

CELCT

Center for the Evaluation of Language
and Communication Technologies
(<http://www.celct.it>)

Amedeo Cappelli
(cappelli@celct.it)



Overview

CELCT (/selekt/):

- joint initiative of **DFKI** and **ITC-Irst**
- autonomous institution located in Trento
- started operating at the beginning of 2004
- funded by the Autonomous Province of Trento
- employees: 6 + 3 consultants + 5 collaborators

Motivations (1)

- **The importance of evaluation**

- Evaluation plays a relevant role in advancing technology
- Evaluation is useful to assess technology (improving software quality)
- Evaluation should collaborate with the research community, but at the same time be independent from technology developers

- **The example of NIST**

NIST designs and enacts evaluation exercises with advice from the research community

Systems that have been participating in the **TREC campaigns** since 1992 have doubled their performances

Most commercial Web search engines nowadays include technologies first developed for the TREC annual evaluation conferences

Motivations (2)

- Contribution to the development of a “strong persistent evaluation infrastructure in Europe” for HLT
- Evaluation Triangle: integration of different competencies
 - **ELDA**
 - ♦ Production and distribution of evaluation corpora
 - ♦ Creation and distribution of evaluation packages
 - ♦ Corpus-based component performance measurement
 - **CELCT**
 - ♦ Component evaluation methodologies and metrics
 - ♦ Component performance analysis
 - ♦ Component performance improvement
 - **DFKI Evaluation Center**
 - ♦ User interface analysis for LT and MM systems
 - ♦ End-to-end evaluation of LT and MM systems
 - ♦ Models and methodologies for the design and evaluation of LT and MM systems

Mission

GOALS:

- setting up all the necessary infrastructure and skills to support evaluation in *Human Language-Multimodal Communication Technologies* (HL-MCTs)
- co-operating with private partners in order to foster the development of HL-MCTs applications for the market

Activities

- studying and devising **evaluation procedures** for HL-MCTs (Question Answering, speech-to-speech translation, spoken language technologies, cross-language Information Retrieval, Word Sense Disambiguation, Multimodal technologies)
- promoting, organizing and editing the proceedings of several evaluation activities in international campaigns, such as:
 - **CLEF** (*Cross Language Evaluation Forum*) 2005 and 2006: main coordinator of the **Cross-Language Question Answering Track**; organiser of the Italian tasks; coordinator of QA track session and editor of the Workshop Proceedings in 2004 and 2005
 - **PASCAL** (*Pattern Analysis, Statistical Modelling, and Computational Learning*): contributor to the creation of the Development and Test Data Sets for the **RTE** (*Recognizing Textual Entailment*) Challenges in 2005 and 2006
 - **DUC** (*Document Understanding Evaluation*): volunteered as an annotator in the Pyramid evaluation method experiment in DUC 2005

Activities (2)

- releasing **guidelines** for HL-MCT evaluation
- **benchmarking** in the field of NLP:
 - ♦ **I-CAB** (*Italian Content Annotation Bank*), a corpus of Italian news annotated with different kinds of semantic information.

The annotation is being carried out manually using Callisto (<http://callisto.mitre.org/>), a tool developed by the MITRE Corporation.

I-CAB: *Italian Content Annotation Bank*

- It is a benchmark intended as a reference work for Information Extraction task, providing annotations of:
 - **temporal expressions**: e.g. *lunedì / Aprile / oggi / 2006*
 - **entities**
 - ◆ persons: e.g. *il presidente / Maria*
 - ◆ organizations: e.g. *NATO / Croce Rossa*
 - ◆ geo-political entities (GPEs): e.g. *Italia / Roma*
 - ◆ locations: e.g. *Via Sommarive 18, 38050 Povo*
 - **relations** among entities: e.g. the relation *date-of-birth* connecting a date to a person

Formalisms adopted

- The activity has been carried out adopting the standards developed within the American **ACE** program (ACE - Automatic Content Extraction, <http://www.nist.gov/speech/tests/ace>)
- The English guidelines provided by the **Linguistic Data Consortium** (LDC, <http://www ldc.upenn.edu>) has been adapted to the Italian

Some Data about I-CAB

- Number of articles = 525
- Number of words = 182,553
- Average number of words per file = 348
- Number of annotated temporal expressions = 4,553
- Number of entities of type person = 7,087

Annotation - scheduling:

- ♦ Organizations (November 2005)
- ♦ GPEs (January 2006)
- ♦ Locations (March 2006)
- ♦ Relations (June 2006)
- ♦ Events (September 2006)

Activities (3)

- **Web site evaluation**

Defining a methodology for:

- Gathering User Requirements
- Task Analysis
- Design

- **Project:** Re-Design and Evaluation of a system for managing the allocation of sports facilities in Trento and Surroundings

Partners:

- **WEBSS** – A Software Development Company
- **ASIS** – An organization that manages sport facilities in Trento for a variety of associations (professional-competitive, educational, governmental, private).

Activities (4)

▪ Eye Tracking

Set up a laboratory with Stationary and Mobile eye trackers

- Eyetracker equipment
- GazeTracker software (for analysis of eye tracker data)
- Sample Study with The Mobile Eye Tracker
- Hands on experience

Italian VoiceAward

- **Italian Chapter of the Voice Business**
- Aim: To evaluate and award voice based applications for Italian language
 - Managing, Testing and analysis
Starting January 2006
 - VoiceDay
June 2006