sJanta: An Open Domain Question Answering System

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
  i) incorporate question context
  ii) automatically find question-specific text
  iii) check both syntactic and semantic similarities in selecting answer
  iv) leverage contemporary NLP tools
  v) 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 ...  

... From 1-4 below, choose the one sentence that correctly describes the underlined portion ...  

...  

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.