

FelisCatusZero: A world history essay question answering system for the University of Tokyo's entrance exam

Kotaro Sakamoto
Yokohama National University
National Institute of Informatics
sakamoto@forest.eis.ynu.ac.jp

Takaaki Matsumoto
Carnegie Mellon University
tmatsumo@andrew.cmu.edu

Madoka Ishioroshi
National Institute of Informatics
ishioroshi@nii.ac.jp

Hideyuki Shibuki
Yokohama National University
shib@forest.eis.ynu.ac.jp

Tatsunori Mori
Yokohama National University
mori@forest.eis.ynu.ac.jp

Noriko Kando
National Institute of Informatics
The Graduate University for
Advanced Studies (SOKENDAI)
kando@nii.ac.jp

Teruko Mitamura
Carnegie Mellon University
teruko@cs.cmu.edu

ABSTRACT

FelisCatusZero is an open-source system to answer world history essay questions of the University of Tokyo's entrance exam in Japanese, which is based on extractive multi-document summarization.

KEYWORDS

Question Answering, Essay, World History, Japanese University Entrance Examination

1 INTRODUCTION

Question answering is widely noticed as an information access technology to answer various complex information needs. However, many previous researches of question answering have focused on comparatively simple questions, such as "Who is Donald Trump?". The real world questions sometimes differ from the previous researches in writing long background/circumstances before the question focus or including abstractive expressions. As the first step of tackling the real world questions, we are investigating question answering technologies of a university entrance exam's world history subject[1][2]. The University of Tokyo's entrance exam of world history has questions that a student has to write an essay with no more than 400 to 600 Japanese characters. Fig 1 shows an English translation of a question from the University of Tokyo's entrance exam. Fig 2 shows an English translation of a gold standard. In this paper, we present FelisCatusZero¹ which answers the essay questions. The system's essay answer generating algorithm is based on extractive multi-document summarization. We tested the system on a few mock exams of the University of Tokyo's entrance exam in Japanese. The results showed that the system got scores close to the average scores of students who were preparing for the exam.

We are currently living in an era of information revolution, and the pace of globalization is accelerating. Not only people and goods are flowing across oceans and borders with increasing frequency; information is being transmitted across the world almost in real time. Underlying these developments is the rapid progress that has been made in transportation and communication technologies.

...

Furthermore, these technological innovations are noteworthy for the parts they played in Western nations' invasions of Asia and Africa. For example, the Reuters news agency gathered information from around the world to help develop the international presence of the British Empire. But, on the other hand, global information sharing and accelerated migration facilitated by the development of transportation were also stimulating factors in the growth of local nationalism.

Write an essay explaining, in 225 English words (550 Japanese characters) or less, how developments in the means of transportation and communication prompted the colonization of Asia and Africa and heightened local nationalism. Use all nine keywords shown below at least once.

Suez Canal, steamship, Baghdad Railway, Morse code, Marconi, the Boxers, Russo-Japanese War, Persian Constitutional Revolution, Gandhi

Figure 1: A translation of an example of an essay question

... These technologies were used in Western advances into Asia and Africa. For example, the opening of the Suez Canal greatly reduced travel time between Europe and Asia, but the British controlled the canal to maintain a route to India. With regard to African policy, Britain supported Cecil Rhodes, who advocated a plan for connecting Cape Town to Cairo via rail and telegraph. Germany implemented a 3B policy, which included the construction of the Baghdad Railway, to advance into the Middle East. ...

Figure 2: A translation of an example of the gold standard

¹github.com/ktr-skmt/FelisCatusZero-multilingual/

2 KNOWLEDGE SOURCES

We use four high school textbooks and one glossary of world history subject as knowledge sources. Fig 3 shows an example of a textbook. Fig 4 shows an example of the glossary.

```
<DOC> <DOCNO>Y-JH-14-01-5</DOCNO>
<TITLE>Imperialism and Asian nationalist movements-Imperialism
and powerful imperialist countries' deployments-Russia</TITLE>
<TEXT>Since the 1890s, Russian capitalism has developed through the
capital import from France, and great industries have grown rapidly in
cities. . . . </TEXT> </DOC>
```

Figure 3: A translation of a paragraph example of a textbook

```
<DOC> <DOCNO>YamakawaWorldHistoryGlossary-5290</DOCNO>
<TITLE>Russia-Japan agreement</TITLE>
<TEXT>An agreement with the intention of the corporate protection
of the interests of both Japan and Russia after the Russo-Japanese War.
. . . </TEXT> </DOC>
```

Figure 4: A translation of an entry example e of the glossary

3 OUTLINE

Fig 5 shows the outline of FelisCatusZero. We input the question as an XML file. The system extracts the keywords which have to be included in the answer, the character limit and the time range from the question. The system retrieves documents including any keywords from the knowledge sources, extracts and groups sentences by each keyword. The system ranks the sentences based on the scores which mean to what extent they should be contained in answer. To generate an answer candidate, the system selects a sentence from each sentence group and sorts the selected sentences chronologically. The system removes answer candidates if they exceed the character limit. The system ranks the answer candidates by the summation of sentence scores. The system chooses an answer candidate with the highest score as the final answer. Finally the system outputs the answer as an XML file.

4 EXPERIMENT

We have let the system answer a question of a mock exam for the University of Tokyo’s entrance exam four times so far. Every time the system answered a question, a human expert in world history evaluated the system answer and gave the system some minor updates.

4.1 Results

Fig 6 shows translations of an example of the system answer. Table 1 shows the four-time expert evaluation results.^{2 3}

Q	SCORE	AVE. SCORE OF STUDENTS	
1st	4	7.3	out of 28
2nd	9	4.3	out of 26
3rd	3	4.6	out of 20
4th	10	9.2	out of 20

Table 1: Evaluation results

²Note that the system could not extract time range from the first question due to an XML tagging error.

³In the third question, the system could get 6 by a minor update.

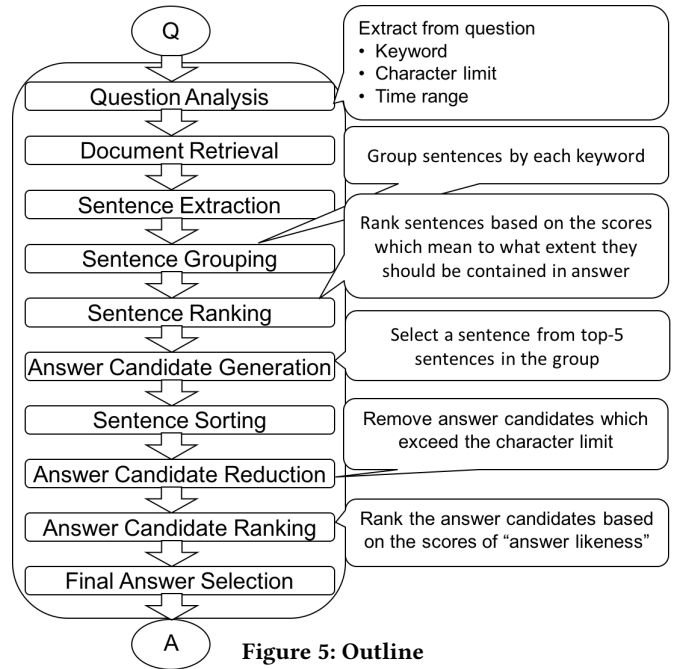


Figure 5: Outline

4.2 Discussions

The system can select many relevant sentences from knowledge sources, though the retrieval accuracy is not good enough to get the perfect score. Also, we think that the system needs an essay structure correction.

```
. . . The opening of the Suez canal in 1869 encouraged the advance-
ment of the Great powers into Africa, which became the target of parti-
tioning disputes. The Baghdad railway was targeted by many countries,
but Germany obtained the rights of construction in 1899, and in 1903 it
founded a society. In the end, the Baghdad railway was only partially
opened.
```

Figure 6: A translation of an example of the system answer

5 CONCLUSIONS

We presented an open-source system FelisCatusZero to answer world history essay questions of the University of Tokyo’s entrance exam in Japanese. It is based on extractive multi-document summarization. We tested the system on a few mock exams of the University of Tokyo’s entrance exam in Japanese. The results showed that the system got scores close to the average score of students who were preparing for the exam. However, to get a higher score, we still have much room for improvement, not limited to accuracies of sentence extraction and extracted sentence combination.

REFERENCES

- [1] Kotaro Sakamoto, Madoka Ishioroshi, Hyogo Matsui, Takahisa Jin, Fuyuki Wada, Shu Nakayama, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. 2016. Overview of the NTCIR-12 QA Lab-2 Task. In *Proceedings of The NTCIR-12 Conference*.
- [2] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. 2016. Overview of the NTCIR-12 QA Lab-2 Task. In *Proceedings of The NTCIR-12 Conference*.