KSU Team's QA System for World History Exams at the NTCIR-13 QA Lab-3 Task

Tasuku Kimura, Ryo Tagami, Hisashi Miyamori (Kyoto Sangyo University, Japan)

Affiliation: 京都産業大学 KYOTO SANGYO UNIVERSITY (むすんで、うみだす。)
Conference: NTCIR Workshop
Task: QA Lab

ABSTRACT

Multiple-choice automatic answering systems in QA Lab-2 were improved by implementing query generation methods in accordance with the answer types.

Named-entity QA systems were designed that focus on the **category prediction using word prediction models** and the **evaluation score based on the graph of dependency relations**.

Essay automatic answering methods were proposed that combines the **document retrieval depending on the instructions of how the essay should be summarized**, and the **knowledge sources constructed from various simple sentences**.

Please notice that this poster only describes the named-entity and essay question subtasks and results for Phase-2 due to the limitations of space.

PROPOSED METHOD

Named-entity

Input questions are classified as either factoid type or slot-filling type.

Question Analysis

Factoid
- Estimating the focus based on the rules
- Estimating the answer category based on the rules
- Dividing of the question

Slot-Filling
- **Estimating the answer category based on the word prediction model** ▶ POINT1

Factoid & Slot-Filling
- Generating the query q for document retrieval

Document Retrieval

Factoid & Slot-Filling
- Document retrieval to the knowledge source (using q)

Answer Candidate Extraction

Factoid & Slot-Filling
- Extracting the answer candidate words from document

Answer Candidate Evaluation

Calculating scores of each answer candidate word using following indicators

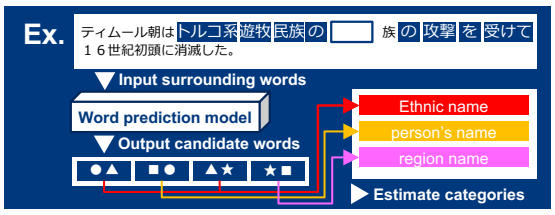
Factoid
- Okapi BM25 rank
- Focus match judgement
- **Graph minimum distance score** ▶ POINT2

Slot-Filling
- Okapi BM25
- Backward match judgement

Factoid & Slot-Filling
- Non-existence word judgement
- Category mismatch judgement

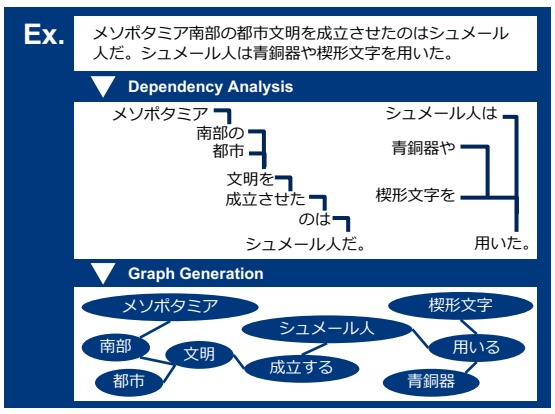
POINT1 Category Estimation using Word Prediction Model

For the slot-filling type questions, a word prediction model was constructed which **estimates the center word from the surrounding words** of the filling part.
When inputting the surrounding words of the slot into the model, the set of the center word candidates is output.
Categories assigned to each word in the set are collated, and all matched is set as the category of the question.



POINT2 Graph Minimum Distance Score

This score is calculated for each candidate word using the graph generated for each question from the **dependency analysis** on the relevant documents obtained by Document Retrieval module.



Specifically, for each node of candidate on the graph, the minimum distance from the node of the word used in the query generated in Question Analysis module is calculated using Dijkstra's algorithm, as this score.
That is, **the score becomes smaller as the words used in candidate word and query exist closer to each other on the graph**.

Essay

Questions are classified as either complex-essay type or simple-essay type.

Question Analysis
- Estimating the instruction type of how the essay should be summarized
- Estimating the category of the question focus
- Extracting a set of query words in accordance with essay types

Document Retrieval
- **(1) Creating the simple-sentence-oriented knowledge source** ▶ POINT

We introduce the **simple-sentence-oriented knowledge sources where the surface expressions are simplified in various ways compared to their original sentences**, so that the system can obtain concise answer candidate sentences containing only the content which should be included in the answer.

- Creating a correspondence table of spelling variations for world history
- Searching the document sets depending on the instructions type

Answer Candidate Extraction
- Extracting the answer candidates based on MMR
- Merging the simple-sentence-oriented answer candidates

Answer Candidate Evaluation
- (2) Creating the answer necessarily including specified phrases
- Sorting the answer candidates in the original order in the knowledge source

POINT Creating the simple-sentence-oriented knowledge source

Normal sentences are often any one of simple sentence, compound sentence and complex sentence including components other than subjects or predicates.
When it is assumed that "one meaning is represented by a pair of a subject and its predicate", **the complex sentence and the compound sentence can be said to be a complicated sentence having plural meanings**.

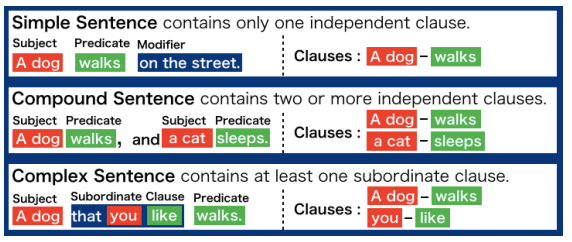


Figure 1: Instructions of sentences by clauses (a relationship between subject and predicate).

By contrast, a **simple sentence tend to be short and simple**, because it basically only contains a subject and a predicate. Therefore, as shown in Fig. 2, we introduce a **simplified surface representation by converting a simple sentence**, complex sentence or compound sentence which include components other than subjects and predicates, into one or more simple sentences which have less components than the original sentence.

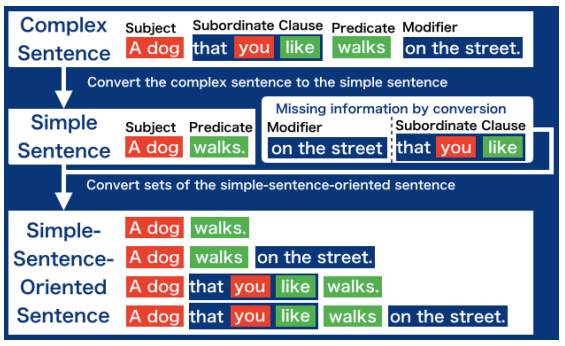


Figure 2: Examples of converting a complex sentence into a simple sentence or simple-sentence- oriented sentences.

RESULT & DISCUSSION

Named-entity

As shown in Table 1, the main difference of each RUN is whether the indicator based on the graph minimum distance score was used.
Table 2 shows the result of their comparison and it indicates that the **correct answer rate of the RUN with the graph minimum distance score becomes slightly higher** that of the RUN using the value of BM25.

Table 1 : Indicators used in each RUN.

Used indicators	Factoid RUN1	Factoid RUN2	Factoid RUN3	Slot-filling[1] RUN1 - 3
Okapi BM25 rank			✓	
Focus match	✓	✓	✓	
Graph minimum distance		✓[2]	✓	
Okapi BM25	✓	✓		✓
Backward match				✓
Non-existence word	✓	✓	✓	✓
Category mismatch	✓	✓	✓	✓

Table 2 : Results.

System Id	Accuracy
RUN1	0.30 (23/77)
RUN2	0.29 (22/77)
RUN3	0.31 (24/77)

1 Slot-filling type questions existed only in the test data for Phase-2.
2 This indicator was used only when score of candidates were equal after being evaluated using other indicators.

Examining the cases where the questions were correctly answered, **it was confirmed that the distance between each named entity and the correct word was relatively small**.

However, it was also confirmed that the system tends to give incorrect answers in the following cases: when there were few named entities in the question, when the named entity in the question does not exist on the knowledge source in the first place, or when the distance between the named entities on the graph happens to be long.
As a cause of these problems, **insufficient correspondence to spelling variations of words of each node is considered**, because the collation is based on the exact match of the surface strings.
Therefore, it is expected that these problems are alleviated by **normalization with thesauruses and/or by introduction of partial match**.

Essay

In Phase-2, three types of systems were constructed depending on whether or not to (1) use the simple-sentence-oriented knowledge source and (2) use the answer generation method necessarily including the specified phrases.

Table 3: The comparison of the system configuration for the essay questions.

System Id	(1)	(2)
KSU-ESSAY-1	Tokyo-Original	No
KSU-ESSAY-2	Tokyo-Original	Yes
KSU-ESSAY-3	Tokyo-Simple	Yes

Table 4: Results of our runs for Phase-2 in essay end-to-end task.

System Id	ROUGE-1	ROUGE-2
KSU-ESSAY-1	0.312	0.060
KSU-ESSAY-2	0.317	0.063
KSU-ESSAY-3	0.348	0.096

Table 4 shows the ROUGE-N of KSU-ESSAY-2 was a little higher than that of KSU-ESSAY-1.

This improvement of ROUGE-N is considered to be achieved because the **sentences including the correct answer were successfully** selected by converting these sentences which could not have been selected due to the character limit of the question, into the simple sentences.

However, it is **necessary to improve the method of converting to the simple-sentence-oriented sentences**, because the proposed knowledge sources contain several unnatural sentences without sufficient semantics.

Table 4 indicates that the ROUGE-N of KSU-ESSAY-3 was higher than that of KSU-ESSAY-2.

It was confirmed that the **essay generated by KSU-ESSAY-3 contained the appropriate sentences as the answer**, because it implemented the method of using the candidate sentences always including the specified phrases.

These results showed the similar tendency to the characteristics of the original method proposed by Sakamoto et al.[1]

[1] K. Sakamoto et al., Forst: Question answering system for second-stage examinations at ntcir-12 qa lab-2 task.