Proposal for NTCIR-11 Pilot task

QA Lab

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

Advisor: Tatsunori Mori (Yokohama National University)

NTCIR-11 introduction session at NTCIR-10 Conference
June 21, 2013

Thanks Madoka for slides
Road map

• Main Idea
• Background
  – QA
  – ACLIA: “Dream” system superior than the top end-to-end, but didn’t know why.
• Evaluation Method
  – Systems
  – What evaluate/ Metrics
  – Corpus
  – Resources
Main Idea

• Challenge for Real World Problem of QA
  – University Entrance Exams - New Q-types not covered by existing QA systems

• 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-Excute-Analysis-Improve Cycle
Background-1

Modules in a QA System

- 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

Question Analysis

[Input: "How tall the Sky-Tree in Tokyo?"]

- [factoid] People, GeoName, Organization, Numeric value, etc.
- [non-factoid] Definition, Reasons, Methods, etc.

Extracting Answer Candidates

Answer Generation

Ranking the answer candidates based on the scores of "answer likeliness"

Answers

[Output: "634 Meters"]
Different teams can exchange and create a “dream-team” QA system.

“Dream” System superior than Best End-to-End, but could not know why.

http://aclia.lti.cs.cmu.edu/ntcir8/
Systems

- QA-Platform
  - UIMA-based Module-base QA system
  - Follows: ACLIA modules
    - Q analysis
    - IR
    - Answer Extraction
    - Answer Generation
  - 2 baseline systems
    - Javelin (CMU)
    - MinerVA (YNU)
  - 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
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 Questions
    - XML, Japanese and English translation
  - Second Exam for the U of Tokyo
    - 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>Multiple Choice (Center Exam)</th>
<th>2nd Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese</td>
<td>English</td>
</tr>
<tr>
<td>NTCIR-11</td>
<td>World History (YN)</td>
</tr>
<tr>
<td>NTCIR-12</td>
<td>Biology (YN)</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></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>runs</td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>runs</td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>runs</td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>runs</td>
<td>return results</td>
<td>R-table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Submitted all rounds are not mandatory.
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!