The SCRIBO Module of the OLENA Platform: a Free Software Framework for Document Image Analysis
Guillaume Lazzara, Roland Levillain, Thierry Géraud, Yann Jacquelet, Julien Marquegnies, Arthur Crépin-Leblond
EPITA Research and Development Laboratory (LRDE), France
olenarlde.epita.fr

The Olena platform

- DIA Applications
  - GUI
  - CLI
  - Web services

- DIA framework (SCRIBO)

- Image Processing C++ Library (Milena)

More information

Online demos
http://olenarlde.epita.fr/Demos
Website
http://olenarlde.epita.fr/
Contact
olenarlde.epita.fr

The SCRIBO Project [1]
- Project conducted in the context of the “System‘tic Paris-Région” Cluster (France).
- 9 Partners: AFP, CEA-List, EPITA, INRIA-Alpaca, Mandriva, Nuxeo, Proxem, Tagmatica, XWiki.
- 3 years of development.
- Budget of 3.5M€

The SCRIBO module: a DIA Framework

- Provides
  - Basic routines
  - Basic DIA toolchains
  - Text in document
  - Document layout analysis
  - Text in picture
  - High-level data structures
  - Novel algorithms and techniques
  - Standard I/O
  - GUI and Command Line Interface (CLI)

- Facts
  - 3 years of development
  - 40K lines of C++
  - Open Source GPL v2
  - Used in Nepomuk/KDE

- Assets
  - End-to-end tools → From digital document to HTML and PDF reconstruction.
  - Based on a well established IP library.


- Provides
  - Data structures
  - Safe data types
  - More than 70 algorithms
  - Memory management

- Facts
  - 10 years of development
  - Version 1.0 released on July 2009
  - 120K lines of C++
  - Open Source GPL v2

References

[1] SCRIBO, Semi-automatic and Collaborative Retrieval of Information Based on Ontologies.
http://www.scribo.ws.

Why and how to design a generic and efficient image processing framework: The case of the Milena library.