News People Search Task @ EVALITA 2011

Cross-document Coreference Resolution of Named Person Entities

Task Guidelines

1. Introduction

The News People Search (NePS) task aims at evaluating cross-document coreference resolution of named person entities in Italian news.

Cross-document coreference of a person entity occurs when the same person is mentioned in more than one text source. It can be defined as a clustering problem, which in principle requires the clustering of name occurrences in a corpus according to the persons they refer to. In the NePS task, we consider *clusters of documents* containing the name occurrences. Cross-document coreference involves two problematic aspects, namely (i) to resolve ambiguities between people having the same name (i.e. when identical mentions refer to distinct persons) and, conversely, (ii) to recognize when different names refer to the same person.

The cross-document coreference resolution task has close links with Word Sense Disambiguation, which consists of deciding the sense of a word in a given context. In both tasks, the problem addressed is the resolution of the ambiguity in a natural language expression. More precisely, the NePS task can be viewed as a case of Word Sense Discrimination, as the number of "senses" (i.e. actual people) is unknown a priori.

2. Task Description

The NePS task consists of clustering a set of Italian newspaper articles that mention a person name according to the different people sharing the name (i.e. one cluster of documents for each different person).

More specifically, for each person name, systems receive in input a set of newspaper articles and the expected output is a clustering of the documents, where each cluster is supposed to contain all and only those documents that refer to the same individual.

The NePS task is limited to documents in which the entities are mentioned by name and takes into account name variability. Different kinds of name variants are considered, such as complete names (Paolo Rossi, Rossi Paolo), abbreviations (P. Rossi, Paolo R.), first names only (Paolo), last names only (Rossi), nicknames (Pablito), and misspellings (Paalo Rossi).

The NePS task is structured along the same lines as the Web People Search evaluation exercise (WePS), which in 2010 was at its third edition. Information about all the WePS tasks are available at: http://nlp.uned.es/weps/weps-3

The main differences with respect to the WePS clustering task are that the NePS task *(i)* addresses Italian language instead of English, *(ii)* takes into account name variability, and *(iii)* uses a corpus of newspaper articles instead of web pages.

3. Data Set Description

The data set used for the NePS task is the Cross-document Italian People Coreference corpus (CRIPCO). Details about the corpus can be found in (Bentivogli et al., 2008).

3.1. Development Set

The NePS 2011 development data are composed of the following items:

a) 105 different person names (Group Names). A Group Name is a complete name, i.e. a pair First Name-Last Name (e.g. Paolo Rossi, Isabella Bossi Fedrigotti, Diego Armando Maradona).

b) for each Group Name, a set of newspaper articles containing at least one mention of the Group Name or of one of its possible variants (e.g. abbreviation, first or last name only, nickname, etc.).

c) the Gold Standard clustering.
For each Group Name, an XML file representing the manual clustering of the newspaper articles according to the different people sharing the name (i.e. one cluster of documents for each different person).

The structure of the Gold Standard clustering file can be seen in the following example:

```
<clustering name="Paolo_Rossi">
  <entity id="6589">
    <doc rank="adige20000702_id252770.txt"/>
    <doc rank="adige20060617_idLET2-01.txt"/>
    <doc rank="adige20060710_idET07-05.txt"/>
    <doc rank="adige20060709_idAS03-01.txt"/>
  </entity>

  <entity id="6590">
    ...
  </entity>
  ...
   <discarded>
    <doc rank="adige19990425_id3342876.txt"/>
    <doc rank="adige19990325_id3332840.txt"/>
    <doc rank="adige20011011_id1422768.txt"/>
  </discarded>
  ...
  </clustering>
```

The <discarded> element (not always present) contains all the documents for which the annotators did not find enough evidence to determine the appropriate cluster (i.e. the appropriate person).

Note that the discarded documents are included in the data set but are discarded from the evaluation process.

The NePS development set is freely available upon acceptance of a license agreement and can be requested at http://www.evalita.it/2011/datadistribution.

3.2. Test Set

The NePS 2011 test data is composed of:

a) 103 Group Names
b) for each Group Name, a set of newspaper articles containing at least one mention of the Group Name or of one of its possible variants.

The Test Set Gold Standard used for the evaluation of the system performances will be freely made available at the end of the evaluation campaign.

4. System Submissions

For each Group Name, systems are expected to return the clustering of the documents, where each cluster is supposed to contain all and only those documents that refer to the same individual.

Note that:

- There can be documents containing mentions to different individuals sharing the same name. These documents must be assigned to as many clusters as necessary.

- All documents must be clustered, with the exception of the "discarded" ones. All discarded (or missing) documents which do not correspond to those tagged as discarded in the Test Set Gold Standard will be aggregated in a new cluster and evaluated.

Participants are allowed to submit up to 3 runs.

4. 1 Submission Format

The system submission format must be the same as the Gold Standard XML format (included the not mandatory "discarded" element), as shown in section 3.1.

Each run must be submitted as a zipped folder containing one XML file for each Group Name. The directory and the XML files must be named according to the requirements below:

a) submission directory (zipped): SystemName_RunNumber.zip (e.g.: yoursystemname_1.zip).
b) XML file: GroupName.clust.xml (eg: paolo_rossi.clust.xml; diego_armando_maradona.clust.xml)

Participant runs will be submitted using a web interface. Details about the submission procedure will be provided when the test data is released. Before completing the submission, a checking routine will be automatically run in order to detect format inconsistencies and common errors in the files. The submission procedure will reject any run which is not compliant with the required format.

Result Evaluation

System results will be compared to the human-annotated gold standard and the metrics used to evaluate system performances will be Extended B-Cubed Precision and Recall (Amigó et al. 2008),

combined with F1 measure. The extended version of B-Cubed was introduced in the WePs-2 task to specifically address the evaluation of overlapping clustering; in case of non-overlapping clustering extended B-Cubed results are identical to those obtained using standard B-Cubed (Artiles et al. 2010). Systems will be officially ranked according to their B-Cubed F1 score.

The evaluation will be carried out using the official scorer distributed for the WePS-2 task, which can be downloaded at www.evalita.it/sites/evalita.fbk.eu/files/doc2011/weps2007_scorer_1.1.tar.gz, or directly at the WePS website (http://nlp.uned.es/weps/weps-1/weps1-data)

References

Enrique Amigó, Julio Gonzalo, Javier Artiles, and Felicia Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. Information Retrieval, 2008.

Javier Artiles, Andrew Borthwick, Julio Gonzalo, Satoshi Sekine, and Enrique Amigó. "WePS-3 Evaluation Campaign: Overview of the Web People Search Clustering and Attribute Extraction Tasks", in CLEF 2010 LABs and Workshops Notebook Papers, Padua, Italy, 22-23 September 2010.

Luisa Bentivogli, Christian Girardi, Emanuele Pianta. "Creating a Gold Standard for Person Cross-Document Coreference Resolution in Italian News". In Proceedings of the LREC 2008 Workshop on Resources and Evaluation for Identity Matching, Entity Resolution and Entity Management. Marrakech, Morocco, 31 May 2008.