Proposal for
NTCIR-11 Pilot task

QA Lab

Madoka Ishioroshi (NII)
Kelly Itakura
Noriko Kando (NII)
Yoshinobu Kano (PRESTO/NII)
Teruko Mitamura (CMU/NII)
Hideyuki Shibuki (Yokohama National University)

Advisor: Tatsunori Mori (Yokohama National University)

NTCIR-11 Kickoff Event
Sept. 2, 2013

Thanks Madoka for slides
Main Idea

• Challenge for Real World Problem of QA
  – University Entrance Exams - New Q-types not covered by existing QA systems. Q is not a single sentence. Need context (ntcir-11), inference (ntcir-12).

• Challenge for a new structure of evaluation – Module-based, building workable system as a joint effort of all the participants
  – QA is very complicated
  – Hard for a research group to be good on every component
  – Your module used and rated by other people

• Continuous Plan-Run-Analysis-Improve Cycle
Background-1

Modules in a QA System

Question Analysis

Document Retrieval

Extracting Answer Candidates

Answer Generation

「How tall the Sky-Tree in Tokyo?」

「634 Meters」
Background-1

Modules in a QA System

- Extracting focus from Question
- Analyzing what is asked (Question Type)

```
How tall the Sky-Tree in Tokyo?
```

[factoid] People, GeoName, Organization, Numeric value, etc.

[non-factoid] Definition, Reasons, Methods, etc.

- Question Analysis
- Document Retrieval
- Extracting Answer Candidates
- Answer Generation

「634 Meters」
Background-1

Modules in a QA System

- Extracting focus from Question
- Analyzing what is asked (Question Type)

Search related documents (or passage) using the extracted focus of questions and Q type

Question Analysis

How tall the Sky-Tree in Tokyo?

Document Retrieval

Extracting Answer Candidates

Answer Generation

634 Meters
Background-1

Modules in a QA System

- Extracting focus from Question
- Analyzing what is asked (Question Type)

Search related documents (or passage) using the extracted focus of questions and Q type

Extracting Answer Candidates from the retrieved documents or passages

Question Analysis

Document Retrieval

Extracting Answer Candidates

Answer Generation

「How tall the Sky-Tree in Tokyo?」

「634 Meters」
Background-1

Modules in a QA System

- Extracting focus from Question
- Analyzing what is asked (Question Type)

Search related documents (or passage) using the extracted focus of questions and Q type

Extracting Answer Candidates from the retrieved documents or passages

Ranking the answer candidates based on the scores of “answer likeliness”
Background-1

Modules in a QA System

- Question Analysis
  - Extracting focus from Question
  - Analyzing what is asked
  - Inf used for scoring:
    • Verify the Q with the sentence that the answer candidate was extracted, Textual Entailment
    • # of KWs around the answer candidate
    • Verify by the Q type and the answer candidate
    • Context where the answer candidate was extracted from

- Document Retrieval

- Extracting Answer Candidates

- Answer Generation
  - Ranking the answer candidates based on the scores of "answer likeliness"

Q

「How tall the Sky-Tree in Tokyo?」

Answers

「634 Meters」
Evaluation Method-1

Systems

• QA-Platform
  – UIMA-based Module-base QA system
  – Follows: ACLIA modules
    • Q analysis
    • IR
    • Answer Extraction
    • Answer Generation
  – 2 baseline systems for Japanese
    • Javelin (CMU)
    • MinerVA (YNU)
  – 1 baseline systems for English (CMU)
  – Systems are usable either
    • local (by Kachaco) or on NII Server “bunbun”
UIMA Native UI for QA Platform to run a pipeline

Upload the Q file

Set various parameters to run

Specifies the index file for Indri (RS)

Document directory

Output directory

Run

Upload the Q file

Set various parameters to run

Specifies the index file for Indri (RS)

Document directory

Output directory

Run
UI for the QA Platform to examine the results of each module

Results of each module

Show the representative results for the first. To see more detailed results by click on the button of “Expanding”

Click to jump to the module

Answer Candidates
Evaluation Method-2

Corpus

• Questions & Answers
  – Japanese National Center for University Admissions Tests (multiple choices)
    • → use as a Yes-No / TF Questions
    • XML, Japanese and English translation
  – Second Exam for the U of Tokyo and some others
    • → factoid & Complex Questions
    • Will be XML tagged

• Knowledge sources provided
  – High school Textbooks, Wikipedia
  – Ontology of World History
  – Ontology-annotated Textbook

• Tools
  – Wide-range of Language annotation tools
  – RITE resource and tools
### Evaluation Method-3

**Tasks**

<table>
<thead>
<tr>
<th></th>
<th>Multiple Choice (Center Exam)</th>
<th>2nd Exam</th>
<th>2nd Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTCIR-11</td>
<td>World History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTCIR-12</td>
<td>World History</td>
<td>Biology, Politics &amp; Economics</td>
<td>Factoid</td>
</tr>
</tbody>
</table>

Evaluate QA end-to-end with **Every Possible combination** of the modules submitted so far and base systems
Tentative Schedule:
Multiple Opportunities to Run

<table>
<thead>
<tr>
<th></th>
<th>Jan-14</th>
<th>Feb-14</th>
<th>Mar-13</th>
<th>Apr-14</th>
<th>May-14</th>
<th>Jun-13</th>
<th>Jul-14</th>
<th>Aug-14</th>
<th>Sep-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Round</td>
<td>runs</td>
<td></td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Round</td>
<td>runs</td>
<td></td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd Round</td>
<td>runs</td>
<td></td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Submitted all rounds are not mandatory.
* R-table = Round-table meeting
Participant’s merits

• Can participate in one phase or multiple phases (flexible participation)

• Can evaluate with Baseline or with other systems and to improve your system for the next phase

• Can see the effectiveness of your modules in various pipelines created by other teams or combination of other teams.
Join Us!