نظام أجابه آلي للنصوص العربية بالاعتماد على الفهم

ا.م.د. يسرى مالك ضمد
هدى عادل
كلية علوم الحاسوب وتكنولوجيا المعلومات
قسم علوم الحاسوب

المستخلص:

بسبب كثرة النصوص الإلكترونية على الشبكة العنكبوتية وازدياد عدد الوثائق فيها صار الحصول على المعلومة التي يحتاجها المستخدم بسرعة متناهية أمرًا صعبًا، من هنا ظهرت الحاجة لتصميم نظم استرجاع معلومات أو إجابة على الأسئلة وغيرها من النظم التي تختزل الوقت والجهد للمستخدم.

في هذا البحث قمنا بتصميم نظام إجابة آلية على الأسئلة والذي يحاكي بعمله قدرة الانسان على فهم النص وإعادة جواب مناسب سواء كان موجود ضمنيا فيه بصورة مباشرة أو غير مباشرة بطرق التنقيب في النص واستخلاص المعلومات او عن طريق المعرفة العالمية والتي تعد خلفية للنص المدخل، وللوصول إلى هدفنا المنشود قمنا بتصميم محلل نحوي دلالي يعد أساس دقة النظام إذ يقوم بتحويل النص غير المهيكل الى نص مهيكل مترابط يسهل عملية البحث واسترجاع الإجابة لمستخدم النظام، إذ اعتمدنا قوانين عدة لتكوين العبارات الأسمية وتصنيفها إلى مجموعة من الانماط الدلالية وربط هذه الأنماط باستخدام شبكة الانتقالات المعززة مع المعرفة المعجمية والصرفية لتقديم هيكل دلالي شبيه بالشجرة، ومن ثم تطبيق نظام الإجابة الآلية على الهيكل الدلالي الناتج عن طريق عمليات بحث وترشيح للجمل التي قد تتضمن الإجابة وعند عدم توفر الإجابة في الجمل المرشحة يقوم بالبحث في المعجم الدلالي الساند للنظام وفي حالة عدم وجودها يعتذر عن إعادة الجواب برسالة نصية مناسبة.

استعمل في تقييم الجواب مقياس الدقة Accuracy فكانت النتيجة (٨٥%) وهي نتيجة جيدة جدًا بالنسبة لهذا المجال مع قلة البحوث في فهم اللغة الطبيعية وخاصة بالنسبة للغة العربية.

الكلمات المفتاحيه: معالجة اللغات الطبيعيه ، فهم اللغات الطبيعيه ،نظم الاجابه الاليه ،تنقيب النص، استخلاص المعلومات.

Automatic question answering system depend on the understanding of Arabic texts

Dr.Yusra Malik **Huda Adel**

yusra.dhommed@uobasrah.edu.iq

College computer sciences & information technology

computer sciences department

Abstract

The increasing in the electronic texts on the Internet and the number of documents which have become difficult to obtain the information needed by the user, to fast that there was a need to design systems to retrieve information or to answer the questions and other systems that reduce the time and effort for the user.

we designed a system to answer the questions, which simulates human capacity and work smarter through the understanding of the text and to answer appropriately, whether that is implicit in the text or indirectly or through the retrieved from world knowledge, which is considered as a background for text input. To reach the desired goal we designed a grammatical semantic Analyzer which is The accuracy of the system that converts unstructured text into coherent structured text which is search and retrieval easy answer to the system user, so we adopted him enhanced with knowledge transfer network lexical and morphological to provide semantic structure similar to tree, and then applying the answer (QA) at the semantic structure by searches and filtering of sentences that might contain the answer and when the answer in sentences candidate search semantic lexicon anvil system and when the answer cannot be provided to a text message.

Our system is evaluated by using the common evaluation system Precision with compression rates (85%), where the results are very good and very encouraging.

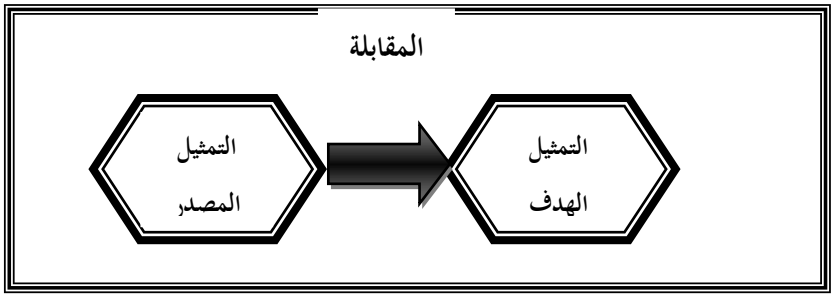
Key words: natural language understanding, Question answering , text mining, natural language processing , information extraction.

١-المقدمه:

فهم اللغات الطبيعية natural language understanding (NLU) في حقل الذكاء الاصطناعي واللسانيات الحاسوبية هوأحد فروع معالجة اللغة الطبيعية التي تتعامل مع آليات القراءة والفهم. الهدف من نظام NLU هوتفسير مقطع النص المدخل. ويمكن النظر لعملية التفسير هذه بوصفها ترجمة للنص من اللغة الطبيعية إلى تمثيل بلغة شكلية خالية من اللبس. هذا التمثيل، من المفترض أن يعبر عن محتوى النص. ويستعمل لإنجاز المهام الملموسة التي ينطوي عليها طلب المستخدم أيضًا [Kai-UweK¨uhnberger,2012].

ويمكن تمثيل عملية الفهم بثلاث نقاط :

١. التمثيل المصدر (الصيغة اللغوية).
٢. التمثيل الهدف (المعنى المقصود).
٣. ووجود آلية للمقابلة (Mapping) للانتقال من التمثيل المصدري الى التمثيل الهدف وكما في الشكل (١) [Eric M.Munson& Joseph F. Murphy,1983,]سلمى ، 2001



الشكل (١) تمثيل عملية المقابلة (mapping)

هنالك عنصر حاسم في الفهم ينطوي على حساب تمثيل لمعنى الجمل والنصوص، دون تحديد مفهوم التمثيل. على سبيل المثال، لماذا لا نستعمل كلمات الجملة نفسها على أنها تمثيل لمعناها ؟ أحد الأسباب هوأن معظم الكلمات لها معان متعددة، والتي تدعى (senses). وهذا الغموض يمنع النظام من صنع الاستدلالات المناسبة اللازمة لفهم أنموذج ما. [Kai–UweK¨uhnberger,2012].

هناك عدة معايير للفهم يتصل كل منها بجانب من جوانب الذكاء، من هذه المعايير:

- القدرة على إجابة الأسئلة بطريقة ملائمة.
- القدرة على إعاده صياغة جمل اللغة، شارحا معناها بطريقة مختلفة.
- القدرة على بناء الاستنتاجات.
- القدرة على الترجمة من لغة طبيعية لأخرى.
- القدرة على التعرف على المسميات (تعيين وتعريف المراجع (References)).[آلان ،١٩٨٩، عبد الحميد،١٩٩٤، William John Teahan، 2010]

٢–صعوبات بناء الفهم :

تكمن صعوبة بناء فهم للغة في كيفية تمثيل المعنى اللغوي كمعرفة، وفيما يلي أبرز تلك الصعوبات:

أ- تعقد التمثيل الهدف الذي تؤدي اليه المقابلة، معظم المحاولات المبكرة لبناء برامج لفهم اللغة الطبيعية استعملت تمثيلات هدفيه بسيطة لأنها اهتمت غالباً بالتحليل السطحي للجملة وأيضاً لأنها غالباً ما كانت تمثل المعرفة في مجال محدد. [Eric M.Munson& Joseph F. Murphy,1983]

ب- الخاصية التعددية للمقابلة، إذ توجد أنواع متعددة من المقابلات واحد لواحد، واحد لمتعدد، متعدد لواحد، ومتعدد لمتعدد فالتقابل واحد لواحد لا يشكل أي مشكلة بالنسبة للمعالجة الآلية للغة في حين تمثل بقية التقابلات كل أسباب التعقيد ومشاكل المعالجة.

ت- مستوى التفاعل بين مكونات التمثيل المصدري، كلما أزداد التفاعل بين المكونات أزداد تعقيد المقابلة.

٣-مستويات الفهم:

لفهم نص مكتوب بلغة فإن ذلك يتطلب المرور بعدة مراحل للوصول في النهاية الى بناء فهم لذلك النص. وهذه المستويات هي:

١. فهم كل كلمة على حده في الجملة.
٢. فهم معنى الجمله المتكونه من هذه الكلمات.
٣. محاولة فهم مجموعة من الجمل سواء فقرة أونص وهي عملية معقدة جداً إذ انها تتطلب فضلًا عن فهم الجمل المنفصلة، فهم العلاقات بين الجمل، كالموضوع المشترك والأسباب والاستنتاجات والأغراض والمواصفات وغيرها من العلاقات.[Allen, James,1995]

٤-المعرفة وتمثيل المعنى:

٤-١ المعرفة ودورها في فهم اللغة الطبيعية:

معرفة العالم والسياق هي جوانب مهمة لفهم اللغة الطبيعية. في أغلب الأوقات تكون المعلومات ضمنية داخل الجمل. ومن شأن الأنظمة التي تدعي فهم اللغة أن تستخرج هذه المعلومات عن طريق استعمال المعرفة العالمية وتستعمل التمثيل المناسب لتلك المعلومات أي إطار يدعم تخزين واسترجاع المعرفة العالمية يمكن أن ينظر إليه كتمثيل معرفة. ويمكن تقسيمها على قسمين: قاعدة بيانات المعرفة ومحرك الاستدلال والتي توفر مجموعة من العمليات على قاعدة المعرفة [يسرى،٢٠١٢]

أحد التحديات الحاسمة تتمثل في إكتساب المعرفة. تستعمل النظم الحالية قواعد البيانات المصنعة يدويا والتي تحتوي على: المعرفة اللغوية (الصرفية والمعجمية والنحوية والدلالية والخطاب) والمعرفة غير لغوية (المعرفة العالمية الدلالية والذرائعية والتخطيط والاستدلال).

المجالات ذات المدى الواسع من الصعب التعبير عنها باستعمال تقنيات تمثيل المعرفة المعروفة حاليا. بل يجب أن يتم تضييق المجال. [Group UE6 ,2012]

٤-٢ تمثيل المعنى Representation of Meaning:

غالبا ما يفهم مصطلح "المعنى" خاصة في اللسانيات الحاسوبية بمعناه الأعم كعلاقة بين علامة لغوية (label) وكيان غير لغوي (object) يمكن أن تكوّن مفهوما في العقل، كائنا في العالم، صيغة منطقية، الخ [Group UE6 ,2012]

في تمثيل محتوى النص، ينبغي أن تعد الاسئلة التالية حول، ما التمثيل المناسب للمعنى اللغوي الذي يجب أن يتم اختياره؟، وما هي المعلومات التي ينبغي ان تأخذ ؟، وكيف يمكن أن يجري بناؤها؟. لا توجد إجابة فريدة من نوعها لهذه الأسئلة، بل تعتمد على النظرية الكامنة وراء المعنى اللغوي. على سبيل المثال، الافتراضات بشأن تمثيل المعنى اللغوي الذي أدلى به شانك (١٩٧٢) وآلن (١٩٨٧) يمكن تلخيصها على النحوالتالي [Kai–UweK¨uhnberger,2012] إذا **كانت جملتين لها نفس المعنى، يجب أن يكون لها نفس التمثيل.**

يجب أن يكون التمثيل دقيقا لا لبس فيه يكفي لتمكيننا من التعبير عن كل قراءة مختلفة للجملة كصيغة مختلفة، المعلومات التي ذكرت ضمنا في الجملة يجب تمثيلها صراحة.لابد لأي معالجة لفهم معنى جملة من تجميع كل أنواع الدلالات بطريقة ما للوصول من معاني مكونات الجملة الى معناها العام. وهنالك أنواع عديدة من الدلالات التي يمكن من خلالها الحصول على المعنى الكلي للجملة وهي [د.أحمد ، ١٩٨٢،يسرى ٢٠٠٦،]

الدلالة الصوتية ، الدلالة الصرفية ، الدلالة النحوية ، الدلالة المعجمية ، الدلالة السياقية ، الدلالة الاجتماعية .

٥- تنقيب النصوص Texts Mining (TM):

يمكن تعريف تنقيب النص كعملية للمعرفة المكثفة، إذ يتفاعل المستخدم مع وثيقة جرى تجميعها مع مرور الوقت عن طريق استعمال مجموعة من أ