

EVALITA 2007

<http://evalita.itc.it>

Temporal Expression Recognition and Normalization Task Guidelines for Participants

Valentina Bartalesi Lenzi, Rachele Sprugnoli
CELCT, Trento 38100 Italy
{*bartalesi | sprugnoli*}@celct.it

Task Definition

In the Temporal Expression Recognition and Normalization task, systems are required to recognize the Temporal Expressions occurring in a text by identifying their extension, and to normalize them, i.e. give a representation of their meaning by assigning values to a pre-defined set of attributes.

The task will be based on the TIMEX2 standard, with some adaptations to Italian (see Magnini 2007). Both development data and test data are part of the Italian Content Annotation Bank (I-CAB) and will be distributed upon acceptance of the agreement terms for a free research licence.

Two evaluation conditions will be supported for the evaluation. When participants submit their system output for scoring, they must identify which of these two conditions they wish to be evaluated on:

- Recognition only
- Recognition + Normalization

It is possible to submit a maximum of two runs for each subtask.

The Corpus (I-CAB)

Both training data and test data are part of the Italian Content Annotation Bank (I-CAB), developed by ITC-irst and CELCT.

I-CAB consists of 525 news documents taken from the local newspaper “L’Adige”. The selected 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. I-CAB is divided into a development part (335 news stories, for a total of around 113,000 words) and a test part (190 news stories, for a total of around 69,000 words).

Data Format

Training data are distributed in the following formats:

- SGML files contain the source text. All SGML files are in UTF-8 (see Appendix A).

- APF (ACE¹ Program Format) files contain the annotation in the form of XML stand-off annotation, which means that the file as a whole conforms to XML encoding standards, and the raw data being annotated resides in a separate file. The annotations “point” to portions of the raw text via indices (see Appendix B for the Recognition only subtask, and Appendix C for the Recognition and Normalization subtask).

In order to have a clear visualization of the annotated APF files, it is possible to import them with Callisto 1.4.0 (<http://callisto.mitre.org>), a freely distributed annotation tool developed by the MITRE Corporation.

To import the APF files follow these indications:

1. Mouse on: “File” -> “Import”. A pop-up window will appear;
2. Browse to get to the file to import;
3. Select “ACE Event Task” for “Available Importers” and “ACE2004 APF v.5.1.1”;
4. Leave the file encoding at UTF-8;
5. Press “Import”.

Test data will be distributed in the SGML format.

Requirements:

- The data format required for system output is the APF. For each source document in the evaluation data set, the system will have to produce a single APF file as output. The XML DTD for APF can be downloaded from the EVALITA web site: <http://evalita.itc.it/tasks/timex.html>.
You can use an XML validator to verify that a system output file conforms to the ACE DTD for APF. This java implementation is downloadable from the NIST ACE web site: <ftp://jaguar.ncsl.nist.gov/ace/resources/XMLValidator.tar.gz>
- Output files are required to be named as follows:
evalita07_timex_participantname_run1 (or *run2*)
- Output files have to be sent by e-mail both to bartalesi@celct.it and to sprugnoli@celct.it
- Deadline for the submission system results is 1st June 2007, midnight local time

Evaluation Metrics

The final ranking will be based on the TERN value score as defined in the ACE 2007 evaluation campaign².

The TERN-Value score for a system is defined to be the sum of the values of all of the system’s output timex2 tokens, normalized by the sum of the values of all of the reference timex2 tokens. The maximum possible timex2 value score is 100 percent.

$$TERN_Value_{sys} = \frac{\sum_i value_of_sys_token_i}{\sum_j value_of_ref_token_j}$$

We will also provide the following measures:

¹ Automatic Content Extraction

² See Appendix A in <http://www.nist.gov/speech/tests/ace/ace07/doc/ace07-evalplan.v1.3a.pdf> for more details on the TERN-Value score.

- Precision: indicates the percentage of correct positive predictions and is computed as the ratio between the number of Temporal Expressions correctly identified by the system (True Positive) and the total number of Temporal Expressions identified by the system (True Positive plus False Positive).

$$Precision = \frac{TP}{TP + FP}$$

- Recall: indicates the percentage of positive cases recognized by the system and is computed as the ratio between the number of Temporal Expressions correctly identified by the system (True Positive) and the number of Temporal Expressions that the system was expected to recognize (True Positive plus False Negative).

$$Recall = \frac{TP}{TP + FN}$$

- F-measure: the weighted harmonic mean of precision and recall.

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

Recognition subtask

As far as the Recognition subtask is concerned, a minimal overlap in the extent of the reference and the system output tags is required. Overlaps are measured in terms of number of characters for text input and the minimum acceptable overlap of matching is 0.300. The cost (weight) for spurious Temporal Expressions is 0.750 and the cost for missed Temporal Expressions is 1.000.

Recognition + Normalization subtask

Parameters for scoring the Recognition in this subtask are the same as above:

- 0.300, minimum acceptable overlap
- 0.750, cost for spurious Temporal Expressions
- 1.000, cost for missed Temporal Expressions

As far as Normalization is concerned, attribute value assignment measures the ability of the system to correctly assign the normalization attribute values of the Temporal Expressions,. These attributes include VAL, MOD, SET, ANCHOR_VAL and ANCHOR_DIR. The weights assigned to each Timex2 attribute are given in Table 1. Notice that we have modified the default weight of the ANCHOR_DIR attribute (we have assigned to ANCHOR_DIR the same value as ANCHOR_VAL). This change seemed reasonable as these two anchoring attributes are always used together.

| Attribute | Attribute Weight |
|------------|------------------|
| VAL | 1.000 |
| ANCHOR_DIR | 0.500 |
| ANCHOR_VAL | 0.500 |
| MOD | 0.100 |
| SET | 0.100 |

Table 1: Default parameters for scoring TERN attributes

Figure 1 shows an example of the evaluation output. Note that we have modified the Perl program in order to use a consistent and not ambiguous terminology: the abbreviations *Rec* (Recognition) and *Norm* (Normalization) are visualized instead of *Detection* and *Rec* (Recognition) displayed in the original output. For a brief description of the fields in each record, see the “Description of the ace-eval output” document on the NIST ACE web site:

ftp://jaguar.ncsl.nist.gov/ace/resources/ace-eval_output_documentation.pdf

| Timex2 Recognition and Normalization statistics: | | | | | | | | | | | | | | | | | | | | | |
|--|--------|-----|------|------|-----------|------|------|------------|-------|------|----------|------|-------|-------------|------------------------|-------|------|--------|-------|-------|------|
| ref | Count | | | | Count (%) | | | | | | Cost (%) | | | | Unconditioned_Cost (%) | | | | | | |
| | source | Ent | Rec | Norm | Rec | | Norm | Unweighted | | Rec | | Norm | Value | Value-based | | Max | Rec | | Norm | | |
| type | Tot | FA | Miss | Err | FA | Miss | Err | Pre-- | Rec-- | F | FA | Miss | Err | (%) | Pre-- | Rec-- | F | Value | FA | Miss | Err |
| unknown | 11 | 3 | 6 | 0 | 27.3 | 54.5 | 0.0 | 62.5 | 45.5 | 52.6 | 20.5 | 54.5 | 0.0 | 25.0 | 69.0 | 45.5 | 54.8 | 100.00 | 20.45 | 54.55 | 0.00 |
| total | 11 | 3 | 6 | 0 | 27.3 | 54.5 | 0.0 | 62.5 | 45.5 | 52.6 | 20.5 | 54.5 | 0.0 | 25.0 | 69.0 | 45.5 | 54.8 | 100.00 | 20.45 | 54.55 | 0.00 |

Figure 1 The evaluation output

Scorer

For the official evaluation we are going to use the TERN scoring part of `ace07-eval-v01-EVALITA.pl`, whose formulas are described in “The ACE 2007 (ACE07) Evaluation Plan” (<http://www.nist.gov/speech/tests/ace/ace07/doc/ace07-evalplan.v1.3a.pdf>).

You can download the scorer from the EVALITA web site:

<http://evalita.itc.it/tasks/timex.html>

Contacts

| | |
|--|--|
| <p>Valentina Bartalesi Lenzi Tel. 0039 0461 405352 bartalesi@celct.it</p> | <p>Rachele Sprugnoli Tel. 0039 0461 405374 sprugnoli@celct.it</p> |
|--|--|

Appendix A

Input format (example available at: <http://evalita.itc.it/tasks/timex-sample.tar.gz>).

<DOC ID="adige20041007_id413726" DATE="20041007">

<TEXT>

Stasera la performance della rassegna « Mesi d'autore »

Il maggio della Cilento

Di ALESSANDRO FRANCESCHINI

TRENTO - « Mesi d'autore », la rassegna ideata da Amedeo Savoia e promossa dall'Istituto Trentino di Cultura, la Provincia Autonoma di Trento e il Comune di Trento, propone per questa sera una performance tematica sul mese di Maggio. Gli affreschi medievali della Torre dell'Aquila del Castello del Buonconsiglio, dedicati al mese degli innamorati, sono stati la musa ispiratrice di tutte le prestazioni artistiche che vedremo in scena. A partire ad un racconto scritto da Antonella Cilento dal titolo "Occhi a Maggio". La voce calda di Michela Embriaco darà vita alle parole scritte dall'autrice partenopea mentre sugli schermi verranno proiettate le immagini di Marco Comuzzi. Il gruppo Quadrivium eseguirà le musiche originali scritte da Fabio Cifariello Ciardo e che cercheranno di riproporre le atmosfere magiche di maggio. La regia sarà curata da Giacomo Anderle e da Alessio Kogoj. L'appuntamento stasera al Teatro Cuminetti a partire dalla 20.30.

Cilento - che sarà la regina della serata - è un giovane ma già affermata scrittrice che ha saputo ritagliarsi uno spazio importante anche nel campo della formazione: la sua scuola di scrittura - dove lei è docente di scrittura creativa - è nata da alcuni anni ed è ormai considerata tra le più interessanti e valide, non solo a Napoli, ma in tutta Italia. I suoi libri sono stati pubblicati da editori come Guanda, Avagliano e Sironi, a testimonianza della qualità di una scrittura ricercata e intensa.

Stamattina presso la Biblioteca Comunale di Trento la scrittrice incontrerà la cittadinanza, ed in particolare gli studenti delle medie superiori, per una conversazione informale sulla scrittura e sul suo lavoro. A partire dalle 10 nella sede di via Roma.

Entrata libera e gratuita a tutti e due gli appuntamenti. E per chi non fosse ancora soddisfatto la serata si concluderà con un brindisi, a base di Marzemino, offerto dal Gruppo Mezzacorona.

</TEXT>

</DOC>

Appendix B

Output format for the Recognition only subtask (example available at: <http://evalita.itc.it/tasks/timex-sample-onlyrec.tar.gz>).

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE source_file PUBLIC "SYSTEM" "apf.v5.1.1.dtd">

<source_file URI="I-CAB/LOC/v.4corretta-timex/I-CAB_V.4.2_timex-person-org-gpe-
loc/training/20041007/Cultura/adige20041007_id413726.shtml" SOURCE="unknown"
TYPE="text" VERSION="5.0" AUTHOR="sprugnoli" ENCODING="UTF-8">
  <document DOCID="adige20041007_id413726">
    <timex2 ID="adige20041007_id413726-T11">
      <timex2_mention ID="adige20041007_id413726-T11-1">
        <extent>
          <charseq START="931" END="936">maggio</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T7">
      <timex2_mention ID="adige20041007_id413726-T7-1">
        <extent>
          <charseq START="1794" END="1801">dalle 10</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T8">
      <timex2_mention ID="adige20041007_id413726-T8-1">
        <extent>
          <charseq START="0" END="6">Stasera</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T1">
      <timex2_mention ID="adige20041007_id413726-T1-1">
        <extent>
          <charseq START="288" END="298">questa sera</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T13">
      <timex2_mention ID="adige20041007_id413726-T13-1">
        <extent>
          <charseq START="57" END="65">Il maggio</charseq>
        </extent>
      </timex2_mention>
    </timex2>
  </document>
</source_file>
```

</timex2>
<timex2 ID="adige20041007_id413726-T2">
 <timex2_mention ID="adige20041007_id413726-T2-1">
 <extent>
 <charseq START="1016" END="1022">stasera</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T3">
 <timex2_mention ID="adige20041007_id413726-T3-1">
 <extent>
 <charseq START="1054" END="1064">dalla 20.30</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T12">
 <timex2_mention ID="adige20041007_id413726-T12-1">
 <extent>
 <charseq START="1924" END="1932">la serata</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T9">
 <timex2_mention ID="adige20041007_id413726-T9-1">
 <extent>
 <charseq START="325" END="342">sul mese di Maggio</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T4">
 <timex2_mention ID="adige20041007_id413726-T4-1">
 <extent>
 <charseq START="1096" END="1107">della serata</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T5">
 <timex2_mention ID="adige20041007_id413726-T5-1">
 <extent>
 <charseq START="1316" END="1326">alcuni anni</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T10">
 <timex2_mention ID="adige20041007_id413726-T10-1">
 <extent>

```
<charseq START="434" END="440">al mese</charseq>
</extent>
</timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T6">
  <timex2_mention ID="adige20041007_id413726-T6-1">
    <extent>
      <charseq START="1571" END="1580">Stamattina</charseq>
    </extent>
  </timex2_mention>
</timex2>
</document>
</source_file>
```


Appendix C

Output format for the Recognition and Normalization subtask (example available at: <http://evalita.itc.it/tasks/timex-sample-norm.tar.gz>).

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE source_file PUBLIC "SYSTEM" "apf.v5.1.1.dtd">

<source_file URI="adige20041007_id413726.sgm" SOURCE="unknown" TYPE="text"
VERSION="5.0" AUTHOR="romano" ENCODING="UTF-8">
  <document DOCID="adige20041007_id413726">
    <timex2 ID="adige20041007_id413726-T11" VAL="XXXX-05">
      <timex2_mention ID="adige20041007_id413726-T11-1">
        <extent>
          <charseq START="933" END="938">maggio</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T7" VAL="2004-10-07T10:00">
      <timex2_mention ID="adige20041007_id413726-T7-1">
        <extent>
          <charseq START="1796" END="1803">dalle 10</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T8" VAL="2004-10-07TEV">
      <timex2_mention ID="adige20041007_id413726-T8-1">
        <extent>
          <charseq START="2" END="8">Stasera</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T1" VAL="2004-10-07TEV">
      <timex2_mention ID="adige20041007_id413726-T1-1">
        <extent>
          <charseq START="290" END="300">questa sera</charseq>
        </extent>
      </timex2_mention>
    </timex2>
    <timex2 ID="adige20041007_id413726-T13" VAL="XXXX-05">
      <timex2_mention ID="adige20041007_id413726-T13-1">
        <extent>
          <charseq START="59" END="67">Il maggio</charseq>
        </extent>
      </timex2_mention>
    </timex2>
```

<timex2 ID="adige20041007_id413726-T2" VAL="2004-10-07TEV">
 <timex2_mention ID="adige20041007_id413726-T2-1">
 <extent>
 <charseq START="1018" END="1024">stasera</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T3" VAL="T20:30">
 <timex2_mention ID="adige20041007_id413726-T3-1">
 <extent>
 <charseq START="1056" END="1066">dalla 20.30</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T12" VAL="2004-10-07TEV">
 <timex2_mention ID="adige20041007_id413726-T12-1">
 <extent>
 <charseq START="1926" END="1934">la serata</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T9" VAL="XXXX-05">
 <timex2_mention ID="adige20041007_id413726-T9-1">
 <extent>
 <charseq START="327" END="344">sul mese di Maggio</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T4" VAL="2004-10-07TEV">
 <timex2_mention ID="adige20041007_id413726-T4-1">
 <extent>
 <charseq START="1098" END="1109">della serata</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T5" VAL="PAST_REF"
ANCHOR_VAL="2004" ANCHOR_DIR="BEFORE">
 <timex2_mention ID="adige20041007_id413726-T5-1">
 <extent>
 <charseq START="1318" END="1328">alcuni anni</charseq>
 </extent>
 </timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T10" VAL="XXXX-05">
 <timex2_mention ID="adige20041007_id413726-T10-1">
 <extent>

```
<charseq START="436" END="442">al mese</charseq>
</extent>
</timex2_mention>
</timex2>
<timex2 ID="adige20041007_id413726-T6" VAL="2004-10-07TMO">
  <timex2_mention ID="adige20041007_id413726-T6-1">
    <extent>
      <charseq START="1573" END="1582">Stamattina</charseq>
    </extent>
  </timex2_mention>
</timex2>
</document>
</source_file>
```

References

Magnini, B., Negri, M., Pianta, E., Speranza, M., Bartalesi Lenzi, V., and Sprugnoli, R. *Italian Content Annotation Bank (I-CAB): Temporal Expressions*, Technical report, ITC-irst, 2007.

On-line: <http://evalita.itc.it/tasks/timex.html>

Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson G., *TIDES 2005 Standard for the Annotation of Temporal Expressions* April 2005, Updated September 2005.

On-line: http://timex2.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf

Web Sites

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

Callisto, <http://callisto.mitre.org>

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

Ontotext, <http://ontotext.itc.it>

Timex2, <http://timex2.mitre.org/>