RECOGNITION AND ANALYSIS OF BINARY QUESTIONS FOR STANDARD ARABIC

Essia Bessaies, Slim Mesfar, Henda Ben Ghezala
University of Manouba, TUNISIA
Introduction
Introduction

• There has been a lot of research in the field of English and some European language Question Answering Systems.
• However, Arabic Question Answering Systems could not match the pace due to some inherent difficulties with the language itself as well as due to lack of tools available to assist the researchers.
For this purpose, the developed question answering system is based on a linguistic approach.

Formalize the automatic recognition rules ➔ apply them to a dynamic corpus composed of medical journalistic articles.
## Related Works

<table>
<thead>
<tr>
<th>System</th>
<th>Objective</th>
<th>Domain</th>
<th>Dataset</th>
<th>Results</th>
<th>Shortcomings</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQuASys (S. BEKHTI and M. AL-HARBI, 2013)</td>
<td>answer unformatted fact-based questions written in Arabic natural language.</td>
<td>Close domain</td>
<td>ANERcorp: 150,000 tagged tokens) as well as few gazetteers (ANERgazet) available online</td>
<td>recall rate of 97.5% and 66.25% as a precision rate</td>
<td>The system focused only on factoid questions</td>
</tr>
<tr>
<td>DefArabicQA (O. Trigui, 2010)</td>
<td>provide short answers for Arabic natural language questions</td>
<td>Close domain</td>
<td>collection of Arabic text documents</td>
<td>not presented</td>
<td>Does not include the other types of question (How and Why)</td>
</tr>
<tr>
<td>Yes/No Arabic Question Answering System (H. Kurdi, et al, 2014)</td>
<td>design a formal model for a semantic based yes/no Arabic question answering system based on paragraph retrieval</td>
<td>Open domain</td>
<td>20 documents which used to test the system and a collection of 100 different yes/no question</td>
<td>The results of using documents technique: 85% when 20 documents are used. The result of using paragraphs technique: 88% when 20 documents are used</td>
<td>The system focused only on yes/no questions. and the corpus size is small (20 documents)</td>
</tr>
<tr>
<td>JAWEB (W. N. Bdour and N. K. Gharaibeh, 2008)</td>
<td>provide short answers for Arabic natural language questions</td>
<td>Close domain</td>
<td>an extended version of the Arabic corpus</td>
<td>The system provided 15-20% higher recall</td>
<td>The system focused only on factoid questions</td>
</tr>
</tbody>
</table>
Our approach

Step 1

Question Analysis
DIC .Nom .Nog

Topic & Focus

Step 2

Text pre-processing
Segmentation

Step 3

Passage Retrieval
Noojapply.exe

Step 4

The answer extraction
Inherit the sentence properties (+Neg or +Affirmative) in order to give the binary response
Our approach

Step 1: Question Analysis

- Make a linguistic analysis of questions ➔ Add annotations associated with all recognized forms (lexical, morpho-logical as well as syntactic information)

- Apply a syntactic grammar to identify and annotate the **topic** and **focus** of question.
Our approach

Step 1: Question Analysis

Example:

Is measles a contagious disease?

هل يعتبر مرض الحصبة معدياً؟

هل, Is = Binary interrogative mark
معدياً, Contagious = Focus
مرض الحصبة, Measles disease = Topic
Our approach

**Step 2 : segmentation tool for Arabic Texts**

- Integration of a segmentation tool for Arabic texts ➔ an enhanced version of (Nadia Ghezaiel and Kais Haddr. 2016)

- The segmentation tool will also identify:
  - The negation status of the sentence: +Negative OR +Affirmative
  - The sentence style: +Declarative, +Imperative, +Interrogative OR +Exclamative
Our approach

Step 2: Segmentation tool for Arabic texts
Our approach

Step 2: Segmentation tool for Arabic texts

Example:

هل يعتبر مرض الحصبة معدٍ؟ الحصبة هو مرض فيروسي حاد ومعدي يصيب الأطفال، ويسبب لهم بعض المضاعفات التي تكون خطيرة في بعض الأحيان. ويعتبر مرض الحصبة من أكثر الأمراض انتشارًا في سن الطفولة بصفة خاصة، ولكنه قد يصيب الكبار أيضاً.
Our approach

Step 2: Segmentation tool for Arabic texts

Example:
Our approach

Step 2: Segmentation tool for Arabic texts

- Apply a NooJ regular expression: `<S-Interro>`
Our approach

Step 2: Segmentation tool for Arabic Texts
Our approach

Step 3: Passage Retrieval
Our approach

Step 3 : Passage Retrieval
Our approach

Step 4 : The answer extraction

• In this 4th step, we inherit the negation status of the sentence (+Neg OR + Affirmative) given by the segmentation tool

• As a preliminary result, we suppose that the given status represents correctly the binary answer of our question.

• Then, we are working on the integration of similarity scores in order to better rank the retrieved passages.
Conclusion and perspectives

- We are developing a question answering system which is based on a linguistic approach.
- The use of the linguistic engine of Nooj in order to formalize the automatic recognition rules and then applying them to a dynamic corpus composed of Arabic medical journalistic articles.
- Question analysis: apply a syntactic grammar to identify and annotate the topic and focus of question.
Conclusion and perspectives

• Segmentation tool: will also identify the negation status of the sentence and the style of the sentence.

• We are working on the integration of similarity scores in order to better rank the retrieved passages.

• Finally, as a long term ambition, we intend to consider studying the processing of the "why" and "how" question types.
Thank you!

Any question?
References

