SLQAL at the NTCIR-13 QA Lab-3 Task

Kou Sato
Waseda University
nn.ks-11897@ruri.waseda.jp

Tetsuya Sakai
Waseda University
tetsuya@acm.org

ABSTRACT

The SLQAL team participated in Phase-1, Phase-2 and Research run of the NTCIR-13 QA Lab-3 Japanese subtask(National Center Tests(Multiple-Choice type questions)) [2]. This paper describes our approraches and results in Phase-1 and Phase-2.

Team Name

SLQAL

Subtasks

Japanese subtask (Phase3:National Center Tests(Multiple-Choice type questions))

Keywords

Question Answering

1. INTRODUCTION

The SLQAL team participated in Phase-1, Phase-2 and Research run of the NTCIR-13 QA Lab-3 Japanese subtask (National Center Tests (Multiple-Choice type questions)) [2]. This paper describes our approraches and results in Phase-1 and Phase-2.

2. THE SLQAL SYSTEM

Figure 1 shows the overview of the SLQAL system. This system consists of three submodules: Keyword extraction, Query Generation and Answer Selection.

2.1 Keyword extraction

This submodule first parses the question passages and choices given as XML data, and performs morphological analysis using MeCab[1].

To let the system select the correct choice, we utilise not only the choices but also the problem statements and the underlined parts in them. This submodule stores all underlined phrases in the passage and then replaces all references to the underlines phrases within the question with the actual phrases. It then extracts a set of nouns from the problem statement, which we refer to as *question_nouns*, as well as a set of nouns from the j-th choice, which we refer to as $choice\_nouns_j$.

2.2 Sentence Retrieval

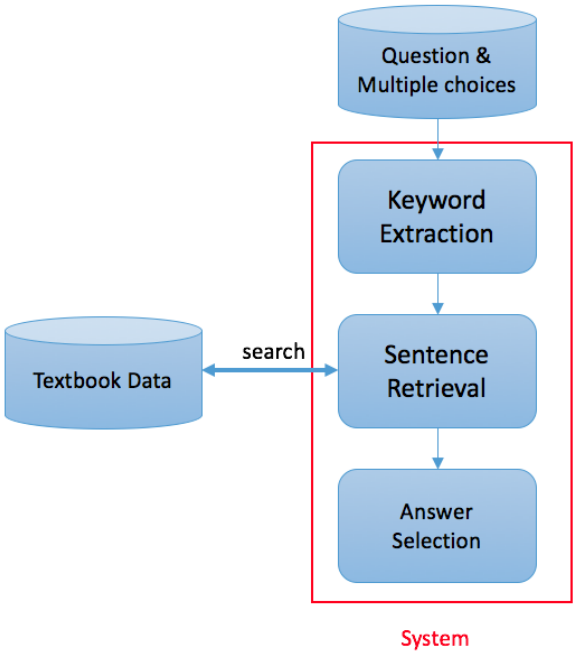


Figure 1: Overview of the SLQAL System

This submodule generates a query to search the textbook data provided by the organisers, and retrieves sentences using ElasticSearch[1]. The textbook data were indexed using the period as the sentence marker, thus treating each sentence as a document.

Given a problem, for each choice j, the submodule formulates a search query of the form $validation\_query = question\_nouns \cup choice\_nouns_j$. ElasticSearch then returns sentences as the search result, where the sentence at rank i has a score which we denote as $score_{i,j}$.

2.3 Answer selection

This submodule selects the estimated answer from the multiple choices by utilising $score_{i,j}$, the score of the sentence retrieved at rank i for the j-th choice. It simply averages the top k scores as follows, and selects the choice with

[1]Elasticsearch: RESTful, Distributed Search Analytics - Elastic. https://www.elastic.co/jp/products/elasticsearch.

Year (Phase)	Score
2012 (Phase-1)	29
2013 (Phase-1)	36
2014 (Phase-2)	23

Table 1: SLQAL's score in Phase-1 and Phase-2

the highest average score:

$$average\_score_j@k = \frac{\sum_{i=1}^{k} score_{i,j}}{k} \ . \quad (1)$$

We let $k = 5$ for our submission.

3. RESULTS AND DISCUSSIONS

Table 1 shows our official Phase-1 and Phase-2 results. Table 2 provides the details of our *average_score*s. Each column means:

ID
Problem ID.

T/F
Whether the system is required to choose the choice that is factually correct (T) or to choose the one that is factually false (F).

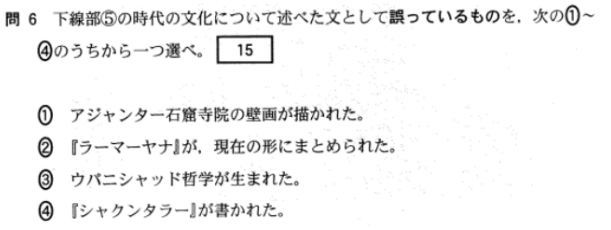
gold
Gold-standard answer ID.

system
The choice returned by the system.

1 ∼ 6
$average\_score_j$ for $j = 1, \ldots, 6$.

absdiff
Absolute difference between the score of the gold choice and that of the system's choice.

Figure 2 shows A15 from the National Center Test 2014. We picked this question because the absdiff value is very high in Table 2. As the table shows, while the gold choice is Choice 3 ($average\_score = 12.93$), our system returned Choice 2 ($average\_score = 5.92$) (This question is type F, so the system returns a choice with the lowest $average\_score$).


問 6　下線部⑤の時代の文化について述べた文として誤っているものを，次の①～④のうちから一つ選べ。 15

① アジャンター石窟寺院の壁画が描かれた。
② 『ラーマーヤナ』が，現在の形にまとめられた。
③ ウパニシャッド哲学が生まれた。
④ 『シャクンタラー』が書かれた。

Figure 2: Problem No.15 of Center test in 2014

One possible problem with our approach is that we use sentences as the indexing unit, which may cause unsuccessful matches with the validation query due to lack of context. We therefore tried using paragraph indexing and considering the proximity between words prior to the Research Run period, but without success.

4. CONCLUSIONS AND FUTURE WORK

The SLQAL team participated in Phase-1, Phase-2 and Research Run of the NTCIR-13 QA Lab-3 Japanese subtask (National Center Tests Multiple-Choice type questions). Our official Phase-2 score is 23.

One possible direction for future work is applying anaphora resolution to each sentence from the textbook for the purpose of overcoming the query-document mismatch problem.

5. REFERENCES

[1] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *EMNLP*, volume 4, pages 230–237, 2004.

[2] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishiroshi, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. Overview of the NTCIR-13 QA Lab-3 task. In *Proceedings of NTCIR-13*, pages 1–7, 2017.

ID	T/F	gold	system	1	2	3	4	5	6	absdiff
A1	T	**4**	**1**	0	0	0	0			**0**
A2	F	**3**	**4**	9.31	7.2	9.31	6.32			**2.99**
A3	T	**2**	**3**	4.87	4.87	5.83	4.87			**0.96**
A4	T	2	2	0	4.87	0	4.87	4.87	4.87	0
A5	F	**4**	**3**	12.16	11.35	8.2	8.53			**0.33**
A6	T	**2**	**3**	4.87	4.87	5.83	4.87			**0.96**
A7	T	1	1	8.58	8.12	8.58	8.12			0
A9	T	**2**	**4**	8.93	9.62	7.29	10.26			**0.64**
A10	T	1	1	0	0	0	0			0
A11	T	**2**	**3**	7.44	9.3	12.02	7.91			**0.96**
A12	T	**4**	**3**	4.87	4.87	5.83	4.87			**0.96**
A13	T	**4**	**1**	0	0	0	0			**0**
A14	T	**2**	**3**	4.87	4.87	5.83	4.87			**7.01**
A15	F	**3**	**2**	12.24	5.92	12.93	9.17			**7.01**
A16	T	**1**	**2**	0	9.02	0	9.02			**9.02**
A17	T	**2**	**3**	4.87	4.87	5.83	4.87			**0.96**
A19	T	**3**	**1**	0	0	0	0			**0**
A20	T	**1**	**3**	4.87	4.87	5.83	4.87			**0.96**
A21	T	**2**	**1**	9.41	9.41	9.41	9.41			**0**
A22	T	**1**	**3**	0	0	7.64	7.64			**7.64**
A23	T	**3**	**4**	6.85	6.01	8.9	11.36			**2.46**
A25	T	**4**	**2**	0	4.87	0	4.87	4.87	4.87	**0**
A26	F	**4**	**2**	8.31	7.79	11.97	12.86			**5.07**
A27	T	1	1	6.98	6.08	6.51	4.27			0
A29	T	**3**	**4**	8.14	7.72	7.42	9.22			**1.8**
A30	T	**4**	**3**	4.87	4.87	5.83	4.87			**0.96**
A31	T	**4**	**2**	0	9.95	0	9.95			**0**
A32	T	**1**	**2**	3.11	7.3	3.11	7.3			**4.19**
A33	T	1	1	8.14	7.8	8.04	4.42			0
A34	F	**2**	**4**	13.66	9.83	8.15	7.13			**2.7**
A35	T	2	2	0	7.73	0	7.73			0

Table 2: Query score for each choice in Phase-2