

# MTMT in QALab-3:

## World History Essay Question Answering System that Utilizes Textbooks and Open Knowledge Bases

Takaaki Matsumoto<sup>†‡</sup>, Teruko Mitamura<sup>†</sup> (<sup>†</sup> Carnegie Mellon University, <sup>‡</sup> SOC Corporation)

### Introduction

English NTCIR QALab tasks are hard since there are no knowledge bases (KB) equivalent to Japanese tasks. This study addresses a new approach for KB building and our essay question answering system.

### Translation of Textbooks

Because NTCIR QALab tasks target the world history subject of Japanese university entrance examinations, they provide Japanese world history textbooks for knowledge bases (KB). In English task, we can use the machine translated (MT) textbooks (statistical MT, in 2015).

We tried the latest neural MT to the Japanese textbooks, however, evaluation score got worse compared with statistical MT (from 0.063 to 0.056 in ROUGE-1, for NTCIR QALab-2 phase-1 dataset). The latest neural MT mistranslated some difficult terms.

Japanese Term	Google Translate	Bing Translator	Wikidata	Answer
林則徐	Hayashi Noriro	the zexu	Lin Zexu	Lín Zéxú
欽差大臣	Minister of Ginza	Minister of the Qin	Imperial Commissioner	Imperial Commissioner
キリスト教綱要	Christianity requirements	Christian elements	Institute of the Christian Religion	Institute of the Christian Religion

We propose the following two-step approach to build knowledge bases by utilizing Wikidata.

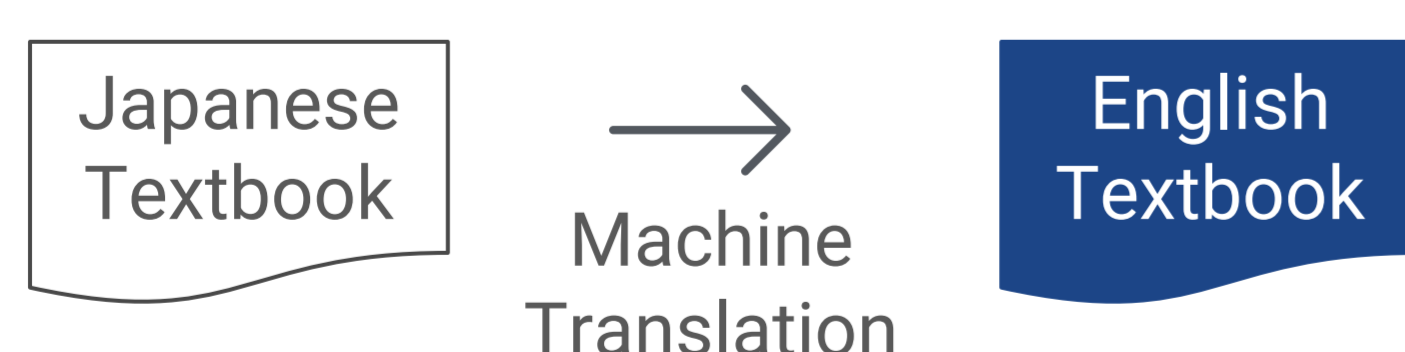
#### 1. Pre-translation

To translate nouns correctly, the nouns in the Japanese textbooks are translated into English by a bilingual world history term corpus which is developed by using Wikidata. As a result, Japanese-English mixed texts are generated.

#### 2. Machine Translation

The pre-translated texts are translated by the latest neural MT (Bing Translator), and full English texts are obtained.

Conventional Method: Direct MT



Proposed Method: Two-Step MT



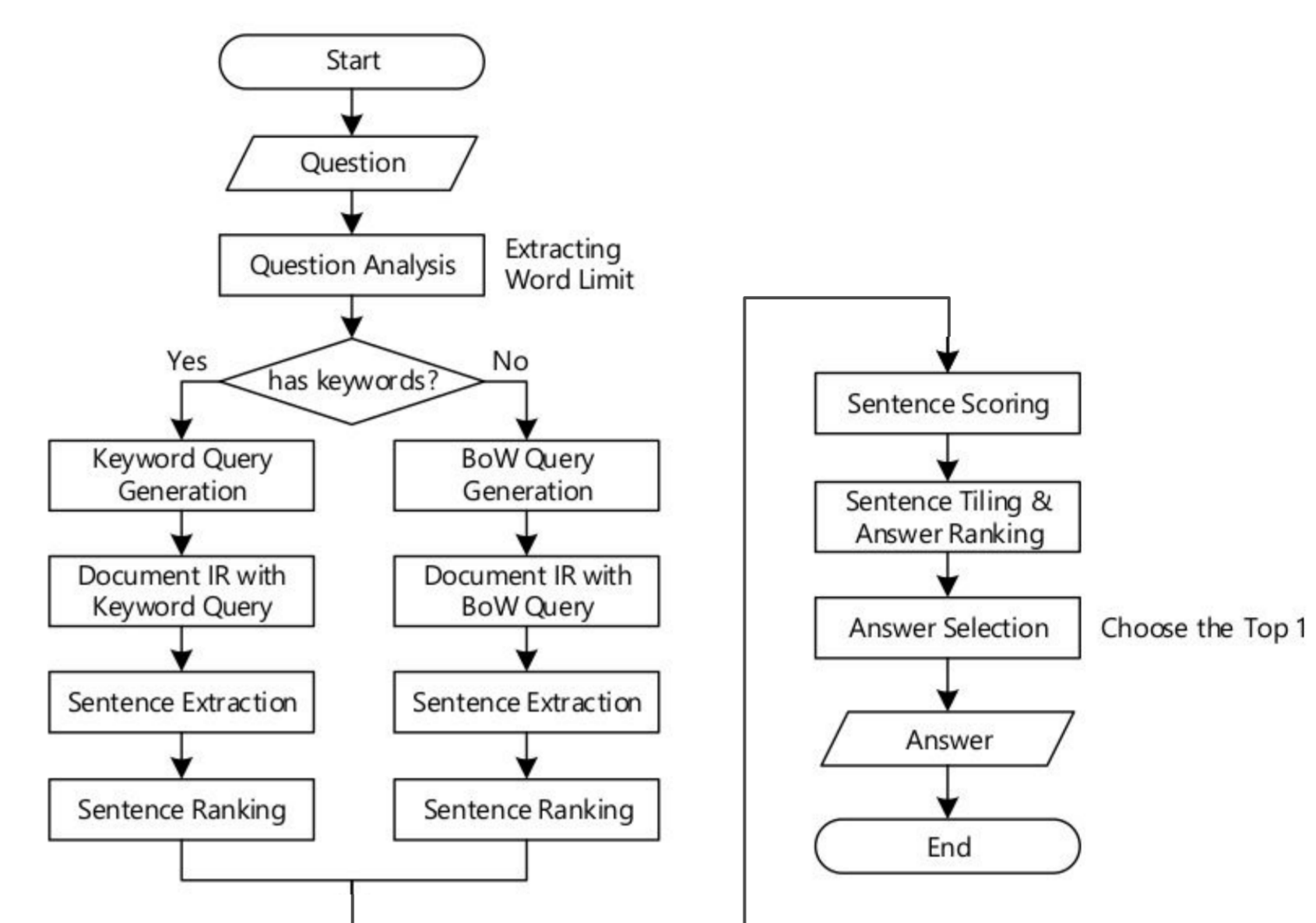
### Adding an English Textbook

Although its content is not equivalent to the Japanese textbooks, we added an English open textbook (Boundless.com) since ROUGE-1 improved.

### Baseline and Sentence Scoring

Our system is a fork of Sakamoto's multilingual QA system (FeliscatusZero-multilingual [1]) for world history essays.

[1] K. Sakamoto, T. Matsumoto, M. Ishioroshi, H. Shibuki, T. Mori, N. Kando, and T. Mitamura. Feliscatuszero: A World History Essay Question Answering for the University of Tokyo's Entrance Exam. In Proceedings of Open Knowledge Base and Question Answering Workshop at SIGIR, 2017.



Extracted sentences from the textbooks are ranked by calculating word level similarity to given keywords and retrieved nouns from the question.

$$\text{Score} = \sum_{i=1}^m \frac{\max(w_i \cdot k_1, w_i \cdot k_2, \dots, w_i \cdot k_n)}{\log m}$$

$m$ : Number of words of a sentence  
 $k_n$ : GloVe vector of the  $n$ -th keyword  
 $w_i$ : GloVe vector of the  $i$ -th word in a sentence

### Evaluation and Discussion

Our system took the first place in ROUGE-1 and ROUGE-2 of the four end-to-end submissions.

However, the number of grammatical and semantic errors were relatively greater than those of other submissions. We consider that Japanese-English mixed texts sometimes cause problems during MT process.

	CASE		STEM		STOPWORD	
	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2	ROUGE-1	ROUGE-2
CMUQA1	0.1231	0.0159	0.1257	0.0164	0.0347	0.0065
CMUQA2	0.1251	0.0160	0.1278	0.0166	0.0371	0.0066
CMUQA3	0.0768	0.0105	0.0968	0.0184	0.0594	<b>0.0129</b>
Forst1	0.1057	0.0086	0.1057	0.0086	0.0183	0.0022
IMTKU1	0.1187	0.0133	0.1239	0.0156	0.0457	0.0013
IMTKU2	0.0065	0.0000	0.0101	0.0001	0.0126	0.0001
<b>MTMT1</b>	<b>0.1717</b>	<b>0.0180</b>	<b>0.1841</b>	<b>0.0207</b>	<b>0.0615</b>	0.0054
MTMT2	0.1546	<b>0.0190</b>	0.1630	0.0197	0.0465	0.0020

### Conclusion

By utilizing the newly translated Japanese textbooks with Wikidata and neural MT, and the English open textbook, our system took the first place in ROUGE 1 and 2 of the four end-to-end submissions.