<https://docs.google.com/document/d/1eRpnYxignPYPfW-6VwA88WOoi8ugxVX3_Kda4V-wDOU/edit>

*Question Answering System, Approaches and*

*Techniques: A Review - provide a summary what you have learned with examples*

The research provides an overview of the various approaches and techniques used in the development of question answering systems. The paper introduced four frameworks for QAS and they are Question Processing Module, Document Processing Module, Paragraph Extraction Module, and Answer extraction Module.

**Question Processing Module**: a component of a question answering system that is responsible for understanding and analyzing the input question. This module also identifies the type of question, for example, if the question is “When was Shivaji born”, we know that this is a “When” type of question that asks for time.

**Document Processing Module**: a component that contains a query generation algorithm that leverages the user’s question for text searching.

**Paragraph Extraction Module**: This module breaks the document into smaller parts and compute score based on the questions. Then, the module will leverage the scores to find the most relevant part of the document for the specific question.

**Answer extraction Module**: the module tries to extract the exact phrase of the answer and expand it from the center until it achieves the requirement of the answer extraction algorithms.

The review also mentioned a few approaches to address the issues with QAS.

**Linguistic Approach:** This approach is grounded in the study of linguistics and focuses on the formal properties of language, such as syntax, semantics, and morphology.

1. Parse sentences to identify their grammatical structure.
2. Use semantic rules to determine the meaning of words and phrases.
3. Use morphological analysis to identify the root forms of words.

**Statistical Approach:** This approach applies statistical technique in QA system to find the most relevant result in the document. Compared to other approaches, the statistical technique can be used on wide range of NLP tasks, but it won’t perform very well on tasks that require a deeper understanding of the underlying structure.

**Pattern Matching Approach:** This approach is based on the idea that a question can be mapped to a set of patterns, and that the answer to the question can be found by matching these patterns to the text. Generally, this approach leverages regular expression, surface pattern, and template.

**Surface Pattern:** Surface patterns can include keywords, syntactic structures, and other features that can be easily identified by pattern matching techniques. For example, if the question is “Where is the Capital of Canada”, then it can be fit into the surface pattern “What is a of b”. By leveraging the pattern, we can locate the answer in the document.

**Template:** a predefined structure or pattern used to represent the intent and meaning of a question. For the same question in Surface Pattern, the template will take a different approach. It will identify the type of question to “Where” which means the answer should be a location, and then it will capture the entities of “Capital” and “Canada”. After doing all the work in question part, it will fir this information into a template to help finding answer in the document.

What I learned from this research:

The part that interests me the most is the Pattern Matching approach. This approach will understand the user’s question first by identify object, verbs, adjectives etc. Then it will fit the information into the pre-defined pattern, for example, “x was born on y year”. By getting the pattern of the question, it can bring the input information and pattern into the document to get the answer, and as it shown on the result, the performance is good and it can be applied to open domain.