A Platform for the Production of Language Resources

PANACEA, a **Platform for the Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for the Human Language Technologies,** is addressing the most critical aspect for Machine Translation and other language technology-based systems : the so-called language-resource bottleneck.

PANACEA is a web service-based production line that automates the stages involved in the acquisition, production, updating and maintenance of the Language Resources required by MT and other Language Technologies.

PANACEA will generate an automatic, dynamic system conceived as an integrated machinery for the production of Language Resources, covering the following processes:
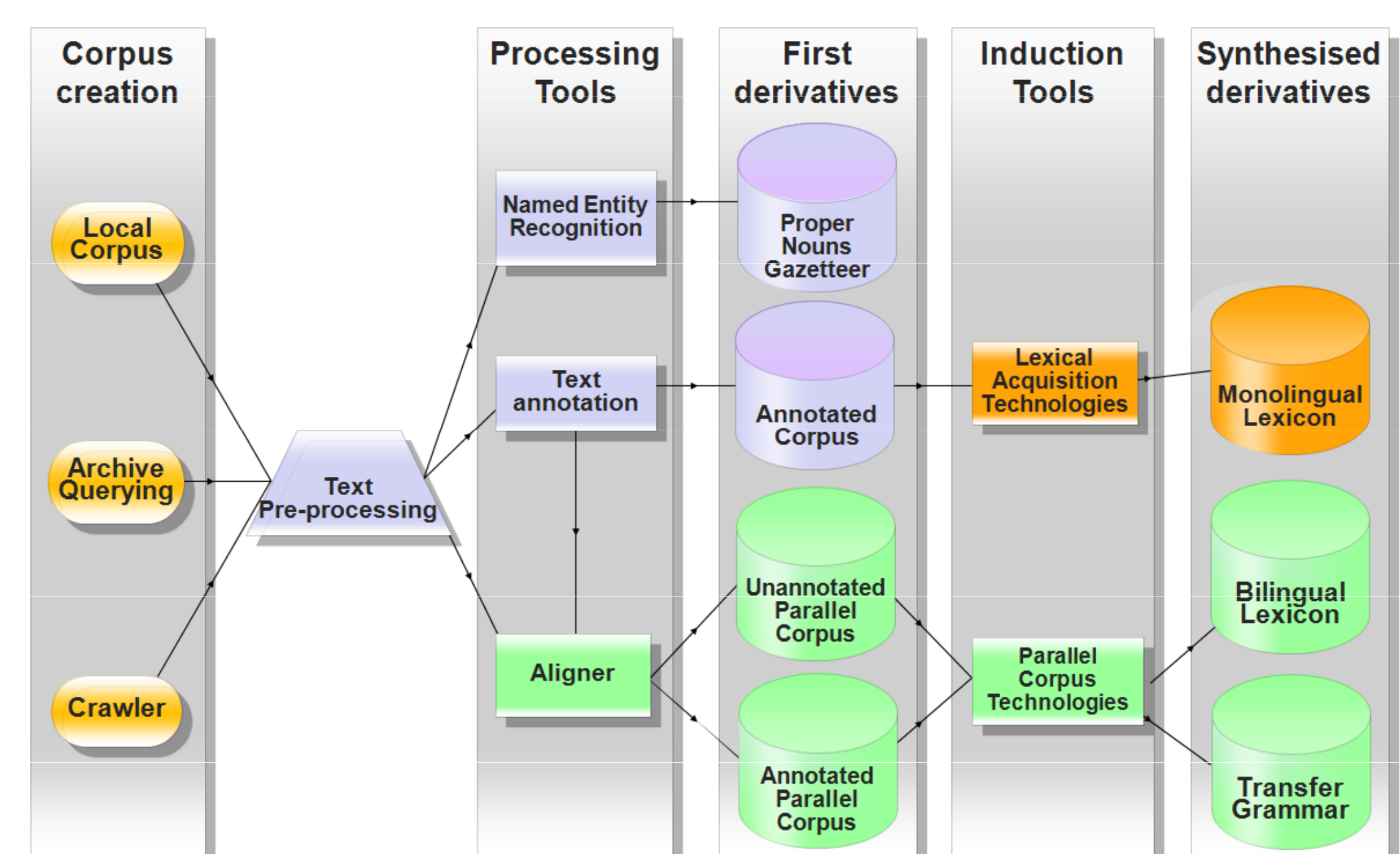
- Monolingual and parallel text acquisition and pre-processing;
- Sentential and sub-sentential alignment;
- Bilingual dictionary and transfer grammar production;
- Lexical information acquisition for rich information dictionary production.

Available and accessible results

The first version of the platform and of the interoperable web services will be available by **February 2011**. These first integrated web services will give access to Corpus Acquisition components, Clean-Up and Normalization and Text Processing components up to PoS Tagging for Spanish, Italian, Greek, English, German and French. By **October 2011** the platform will give access to Sub-sentential and Sentence alignment components and to Bilingual Dictionary and Transfer Grammar production Components. By **June 2012** the components for Rich Dictionary production will be accessible as well.

Project Challenges

- To propose the definition of an interoperability space for the integration of robust, scalable, web service-deployed components;

- To handle the impact of dealing with massive data;

- To be convincing on the industrial use of the available automatic acquisition technologies by introducing ready-to-use tools, with confidence indicators, which give priority to high precision;

- To carry out further research for increasing accuracy in the automatic acquisition and production of Language Resources.

Impact on Human Language Technologies

PANACEA wants to build a Language Resource Factory that can ensure the continuous supply needed by the Language Technology industry to meet current and future needs for Machine Translation systems covering all languages and domains in the time required by the market.

Keep visiting www.panacea-lr.eu to access PANACEA's results and available web services

Project Information

Funding Scheme:	**STREP project**
Project Acronym:	**PANACEA**
EU funding:	**2.68 M€**
Start date:	**1 January 2010**
Duration:	**36 months**
Contract Nr.	**FP7-ICT-2009-4-248064**