

Intermediate Scientific Report
Hasler Project no. 2234
*“Enabling the evolution of J2EE applications through
reverse engineering and quality assurance”*

October 23, 2008

a) Summary of results

This project focuses on reverse engineering enterprise applications that are implemented in various languages. Some of the most significant results include:

- Meta-modeling — We designed a modeling platform based on a minimal meta-meta-model. We developed a new approach for integrating annotations in the reverse engineering process.
- Parsing — We started to investigate a lightweight approach for parsing programs written in unknown languages.
- Visualization — We have developed a novel software visualization technique by employing a cartography metaphor. While the technique is generic, it also proved to be useful for representing J2EE systems.

Results

We present the results obtained during the period from 2007-10-01 to 2008-09-30.

Meta-modeling

Enterprise applications contain large amounts of heterogeneous data. We have developed Fame, a modeling platform that is based on a minimal meta-meta-model [KV08]. The platform is implemented in several languages and it will allow us to leverage data imported from these various languages.

The complexity and size of enterprise systems hamper the understanding process. This situation is even more accentuated by relevant information being only implicitly present in the system, thus difficult to recover by purely automated techniques. We developed an approach to capture such external information in form of annotations that enrich the model and that can be taken into account in further analyses [Br08, BGGN08]. The approach is implemented on top of Moose and will allow us to integrate external information gathered from studying large case studies into our research.

We have also investigated synergies between other programming languages at the level of models. PyPy is a collection of tools to support the implementation of the Python language in Python itself. A key feature of this toolchain is the use of an interpreter as a high-level language specification which can be specialized to a compiler. We have participated in an experiment to apply the PyPy toolchain to the implementation of a virtual machine for Smalltalk [BKL⁺08]. This was a larger effort that also included contributions from our recently completed SNF project, “Analyzing, Capturing and Taming Software Change” (SNF 200020-113342).

Parsing

Parsing is always a bottleneck in the reverse engineering process as languages evolve and parsers are left behind. To cope with this problem we started to research an approach that leverages the knowledge hidden in the IDE. Typically, IDEs represent the programs abstract syntax trees. Our approach allows us to extract this information and include it in the analysis environment [LV08]. We intend to continue on this path and complement it with other approaches that map the abstract syntax trees to higher level models.

Visualization

Visualization compresses the complexity of software systems into pictures to ease understanding. Because software has no physical shape, there is no direct mapping to a visual representation. We built a novel software visualization technique by employing a cartography metaphor. Using our approach, we laid out the software systems based on the vocabulary used [KLN08]. While the technique is generic, it also proved to be useful for representing J2EE systems to differentiate between artifacts written in Java, XML or JSP.

On a related track, we have also explored several other layouts for representing software systems. In particular, we have proposed several visualizations of Lisp systems to show how they look different than object-oriented programs [DGM08].

Case Studies

Enterprise systems are typically proprietary. Thus, to study them we need access to case studies provided by industrial partners. Over the past months, we acquired J2EE industrial case studies from two companies, namely Eidgenössischen Institut für Geistiges Eigentum (IGE) and Glue Software Engineering AG, and we started to analyze the specifics of these systems in cooperation with the developers.

Workshops and Tutorials

Queries are effective tools to deal with large amounts of data. We provided an overview of various techniques and co-organized a working session on how query technologies can be used for program comprehension [VGG08].

In this project, the implementation and research effort is based on the Moose platform. Moose was originally built at the University of Berne and it recently started to attract interest from a number of other researchers. To strengthen the community and to exchange ideas we organized a second workshop on FAMIX and Moose in Software Reengineering [DGG⁺08].

Staff contributions

- Toon Verwaest is in the first year of the PhD. He has implemented Fame, a polyglot meta-modeling platform of representing various languages and he has reported on this experience [KLN08]. Fame stays at the core of the Moose analysis platform. He started to investigate techniques for developing parsing techniques for unknown languages [LV08]. He also participated in the PyPy experiment mentioned earlier [BKL⁺08].
- Tudor Gîrba serves as PostDoc. He is investigating problems specific to quality assurance. He started to work on large J2EE case studies provided by industrial partners like Eidgenössischen Institut für Geistiges Eigentum (IGE) and Glue Software Engineering AG. He is also investigating the integration of 3rd party information into the reverse engineering process [BGGN08].

Changes to the research plan

The project is largely proceeding as planned, however there have been some shifts in focus:

- For the meta-modeling part we focused first on a high-level infrastructure for handling models as we strongly believe it will ease the manipulation of further models [KV08].

- The metrics effort was redistributed to parsing and it has further led to a novel approach to getting information out of the IDE [LV08]. We believe the track is promising and we intend to continue in this direction. Further work on metrics will be largely driven by the case studies.
- The visualization track remained within the bounds of the initial plan.

Due to the size of the project, we have attracted several master and bachelor students to work on topics related to it. Furthermore, Fabrizio Perin has recently joined our group as a PhD student and he will also contribute to the project.

Important events

- Toon Verwaest presented the following papers:
 - “Back to the future in one week — implementing a Smalltalk VM in PyPy” at the Workshop on Self-sustaining Systems (S3) 2008
 - “FAME, A Polyglot Library for Metamodeling at Runtime” at the Workshop on Models at Runtime co-located with Models 2008
- Tudor Gîrba presented the following paper:
 - “Enriching Reverse Engineering with Annotations” at International Conference on Model Driven Engineering Languages and Systems (Models 2008)
- Tudor Gîrba gave the following tutorials:
 - “Pragmatic Design Quality Assessment” at the International Conference on Software Engineering (ICSE 2008)
 - “Moose Tutorial” at the Working Conference on Reverse Engineering (WCRE 2008)
- Oscar Nierstrasz presented the following papers:
 - “Consistent Layout for Thematic Software Maps” at Working Conference on Reverse Engineering (WCRE 2008)
 - “Extracting models from IDEs” at Workshop on FAMIX and Moose in Software Reengineering co-located with WCRE 2008
- Oscar Nierstrasz was an invited speaker at SVPP 2008 (Software Variability: a Programmers’ Perspective — Brussels, Belgium, August 8-9, 2008).
- Tudor Gîrba was an invited speaker at SATToSE 2008 (Seminar on Advanced Tools and Techniques for Software Evolution)
- Oscar Nierstrasz was PC Member:
 - WASDeTT 2008 (International Workshop on Advanced Software Development Tools and Techniques — Paphos, Cyprus, July 8, 2008)
 - TOOLS Europe 2008 (International Conference on Objects, Models, Components, Patterns — Zurich, Switzerland, June 30 - July 4, 2008)
 - ICPC 2008 (International Conference on Program Comprehension — Amsterdam, The Netherlands, 10-13 June, 2008)
- Tudor Gîrba was PC Member:
 - Models 2008 (International Conference on Model Driven Engineering Languages and Systems — Toulouse, Sept 30 - Oct 2, 2008)

- ICPC 2008 (International Conference on Program Comprehension — Amsterdam, The Netherlands, June 10-13, 2008)
- MSR 2008 (Working Conference on Mining Software Repositories — collocated with ICSE 2008, Leipzig, Germany, May 10-11, 2008)
- PCODA 2007 (International Workshop on Program Analysis through Dynamic Analysis — co-located with WCRE 2007, Oct. 29, 2007)
- ICSM 2007 (International Conference on Software Maintenance — Paris, France, Oct. 2-5, 2007)
- IWPSE 2007 (International Workshop on Principles of Software Evolution — co-located with ESEC/FSE, Dubrovnik, Croatia, Sept, 3-4, 2007)
- Oscar Nierstrasz was co-organizer of FAMOOSr 2008 (Workshop on FAMIX and Moose in Reengineering — co-located with WCRE 2008).
- Tudor Gîrba was co-organizer:
 - FAMOOSr 2008 (Workshop on FAMIX and Moose in Reengineering — co-located with WCRE 2008).
 - QTAPC 2008 (Working Session on Query Technologies and Applications for Program Comprehension — co-located with ICPC 2008).

b) Publications

Published papers are annexed to this report. They are all available electronically as PDF files at the following url:

<http://www.iam.unibe.ch/~scg/cgi-bin/scgbib.cgi?hasler08>

Please note that theses and student projects are *not* included with this report, but are nevertheless available electronically from the above URL.

Published papers

- [BGGN08] Andrea Brühlmann, Tudor Gîrba, Orla Greevy, and Oscar Nierstrasz. Enriching reverse engineering with annotations. In Krzysztof Czarnecki et al., editor, *International Conference on Model Driven Engineering Languages and Systems (Models 2008)*, volume 5301 of *LNCS*, pages 660–674. Springer-Verlag, 2008.
- [DGG⁺08] Stéphane Ducasse, Tudor Gîrba, Orla Greevy, Michele Lanza, and Oscar Nierstrasz. Workshop on FAMIX and Moose in software reengineering (FAMOOSr 2008). In *15th Working Conference on Software Maintenance and Reengineering (WCRE 2008)*, pages 343–344, October 2008.
- [DGM08] Adrian Doza, Tudor Gîrba, and Radu Marinescu. How Lisp systems look different. In *European Conference on Software Maintenance and Re-Engineering (CSMR 2008)*, pages 223–232. IEEE Computer Society Press, 2008.
- [GLM08] Tudor Gîrba, Michele Lanza, and Radu Marinescu. Pragmatic software quality detection, 2008. Tutorial held at ICSE 2008.
- [KLN08] Adrian Kuhn, Peter Loretan, and Oscar Nierstrasz. Consistent layout for thematic software maps. In *Proceedings of 15th Working Conference on Reverse Engineering (WCRE'08)*, pages 209–218, Los Alamitos CA, October 2008. IEEE Computer Society Press.
- [KV08] Adrian Kuhn and Toon Verwaest. FAME, a polyglot library for metamodeling at runtime. In *Workshop on Models at Runtime*, page n10, 2008.
- [LV08] Daniel Langone and Toon Verwaest. Extracting models from IDEs. In *2nd Workshop on FAMIX and Moose in Software Reengineering (FAMOOSr 2008)*, pages 32–35, October 2008.
- [VGG08] Mathieu Verbaere, Michael W. Godfrey, and Tudor Gîrba. Query technologies and applications for program comprehension. In *Proceedings of International Conference on Program Comprehension (ICPC 2008)*, pages 285–288, 2008.

Theses and Student projects

- [Br08] Andrea Brühlmann. Enriching reverse engineering with annotations. Master's thesis, University of Bern, April 2008.

c) Publications in press

Publications to appear

- [BKL⁺08] Carl Friedrich Bolz, Adrian Kuhn, Adrian Lienhard, Nicholas D. Matsakis, Oscar Nierstrasz, Lukas Renggli, Armin Rigo, and Toon Verwaest. Back to the future in one week — implementing a Smalltalk VM in PyPy. In *Workshop on Self-sustaining Systems (S3) 2008*, 2008. To appear.