

Natural Language Processing (自然言語処理学)

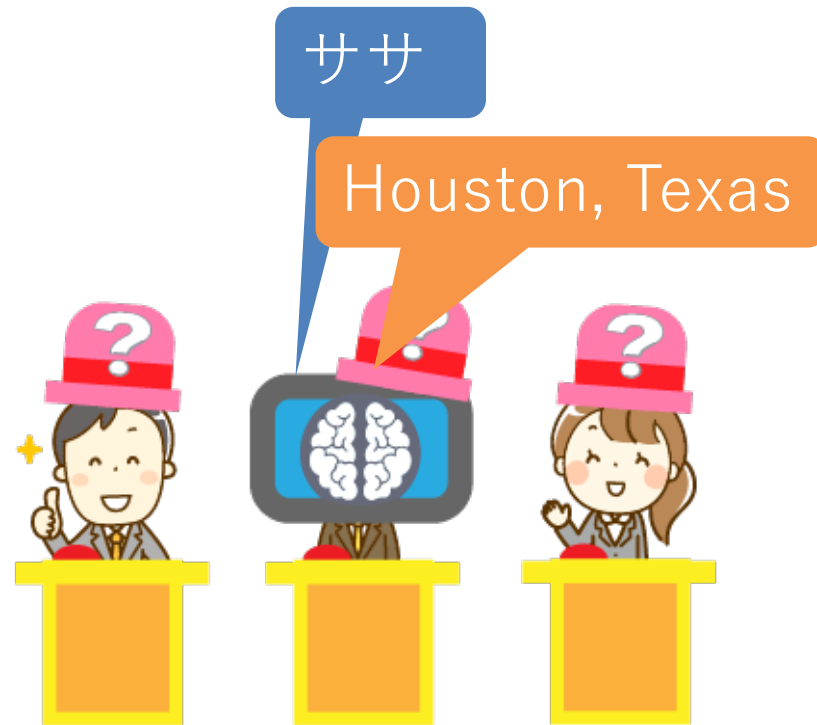
10-15. Lecture Project (Programming Project)

Tohoku University
Jun Suzuki

Making Question Answering System

童謡『たなばたさま』の
歌詞で、「さらさら」と
歌われる植物は何の葉？

In what city and state did
Beyonce grow up?



Brief Problem Setting

- Use natural questions that are often used in quiz TV shows and quiz events

Q : 童謡『たなは
たさま』の歌
詞で、「さらさ
ら」と歌われる
植物は何の葉？

Q2 : In what city and state
did Beyonce grow up?

- Generate the correct answer

A: "ササ"

Houston, Texas



OK?

Goal of this project

- Build question-answering systems that can answer natural questions in both English and Japanese
 - Build systems in a real setting
 - Experiencing (seeing) is believing!
- Learn difficulty in building NLP systems

Regulations (abstract)

- Develop your code on the provided VertexAI terminal
 - No support for other computational environments
 - Only accept systems in this environment as the final submission
- Need to generate answers by using provided LMs with “prompt” method
 - Asked to generate “GOOD” prompt for LMs
- Need to develop both English and Japanese systems

Final Report: Contents

- Your strategy (for improving the system)
 - What kinds of ideas did you implement?
 - How did you implement them (show us your code!)
- Result
 - Which idea was effective?
 - Why do you think the idea was effective?
- Analysis
 - Collect and categorize false instances into several types and explain each category

- Detailed Regulations

- Allow using any libraries, codes, and datasets except other QA systems (including other students' codes...) to generate “prompt”
 - Need to disclose all the citation information
 - Need to submit the entire system you developed

- Not allowed to include the correct answers in prompt
=> Not allowed to predict answers by using any other methods before asking answers to the provided LMs

- Not allowed to change or update the parameters for the provided LMs

- Not allowed to regenerate answers after the LMs generate texts (answers)
- Only allow to select correct answer from the generated text

- Need to answer the question in an off-line setting
=> Not allowed to connect internet while
answering the question

- Answer all the questions (in Test data) within 30 minutes on the provided VertexAI platform

- Confirm with TAs or teacher in advance if you would like to do something that you are not sure you are allowed to do under the regulations
=> This project is not a game of finding loopholes in the rules