KSU Team's Multiple Choice QA System at the NTCIR-12 QA Lab-2 Task

Tasuku Kimura, Ryosuke Nakata, Hisashi Miyamori

Kyoto Sangyo University, Japan

Abstract

This poster describes the systems and results of the team KSU for QA Lab-2 task in NTCIR-12. In each phase of the task, we developed three automatic answering systems for world history questions in the National Center Test for University Admissions. In order for QA systems using document retrieval to answer questions correctly, it is important to estimate exact question types, and to utilize knowledge sources and query generation methods in accordance with these types. Therefore, we designed systems that focus on knowledge sources and query generations using the underlined texts in given exams. Scores of the formal runs were 20 correct answers(49%) and 68 points with KSU-JA-02@PH1 system in Phase-1, 26 correct answers(41%) and 70 points with KSU-JA-01@PH2 system in phase-2 and 14 correct answers(39%) and 38 points with KSU-JA-01@PH3 system in phase-3.

Please note that this poster only describes the systems and results for Phase-3 due to the limitations of space.

Our QA systems for Phase-3

Table 1 shows subtasks in which we participated in QA Lab-2.

Table1: Summary of subtasks in which our team participated

Test	Questions	Lang
Phase-1	National Center Test	Japanese
Phase-2	Mock Examination of National Center Test	Japanese
Phase-3	National Center Test	Japanese

QA systems were implemented by modifying and improving the baseline system provided by the organizers for QA Lab-2. In order for QA systems using document retrieval to answer questions correctly, it is important to estimate exact question types, and to utilize knowledge sources and query generation methods in accordance with these types. Figure2 shows Approaches improved in Phase-3.

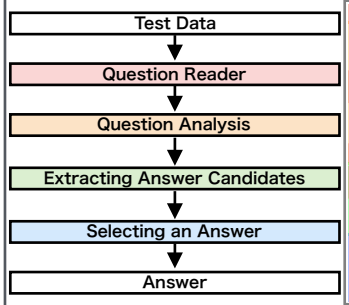


Figure1: Basic system configuration

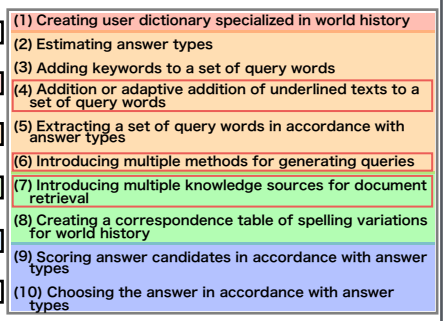
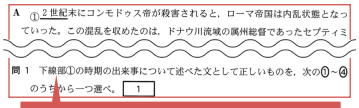


Figure2: Approaches improved in Phase-3

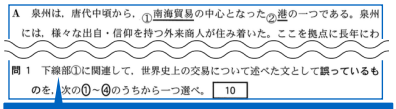
Adaptive utilization of underlined texts ((4) in Fig.2)

Some underlined texts were unnecessary to answer sub-questions (Figure 4). Therefore, we developed the classifier which determined whether the underlined texts were necessary or not. Those were adaptively added to the set of query words



The underlined text "2nd century" is necessary to answer it.

Figure3: When the underlined texts are necessary to answer the question



The underlined text "the trade in the southern sea" is not necessary to answer it.

Figure4: When the underlined texts are unnecessary to answer the question

Query generation and knowledge sources

Query generation ((6) in Fig.2)

In addition to basic query methods, the CLASS query method transforms the query words to their super-ordinate concept labels and searches only from documents having particular labels.

A set of query words

Mughal Empire, Persia, official language

The generated query in each approach

OR query

Mughal Empire OR Persia OR official language

AND query

+(Mughal Empire AND Persia) OR official language

VS query

Query-1: OR query
Query-2: AND query

CLASS query

(nation) AND (Mughal Empire OR Persia OR official language

Knowledge sources ((7) in Fig.2)

It is often the case that texts of sub-questions and choices are sentence. Therefore, It is assumed that using the article-base documents is appropriate. Then, knowledge sources were built by registering one article, paragraph and sentence as a document.

Table2: Knowledge source configurations

Knowledge source	Configuration method
WD	Wikipedia per article
WP	Wikipedia per paragraph
WS	Wikipedia per sentence
T	Textbook per paragraph
TWD	WD + T
TWP	WP + T
TWS	WS + T
Evt	Event ontology EVT per NE

Results and discussions in Phase-3

Table 3 shows the difference of the system configurations for Phase-3 in (4) adaptive utilization of underlined texts, (6) query generation and (7) knowledge sources. The systems 3-3, 4-1 and 4-3 are identical to priority-01, priority-02 and priority-03 in Fig 5, respectively.

Table3: Systems developed with different configurations.

		(7) Knowledge sources							
		T	WD	WP	WS	TWD	TWP	TWS	Evt
(6) Query generation	OR	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8
	AND	2-1	2-2	2-3	2-4	2-5	2-6	2-7	2-8
	VS	3-1	3-2	3-3, 4-3	3-4	3-5	3-6	3-7	3-8
	CLASS			(4) Adaptive utilization of underlined texts					4-1

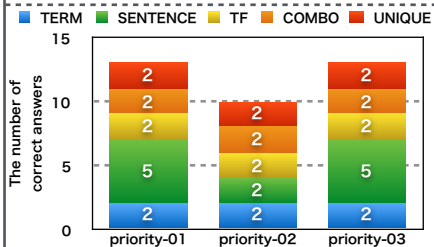


Figure5: Results of our runs for Phase-3

- Comparison between priority-01 and priority-02

  The accuracy of priority-02 was low. Also, for priority-02, the retrieval results often showed no difference for each of the choices.

- Comparison between priority-01 and priority-03

  The accuracy of priority-01 and priority-03 showed no difference. However, for priority-03, there was a tendency for the score of the queries in priority-03 to become higher than that in priority-01.

Systems with other configuration in Phase-3

We investigated into systems with other configuration for Phase-3. In table 3, the system 3-3 is identical to priority-01. The systems 4-1 and 4-3 are identical to priority-02 and priority-03, respectively. For an experiment, each system answered the test data for Phase-3.

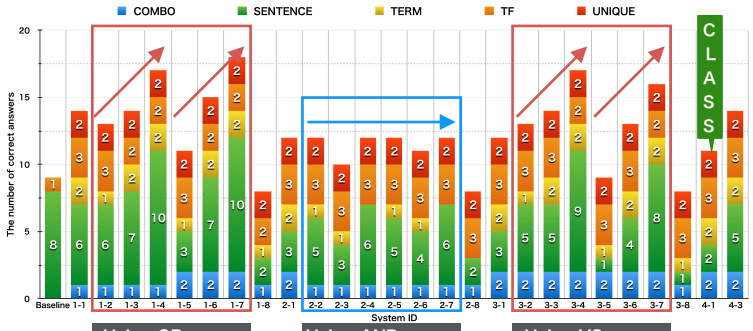


Figure6: Comparison of the number of correct answers between the systems with different configurations

- The smaller the granularity of documents became, the higher the accuracy of systems got.
  → The sentence-based documents seem to have more documents which are more similar to those descriptions used in the answer choices
- The difference of the accuracy between the knowledge sources are small.
  → The systems using the query seem to have more often selected the same answer as a result

Conclusion

- We designed QA systems that focus on knowledge sources and query generations.
- The set of sentence-based documents was effective on the sub-questions having the answer type of SENTENCE.
- Appropriate introduction of the dependency analysis and the semantic analysis remain for future work.