



QA Lab-3: QA Lab for Entrance Exam

KOTARO SAKAMOTO*1, *2, *3, HIDEYUKI SHIBUKI*1, MADOKA ISHIOROSHI*2, AKIRA FUJITA*2, TERUKO MITAMURA*3, NORIKO KANDŌ*2, *4, TATSUNORI MORI*1, YOSHINOBU KANO*5

*1: YOKOHAMA NATIONAL UNIVERSITY, *2: NATIONAL INSTITUTE OF INFORMATICS, *3: CARNEGIE MELLON UNIVERSITY, *4: THE GRADUATE UNIVERSITY FOR ADVANCED STUDIES (SOKENDAI), *5: SHIZUOKA UNIVERSITY

Goal

Using Japanese **university entrance exams** and their English translations on the subject of “World history”

Investigation of the **real-world complex QA** technologies

- Question Analysis
- Information Retrieval
- Information Extraction
- Automatic Summarization

Overview of QA Lab-2

Tasks

- National Center Test (multiple-choice type, EN & JA)
- Second-stage examinations of 5 universities (free-description type, EN & JA)
- Benesse mock exam (multiple-choice type, JA)
- Yozemi mock exam (multiple-choice type, JA)
- Sundai mock exam (free-description type, JA)

Free to participate
either of exams

Multiple-choice type questions

- 12 teams participated
- Good results

Free-description type questions

- 3 teams participated

Named-Entity questions

- Similar to factoid QA
- Relatively good results

Essay questions

- Complex and laborious task
- Similar to query-biased multi-document summarization
- **Not good** results

Focus on essay
questions at QA Lab-3

QA Lab-3 tasks

- Multiple-Choice Question Task
- Named-Entity Question Task
- Essay Question Task

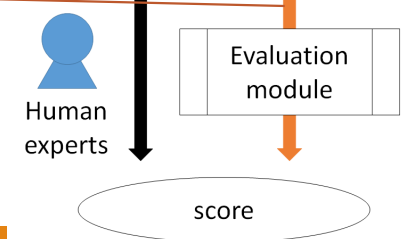
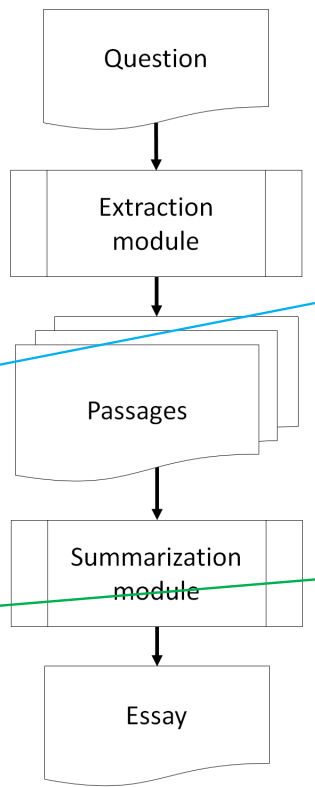
x English / Japanese

separate

- o **End-to-End** subtask
- o **Extraction** subtask
 - o IN: Question OUT: Passages including texts of Gold Standard Data
- o **Summarization** subtask
 - o IN: Question + Passages OUT: Essay
- o **Evaluation Method** subtask
 - o IN: Question + Essays + Gold Standard Data OUT: Ranking of Essays (with score)



For essay questions





Research run

How much did the QA technologies improved from QA Lab-1?

- Using the same training/test sets as the past QA Lab runs, comparison with the past results
- Using the systems participating in the past QA Lab runs, comparison with the present systems

What's new in NTCIR-13

Restructuring tasks based on question types

- Multiple-Choice Question Task
- Named-Entity Question Task
- Essay Question Task

Breaking down **Essay Question Task** into manageable subtasks

- **Extraction subtask** and **Summarization subtask** in order to be highly participative
- **Evaluation method subtask** for automated evaluation of essays

Research run for the progress so far

Challenges in QA Lab-3

- 1) essay questions that require **logical summaries** along a historical theme
- 2) **competition with more than 3,500 students**, examinees, from all over Japan (JA only)
- 3) questions with context
- 4) answer by text as high-compress-ratio **query-biased summarization**
- 5) advanced **entity-focused passage retrieval**
- 6) enhance knowledge resources
- 7) semantic representation and sophisticated learning
- 8) **appropriate evaluation measure** for essay
- 9) **research run** using the past QA Lab data/systems



Question and Answer Data

Training Data

- 8 years of National Center Test (JA and EN)
- 5 years of Second-stage Examinations for 5 univs (JA and EN)
- 11 years of Essay questions of 2 univs' entrance exams (JA)
- Nugget data for 11 years of Essay questions of 2 univs' entrance exams (JA)

Test Data

- 2 sets (JA and EN)

From
QA Lab-1
and -2
training and
test data

Knowledge Sources

Four Japanese high school textbooks

- Available in Japanese

A snapshot of Wikipedia

- Available in English and Japanese

World History Ontology

- Available in Japanese

Participants are free to use any resources

Tools

Two Baseline Systems (Japanese and English)

- UIMA module-based end-to-end QA systems

One Passage Retrieval Module

- to enhance the module-based collaboration.

Scorer for National Center Test

Format Checker for National Center Test and Second-stage Examination

Proposed Schedule (1/2)

Aug 2016	First version of task description released
Sep 2016	Corpus, Training set, Baseline systems will be released to the participants
Aug – Sep 2016	Discussion on task description
End Sep 2016	Fixed task detailed task description
Oct 2016 only)	1st Competitive runs with more than 3,500 high-school students, examinees, from all over Japan (JA
<Phase 1>	
Feb 2017	1st runs (End-to-End, Extraction, Summarization, Evaluation method)
Mar 2017	Results of the 1st runs return to the participants
Apr 2017	Round-table discussion on the above results

Proposed Schedule (2/2)

<Phase 2>

Jun 2017 **2nd runs (End-to-End, Extraction, Summarization, Evaluation method)**

Jul 2017 Results of the 2nd runs return to the participants

<Research Run>

Jul 2017 **Research runs**

Aug 2017 Round-table discussion on the above results

<Towards Conference>

Oct 2017 Draft paper submission

Oct 2017 **2nd Competitive runs** with more than 3,500 high-school students, examinees, from all over Japan (JA only)

Nov 1: 2017 Camera-ready Paper submission

Dec 2017 NTCIR-13 Conference



Mailing List for Discussion

English

qalab@nii.ac.jp

Japanese

qalab-ja@nii.ac.jp

Join Us !!!

Email: qalab-admin@nii.ac.jp