SECONDA
(A tool for analyzing software ecosystems)

Software Engineering Lab.

Service de Génie Logiciel, Institut d'Informatique
Université de Mons

UMONS   Faculté
            des Sciences

February 3, 2012

1 Introduction

2 Architecture

3 Data extraction

4 Identity merging

5 Ecosystem visualisation

6 Project visualisation
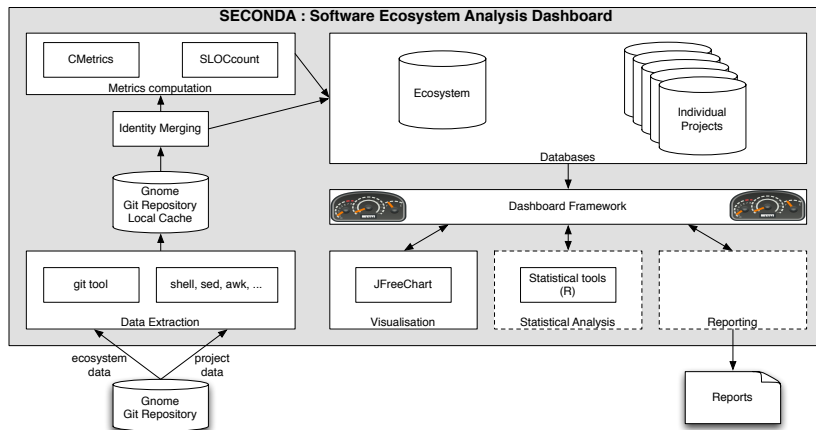
7 Future work

Software Ecosystems

- Lungu 2009: *Reverse Engineering Software Ecosystems*, PhD Thesis
  - Software **ecosystems** are coherent collections of software projects that evolve together
  - Software **communities** are groups of software developers that work together to maintain the software ecosystem's projects

- The quality and "health" of a project is related to its ecosystem and community.

- We need to study:
  - The relations between the projects of the ecosystem
  - The relations between the developers of the community

SECONDA

- Software analysis
    - Ecosystem grouped analysis
    - Single project analysis

- Developer's Analysis
    - Community analysis
    - Single developer analysis

- Tools:
    - Data extraction
    - Identity merging
    - Metrics
    - Visualisation
    - Statistics
    - Reporting

SECONDA architecture overview

Gnome Ecosystem

- Tool is currently tailored to analyze Gnome ecosystem:
    - > 1300 projects
    - > 5000 contributors
    - $>$ one decade (since 1997) development history
    - 2 months $<$ project's lifetime < 14 years

	authors	committers	commits	files
minimum	1	1	1	25
Q1	3	2	23	61
median	12	9	131	112
Q3	59	46	517	237
maximum	1142	692	35191	7097
mean	62.07	45.78	760.2	252.3

Identity Merging

- Identification of the different identities of a developer to provide:
    - Better identification of each developer activities
    - More accurate analyses

- 2 different merging algorithms implemented so far
    - easy integration of other algorithms
    - Goeminne, Mens 2011: *A comparison of identity merge algorithms for software repositories*

- Manual interaction for fixing false positives / negatives

Identity Merging Panel

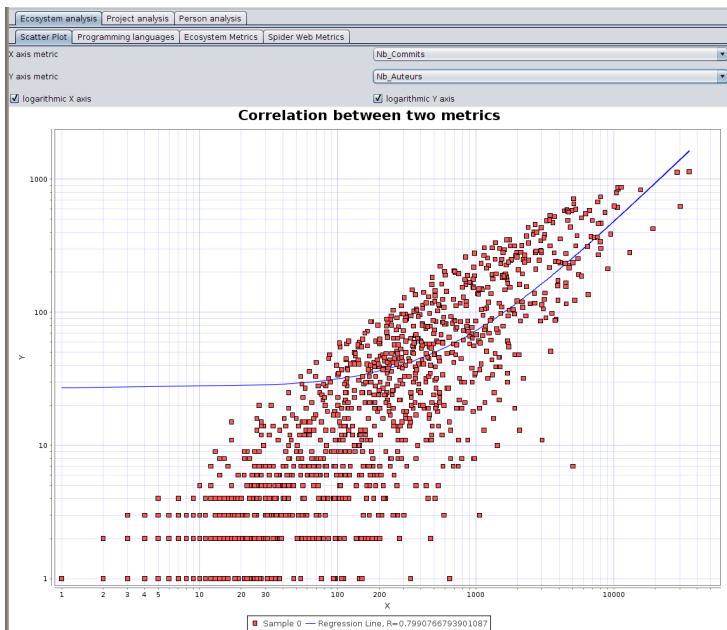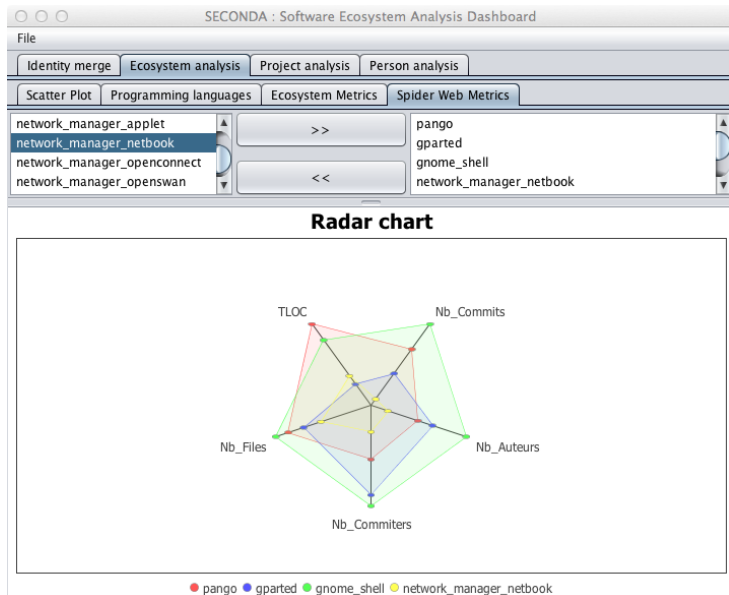| Ecosystem analysis	Project analysis	Person analysis	Identity merge

By identity	By project

-- All projects -- ▼ ☐ grouped by identity   merge

Id	Content
6	(bakery,Claude Paroz < _____ @2xlibre.net>)
6	(bakery,Claude Paroz < _____ @src.gnome.org>)
7	(gnome-subtitles,Daniel Nylander < _____ >)
8	(bakery,Chao-Hsiung Liao < _____ @src.gnome.org>)
9	(gnome-subtitles,Claude Paroz < _____ @2xlibre.net>)
9	(gnome-subtitles,Claude Paroz < _____ @src.gnome.org>)
10	(bakery,Daniel Nylander < _____ @src.gnome.org>)
10	(gnome-documents,Daniel Nylander < _____ >)
11	(bakery,Djihed Afifi < _____ @src.gnome.org>)
12	(bakery,Petr Kovar < _____ @volny.cz>)
13	(bakery,Matej Urbančič < _____ @svn.gnome.org>)
14	(bakery,Kjartan Maraas < _____ @gnome.org>)
14	(bakery,Kjartan Maraas < _____ @src.gnome.org>)
15	(gnome-subtitles,Olav Vitters < _____ @src.gnome.org>)
16	(gnome-documents,Chao-Hsiung Liao < _____ @yahoo.com.tw>)
17	(bakery,Olav Vitters < _____ >)

Ecosystem Visualisation

- General metrics (number of authors, committers, commits, number of files, LOC, ...)

- Project's metrics correlation (scatterplots)

- Programming languages usage distribution (boxplots)

- Ecosystem metrics distribution (boxplots)

- Projects's metrics comparison (spider web charts)

Ecosystem Visualisation

- General metrics + complexity metrics for each project revision

- Visualisation of a project's file size evolution (histogram + rev. selection)

- Visualisation of a project's metrics evolution (boxplot + rev. selection)

Future Work

- Statistics module
- Reporting module
- More individual developer metrics
- Other ecosystems
- Incremental and automatic data extraction and analysis
- Support for other data sources (repositories, bug trackers, mailing lists, dev. fora)
- . . .
- Suggestions?

SECONDA
(A tool for analyzing software ecosystems)

Software Engineering Lab.

Service de Génie Logiciel, Institut d'Informatique
Université de Mons

UMONS   Faculté
des Sciences

February 3, 2012