SLQAL at the NTCIR-12 QALab-2 Task

Shin Higuchi
Waseda University
shin1213@ruri.waseda.jp

Tetsuya Sakai
Waseda University
tetsuya@waseda.jp

ABSTRACT

SLQAL (Waseda University **S**akai **L**aboratory **QAL**ab team) participated in Phase-1 and Phase-3 of the NTCIR-12 QALab-2 Japanese subtask. This paper briefly describes our approaches. Our runs scored 25 points in Phase 1 and 35 points in Phase 3. An initial failure analysis shows that our system performs particularly poorly for Type-T questions as well as questions that require time expression processing. This work was done as a bachelor's thesis of the first author of this paper.

Team Name

SLQAL

Subtasks

Japanese Subtask (Phase-1 and Phase-3)

Keywords

answer selection; question analysis; question answering; knowledge base.

1. INTRODUCTION

SLQAL (Waseda University **S**akai **L**aboratory **QAL**ab team) participated in Phase-1 and Phase-3 of the Japanese subtask of NTCIR-12 QALab-2 [1]. This paper briefly describes our approaches. Our runs scored 25 points in Phase 1 and 35 points in Phase 3. This work was done as a bachelor's thesis of the first author of this paper.

The remainder of this paper is organised as follows. Section 2 briefly describes the SLQAL system, and Section 3 presents our official results. Finally, Section 4 concludes this paper.

2. THE SLQAL SYSTEM

Figure 1 shows the configuration of the SLQAL system. Below, we briefly describe how some of the modules work.

2.1 Question Analysis

The Question Analysis module is composed of two submodules: question and answer type analysis, and key term extraction.

2.1.1 Question and Answer Type Analysis

Given a question XML file, this submodule determines the question type and the answer type. The question type is one of the following:

Type-T "Which one is true?"-type question;

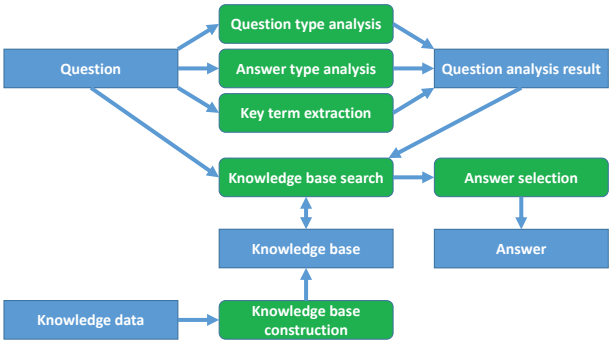

Figure 1: Components of the SLQAL system.

Type-F "Which one is false?"-type question;

Type-P Questions of the following type: "Regarding a pair of statements <A,B>, which of the following is correct? <true, true>, <true, false>, <false, true>, <false, false>?

As for the answer types, we consider six types: DEFINITION, PERSON, LOCATION, NATION, REASON, TIME.

This submodule is rule-based. For example, if the question contains expressions such as "correct" and "choice, " the question type will be determined as Type-T.

2.1.2 Key Term Extraction

This submodule detects important terms in the question, based on a named entity dictionary that we constructed in advance from world history textbooks. It also computes an idf score for each key term for the purpose of answer selection (Section 2.3).

2.2 Knowledge Base Construction

This is a module that creates a world history knowledge base offline, using two world history textbooks. We employ Mecab[1] for analysing the textbooks and obtain *fact tuples* of the form: <subject, verb, object, time, location>. To allow fuzzy matching between the question text and the fact tuples, we implemented the following two types of expansion.

2.2.1 Time Expression Expansion

Whenever we detect time expressions such as "first half of the century," "around Year XXXX," and "Era," we convert them to specific time ranges using pre-defined rules.

[1] `http://taku910.github.io/mecab/`

2.2.2 *Verb Expansion*

Using a verb thesaurus that we constructed from Weblio Synonym Dictionary[2], we also perform verb expansion. Figure 2 shows how the verb *katsu* (to win) can be expanded using our approach.

勝つ

意義素	類語
コンテストまたは競争の勝者である	勝利・得る・受賞・優勝・勝ち取る・制覇・勝ちとる・勝ちえる・捷利・獲る・勝ち得る・勝取る・贏ち得る
勝利を収める	打負かす・仆す・打倒す・負かす・討ち破る・剋する・打ち負かす・打ち克つ・討破る・打ち倒す・克する・克つ・撃ち破る・打破る・打ち勝つ・破る・打勝つ・倒す

Figure 2: Example of verb expansion

2.3 Answer Selection

Our answer selection module is based on two strategies. The first strategy tries to select the correct answer based on matching between fact tuples extracted from the choices and those from the knowledge base. The second strategy tries to select the correct answer based on matching between key terms extracted from the choices and key terms extracted from the textbooks. If the two strategies disagree, the system makes the following decisions: if the question type is Type-T or Type P, then follow the decision of the first strategy; otherwise, follow the decision of the second strategy.

3. RESULTS

Figure 3 visualises the overall performance of our system in comparison with other Japanese runs. The blue bars represent the total scores, while the orange bars represent the number of correctly answered questions. The top performer (NUL) scored 68 points by answering 24 questions correctly; our system (SLQAL) scored 35 points by answering 13 questions correctly.

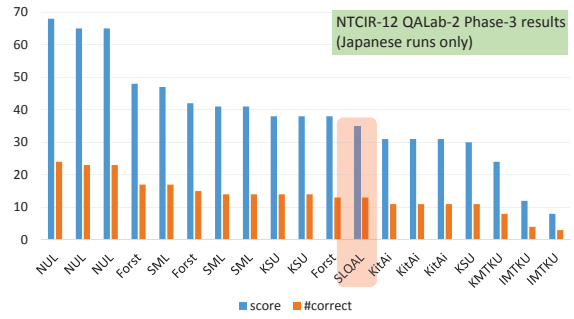

Figure 3: Official Phase-3 results (Japanese runs only).

Table 1 compares our system with the top performer shown in Figure 3, namely, a run from NUL. The first column shows the questions IDs; the second column shows the question type (Type-T, Type-F or Type-F); the third and the fourth columns show whether the question requires the system to process time expressions and graphical figures. The NUL and SLQAL columns show which multiple choice questions were answered correctly by these systems.

[2]http://thesaurus.weblio.jp/category/wrugj

The main results of the above comparison are as follows:

- Of the 23 Type-T questions, NUL answered 17 questions correctly while SLQAL answered only 7 questions correctly;
- Of the 7 Type-F questions, NUL answered 4 questions correctly while SLQAL answered 3 questions correctly;
- Of the 6 Type-P questions, both NUL and SLQAL answered 3 questions correctly;
- Of the 15 questions that require time expression processing, NULL answered 10 questions correctly, while SLQAL answered only 5 questions correctly;
- As for the one question that requires figure understanding, NULL managed to answer it correctly while SLQAL did not.

Table 1: Per-question Comparison with the top performer.

	T/F/P	time	figure	NUL	SLQAL
A02	T			CORRECT	
A08	F			CORRECT	CORRECT
A07	T	YES		CORRECT	CORRECT
A36	P	YES			CORRECT
A23	T			CORRECT	CORRECT
A04	T	YES		CORRECT	CORRECT
A13	F			CORRECT	
A20	T				
A34	T			CORRECT	
A25	P			CORRECT	
A03	F	YES			
A09	T				
A27	T			CORRECT	CORRECT
A32	F			CORRECT	
A28	T			CORRECT	
A14	T			CORRECT	
A12	T	YES		CORRECT	
A11	T			CORRECT	
A35	T	YES		CORRECT	
A18	T	YES		CORRECT	CORRECT
A26	T			CORRECT	
A10	P			CORRECT	CORRECT
A05	P	YES		CORRECT	
A16	T				
A15	T	YES		CORRECT	
A31	F	YES		CORRECT	
A29	F				CORRECT
A33	P				CORRECT
A21	T				CORRECT
A06	T	YES			
A01	T	YES		CORRECT	
A19	T			CORRECT	CORRECT
A17	P	YES			
A22	T				
A24	F	YES			CORRECT
A30	T	YES	YES	CORRECT	
		#correct		24	13
		#correct T		17	7
		#correct F		4	3
		#correct P		3	3
		#correct time		10	5
		#correct figure		1	0

NUL outperforms SLQAL for 16 questions, while SLQAL outperforms **NUL** for 5 questions. The difference between these two runs is statistically significant according to the sign test (exact p-value= 0.027; if a Z-test approximation is used, $Z_0 = 2.182$, p-value= 0.029).

4. CONCLUSIONS

This paper briefly describe our approach to Phase-1 and Phase-3 of the NTCIR-12 QALab-2 Japanese subtask. Unfortunately, our official results were not satisfactory. An initial failure analysis that compared our system against a top performer shows that our system performs particularly poorly for Type-T questions and question that require time expression processing.

This work was done as a bachelor's thesis of the first author of this paper.

5. REFERENCES

[1] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshionobu Kano, Teruko Mitamura, Tatsunori Mori, Noriko Kando. Task Overview for NTCIR-12 QA Lab-2, Proceedings of NTCIR-12, 2016.