

~~~~~

# ResPubliQA 2010

## Question Answering Evaluation over European Legislation

~~~~~

-----

### Preliminary Call for Participation

-----

Following the success of ResPubliQA 2009, we are pleased to announce ResPubliQA 2010, the second evaluation campaign of Question Answering systems over European Legislation to be held within the framework of CLEF 2010 conference<sup>1</sup>

For more information and updates visit the ResPubliQA website at:

<http://celct.isti.cnr.it/ResPubliQA/>

We invite participation from IR and NLP practitioners and potential users of QA systems concerned with European texts.

Detailed guidelines describing the task will be distributed among the participants and downloadable from the ResPubliQA website.

The results of the evaluation campaign will be disseminated at the final workshop which will be organized, in conjunction with the CLEF 2010 conference, 20-23 September in Padua, Italy.

### ResPubliQA 2010: TASK OVERVIEW

The aim of ResPubliQA 2010 is to capitalize on what has been achieved in the previous evaluation campaign while at the same time adding a number of refinements:

- The addition of new question types and the refinement of old ones;
- The opportunity to return both paragraph and exact answer;
- The addition of a new collection: EUROPARL

Two separate tasks are proposed for the ResPubliQA 2010 evaluation campaign:

1. PARAGRAPH SELECTION (PS) TASK: to retrieve one paragraph containing the answer to a question in natural language. One of the following responses must be returned:

---

<sup>1</sup><http://www.clef2010.org/>

- a) ONE single paragraph containing the candidate answer
- b) the string NOA to indicate that the system prefers not to answer the question.

2. ANSWER SELECTION (AS) TASK: beyond retrieving a paragraph containing the answer to a question in natural language, systems are required to demarcate also the exact answer. One of the following responses must be returned:

- a) the exact answer highlighted inside one paragraph
- b) the string NOA to indicate that the system prefers not to answer the question.

N.B. Systems that prefer to leave some questions unanswered, can OPTIONALLY decide to submit also a candidate paragraph/answer with the aim of evaluating the validation performance.

The two tasks are only different in the output required. Document collection and test data for both tasks are the same.

DOCUMENT COLLECTION: the following multilingual parallel-aligned document collections are used:

- The ResPubliQA collection: a subset of JRC-Acquis with parallel-aligned documents in 9 languages
- A small subset of the EUROPARL collection with parallel-aligned documents in 9 languages has been created by crawling the web to get the data from the website of the European Parliament<sup>2</sup> (starting from January 2009).

Both collections will be available at the ResPubliQA website<sup>3</sup>.

The subject of the Acquis documents is European legislation while EUROPARL deals with the parliamentary domain. The two collections are different in style and content while being fully compatible at the same time.

LANGUAGES: parallel-aligned documents are available in 9 languages, i.e: Bulgarian, Dutch, English, French, German, Italian, Portuguese, Romanian and Spanish.

Only the tasks in which there will be at least one registered participant will be activated.

TEST DATA: a pool of 200 questions will be provided

- independent questions that can be answered by a paragraph
- question types: factoid, definition, purpose, reason, opinion, other
- NO NIL; NO LIST

---

<sup>2</sup> <http://www.europarl.europa.eu/>

<sup>3</sup> <http://celct.isti.cnr.it/ResPubliQA/Downloads>

EVALUATION: each output of both the PS and AS tasks are automatically evaluated against the GoldStandard manually produced. Non-matching paragraphs and answers are manually evaluated by native speakers assessors.

The adoption of the c@1 evaluation metric encourages systems to maintain the number of correct answers while reducing the amount of incorrect ones by leaving some questions unanswered (NOA). Answer Validation techniques (including Machine Learning) are expected to be used for taking this final decision. For more details, please read the ResPubliQA 2009 Overview, available at the campaign website.

RUNS: systems are allowed to participate in one or both tasks which will operate simultaneously on the same input questions. A maximum of two runs in total can be submitted, i.e. two PS runs, two AS runs or one PS plus one AS run.

#### **PRELIMINARY TIMELINE**

- Track guidelines: January 25
- Registration at the ResPubliQA website: by March
- Test set release: May 17
- Run submissions: May 27\*
- Results to the participants: July 9
- Submission of Papers: August 15
- Workshop: 20-23 September 2010, in Padua, Italy

\*Participants will have 5 DAYS to upload their submissions, starting from the moment when the questions are downloaded.

#### **LAB ORGANIZERS**

- Anselmo Peñas, E.T.S.I. Informática de la UNED, Madrid, Spain
- Pamela Forner, CELCT, Trento, Italy
- Richard Sutcliffe, Dept. of Computer Science, University of Limerick, Limerick, Ireland

#### **ADVISORY BOARD**

- Donna Harman (National Institute for Standards and Technology (NIST), USA)
- Maarten de Rijke (University of Amsterdam, The Netherlands)
- Dominique Laurent (Synapse Développement, France)