

## Entrance Exams

### Guidelines (27/2/2015)

#### INTRODUCTION

The Japanese University Entrance Exams include questions formulated at various levels of complexity and test a wide range of capabilities. The challenge of "Entrance Exams" aims at evaluating systems under the same conditions humans are evaluated to enter the University. Following the first campaigns in 2013 and 2014, we will reduce the challenge to Reading Comprehension exercises contained in the English exams. Exams are created by the Japanese National Center for University Admissions Tests. The "Entrance Exams" corpus is provided by [NII's Todai Robot Project](#) and [NTCIR RITE](#).

#### TASK DESCRIPTION

Participants are asked to read a given document and answer some questions. Questions are given in multiple-choice format, with four options from which a single answer must be selected. Questions are taken from reading tests of real university entrance exams, and may include a variety of question types that require a wide range of inferences. Systems have to answer questions by referring to "common sense knowledge" that high school students who aim to enter the university are expected to have. An important difference with previous exercises is that we do not intend to restrict question types. Any types of reading comprehension questions in real entrance exams will be included in the test data.

#### TRIAL AND TEST DATA

The Entrance Exams 2015 test set will be composed of reading comprehension tests taken from the *Center Test*, which is a nation-wide achievement test for Japanese university admissions.

- 19 test documents
- 89 questions with
- 356 choices/options (4 for each question)

For trial data there are available 24 tests translated into several languages.

To obtain the test data, it is necessary to **sign an agreement form** (which is available at the Entrance Exams website <http://nlp.uned.es/entrance-exams/>) and **send it to the NTCIR project office:**

NTCIR project office (Rm.1309)  
National Institute of Informatics  
2-1-2 Hitotsubashi Chiyoda-ku, Tokyo 101-8430, JAPAN  
PHONE: +81-3-4212-2750  
FAX: +81-3-4212-2751  
Email: ntc-secretariat@nii.ac.jp

## LANGUAGE

The original test documents, questions, and options are available in *English*. Besides, the data has been translated into Russian, French, Spanish, Italian and German.

## QUESTIONS

Questions are in the form of multiple-choice:

- for each question, 4 possible answers are given;
- they focus on testing the comprehension of one single document;
- they are extracted from real entrance exams without any modification.

Therefore, the results are directly comparable with human records.

## ANSWERS

The system is not required to answer every question, as the c@1 measure will be used for evaluation (see below). Therefore, there are three possibilities:

- (1) To submit an answer and ask for it to be evaluated,
- (2) Not to submit an answer,
- (3) To submit an answer and ask for it not to be evaluated.

Options (2) and (3) will result in identical C@1 scores but the differences could be interesting for system analysis (e.g., to provide additional feedback about a system's self-validation). The Output Format section below gives more details.

## RUNS

Participants are allowed to submit a maximum of 10 runs. Each run must be categorized as one of the following types, depending on the resources that have been used to assist in answering the questions:

- No external resource is used (only the test document);
- The test document and other resources are used

## FORMATS

The data format is exactly same as for the previous campaign. The DTDs for the input and output format can be downloaded from the Entrance Exams website (<http://nlp.uned.es/entrance-exams/>).

The test set will be formatted as an xml file (UTF-8 encoded). The xml will be structured with elements containing the following information:

### INPUT FORMAT

**topic t\_name t\_id reading-test r\_id doc d\_id question q\_id q\_str answer a\_id**

where:

- **topic** includes the following elements: topic, questions, and multiple choice answers;
- **t\_id** is the id of the topic (in this task this is always 0);
- **t\_name** is the title of the topic (always "Entrance Exam");
- **reading-test** includes the test document and all questions with the multiple choice answers which refer to the test document;
- **r\_id** is the unique id of the reading test;
- **doc** is the test document which the questions are being asked against;

- **d\_id** is the id of the test documents;
- **question** is the question and the multiple choice answer;
- **q\_id** is the id of the question (a number from 1 to 10);
- **q\_str** is the question string (UTF-8 encoded);
- **answer** contains one of the four multiple choice options;
- **a\_id** is the id of the answer (a number from 1 to 4).

e.g.:

```
<test-set>

<topic t_id="0" t_name="Entrance Exam">
  <reading-test r_id="1">
    <doc d_id="1">
      Some Text here.....Some Text here....Some Text
      here....Some Text here....Some Text here....Some
      Text here.....Some Text here.....Some Text
      here....Some Text here
    </doc>
    <question q_id="1">
      <q_str> Text of question </q_str>
      <answer a_id = "1">Text of answer </answer>
      <answer a_id = "2">Text of answer </answer>
      <answer a_id = "3">Text of answer </answer>
      <answer a_id = "4">Text of answer </answer>
    </question>
    <question q_id="2">
      <q_str> Text of question </q_str>
      <answer a_id = "1">Text of answer </answer>
      <answer a_id = "2">Text of answer </answer>
      <answer a_id = "3">Text of answer </answer>
      <answer a_id = "4">Text of answer </answer>
    </question>
    ...
    ...

  </question>
</reading-test>

.
.
</topic>
</test-set>
```

### OUTPUT FORMAT

A submission file must be an xml file (UTF-8 encoded). The xml will be structured with elements containing the following information:

**run\_id topic t\_id reading-test r\_id q\_id answered answer a\_id**

- The **run\_id** attribute of the root element is an alphanumeric string which identifies the runs of each participant. It should be the concatenation of the following elements:
  - **the team ID** (sequence of four lower case ASCII characters),
  - **the current year** (15 stands for 2015),

- **the number of the run** (01 for the first one, or 02 for the second one, and so on. The maximum number of runs which can be submitted by each system is 10),
- **the language pairs** including both source and target languages; Clearly, the content of this field never changes within the same submission file. Each submission file must be given an .xml extension, e.g. "clct1501itit.xml";
- **topic** includes the following elements: topic, questions, and multiple choice answers;
- **t\_id** is the id of the topic as given in the test-set (should be 0);
- **reading-test** includes questions relating to the test document, together with their multiple choice answers ;
- **r\_id** is the id of the reading test as given in the test-set;
- **question** is the question and its multiple choice answers;
- **q\_id** is the question number as given in the test set. Questions must be returned in the same ascending (increasing) order in which they appear in the test-set;
- **answered** indicates if the question has been answered or not (YES or NO must be set);
- **answer** contains the candidate answer, as determined by the system. For each question, only one option (answer tag) must be selected.

In the case where the attribute "answered" is set to "NO" the system has two options:

- the "answer" element is not returned, meaning that the system prefers not to answer;
- The system returns in the "answer" element the candidate answer which the system would have given;
- **a\_id** is the id of the answer as given in the test-set;

e.g.:

```
<output run_id="XXXX1501YXXXX">
<topic t_id="0" >
  <reading-test r_id="1">
    <question q_id="1" answered="YES">
      <answer a_id="1"/>
    </question>

    <question q_id ="2" answered="NO" />

    <question q_id ="3" answered="YES" >
      <answer a_id ="5"/>
    </question>

    <question q_id ="4" answered="NO">
      <answer a_id ="1"/>
    </question>
    .
    .
  </reading-test>
  .
  .
  .
  .
  .
</topic>

</output>
```

## EVALUATION

Scoring of the output produced by participant systems will be performed automatically by comparing the answers of systems against the gold standard collection with annotations made by humans. No manual assessment will be performed.

Each test will receive an evaluation score between 0 and 1 using  $c@11$ . This measure, used in previous CLEF QA Tracks, encourages systems to reduce the number of incorrect answers while maintaining the number of correct ones by leaving some questions unanswered.

Systems will receive evaluation scores from two different perspectives:

1. at the question-answering level: correct answers are counted individually without grouping them;
2. at the reading-test level: figures both for each reading test as a whole, and for each separate topic are given.

## IMPORTANT DATES

Send the agreement form	by April 30, 2015
Test set release	May 6, 2015
Run submissions	May 15, 2015
Individual results to participants	May 20, 2015
Submission of Working Notes Papers	June 7, 2015
CLEF Workshop	September 15-18, Toulouse, FR

Test questions will be posted on a web site that will be announced to **the teams who have submitted the signed agreement form to the NTCIR office**. The test set will be available on **May 6** and submissions will be due **within 5 days** from the first test set download and not later than **May 15** by 11:59 p.m. (CEST). Late submissions will not be considered.

Participant runs will be submitted using an automatic submission procedure. Details about the submission procedure will be provided when the test data is released. Before completing the submission, a checking routine will automatically be run in order to detect format inconsistencies and common errors in the files (invalid document numbers, wrong formats, missing data, etc.). The automatic submission procedure will reject any run that does not comply with the required format.

## ORGANIZING COMMITTEE

- Anselmo Peñas (UNED, Spain)
- Yusuke Miyao (National Institute of Informatics, Japan)
- Alvaro Rodrigo (UNED, Spain)
- Eduard Hovy (Carnegie Mellon University, USA)
- Noriko Kando (National Institute of Informatics, Japan)
- Teruko Mitamura (Carnegie Mellon University, USA)

---

<sup>1</sup> Anselmo Peñas and Alvaro Rodrigo. A Simple Measure to Assess Non-response. In Proceedings of 49th Annual Meeting of the Association for Computational Linguistics - Human Language Technologies (ACL-HLT 2011), Portland, Oregon, USA, June 19-24, 2011