

sJanta: An Open Domain Question Answering System



NTCIR11
QA-Lab

Md. Arafat Rahman, BRU, R
arafatrahmanbrur@gmail.com
Md-Mizanur Rahoman
Sokendai, mizan@nii.ac.jp

Abstract

- It is a report of participation of system called sJanta at NTCIR11 QA-Lab English sub-task
- sJanta is a modular question answering system that can
 - incorporate question context
 - automatically find question-specific text
 - check both syntactic and semantic similarities in selecting answer
 - leverage contemporary NLP tools
 - implement simple but promising framework

Overview of QA-Lab task

- **Questions are from real entrance-exam**
 - Multiple Choice Question (MCQ)
 - Blank-space Filling
 - Automatic Answer Generation etc.
- **An Exemplary Question**

```
... an important place of pilgrimage in
<uText id="U1"><label>(1)</label>
Christianity
</uText>.
In the 9th century ...
```

```
<instruction>
From 1-4 below, choose the one sentence that
correctly describes the underlined portion
<ref comment="" target="U1">(1)</ref>.
</instruction>
...
<choice ansnum="1">
<cNum>(1)</cNum>It was established based
on the teachings of Jesus, who criticized Islam.
</choice>
<choice ansnum="2">
<cNum>(2)</cNum>Its holy book is the New Testament.
</choice>
<choice ansnum="3">
<cNum>(3)</cNum>It was made the state religion by
the Emperor Diocletian.
</choice>
<choice ansnum="4">
<cNum>(4)</cNum>Wycliffe was declared a heretic at
the Council of Clermont (ecumenical council).
</choice>
...
```

Experiments

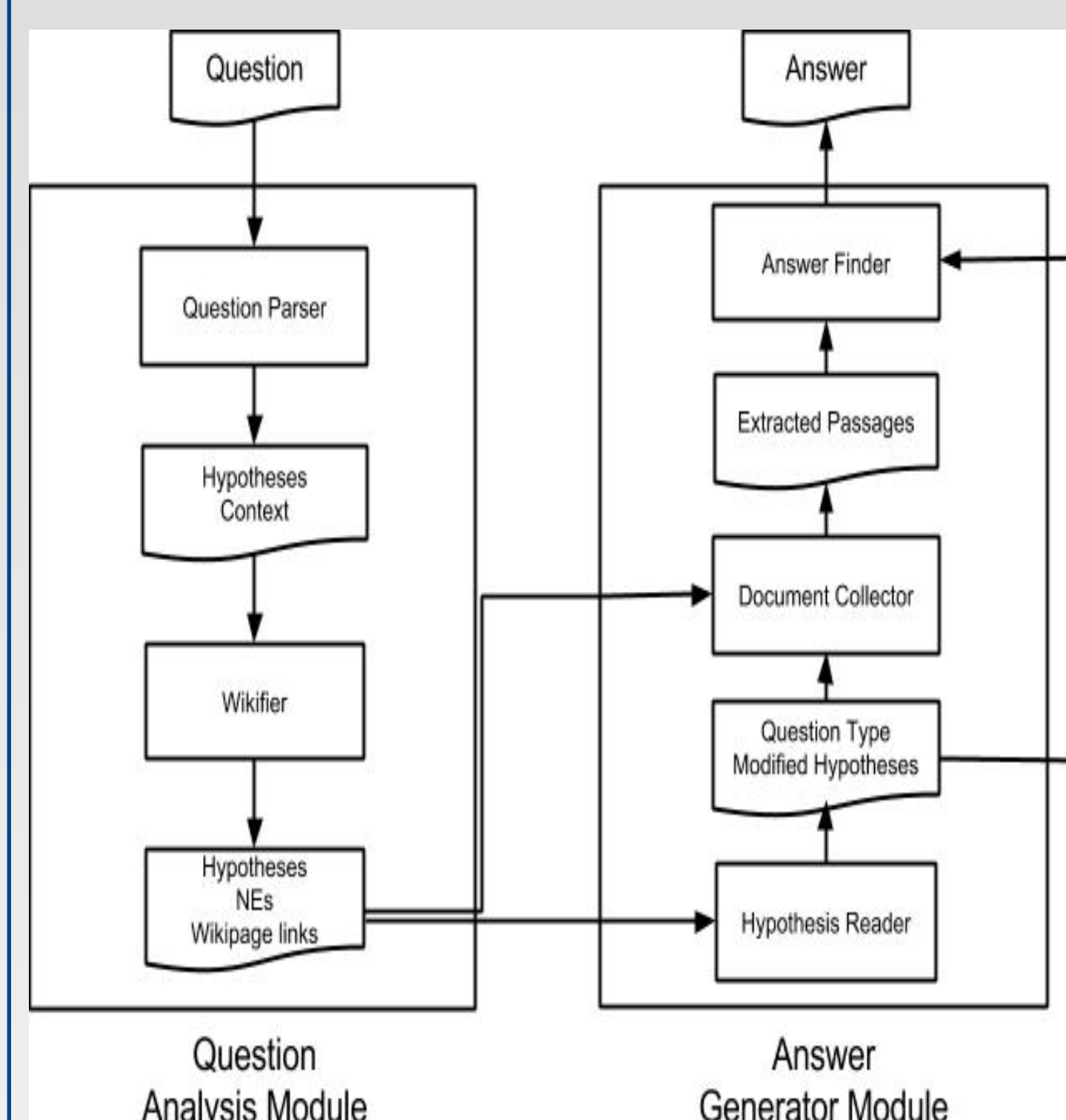
Performance of sJanta Over Phase 1 Question Set

Total Qs	# of Qs Answered	# of Correct Answer	% of Correct Answer w.r.t. Answered Qs (i.e., NTCIR Score)	Score w.r.t. Total Qs	Score w.r.t. Answered Qs
35	26	13	50.00	36 out of 97	36 out of 71

Performance of sJanta Over Phase 2 Question Set

Total Qs	# of Qs Answered	# of Correct Answer	% of Correct Answer w.r.t. Answered Qs (i.e., NTCIR Score)	Score w.r.t. Total Qs	Score w.r.t. Answered Qs
41	30	10	33.33	23 out of 100	23 out of 73

Framework



Architecture of sJanta

System Architecture

- **Question Analysis Module**
 - Question Parser
 - Wikifier
- **Answer Generator Module**
 - Hypotheses Reader
 - Document Collector
 - Answer Finder

Question Analysis Module

Question Parser:

- Parse Question Instruction, Context & Answer Choices.
- Generate Context Aware Hypotheses (i.e., Context + Answer Choices)

Wikifier:

- Run Wikification on Hypotheses.
- Generate Named Entities & Wikipage Links

Answer Generator Module

Hypothesis Reader:

- Detect Question Type: Affirmation or Negation.
- Generate Modified Hypotheses.

Document Collector:

- Extract Wikipage Articles that Contain Hypotheses.

Answer Finder:

- Extract Nouns & Verbs.
- Extract Relationships of Nouns & Verbs with Others Words in the Sentence.
- Generate Scores for Hypotheses.
- Rank Hypotheses & Select Answer.

Conclusion

- sJanta is a QA-Lab English sub-task system that used linguistic-based approach in question understanding and answer extracting.
- It implemented upon contemporary NLP tools and scalable for big knowledge source.

Future Work

- Extend sJanta for more types of questions.
- Extend hypotheses generation so that it can check multiple hypotheses.