

# NLPCC 2014 Shared Tasks Guidelines

## Large-Scale English Question Answering

### 1. Introduction

This task aims to evaluate the techniques for answering open-domain questions, namely Large-Scale Question Answering over Open-domain Data.

### 2. The Description of the Task

Given a natural language question, the purpose of the task is to automatically find

- a. Whether the question has answers;
- b. If the question has answer(s), find the correct answer(s) for it.

The questions and answers are limited within English. We will not involve multi-lingual question answering. There is no restriction of the data sources to find answers. Both structured data like knowledge base and unstructured data like plain text can be used. In the task, if the question has answer(s), participators are required to find the top N probable answers.

### 3. Data Format

The training data and testing data is xml format, for an instance:

```
<QAPair id="1">
<question id="1" >Who wrote the Declaration of Independence</question>
<answer id="1">Thomas Jefferson</answer>
<answer id="2">Benjamin Franklin</answer>
<answer id="3">John Adams</answer>
<answer id="4">Roger Sherman</answer>
<answer id="5">Robert R. Livingston</answer>
</QAPair>
```

Only one submission file is allowed for each team. In the submission file, each line corresponds to one of the questions. It should include the following information

and use ‘\t’ as the separator:

System-ID: name of the system

Question-ID: ID of the question

Has-Answer: “True” or “False” indicates whether the question has answer(s)

N: number of answers found by the system (no more than 5)

Top-N-Answers: top N probable answers found by the system

For the above example, the submission should be:

```
System_Name\t1\tTrue\t5\tThomas    Jefferson\tBenjamin    Franklin\tJohn    Adams\tRoger  
Sherman\tRobert R. Livingston
```

#### 4. Evaluation Metric

In the open-domain question answering task, metric is based on MRR and accuracy@N. For all the questions in the test data set,

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Where Q is the test set, where |Q| denotes the number of questions in Q. rank<sub>i</sub> represents the position of the first correct answer for Q<sub>i</sub>

$$\text{accuracy@N} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} C_i(N)$$

$$C_i(N) = \begin{cases} 1, & \text{there are at least 1 correct answer in top N results} \\ 0, & \text{otherwise} \end{cases}$$

#### 5. Data

We provide 20,000 question-answer pairs from query log in this evaluation data set for training and testing. The data is notated in xml format, Unicode encoding (utf-8).