CELI Language & Information Technology
CELI was established in 1999 in Turin, Italy by a group of language technology researchers with the aim to address the growing market in language engineering.

Today, CELI is a provider of software solutions in the field of Natural Language Processing (NLP) in mono-, multi-, and cross-language perspective.
Natural Language Processing objectives and fields of application

NLP Definitions:
- field concerned with the interactions between computers and human (natural) languages
- algorithms and methods to process texts and to make their information accessible to computer applications, i.e. allow computers to understand human language
- computer science enabling computers to use human languages both as input and as output

NLP Objectives:
design and build software that will analyze, understand, and generate languages that humans use naturally, so that eventually we will be able to address our computers as though we were addressing another person

NLP Application:
Customer and research projects
Our mission is to improve the process of extracting meaningful intelligence from unstructured multilingual texts.
Solutions

Voice of the customer

Semantic annotations, indexing and search on unstructured information for Knowledge Management and (social) CRM

Knowledge

Speech Technology

Text processing components for speech applications on mobile terminals
What we do

CELI NLP solutions provide knowledge

**Input:** multilingual unstructured text information

**Output:** knowledge (examples)

- **Names**
  - Products, people, places, etc.,

- **Opinions**
  - Preferences of your consumers

- **Discoveries**
  - New unpredicted information

- **Phonetic transcription**
  - Pronunciation representation for speech products

- **Classified texts**
  - Classification according to topics/problems: credit card problems, unsatisfied caller, access to an account, etc.
What we do

CELI NLP solutions provide business insight

CELI NLP Solution

NLP Components

Process & Analyze Text

Generate Business Insight

Auto Classification
Morphological Analysis
Named Entity Recognition
Language Identification
Opinion Mining
Phonetic Transcription
CELI’s NLP solutions - sources of innovation

CELI’s solutions takes origins in deep scientific research that CELI carries out since 1996 in partnership with important universities and research institutions all over the world. All the research projects relate to the multilingual information processing and access.

Some of our research partners:

- German Research Center for Artificial Intelligence (DFKI), Germany
- National Institute of Information and Communications Technology, Japan
- Polish Academy of Science, Poland
- Scuola Normale Superiore di Pisa, Italy
- University of Turin, Italy
- University of Trento, Italy
- Free University of Bozen, Italy
- Fondazione Bruno Kessler (FBK), Italy
- Brandeis University, USA
- Wien University, Austria
- Institute of Legal Information Theory and Techniques, Italy
- Istituto di Linguistica Computazionale “Antonio Zampolli”, Italy
Research is a strategic strength of CELI allowing the transfer of most recent achievements to commercial applications.

Customized language technology for multiple applications

Customized language resources and software modules

Consulting and expertise in the development of linguistic technologies
## What we do

### Delivery models

<table>
<thead>
<tr>
<th>CELI solutions</th>
<th>Delivery models</th>
<th>Types of products</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Language Processing</td>
<td>Web services API</td>
<td>NLP components</td>
<td>Text classification, opinion mining, sentiment analysis, named entities extraction, semantic clustering, word similarity linguistic analysis ...</td>
</tr>
<tr>
<td></td>
<td>Licences</td>
<td>Language Resources</td>
<td>Text corpus, speech corpus, morphology lexicons, special purposes dictionaries, semantic networks....</td>
</tr>
<tr>
<td></td>
<td>Projects (Consulting)</td>
<td>Design, development and integration</td>
<td>Development and integration of CELI language analysis solutions for particular applications: TTS, ASR, Contact Centers....</td>
</tr>
</tbody>
</table>
Our language processing and language analysis module can operate simultaneously in wide range of languages

- Italian
- English
- French
- Spanish
- Portuguese
- German
- Dutch
- Swedish
- Norwegian
- Finnish
- Danish
- Polish
- Russian
- Ukrainian
- Czech
- Slovak
- Slovenian
- Turkish
- Modern Standard Arabic
- Romanian
- Bulgarian
- Japanese
- Greek

- CELI is one of very few companies that can offer such wide language coverage
- Resources and technologies for all languages rely on one common data format and structure
- New languages are under development
What we do

What distinguishes us from our competitors

- Information processing in mono-, multi- and cross-language perspective
- Wide coverage of languages and NLP technologies
- Multiple delivery models:
  - **Web service delivery model:**
    - Flexible budget organization
    - Fees based on data volumes/number of licenses released, etc.
    - Pay as you go - you pay only when you use it
    - All capital expenditure remain on provider side (hw and sw)
    - Low integration effort in terms of time and costs
    - Low switching costs - don’t cause dependence to one provider
    - Fast and easy customization without geographical barriers
    - Maintenance on provider side
  - **Project based model:**
    - Not only competence but also infrastructure and language resources
  - **Licences:**
    - Non standard packages released with standard non terminable licenses
Projects and Products Examples
NLP applications
Multilinguism is a crucial part in communications in the global market and must be addressed to effectively manage data.

### OBJECTIVES

- Provide classified information received in many different languages from customer care centers located in different countries to a marketing department.

### APPROACH

- Create common taxonomy for all languages taking into account cultural differences.
- Develop software and lingware for analysis of high volume of data in different languages.
- Organize a team of language experts for development of the multilingual resources and quality check of results.

### RESULTS

- Development of infrastructure integrated with the client’s CRM system that manages continuous flux of multilingual information and its automatic classification.
- The client receives automatically classified information for all requested languages and has a view of its customer satisfaction in significantly reduced time and costs of translation.
- The automatic system elaborates 20,000 messages per month.
Multilinguism is a crucial part in communications in the global market and must be addressed to effectively manage data.

With Celi’s linguistic technologies:
- Multilingual call-centers
- Notes and tickets in multiple languages
- Multilingual language technology integrated with Enterprise CRM System
- Raw text messages classified
- Classification data integrated
- Unified Reports

Without Celi’s technologies:
- Multilingual call-centers
- Translators
- Inconsistent Reports

Project example 1 - Solution for CRM
Synthetic voices are now expected to be the quality of a natural voices - this requires full understanding of human language habits

Project example - Solutions for multilingual text-to-speech system

OBJECTIVES

- support the development of text-to-speech systems for new languages
- improve the performance of the existing languages
- organize comparative and independent evaluations

APPROACH

- multiple language coverage
- survey languages with focus on particularities to be taken into account for TTS development
- pronunciation analysis
- lexical resources w/ phonetic transcription
- complete module development for TTS
- comprehensive quality control both as technology and approach

RESULTS

- TTS systems for 6 new languages
- Improved significantly performance of 7 other languages
- The final products, the TTS system, have been ranked among the best solutions currently available on the market
Synthetic voices are now expected to be the quality of a natural voices - this requires full understanding of human language habits

**CEL I contribution of consulting and technologies**

- Overall linguistic feasibility study and design
- Multilingual text preprocessing methods
- Morphological / Syntactic analyzer
- Lexical resources, text corpora
- Phonetic transcription grammars
- Prosody-annotated corpus
- Voice recording assistance
- Acoustic database annotation assistance
- Quality assessment
- Evaluation/comparison of competing products

---

**Project example 2 - text processing for Text-to-Speech**

**concatenative TTS modules**

- **NLP module**
  - text preprocessing
  - morphological / syntactic analysis
- letter-to-sound phonetic transcription
- prosody generation

- **Voice generation Module**
  - acoustic database creation/annotation
  - unit selection algorithms
  - acoustic processor

**Multilingual input text**

**Synthesized speech**
Objective:

measure the online presence, analyze performance, monitor reputation, competition and engagement in order to prepare focused strategies for online and offline communication.

The digitization of consumption has radically changed consumer habits. It became natural that the consumer seeks the views and reviews about the desired product on the web before making a purchase.
Project example 3 - Social Media

Social Media Monitoring & Analysis

BRAND PRESENCE
Share of buzz
How much you are talked about on the web;
Count of posts and citations;
Competitive analysis
Trends
What is the impact of an online or offline campaign

SENTIMENT ANALYSIS
Share of sentiment
In what way people talk about you on the web; positively, negatively or with neutral tone? Are there any questions about you?
Sentiment for specific arguments
What are the main strengths and weaknesses?

TOPIC ANALYSIS
Share of topic
What are the most important arguments;
What is the most discussed topic?
Concept cloud
What are the most frequent concepts;
User generated lexicons

SOURCES
Share of buzz for each domain
What are the most important places that people talk about you. Blog, forum, QA?
I promessi sposi 2.0

la felicità di comprendere
Vittorio Di Tomaso
ditomaso@celi.it

Via San Quintino 31 - 10121 Torino
Tel. +39 011.562.71.15 - Fax +39 011.506.40.86
info@celi
www.celi.it