EVALITA 2009

Named Entity Recognition Task
Guidelines for Participants

Manuela Speranza
FBK-irst, Povo 38100 (Trento) Italy
manspera@fbk.eu

Task Description

In the Named Entity Recognition task, systems are required to recognize the Named Entities occurring in a text by identifying their extension and type, i.e. Person (PER), Organization (ORG), Geo-Political Entity (GPE) or Location (LOC).

The task will follow the same guidelines of the Named Entity Recognition task at EVALITA 2007, reported in this document.

The task will be based on the ACE-LDC standards for the Entity Recognition and Normalization Task (LDC 2005), with all the adaptations needed to limit the task to the recognition of Named Entities (see Magnini et al., 2007, available for download from the Named Entity task webpage at http://evalita.fbk.eu/entity.html).

The Corpus (I-CAB)

Both development data and test data are part of the Italian Content Annotation Bank (I-CAB), developed in the context of the Ontotext Project, and will be distributed upon acceptance of the agreement terms for a free research license.

Development data, consisting of 525 news stories taken from the local newspaper "L'Adige", include development and test data distributed for the Named Entity Recognition task at Evalita 2007 (for a total of almosto 180,000 words). The news stories belong to four different days (September, 7th and 8th 2004 and October, 7th and 8th 2004) and are grouped into five categories: News Stories, Cultural News, Economic News, Sports News and Local News. Test data, on the other hand, consist of completely new data (taken from the same newspaper).

Data Format

Development data and test data consist of two separate text files, with one token per line and an empty line after each sentence.

Both development and test data have been further annotated with Part of Speech information[1] using the Elsnet tagset for Italian (available for download at: http://evalita.fbk.eu/doc/elsnet-tagset-IT.pdf).

Each file consists of a number of columns separated by a blank, containing respectively:
- the token (first column);
- the Elsnet PoS-tag (second column);
- the Adige news story to which the token belongs (third column)

Only development data are also annotated with Named Entities in the IOB2 format: the entity tag is contained in the fourth column and consists of two parts:
- the IOB2 tag: "B" (for "begin") denotes the first token of a Named Entity, I (for "inside") is used for all other tokens in a Named Entity, and O (for "outside") is used for all other words;
- the Entity type tag: PER (for Person), ORG (for Organization), GPE (for Geo-Political Entity), or LOC(for Location).

Example of development data format:
il RS adige20041008_id414157 O
capitano SS adige20041008_id414157 O
della ES adige20041008_id414157 O
Gerolsteiner SPN adige20041008_id414157 B-ORG
Davide SPN adige20041008_id414157 B-PER
Rebellin SPN adige20041008_id414157 I-PER

To sum up, systems participating to the task will have to take as input a three-column file (see Appendix A) and produce as output a four-column file annotated with Named Entities in the IOB2 format (see Appendix B).

Submission of system results:

- § Deadline: September 20th, 2009, midnight (GMT + 1 hour)
- § System results have to be sent by e-mail to Manuela Speranza (manspera@fbk.eu)
- § Each participant has the possibility to submit a maximum of two runs
- § System results will consist of a single data file and will have to be named as follows: evalita09_NER_participant_run

---

[1] Please notice that the corpus has been PoS-tagged automatically with no manual correction; as a consequence, participants using such information must be aware that they will have to cope with incorrect tags.

Evaluation Metrics

With respect to the results submitted by the participants, we will provide the following measures: Precision, Recall, F-Measure and Accuracy.

Precision indicates the percentage of correct positive predictions and is computed as the ratio between the number of Named Entities correctly identified by the system (True Positive) and the total number of Named Entities identified by the system (True Positive plus False Positive), as shown in (1).

Recall indicates the percentage of positive cases recognized by the system and is computed as the ratio between the number of Named Entities correctly identified by the system (True Positive) and the number of Named Entities that the system was expected to recognize (True Positive plus False Negative), as shown in (2).

$$(1)\ Prec. = \frac{TP}{TP + FP} \qquad\qquad (2)\ Recall = \frac{TP}{TP + FN}$$

Table 1 reports three sub-sections of the example presented in Appendix A. The first column contains the tokens, the second and the third contain respectively the correct tags and the tags assigned by a system.

token	expected tag	system tag
Il	O	O
capitano	O	O
della	O	O
Gerolsteiner	B-ORG	B-ORG
Davide	B-PER	B-PER
Rebellin	I-PER	O
ha	O	O
allungato	O	O
frazionando	O	O
il	O	O
gruppo	O	B-PER
.	O	O
sul	O	O
traguardo	O	O
di	O	O
Bourges	B-GPE	B-LOC
.	O	O

Table1. Example of system results compared to expected annotation

From Table 1 we have that:

- the total number of Named Entities that the system was expected to recognize (i.e. True Positive + False Negative) is 3: "Gerolsteiner" (ORG), "Davide Rebellin" (PER), and "Bourges" (GPE).
- the number of Named Entities that the system has identified (True Positive + False Positive), is 4: "Gerolsteiner" (ORG), "Davide" (PER), "gruppo" (PER), and "Bourges" (LOC).

- the number of Named Entities that the system has identified correctly (True Positive) is 1, as the only correctly identified Named Entity is "Gerolsteiner" (ORG).[2]

As a result, the system obtains a Precision of 25% (given by 1/4) and a Recall of 33.33% (given by 1/3).

F-Measure, the weighted harmonic mean of precision and recall (see 3), will be used for the official ranking of systems.

$$(3)\ F = \frac{2 \times (precision \times recall)}{(precision + recall)}$$

In the example above, the system would obtain a value of F-Measure equal to 28.57%.

Accuracy indicates the percentage of correct predictions and is computed with respect to tokens as shown in (4).

$$(4)\ Acc. = \frac{CorrectTokens}{TotalTokens}$$

In the example above, we have a total number of 47 tokens (see Appendix A) 3 of which have wrong tags:
1. `Rebellin` ("O" instead of "I-PER");
2. `gruppo` ("B-PER" instead of "O");
3. `Bourges` ("B-LOC" instead of "B-GPE").

The Accuracy obtained by the system would be equal to 93.62% (given by 44/47).

Scorer

For the official evaluation of system results, we will use the scorer made available by CONLL for the 2002 Shared Task (the scorer can be freely downloaded from the CONLL website: http://www.cnts.ua.ac.be/conll2002/ner.tgz).

With respect to the example above, the CONLL scorer would return the results in the following format:

```
processed 47 tokens with 3 phrases; found: 4 phrases; correct: 1.
accuracy:  93.62%;
precision:  25.00%;
recall:  33.33%;
FB1:  28.57
```

Contact Person

Manuela Speranza (manspera@itc.it)

---

[2] "Davide" (PER) is not correct because the Named Entity to be recognized was actually "Davide Rebellin" (PER), "gruppo" (PER) is not correct because it is not a Named Entity, and "Bourges" (LOC) is not correct because it is in fact a Named Entity of type GPE.

Appendix A

Input format (example available at: http://evalita.fbk.eu/doc/ner-input-sample.txt)

A E adige20041008_id414157
circa B adige20041008_id414157
90 N adige20041008_id414157
chilometri SP adige20041008_id414157
dall' ES adige20041008_id414157
arrivo SS adige20041008_id414157
, XPW adige20041008_id414157
il RS adige20041008_id414157
capitano SS adige20041008_id414157
della ES adige20041008_id414157
Gerolsteiner SPN adige20041008_id414157
Davide SPN adige20041008_id414157
Rebellin SPN adige20041008_id414157
ha VIY adige20041008_id414157
allungato VSP adige20041008_id414157
su E adige20041008_id414157
uno RS adige20041008_id414157
dei EP adige20041008_id414157
pochi DP adige20041008_id414157
tratti SP adige20041008_id414157
in E adige20041008_id414157
salita SS adige20041008_id414157
, XPW adige20041008_id414157
frazionando VG adige20041008_id414157
il RS adige20041008_id414157
gruppo SS adige20041008_id414157
. XPS adige20041008_id414157

Alla ES adige20041008_id414157
sua DS adige20041008_id414157
ruota SS adige20041008_id414157
si PN adige20041008_id414157
sono VIY adige20041008_id414157
portati VPP adige20041008_id414157
altri DP adige20041008_id414157
sei N adige20041008_id414157
corridori SP adige20041008_id414157
che CCHE adige20041008_id414157
hanno VIY adige20041008_id414157
poi B adige20041008_id414157
disputato VSP adige20041008_id414157

lo RS adige20041008_id414157
sprint SN adige20041008_id414157
sul ES adige20041008_id414157
traguardo SS adige20041008_id414157
di E adige20041008_id414157
Bourges SPN adige20041008_id414157
. XPS adige20041008_id414157

Appendix B

Output format (example available at: http://evalita.fbk.eu/doc/ner-output-sample.txt)

A E adige20041008_id414157 O
circa B adige20041008_id414157 O
90 N adige20041008_id414157 O
chilometri SP adige20041008_id414157 O
dall' ES adige20041008_id414157 O
arrivo SS adige20041008_id414157 O
, XPW adige20041008_id414157 O
il RS adige20041008_id414157 O
capitano SS adige20041008_id414157 O
della ES adige20041008_id414157 O
Gerolsteiner SPN adige20041008_id414157 B-ORG
Davide SPN adige20041008_id414157 B-PER
Rebellin SPN adige20041008_id414157 I-PER
ha VIY adige20041008_id414157 O
allungato VSP adige20041008_id414157 O
su E adige20041008_id414157 O
uno RS adige20041008_id414157 O
dei EP adige20041008_id414157 O
pochi DP adige20041008_id414157 O
tratti SP adige20041008_id414157 O
in E adige20041008_id414157 O
salita SS adige20041008_id414157 O
, XPW adige20041008_id414157 O
frazionando VG adige20041008_id414157 O
il RS adige20041008_id414157 O
gruppo SS adige20041008_id414157 O
. XPS adige20041008_id414157 O

Alla ES adige20041008_id414157 O
sua DS adige20041008_id414157 O
ruota SS adige20041008_id414157 O
si PN adige20041008_id414157 O
sono VIY adige20041008_id414157 O

portati VPP adige20041008_id414157 O
altri DP adige20041008_id414157 O
sei N adige20041008_id414157 O
corridori SP adige20041008_id414157 O
che CCHE adige20041008_id414157 O
hanno VIY adige20041008_id414157 O
poi B adige20041008_id414157 O
disputato VSP adige20041008_id414157 O
lo RS adige20041008_id414157 O
sprint SN adige20041008_id414157 O
sul ES adige20041008_id414157 O
traguardo SS adige20041008_id414157 O
di E adige20041008_id414157 O
Bourges SPN adige20041008_id414157 B-GPE
. XPS adige20041008_id414157 O

References

Magnini, B., Pianta, E., Speranza, M., Balrtalesi Lenzi, V., and Sprugnoli, R. Italian Content Annotation Bank (I-CAB): Named Entities, Technical report, ITC-irst, 2007.
On-line: http://evalita.itc.it/doc/I-CAB-Report-Named-Entities.pdf

Linguistic Data Consortium (LDC), *Automatic Content Extraction English Annotation Guidelines for Entities*, version 5.6.1 2005.05.23.
On-line: http://projects.ldc.upenn.edu/ace/docs/English-Entities-Guidelines_v5.6.1.pdf

Web Sites

ACE, http://www.nist.gov/speech/tests/ace/index.htm
http://www.ldc.upenn.edu/Projects/ACE/

Ontotext, http://ontotext.fbk.eu

L'Adige, http://www.ladige.it/

CONLL 2002, http://www.cnts.ua.ac.be/conll2002/ner/