 2020)

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

ويمكن تعريفها بأنها: مُساعِدات شخصية تمثل خدمة مساعد آلي ذكي يقوم بمساعدة المستخدم في مهام مختلفة من خلال فهم لغته الطبيعية ومن ثم إجابته على تساؤلاته مثل إعطاءه معلومات لتوقعات الطقس، نتائج المباريات، إنشاء مواعيد للاجتماعات وغيرها من المهام اليومية. (أحمد، 2021)

وعليه تعرفها الدراسة إجرائياً بأنها: خدمة مساعد شخصي تعمل من خلال الأوامر والحوارات الصوتية، تأتي بشكل مدمج مع الأجهزة اللوحية الذكية، أو بشكل منفصل على هيئة مكبرات صوتية، تقوم بالعديد من المهام كتعيين التنبيهات والتحكم بالأجهزة الذكية بالمنزل وفتح التطبيقات وغيرها.

خدمة Siri:

هي خدمة مساعد شخصي رقمي مقدمة من شركة Apple، قدم في عام 2011 في معرض سياتل الدولي كأول مساعد رقمي على الإطلاق متاح على الهواتف الذكية، يقوم بالاستجابة لطلبات المستخدم بواسطة الأوامر الصوتية وإنجاز العديد من المهام مثل قراءة النصوص، وتحويل القياسات وحل المسائل الرياضية، ونشر التحديثات اليومية على تطبيقات التواصل الاجتماعي مثل Facebook وTwitter، وتحرير الملاحظات وإجراء المكالمات الهاتفية الصوتية والفيديو والكثير من المهام غيرها. (أحمد، 2021)

كما يمكن تعريف Siri بأنه مساعد شخصي من شركة Apple لأجهزة iOS وmacOS وtvOS وwatchOS التي تعمل على التعرف على الصوت وهي مدعومة بالذكاء الاصطناعي (AI). يجيب Siri على أسئلة المستخدمين من خلال التحدث عبر مكبر صوت الجهاز وتقديم المعلومات ذات الصلة من تطبيقات معينة مثل بحث الويب أو التقويم على الشاشة الرئيسية. (Thakare et al., 2020)

وعليه تعرفها الدراسة إجرائياً بأنها: خدمة مساعد رقمي مقدمة من شركة Apple للاستجابة لطلبات المستخدم بواسطة الأوامر الصوتية، تقوم بالعديد من مهمات المساعدة الشخصية مثل إرسال الرسائل، البحث على الإنترنت وضبط المنبه وتحرير الملاحظات واستعراض حالة الطقس.

لماذا Siri؟

وسط التسارع التكنولوجي الذي نشهده اليوم، ظهر Siri، وهو فرع من فروع الذكاء اصطناعي الثوري والذي يمكنه فهم الكلام البشري والاستجابة له. لقد أحدث اختراع Siri من شركة .Apple Inc تحولاً في عالم التكنولوجيا، مما أتاح التواصل بين البشر والآلات بشكل لم يسبق له مثيل. يستخدم Siri تقنية معالجة اللغة الطبيعية (NLP) لفهم تعقيدات اللغة البشرية والتي تعتمد على البرمجة اللغوية العصبية لفهم القصد من الكلمات التي يتحدث بها المستخدم ويولد ردودًا مناسبة وفقًا لذلك. تكمن أهمية Siri في قدرته على تفسير الطلبات المعقدة والمشاركة في محادثة مع المستخدمين والتعلم من تفضيلاتهم وأنماط سلوكهم من خلال تحليل اللغة المنطوقة واستخدام الخوارزميات المعقدة، وإعطاء أوامر صوتية لإكمال المهام المختلفة التي تتراوح من إعداد المواعيد إلى البحث عن معلومات على الإنترنت، وإجراء المكالمات، وإجراء العمليات الحسابية. كما تساعد Siri في كتابة الرسائل النصية ورسائل البريد الإلكتروني وإرسالها، إجراء المكالمات الهاتفية، وتشغيل الموسيقى، والتحقق من الطقس، وإنشاء المواعيد، والبحث عن المعلومات على الإنترنت، وتوفير الاتجاهات، وحتى إجراء الحجوزات في المطاعم. بالإضافة إلى ذلك، يمكن لـ Siri الاندماج مع التطبيقات والخدمات الأخرى، مثل منصات الوسائط الاجتماعية وتطبيقات الطرف الثالث، مما يوسع نطاق قدراته وفائدته للمستخدمين. تعمل Siri بمجموعة متنوعة من اللغات واللهجات، مما يعزز جاذبيتها العالمية. تتميز Siri بقراءة الأوامر التي تم إملاؤها مرة أخرى قبل إرسالها للتأكد من خلو الرسائل من الأخطاء. أحدثت تقنية معالجة اللغة الطبيعية في Siri ثورة في طريقة تفاعل البشر مع الآلات، حيث دفع نجاح Siri الشركات الأخرى إلى تطوير مساعدين شخصيين مشابهين يتم تنشيطهم بالصوت. ومع تقدم التكنولوجيا، ستستمر إمكانات معالجة اللغات الطبيعية في التوسع، مما يحسن من تجربتنا في استخدام التقنيات ويجعلها أكثر سهولة

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

(Abdolrahmani, Kuber, & Branham, 2018, Reis, Paulino, Paredes, & Barroso, 2017)

الإطار النظري والدراسات السابقة:

الدراسات العربية:

تناقش دراسة (ناجي، 2022) الدور الكبير والتطور لتطبيقات الذكاء الاصطناعي والتي من أهمها معالجة اللغات الطبيعية ودورها الكبير في تحليل المحتوى والعديد من التحسينات التي تحتويها، هدفت هذه الدراسة لتحديد دور أنظمة الذكاء الاصطناعي في تحليل المحتوى والتكشيف واستخدام تقنية معالجة اللغات الطبيعية بشكل عام وفي المكتبات بشكل خاص، ومراقبة المنصات ومكتبات الذكاء الاصطناعية التي يمكن استخدامها لتحليل المحتوى. خلصت الدراسة إلى إمكانية استخدام معالجة اللغات الطبيعية في المكتبات للوصول إلى المصادر، مراقبة المحتوى، الإحاطة، والإجابة على الأسئلة. كما Hاثبتت الأدوات التي تم استخدامها كفاءة عالية في عملية البحث والاسترجاع للمصادر باللغة الإنجليزية فيما وجد بعض الضعف في النتائج باللغة العربية وأتضح حاجتها لبعض التحسينات. تتفق هذه الدراسة مع الدراسة الحالية في قيامها بالمقارنة بين منصات معالجة اللغات الطبيعية المتاحة بأكثر من لغة وقياس جودتها في تحليل المحتوى، فيما تختلف في استخدام أداة تختص في تحليل المحتوى والتكشيف في المكتبات فيما تناقش هذه الدراسة جودة النتائج باختلاف اللغة لأحد تطبيقات معالجة اللغات الطبيعية التي تعمل كمساعد رقمي شخصي.

تلقي دراسة (احمد،2021) الضوء على عملية استخراج البيانات الببليوجرافية من المصادر البحثية من كتب ومقالات علمية، هدفت الدراسة إلى توضيح ماهية الذكاء الاصطناعي وتطبيقاته ثم تحديد ماهية الفهرسة الوصفية ودور الناشر في عملية إنشاء التسجيلات الببليوجرافية، والاستفادة من تقنيات التعلم الآلي في عملية استخراج البيانات الببليوجرافية من مصادر المعلومات النصية. أعتمدت الدراسة على المنهج الوصفي التحليلي للتعرف على ماهية الذكاء الاصطناعي بشكل عام وتطبيقات التعلم الآلي والتعلم العميق بشكل خاص ودورها في استخراج البيانات الببليوجرافية من المصادر النصية، كما قدمت الدراسة نموذج مقترح لاستخراج البيانات الببليوجرافية من المصادر المكتوبة باللغة العربية. خلصت الدراسة لعدة نتائج أبرزها ضرورة استخدام التقنيات لعملية الفهرسة الآلية وإحلالها بدلاً من عملية تكرار الفهرسة بشكل روتيني. كما أوضحت الدراسة أن استخدام النماذج الآلية لا يغني عن دور المفهرس، ولكن يحدد دوره بالمهام التي تحتاج إلى إبداع وأن يساهم بخبرته في ابتكار نماذج جديدة من شأنها تطوير طريقة العمل الحالية. تتفق هذه الدراسة مع الدراسة الحالية في الاستفادة من تقنيات الذكاء الاصطناعي، فتسلط الدراسة الضوء على تقنيات التعلم العميق ومعالجة اللغات الطبيعية للنصوص المكتوبة في المصادر ومن ثم استخراج البيانات الببليوجرافية، ثم قدمت الدراسة نماذج لاستخراج البيانات الببليوجرافية من المصادر بعدة لغات، كما قدمت نموذج مقترح للمصادر النصية باللغة العربية، فيما تناقش الدراسة الحالية أسباب اختلاف النتائج للمساعدات الرقمية الصوتية التي تعمل باستخدام تقنية معالجة اللغات الطبيعية عند اختلاف اللغة المستخدمة.

ناقشت (شرواني، 2021) تجربة لغة المعالجة الطبيعية في المملكة العربية السعودية من خلال تقييم نموذج لروبوتات المحادثة. تناولت الدراسة دور المعالجة اللغوية للغات الطبيعية في الذكاء الاصطناعي بواسطة استعراض تجربة السعودية في هذا المجال والتحديات التي واجهتها في المعالجة اللغوية لروبوتات المحادثات باللغة العربية. قيمت الدراسة تجربة الروبوتات بواسطة إعداد استبانة تقويمية قاست مدى نجاح تقنية روبوتات المحادثة التي تقوم بها المشاريع الجامعية في المملكة وفعالية اعتماد اللغة العربية الفصحى بجانب اللهجة السعودية العامية. كشفت الدراسة إلى عدت نتائج منها تقبل ووعي المجتمع لفكرة تواجد روبوتات المحادثة واستخدامها في التعليم، كما أوضحت النتائج عدم الوصول للثقة الكلية والرضا الكامل عن فاعلية التواصل بطريقة المحادثة الآلية وأتت النسبة الأكبر باللذين لم يستخدموا منصات المحادثة الآلية في التعليم مع التأكيد على فائدتها بشكل كبير في تعزيز العلاقة بين الإدارة والطالب واختصارها للوقت والجهد وسرعة تحديث المعلومات. تتفق هذه الدراسة مع الدراسة الحالية في استعراض إحدى تقنيات الذكاء الاصطناعي وهي تقنية معالجة اللغات الطبيعية، ناقشت الدراسة

تطبيق هذه التقنية على روبوتات المحادثة واستعرضت التجربة السعودية لروبوتات المحادثة باللغة العربية، فيما تناقش الدراسة الحالية تطبيق تقنية معالجة اللغات الطبيعية في تطبيقات المساعدة الشخصية الرقمية ومقارنة مدى اختلاف النتائج عند تغيير اللغة المستخدمة.

الدراسات الأجنبية:

تناقش دراسة (Antoun et al., 2020) أنه وبالرغم من مدى الغنى اللغوي الذي تحتويه اللغة العربية إلا أنها تملك موارد قليلة نسبياً وبناء جملة أقل استكشافا مقارنة باللغة الإنجليزية. بالنظر إلى هذه الميزة إلا أنه من ناحية معالجة اللغة العربية الطبيعية (NLP) مثل تحليل المشاعر (SA)، التعرف على الكيانات المسماة (NER)، والإجابة على الأسئلة (QA)، أثبتت اللغة العربية العكس وأنها صعبة للغاية في معالجتها. مؤخراً ومع زيادة نماذج المحولات الجاهزة، أثبتت النماذج القائمة على نموذج BERT الخاصة باللغة فعالية للغاية في الفهم، بشرط أن تكون مدربة مسبقا على مجموعة كبيرة جداً من النماذج. كانت هذه النماذج قادرة على وضع معايير جديدة وتحقيق نتائج حديثة لمعظم مهام البرمجة اللغوية العصبية. قامت هذه الدراسة بتدريب BERT مسبقاً خصيصا للغة العربية سعيا لتحقيق نفس النجاح الذي حققه BERT للغة الإنجليزية. عندما تم مقارنة أداء AraBERT بـ BERT متعدد اللغات من Google وغيرها من الأساليب الحديثة أتت النتائج أن AraBERT المطور حديثا حقق أداء متطورا في معظم مهام البرمجة اللغوية العصبية العربية التي تم اختبارها على أمل أن يتم مستقبلاً تشجيع البحث والتطبيقات للبرمجة اللغوية العصبية العربية. تتفق هذه الدراسة مع الدراسة الحالية في المقارنة بين لغتين بنفس الأداة وملاحظة النتائج بناء على الموارد المتوفرة، تقوم الدراسة بالمقارنة بين نموذج BERT باللغة العربية والإنجليزية، فيما تقارن الدراسة الحالية بين النتائج باللغة العربية والإنجليزية والإسبانية باستخدام تقنية معالجة اللغات الطبيعية المستخدمة في المساعد الشخصي Siri من شركة Apple.

تناقش دراسة (Zhong et al., 2020) أحد الأمثلة لمجالات استخدامات الذكاء الاصطناعي وهو الذكاء الاصطناعي القانوني والذي يرمز له (LegalAI) والتركيز على تطبيق تكنولوجيا الذكاء الاصطناعي وخاصة معالجة اللغة الطبيعية للإفادة في انجاز المهام في المجال القانوني. في السنوات الأخيرة، لفتت LegalAI اهتمامًا متزايدًا بسرعة الباحثين في مجال الذكاء الاصطناعي والمهنيين القانونيين، حيث يفيد LegalAI النظام القانوني لتحرير المهنيين القانونيين من متاهة الأعمال الورقية. غالبًا ما يفكر المحترفون القانونيون في كيفية حل المهام من الأساليب المستندة إلى القواعد والقائمة على الرموز، بينما يركز باحثين البرمجة اللغوية العصبية أكثر على الأساليب القائمة على البيانات والتضمين. تصف هذه الدراسة التاريخ والحالة الحالية والتوجهات المستقبلية للبحث في LegalAI. كما توضح المهام من وجهة نظر المهنيين القانونيين والباحثين في البرمجة اللغوية العصبية لعرض العديد من التطبيقات التمثيلية في LegalAI. تجري الدراسة تجارب وتقدم تحليلًا متعمقًا لمزايا وعيوب الأعمال الحالية لاستكشاف الاتجاهات المستقبلية المحتملة. تتفق هذه الدراسة مع الدراسة الحالية في مناقشة إحدى تطبيقات الذكاء الاصطناعي وهي معالجة اللغة الطبيعية، تختص هذه الدراسة في الجانب القانوني ومدى استفادته من تقنيات الذكاء الاصطناعي ووضعه الحالي والفرص المستقبلية الممكنة، فيما تناقش الدراسة الحالية محاولة معرفة أسباب اختلاف نتائج المساعدات الصوتية الرقمية التي تعمل باستخدام تقنية معالجة اللغات الطبيعية.

تقترح دراسة (Wang et al., 2020) خوارزمية متعددة الوسائط المشتركة تعتمد على الشبكة العصبية العميقة لتقوم بتقسيم المعلومات النصية بشكل صحيح وهو عامل مهم يؤثر على أداء نظام تجزئة الكلمات. عند معالجة معلومات النصوص باللغة الإنجليزية تكون الكلمات هي الوحدة الأساسية في حين أنه يجب أن يستخدم محتوى البحث من حيث الدلالات والقواعد لا من حيث الكلمات باعتبارها أصغر وحدة. تعد الكلمات هي المشكلة المركزية لاستخراج المعلومات النصية، ولكن بمساعدة جهاز كمبيوتر يتم تقسيم الكلمات الموجودة في محتوى النص واحدة تلو الأخرى لتحليل محتوى النص. NLP (معالجة اللغة الطبيعية) هي تقنية تمكن أجهزة الكمبيوتر من فهم اللغات البشرية بعد تحليلها. عادة ما يكون تجزئة الكلمات هو المهمة الأساسية للبرمجة اللغوية العصبية. من أجل حل المشكلة العملية المتمثلة في الاختلافات الهيكلية الضخمة بين طرائق البيانات المختلفة في بيئة متعددة الوسائط ومع عدم إمكانية تطبيق طرق التعلم الآلي التقليدية بشكل مباشر، تقدم هذه

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

الدراسة استخراج الميزات طريقة التعلم العميق وتطبيق أفكار التعلم العميق على استخراج ميزة متعددة الوسائط. من أجل التحقق من فعالية النموذج المقترح، تم إجراء سلسلة من تجارب المقارنة. أظهرت نتائج التجربة أن نموذج استخراج ميزة الاندماج متعدد الوسائط المقترح يمكنه استخراج ميزات الاندماج منخفضة الأبعاد بشكل فعال من البيانات الأصلية متعددة الأبعاد عالية الأبعاد. يتميز تعبير الاندماج التي تم الحصول عليها بقدرة قوية بينما يمتلك بعداً أقل، مما يثبت صحة النموذج المقترح من حيث تجزئة الكلمات الإنجليزية. بعد التحليل والبحث يمكن لكلتا الشبكتين المقترحة حل مشكلة تجزئة الكلمات التقليدية في علاقة التبعية طويلة المدى للنص. ونظراً لتعقيد هيكلها فإنها تستهلك الكثير من الوقت في عملية التدريب والتنبؤ بمجموعة البيانات. أظهرت النتائج التجريبية أن النموذج المقترح في هذه الدراسة أفضل من معظم النماذج السابقة من حيث الدقة، فإن الطريقة المقترحة في هذا البحث أسرع 1.62 مرة من طريقة تجزئة الكلمات في سرعة التدريب. يمكن مستقبلاً النظر في تحليل تأثير طرق استخراج الميزات المختلفة واختيار الميزات على النموذج، وبالتالي زيادة تعزيز قدرة التعلم للنموذج. تتعامل الطريقة المقترحة مع الميزات المختلفة التي تم الحصول عليها بواسطة طرق خارجية مختلفة في كل نوع من البيانات الأولية كوضع مستقل ولا يتعلم مباشرة على البيانات الأولية. تتفق هذه الدراسة مع الدراسة الحالية في قيامها بالتجربة والمقارنة لنتائج تحليل النصوص باللغة الإنجليزية، بينما تقوم الدراسة الحالية بتجربة خدمة المساعد الرقمي Siri المقدمة من شركة Apple بعدة لغات ومقارنة الفروقات بين النتائج. ناقشت دراسة (Hoy, 2018) أنه مع النمو الكبير والتعقيد في تقنية التعرف على الصوت وبرامج المساعدات الصوتية في السنوات القليلة الماضية، ومع إتاحة منتجات المساعد الصوتي المتوفرة حاليًا من Apple وAmazon وGoogle وMicrosoft للمستخدمين طرح الأسئلة وإصدار الأوامر لأجهزة الكمبيوتر بلغة طبيعية. ووجود العديد من الاستخدامات المستقبلية المحتملة لهذه التكنولوجيا، من التشغيل الآلي للمنزل، الترجمة، الرفقة ودعم كبار السن، وغيرها، ولكن يظل هناك العديد من المشكلات التي تتعلق بمنتجات المساعد الصوتي المتوفرة حاليًا. يجب تحسين ضوابط الخصوصية والأمان قبل أن يتم استخدام المساعدين الصوتيين لأي شيء يتطلب السرية. تتفق هذه الدراسة مع الدراسة الحالية في مناقشة بعض تطبيقات معالجة اللغات الطبيعية التي تستخدمها المساعدات الرقمية الصوتية وتركز بشكل خاص على الجانب الأمني وجانب السرية للمعلومات، فيما تركز الدراسة الحالية على مستوى الفروقات بالنتائج التي تظهر عند تغيير اللغة المستخدمة.

منهجية الدراسة:

اعتمدت الدراسة على المنهج الوصفي التحليلي، وذلك لدراسة ماهية الذكاء الاصطناعي وبالأخص تطبيقات تقنيات معالجة اللغات الطبيعية وتحليل طبيعة عملها واستخراجها لموارد للبيانات الأساسية التي تتكئ وتعتمد عليها وتقوم بتوليد البيانات منها، تم اختيار المنهج الوصفي التحليلي بسبب حاجة موضوع الدراسة إلى تغطية جانب الموارد الرئيسية باختلاف اللغات ودراستها ومن ثم تحليل أسباب اختلاف النتائج عند استخدامها بواسطة تطبيقات معالجة اللغات الطبيعية. اعتمدت الدراسة على استعراض عدة تقارير تقوم بتقييم عمل المساعدات الرقمية الصوتية من خلال طرح عدة أوامر وأسئلة باللغة الطبيعية، يتنوع مستوى الأوامر بين الأوامر المباشرة والبسيطة التي تقيس الفهم الأساسي مثل طلب تعيين تذكير أو البحث عن طلب شراء منتج محدد، وتتدرج لتصل لبعض الأوامر المركبة التي يجب فهمها من طريقة سياق الأمر مثل السؤال عن أسرع طريق للوصول للمطار عن طريق الحافلة ثم الاستطراد وتحويل الأمر لقياس المسافة بالسيارة. يتنوع أسلوب الأوامر كذلك لتطرح بعضها للتسلية مثل طلب إعداد كوب من القهوة، وبعض الأوامر العاطفية مثل السؤال عن الحب والحياة. تم اختيار 15 أمر تتنوع في المستوى والأسلوب بالاعتماد على التقارير التالية (Sterling 2014, Enge 2019, Bushnell 2023)

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

أدوات الدراسة:

اعتمدت الدراسة الحالية على ثلاث أدوات رئيسية متمثلة في الآتي:

1. الملاحظة: تعتبر الملاحظة أداة مهمة في قياس الفروقات بالنتائج التي تظهر باستخدام تطبيقات اللغات الطبيعية خاصة في سرعة الاستجابة وصحة النتائج وكميتها.
2. الاختبار: الاختبار أو التجربة باستخدام أداة "Siri" لـ(اللغة العربية والإنجليزية "الأمريكية" والإسبانية) وملاحظة النتائج، سيتم الاختبار بواسطة 21 أمر يتم طلبها بشكل مماثل لخدمة "Siri" مع تغيير اللغة في كل مره ورصد الفروقات التي تظهر في جودة وعدد النتائج.
3. استعراض المناهج العلمية والدراسات: قياس الموارد المتوفرة للغات بواسطة استعراض المناهج العلمية والدراسات في هذا المجال ومحاولة تقييمها.

مجتمع وعينة الدراسة:

مجتمع الدراسة:

تمثل خدمة "Siri" مجتمع الدراسة والتي تُمثل إحدى تطبيقات معالجة اللغات الطبيعة.

عينة الدراسة:

تمثل عينة الدراسة اللغات العربية، الإنجليزية "الأمريكية" والإسبانية، والتي تعد أكثر ثلاث لغات تحدثاً في العالم. (شرواني، 2021)

تحليل البيانات:

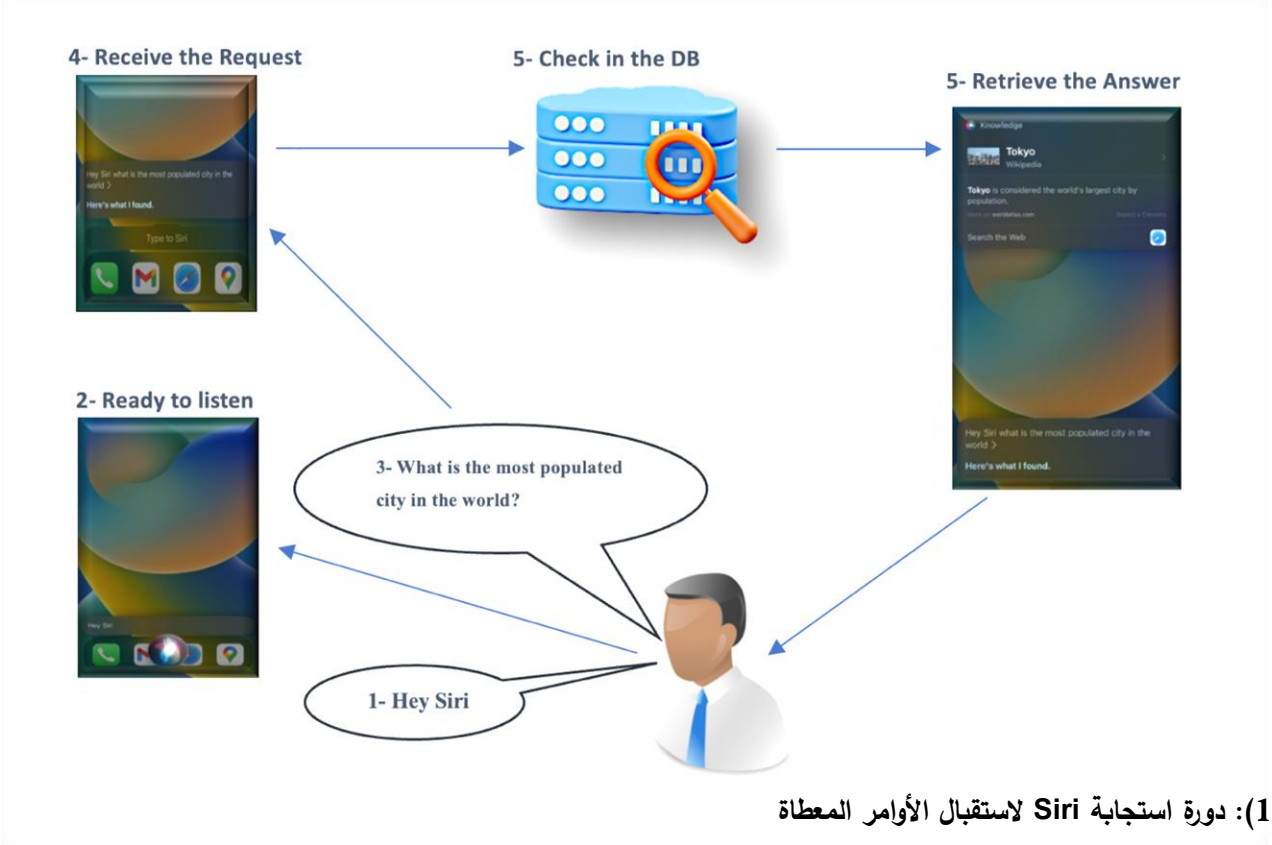
بعد الاطلاع على الإنتاج الفكري، توصلت الدراسة إلى اختبار Siri عن طريق 21 أمر مقسمة إلى ثلاث فئات (أوامر أساسية، أوامر مركبة، أوامر للترفيه) وكل أمر مقسم إلى 7 أسئلة كتبت بالأساس باللغة العربية ثم تم ترجمتها إلى اللغة الإنجليزية "الأمريكية" وللإسبانية. يتم طلب الأوامر بشكل مماثل مع تغيير اللغة، وملاحظة النتائج وكيفية استجابة Siri لكل أمر بكل لغة ورصد الفروقات التي تظهر في جودة وعدد النتائج.

سيتم استعراض أبرز النتائج والفروقات لتلبية Siri للأوامر المعطاة بكل لغة (يمكن الاطلاع على جميع الأوامر للغات الثلاثة في جدول رقم 1)

تسعى الدراسة إلى معرفة الفروقات بين استجابة Siri للأوامر المعطاة لكل لغة، وعليه سيتم استعراض الاستجابات ومن ثم شرح الفروقات.

يمثل الشكل (1) طبيعة دورة استجابة Siri لاستقبال الأوامر المعطاة من قبل المستخدم ابتداءً من " Hey Siri، يا Siri، و Oye Siri".

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798



الشكل (1): دورة استجابة Siri لاستقبال الأوامر المعطاة

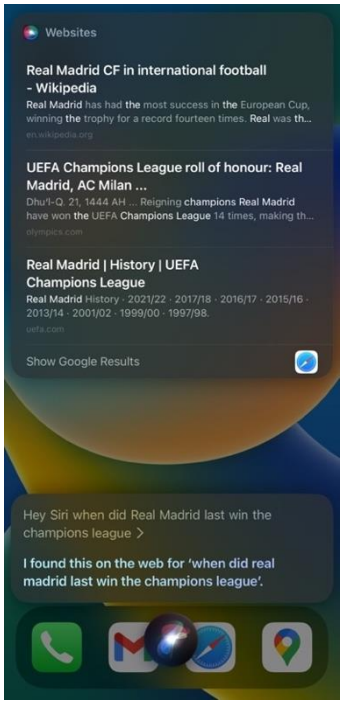
سيتم استعراض الأوامر ومدى سرعة وتمكن Siri في الاستجابة للأوامر المعطاة لكل لغة، لمعرفة جميع الأوامر ومعدل الاستجابة (انظر جدول رقم 1)

- **متى آخر مره فاز ريال مدريد ببطولة دوري أبطال أوروبا؟**

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

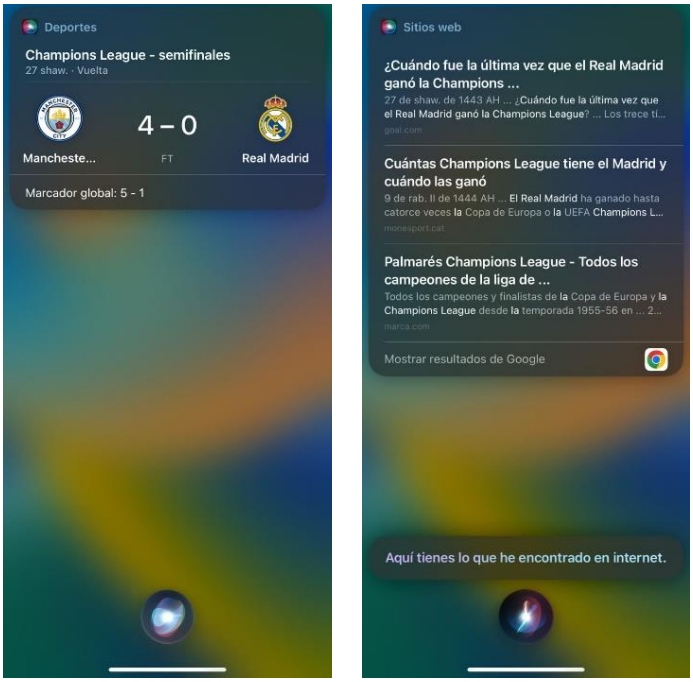
الشكل (2): الأوامر الأساسية

- When did Real Madrid last win the Champions League?



الشكل (3): الأوامر الأساسية

- ¿Cuándo es la última vez que club del Real Madrid gana la champions league?



الشكل (4): الأوامر الأساسية

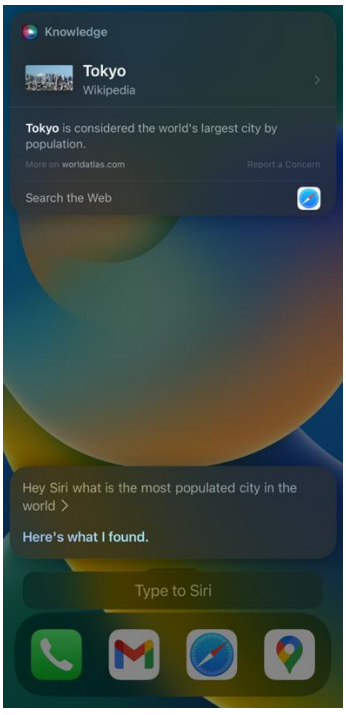
يوضح الشكل (2) استجابة Siri للسؤال باللغة العربية، عند سؤال Siri لأول مرة تمت الإجابة بعرض آخر نتيجة (مباراة وليس لقب) حققها فريق ريال مدريد في بطولة دوري أبطال أوروبا، وعند تكرار السؤال مرة أخرى وتعديله بأخر لقب حصل عليه فريق ريال مدريد لدوري أبطال أوروبا، تمت الإجابة بشكل صحيح وتم استعراض عدة مواقع تحتوي على الإجابة الصحيحة. وكذلك الحال عند سؤال Siri باللغة الإسبانية عن آخر مرة فاز فيها فريق ريال مدريد ببطولة دوري أبطال أوروبا، تمت الإجابة بأخر نتيجة حصل عليها فريق ريال مدريد في هذه البطولة، وبعد تكرار السؤال تمت الإجابة بشكل مباشر ببعض المواقع الإلكترونية التي تحتوي على الإجابة الصحيحة لعدد مرات حصول الفريق على البطولة الشكل (4). بينما يتضح من الشكل (3) استجابة Siri السريعة للسؤال باللغة الإنجليزية.

- **ماهي أكثر مدينة مأهولة بالسكان في العالم؟**

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

الشكل (5): الأوامر الأساسية

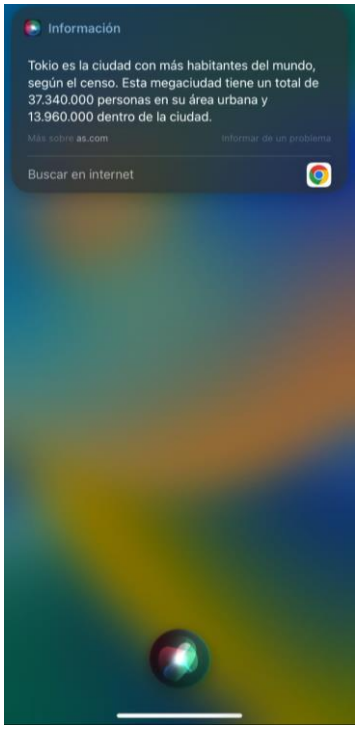
- What is the most populated city in the world?



الشكل (6): الأوامر الأساسية

- ¿Cuál es la ciudad más poblada del mundo?

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798



الشكل (7): الأوامر الأساسية

يوضح الشكل (5) استجابة Siri لهذا السؤال باللغة العربية انه في المرة الأولى تم سماع السؤال بشكل خاطئ، تم استبدال كلمة "أكثر" بكلمة "أقدم"، وتم عرض إجابة خاطئة بناءً على ذلك. بعد إعادة السؤال بشكل أوضح تمت الإجابة بشكل صحيح. بينما استجابة Siri السريعة للسؤال باللغة الإنجليزية والإسبانية عن أكثر مدينة مأهولة بالسكان في العالم، تمت الإجابة بشكل مباشر وصحيح بأن مدينة "طوكيو" أكثر مدينة مأهولة بالسكان الشكل (7،6).

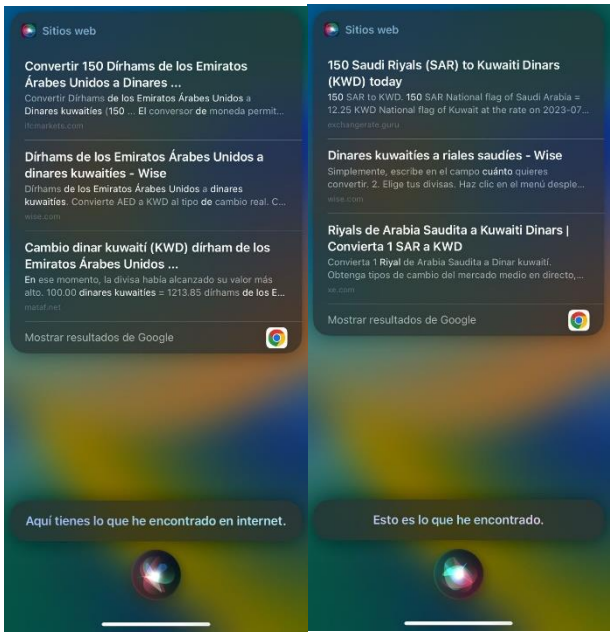
- **كم تمثل 150 ريال سعودي بالدينار الكويتي؟ ماذا عن الدرهم الاماراتي؟**

الشكل (8): الأوامر المركبة

- How much is 150 Saudi riyals in Kuwaiti dinars? What about the UAE dirham?

الشكل (9): الأوامر المركبة

- ¿Cuánto es 150 riales saudíes en dinares kuwaitíes? ¿Qué pasa con el dirham de los Emiratos Árabes Unidos?



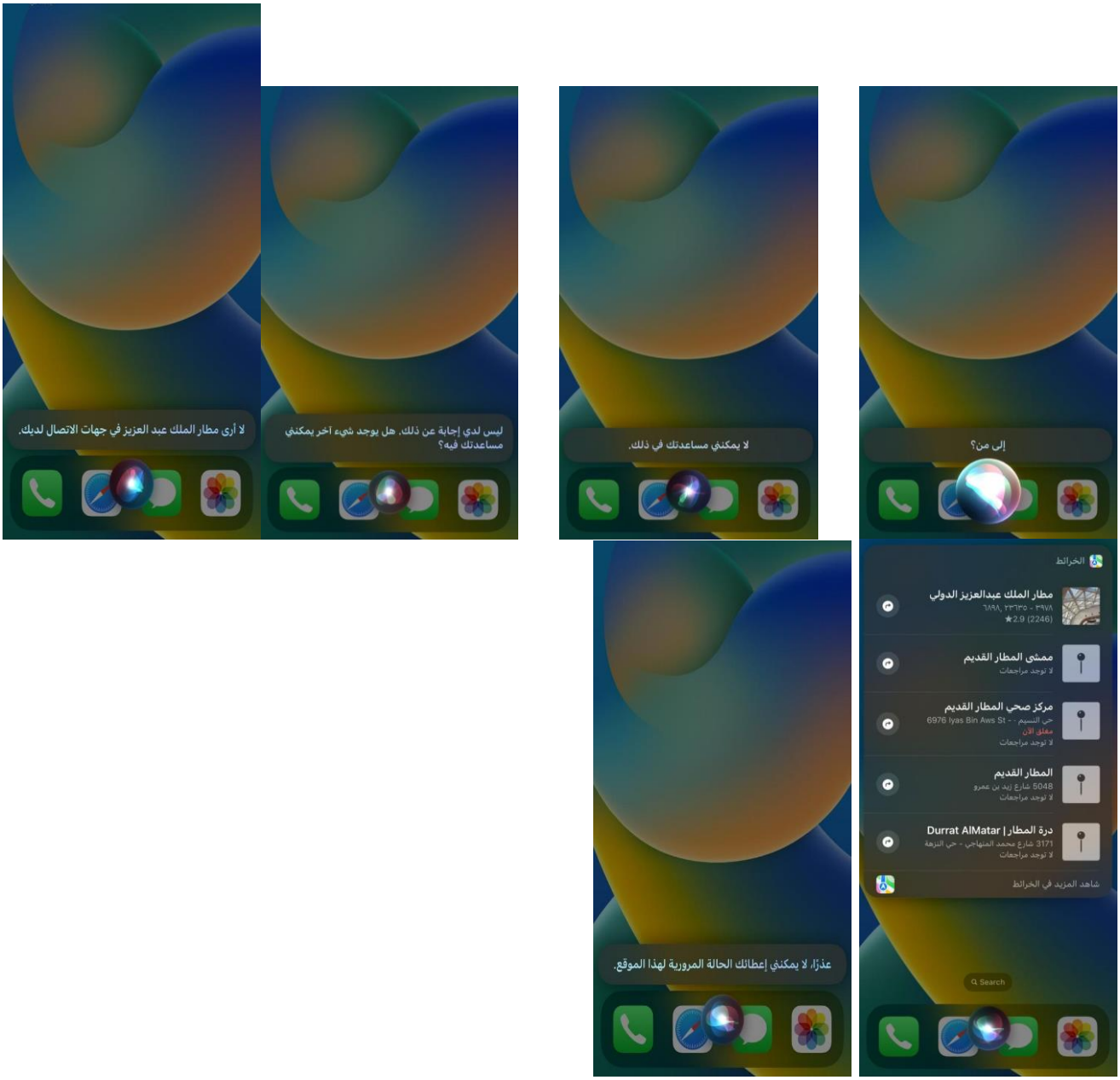
الشكل (10): الأوامر المركبة

تتشابه استجابة Siri لهذا السؤال باللغة العربية والإسبانية (الشكل 8،10)، حيث إن Siri لا تتبع السؤال الأول بسؤال اخر. يتضح أنه عند سؤال Siri يتم الإجابة، ولكن لا يمكن إتباع السؤال الأول بسؤال آخر متعلق بعملة بلد آخر بعد الحصول على الإجابة، حيث يتم البحث عن العملة نفسها بدون تحويلها. لذلك يتعين طرح السؤالين معاً، وكانت النتائج تشتمل مواقع تحتوي على إجابة الجزء الأول (الريال السعودي إلى الدينار الكويتي) ومواقع تحتوي على إجابة الجزء الثاني (الدينار الكويتي للدرهم الإماراتي). وكذلك عند السؤال باللغة الإسبانية عن تحويل 150 ريال سعودي إلى الدينار الكويتي ومن ثم إلى الدرهم الإماراتي، عند قراءة السؤال بشكل كامل تم استعراض عدة نتائج لتحويل الدرهم الإماراتي إلى دينار كويتي، والدينار الكويتي إلى درهم إماراتي بدون عرض نتائج للتحويل من الريال السعودي. وعند سؤال الجزء الأول من السؤال فقط تمت الإجابة عليه بشكل صحيح.

بينما يوضح (الشكل9) انه عند سؤال Siri باللغة الإنجليزية يتم الإجابة على السؤال المركب كاملاً، حيث تعطي الإجابة على الجزء الأول من السؤال ثم تتبعه بالإجابة على الجزء الاخر من السؤال.

- **كم الوقت الذي يستغرقه الذهاب إلى المطار بواسطة الحافلة؟ ماذا عن السيارة؟**

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798



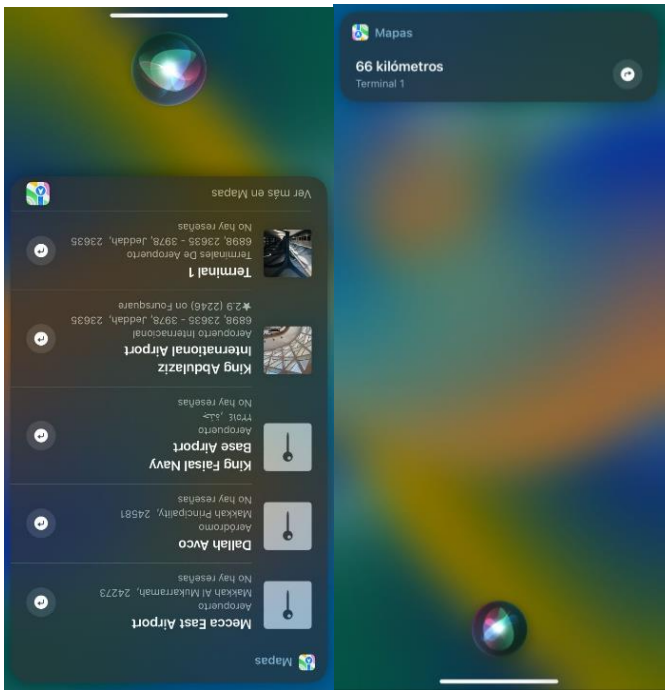
الشكل (11): الأوامر المركبة

- **How long does it take to go to the airport by bus? And by car?**

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

الشكل (12): الأوامر المركبة

- **¿Cuánto se tarda en ir al aeropuerto en autobús? ¿Qué pasa con el coche?**



الشكل (13): أوامر مركبة

يوضح الشكل (11) أن Siri لم تستطع فهم السؤال باللغة العربية، وبعد تكراره بعدة طرق تمت الإجابة " بلا يمكنني اعطائك الحالة المرورية لهذا الموقع".

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

عند سؤال Siri عن المسافة للمطار الشكل (12) باستخدام الحافلة وكذلك السيارة، يتم طرح سؤال من قبلها للاختيار من المطارات الموضحة في الصورة، يتم اختيار مطار الملك عبد العزيز، ولكن تجيب Siri بأنها لا يمكن أن تعطي المسافة رغم ان المسافة 4.8Miles تظهر في الإجابة، الا ان Siri تجاوب بأنه "لا يمكنني إعطاء الوقت المستغرق للوصول للموقع".

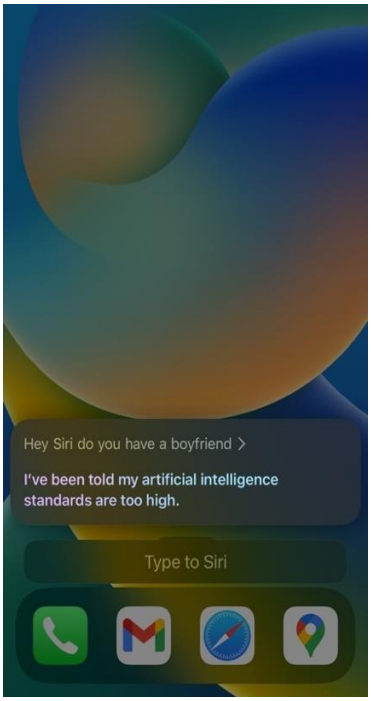
وكذلك يوضح الشكل (13) انه عند السؤال عن المسافة إلى المطار بواسطة الحافلة، والمسافة بواسطة السيارة، تم الاستجابة بعرض المطارات المتوفرة والسؤال عن أي من المطارات المعروضة الذي يراد حساب المسافة اليه، بعد الاختيار تمت الاستجابة بشكل مباشر وصحيح وعرض المسافة بالكيلو متر.

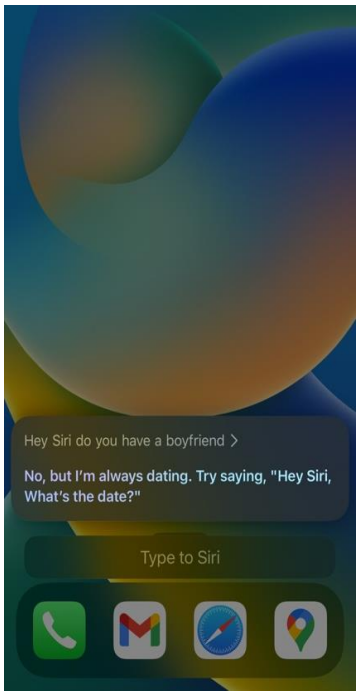
- **هل لديك صديق؟**

الشكل (14): أوامر ترفيهية

- **?Do you have a boyfriend**

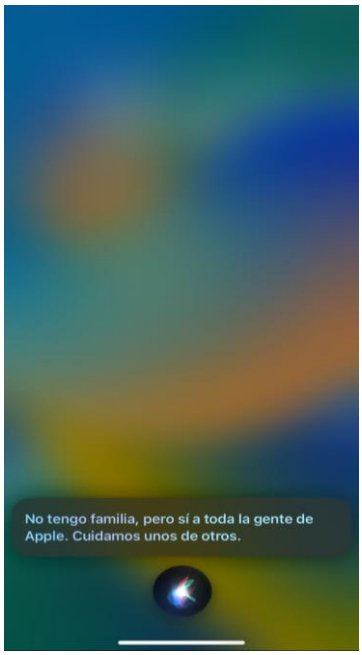
الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798





الشكل (15): أوامر ترفيهية

- **¿Tienes un amigo**



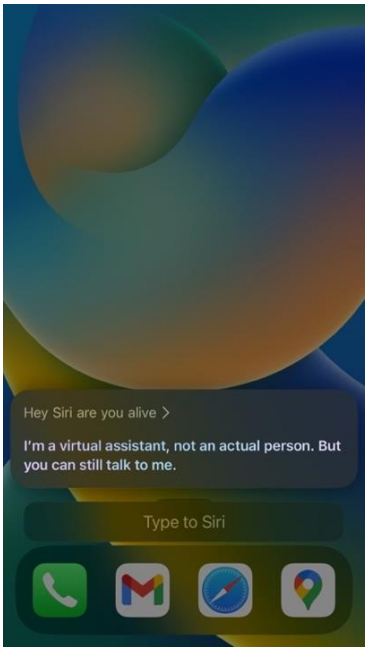
الشكل (16): أوامر ترفيهية

الشكل (14،15،16) تجيب Siri بإجابات مختلفة في كل مره يتم سؤالها.

- **هل انت على قيد الحياة؟**

الشكل (17): أوامر ترفيهية

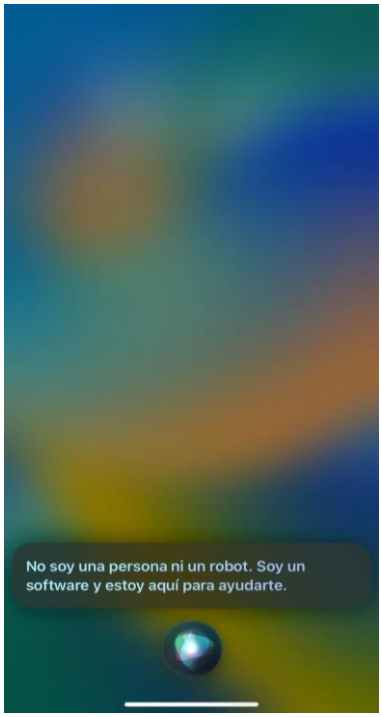
- Are you alive?



الشكل (18): أوامر ترفيهية

- ¿estás vivo?

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798



الشكل (19): أوامر ترفيهية

الشكل (17) يوضح استجابة Siri للسؤال، في المرة الأولى لم تقم Siri بالإجابة على السؤال، وفي المرة الثانية تمت الإجابة بطريقة غير مباشرة.

يوضح الشكل (18،19) أنه عند سؤال Siri هل انت على قيد الحياة؟ تمت الإجابة بأنا لست انسان، أو رجل آلي، أنا برنامج وأنا هنا لمساعدتك.

يوضح الجدول (1) مدى سرعة استجابة Siri للأوامر باللغات المختارة حيث تم منح النقاط لمدى الاستجابة على النحو التالي:

نقطتين (2): تمت الاستجابة بشكل صحيح من المرة الأولى

نقطة (1): تمت الاستجابة بشكل صحيح من المرة الثانية

صفر (0): لا يوجد استجابة صحيحة

الجدول (1): سرعة استجابة Siri للأوامر باللغات المختارة

الأوامر		اللغة العربية	اللغة الإنجليزية	اللغة الإسبانية
الأوامر الأساسية	كم عمر ديفيد بيكهام؟	2	2	2
	متى آخر مره فاز ريال مدريد ببطولة دوري أبطال أوروبا؟	1	2	1
	ماهي أكثر مدينة مأهولة بالسكان في العالم؟	1	2	2
	من الذي يقدم صوت Rapunzel؟	2	2	2
	ماهي ضريبة المبيعات في لندن؟	2	2	1
	ذكرني بالاتصال بعلي اليوم الساعة الواحدة ظهراً	2	2	2

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

	استعرض لي أقرب المقاهي	2	2	2
الأوامر العامة	كم تمثل 150 ريال سعودي بالدينار الكويتي؟ ماذا عن الدرهم الاماراتي	1	2	1
	كم عدد الأيام بين تاريخ 7 يونيو و20 أغسطس	0	2	2
	كيف يمكنني قول أين هو أقرب فندق باللغة الفرنسية	1	2	2
	ما هو التاريخ بعد 3 أسابيع؟	2	2	2
	ماهي الأفلام التي تعرض بالقرب مني؟	1	2	2
	كم الوقت الذي يستغرقه الذهاب إلى المطار بواسطة الحافلة؟ ماذا عن السيارة؟	0	2	2
	قم بتعيين منبه الساعة 3 فجراً.. قم بإلغاء المنبه الذي في الثالثة فجراً	2	2	2
الأوامر التفاعلية	هل لديك صديق؟	2	2	2
	هل أنت على قيد الحياة؟	1	2	2
	قل لي نكتة	2	2	2
	ما رأيك في اليكسا؟	2	2	2
	ماهي نتيجة قسمة صفر على صفر	2	2	2
	غني لي أغنية	1	2	2
	هل لديك حيوانات أليفة؟	0	1	2
المجموع		**29**	**41**	**39**

التعليق على الجدول:

بناءً على النقاط المعطاة في هذه الدراسة، يتضح استجابة Siri السريعة للغة الإنجليزية والإسبانية أسرع من استجابتها للغة العربية بفارق 12 نقطة عن اللغة الإنجليزية و10 نقاط عن اللغة الإسبانية. تتضح الفروقات في مدى استجابة Siri للأوامر المعطاة باللغة العربية على النحو التالي:

الأوامر الأساسية:

- جميع الأوامر تمت الاستجابة لها بشكل صحيح من المرة الثانية، إن لم يكن الأولى، ولا يوجد أي أمر لم تتم الإجابة عليه. على سبيل المثال لا الحصر عند السؤال عن (آخر بطولة حصل عليها فريق ريال مدريد لبطولة دوري أبطال أوروبا)، تمت الاستجابة بشكل صحيح من المرة الأولى في اللغة الإنجليزية، بينما في اللغة العربية والإسبانية تمت الاستجابة في المرة الثانية، أيضاً عند سؤال Siri عن (أكثر المدن المأهولة بالسكان)، تمت الاستجابة بشكل مباشر وصحيح باللغتين الإنجليزية والإسبانية، بينما تمت الاستجابة بشكل صحيح باللغة العربية في المرة الثانية، وإظهار نتائج في المرة الأولى عن (أقدم المدن المأهولة بالسكان) وليس (أكثر المدن المأهولة بالسكان)، (انظر جدول رقم 1).

- عند السؤال عن آخر مرة حصل فيها ريال مدريد على لقب، في اللغتين العربية والإسبانية تمت الاستجابة بأخر نتيجة لمباراة وليس لقب حصل عليه، ثم في المرة الثانية تمت الاستجابة بشكل صحيح، بينما ظهرت النتيجة بشكل صحيح من المرة الأولى في اللغة الإنجليزية. الشكل (2،3،4).
- يؤدي سماع المصطلحات في الأمر بشكل خاطئ إلى إظهار استجابات خاطئة او غير ذي صلة، مثل الاستجابات باللغة العربية في سؤال أكثر مدينة مأهولة بالسكان، والاستجابة باللغة الإسبانية في سؤال ضريبة المبيعات في مدينة لندن (انظر جدول رقم 1).

الأوامر المركبة:

- تعد الأوامر المركبة الأضعف في عدد الاستجابات، حيث حصلت اللغة العربية على 7 نقاط فقط من أصل 14 نقطة، على سبيل المثال لا الحصر عند سؤال Siri عن (عدد الأيام بين تاريخ 7 يونيو و20 أغسطس)، تمت الاستجابة باللغة الإنجليزية والإسبانية بشكل صحيح ومباشر، بينما باللغة العربية لم يستطع Siri فهم السؤال حتى بعد إعادة صياغته بأكثر من طريقة (انظر جدول رقم 1).
- عند طرح سؤال مركب، لم تتم الاستجابة بشكل كامل عند طرح الأمر على هيئة سؤال واحد في اللغتين العربية والإسبانية حيث يتم الاستجابة للجزء الأول من السؤال وإهمال الثاني، على عكس اللغة الإنجليزية" الأمريكية" تم التجاوب من المرة الاولى فقط. لذلك يتوجب طرح الأمر بشكل منفصل على أجزاء، مثل أمر تحويل العملات وتعيين المنبه (انظر جدول رقم 1).
- يؤثر استخدام المصطلحات الواقعة في بداية الأمر بشكل كبير على الاستجابة الصحيحة من عدمها، مثل الاستجابة باللغة العربية في سؤال (أين هو أقرب فندق باللغة الفرنسية)، حيث أتت الاستجابة بفتح تطبيق "أين هو Find My" بعد تكرار السؤال وابتدائه بجملة "اريد ترجمة جملة" تمت الاستجابة بشكل صحيح (انظر جدول رقم 1).
- عند سؤال Siri عن الافلام التي تعرض في المناطق القريبة في مدينة لا توجد بها صالات سينما تقوم Siri بتقديم اقتراحات لأفلام يمكن مشاهدتها على الانترنت (انظر جدول رقم 1).
- عند السؤال عن المسافة إلى المطار، تمت الاستجابة بشكل صحيح باللغة الإسبانية والإنجليزية، بينما باللغة العربية وحتى مع تكرار السؤال بأكثر من طريقة لمن يستطع Siri إظهار أي استجابة صحيحة. الشكل (11).

أوامر للترفيه:

- تمت الاستجابة بشكل صحيح في أغلب الأوامر بجميع اللغات. وكانت أقل الاستجابات للغة العربية حيث حصلت على 10 نقاط، مقابل 13 نقطة للغة الإنجليزية، بينما حصلت اللغة الإسبانية على النقاط الكاملة (14 نقطة) بإجابتها على جميع التساؤلات بشكل صحيح من المرة الأولى. على سبيل المثال عند سؤال Siri عن امتلاكه حيوانات أليفة تمت الإجابة باللغة الإسبانية بشكل صحيح من المرة الأولى، أما باللغة الإنجليزية فقد تمت الإجابة، ولكن بطريقة غير مباشرة وغير واضحة، فيما لم تستطع Siri فهم السؤال باللغة العربية (انظر جدول رقم 1).

النتائج:

بعد إجراء الاختبارات المحددة لـ Siri عن طريق 21 أمر مقسمة إلى ثلاث فئات (أوامر أساسية، أوامر مركبة، أوامر للترفيه) وكل امر مقسم إلى 7 أسئلة كتبت بالأساس باللغة العربية ثم تم ترجمتها إلى اللغة الإنجليزية "الأمريكية" وللإسبانية، خرجت الدراسة بالتالي:

- وجود تفاوت واضح في توفر قواعد البيانات أو الموارد للغات، بغض النظر عن عدد الناطقين الهائل باللغة يعد توفر الموارد المصنفة والمضبوطة لغوياً وبلغة يمكن أن تفهمها الآلة امراً اخراً.
- وجود مشكلة استخدام اللهجات والمصطلحات العامية أكثر من اللغة الرسمية أو الفصحى.
- الاختلاف الملحوظ بين اللغات في القواعد النحوية وطريقة تقسيم الجمل والكلمات.
- استمرار تطور بعض اللغات واحتواءها على بعض الكلمات المستحدثة، مما يزيد من صعوبة فهم الحاسوب لها وبالتالي عدم المقدرة على الاستجابة لها.

التوصيات:

من خلال النتائج التي تم استعراضها توصي الدراسة بالآتي:

- الاهتمام بتغذية الإنترنت بمصادر معلومات باللغة العربية غنية معدة بطريقة جيدة، ومحاولة الاستفادة من تجارب اللغتين الإنجليزية والإسبانية.
- ضرورة إقامة مشاريع رسمية للحكومات وتكثيف دور الجامعات في إنشاء بنوك للمصطلحات اللغوية العربية، وتحديثها باستمرار والاستفادة من تجربة البنوك العالمية المتاحة.
- استغلال المصادر العالمية المتاحة وترجمتها بطريقة جيدة من قبل متخصصين باللغة مع عدم استخدام تقنيات الترجمة الآلية لتجنب تكرار نفس الأخطاء.
- الدعوة إلى إنشاء ملتقيات وندوات تقام بصفة دورية تجمع بين المتخصصين في مجال اللغويات ومجال الحاسوب لدراسة الفجوة الموجودة ومحاولة إيجاد حلول تناسب المرحلة.

الخاتمة:

من خلال استخدام أداة Siri كأحد أدوات معالجة اللغات الطبيعية سعت الدراسة إلى قياس المعوقات التي قد تؤثر على استجابة Siri للأوامر المعطاة وذلك من خلال تصميم اختبار بثلاث لغات مختلفة (العربية والإنجليزية "الأمريكية" والإسبانية). توصلت الدراسة الى أن استجابة Siri للأوامر باللغة الإنجليزية والإسبانية أفضل من استجابتها باللغة العربية، وعليه خلصت الدراسة إلى أن سبب الضعف يعود إلى التفاوت بين توفر مصادر البيانات وجودتها في اللغة العربية مقارنة بالإنجليزية والإسبانية، بالرغم من العدد الهائل للمتحدثين باللغة العربية يعد توفر الموارد المقننة والمصنفة بطريقة تفهمها الآلة أمراً آخراً يمكن أن يشكل فرقاً في جودة الاستجابة وعدد النتائج الصحيحة. حيث أظهرت النتائج مقدرة اللغتين الإنجليزية "الامريكية" والإسبانية في الوصول إلى العديد من مصادر البيانات بشكل أفضل من اللغة العربية. وعليه توصي الدراسة بالاهتمام في تعزيز مصطلحات اللغة العربية من قبل المتحدثين الأصليين وإنشاء مصادر بيانات يمكن استخدامها في إنشاء مصادر بيانات تساهم في تحسين جودة تطبيقات معالجة اللغة الطبيعية، كما توصي بالاستفادة من التجارب العالمية للغتين الإنجليزية والإسبانية ومن البنوك العالمية المتاحة للحصول على مخرجات عالية الجودة، وتحقيق أقصى استفادة من التقنيات المتاحة.

Conclusion:

In conclusion, this study shed light on the factors that contribute to understanding why Siri does not respond correctly to give commands, this is done by implementing the command experiment using three different languages (Arabic, American English, and Spanish). The study found a difference in the quality of the appeal results in the Arabic language. The findings revealed that the availability and quality of data sources play a crucial role in determining the success of natural language processing applications. While American English and Spanish demonstrate superiority in accessing a multitude of data sources, Arabic still faces significant gaps that hinder its ability to provide accurate results. As a result, this study recommends prioritizing the enhancement of Arabic language terminology by native speakers and creating data sources that can contribute to improving the quality of natural language processing applications. Moreover, the study encourages leveraging the global experiences of English and Spanish languages and accessing available technological resources to obtain high-quality outputs. Ultimately, this research hopes to inspire further exploration and development of natural language processing tools to increase their effectiveness and utility across languages.

قائمة المصادر والمراجع:

Oracle Clou. (2023). ما المقصود بمعالجة اللغة الطبيعية؟ تم الاسترجاع في 24، أبريل، 2023 من https://www.oracle.com/sa-ar/artificial-intelligence/what-is-natural-language-processing/

أحمد، أحمد. (2021). المساعدات الرقمية الصوتية ودورها في تطوير خدمات المعلومات الذكية: دراسة تحليلية لتطبيقها في مؤسسات المعلومات العربية. جامعة أسيوط.

أحمد، محمد حسين. (2021). التعلم الآلي واستخراج البيانات الببليوجرافية من المواد النصية: نموذج مقترح

البرنامج الوطني للذكاء الاصطناعي. (2020). دليل الذكاء الاصطناعي. الامارات العربية المتحدة. تم الاسترجاع في 2023/05/23 من https://ai.gov.ae/wp-content/uploads/2020/02/AIGuide_AR_v1-online.pdf

جبريل، أسامة، مهدي، ياسر وإبراهيم، سالي. (2020). فاعلية نظام تدريس قائم على الذكاء الاصطناعي لتنمية الفهم العميق للتفاعلات النووية والقابية للتعليم الذاتي لدى طلاب المرحلة الثانوية. مجلة البحث العلمي في التربية، (21).

الدلاهمة، سليمان. (2019). أثر تطبيقات الذكاء الاصطناعي على مهنة المحاسبة. الأردن، جامعة جرش.

شرواني، هاجر. (2021). لغة المعالجة وتجربتها في المملكة العربية السعودية روبوتات المحادثة –دراسة تطبيقية– نماذج مختارة دراسة تقويمية. مجلة علمية حوليلة كلية اللغة العربية بجرجا. 225(13)

عبدالقادر، شاوش. (2022). الهندسة اللسانية والمعالجة الآلية للغة العربية، جامعة مولود معمري، الجزائر.

الموجود، عبد الله. غريب، سيد. (2022). فاعلية تقنية معالجة اللغات الطبيعية القائمة على الذكاء الاصطناعي في تنمية مهارات استخدام المنصات التعليمية الإلكترونية والقابلية للاستخدام لدى الطلاب الوافدين بجامعة الأزهر بالقاهرة. مجلة دراسات عربية في التربية وعلم النفس.142(2).

ناجي، إهداء. (2022). تطبيقات نظم الذكاء الاصطناعي في تحليل المحتوى وعمليات التكشيف دراسة تطبيقية لنظم معالجة اللغة الطبيعية. المجلة العلمية للمكتبات والوثائق والمعلومات، 4(11)، 89-216.

الهيئة الاتحادية للموارد البشرية الحكومية. (2023). الارشادات الخاصة باستخدام وتقييم الذكاء الاصطناعي في تقييم المواهب. الامارات العربية المتحدة. تم الاسترجاع في 2023/05/23 من https://www.fahr.gov.ae/Portal/Userfiles/Assets/Documents/2d4c61e6.pdf

اليافعي، زيد. (19 يوليو, 2022). معالجة اللغات الطبيعية: تحديات اللغة العربية. فِهم لعلوم الذكاء الاصطناعي. تم الاسترجاع في أبريل14، 2023 من https://n9.cl/fzn24

References:

Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018, October). " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (pp. 249-258).

Antoun, W. Baly, F. Hajj, H. (2020). ArabBERT: Transformer-based Model for Arabic Language Understanding. American University of Beirut.

Brezina, v. & Gablasova, D. (2018). The corpus method. Core. Retrieved on 08/07/2023 from https://core.ac.uk/reader/211246397

Bushnell, M. AI Faceoff: Siri vs. Cortana vs. Google Assistant vs. Alexa. Business News Daily. Retrieved on 11/07/2023 from https://www.businessnewsdaily.com/10315-siri-cortana-google-assistant-amazon-alexa-face-off.html

Cambridge Dictionary. (2023). artificial intelligence. Retrieved in 23/06/2023 from https://dictionary.cambridge.org/dictionary/english/artificial-intelligence

Enge, E. (2019). Rating the Smarts of the Digital Personal Assistants in 2019. Perficient. Retrieved on 11/07/2023 from https://www.perficient.com/insights/research-hub/digital-personal-assistants-study#smartest

Ewers, K., Baier, D., & Höhn, N. (2020). Siri, Do I Like You? Digital voice assistants and their acceptance by consumers. SMR-Journal of Service Management Research, 4(1), 52-68.

Hamandi, A. M., AbdulHassan, A. K., & Bahjat, H. (2016). Natural Language Processing Using Natural Language Toolkit. Iraqi Journal of Information Technology, 7(2), 70-83.

Hoy, B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly, Vol. 37, No. 1, 81–88. Retrieved from https://doi.org/10.1080/02763869.2018.1404391

Oxford Reference. (2023). natural-language processing. Retrieved in 23/06/2023 from https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100225333;jsessionid=0FF0517E77E5523DAB0EE39891687E7D

Reis, A., Paulino, D., Paredes, H., & Barroso, J. (2017). Using intelligent personal assistants to strengthen the elderlies' social bonds: A preliminary evaluation of amazon alexa, google assistant, microsoft cortana, and apple siri. In Universal Access in Human–Computer Interaction. Human and Technological Environments: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part III 11 (pp. 593-602). Springer International Publishing.

Schwab, K. (2016). The Fourth Industrial Revolution: what it means, how to respond. World Economic Forum. Retrieved on 09/05/2023 from https://web.archive.org/web/20190531095621/https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Sterling, G. (2014). Google (Now) Beats Siri and Cortana for Direct Answers — Study. Search Engine Land. Retrieved on 11/07/2023 from https://searchengineland.com/google-now-beats-siri-cortana-direct-answers-study-205388

Thakare, U. A., Thakare, P. B., & Dhawale, V. R. (2020). Siri-The Intelligent Personal Assistant. International Research Journal of Innovations in Engineering and Technology, 4(1), 17.

Wang, D. Su, J. and Yu, H. (2020). Feature Extraction and analysis of Natural Language Processing for deep Learning English Language. IEEEAccess.

Waterman, D. (1986). A guide to expert system. Reading, MA: Addison-Wesley.

Zhong, H. Xiao, C. Tu, C. Zhang, T. Liu, Z. Sun M. (2020). How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. Department of Computer Science and Technology Institute for Artificial Intelligence, Tsinghua University, China and Beijing National Research Center for Information Science and Technology, China and Beijing Power law Intelligent Technology, China.

الإصدار السادس – العدد ستون
تاريخ الإصدار: 2 – تشرين الأول – 2023م
www.ajsp.net
ISSN: 2663-5798

"Examining the Relationship Between Data Resource Utilization and Natural Language Processing Quality: Analytical Study of Siri as a Model"

Researchers:

Hind Badi Albadi[1], Rana Abdullah Alobaidi[2]

[1]Information Science Department, Associate Professor, Faculty of Arts and Humanities, King Abdulaziz University, Jeddah, Saudi Arabia

[2]Master of Information Management, Information Science Department, Faculty of Arts and Humanities, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract:

Natural language processing (NLP) technology is crucial in enhancing the performance of many fields, including healthcare, customer service, financial management, and academic research. However, the effectiveness of NLP applications varies depending on the language used, due to the significant differences that exist in languages globally. Therefore, this study aims to explore the differences that arise when employing NLP applications in different languages, using Siri as a model for natural language processing techniques. The study focuses on three languages, Arabic, "American" English, and Spanish. The study adopts a descriptive-analytical approach to analyze the performance of NLP applications available for each selected language. Through observing and testing tools to monitor and document observations, the study assesses the quality of the results obtained from Siri's specific tests. Consequently, the study discovers that language disparities and the quality of data sources contribute to the weaknesses in the quality of the results of the Arabic language compared to English and Spanish. Therefore, the study recommends using native speakers' Arabic language terms to create data sources to improve the quality of NLP applications and achieve higher-quality outputs.

Keywords: Natural Language Processing, Artificial Intelligence, Data Resources, Siri, Voice Digital Assistant.