Activity Report 2015

Project-Team ATEAMS

Analysis and Transformation based on rEliaBle tool coMpositionS
# Table of contents

1. Members ................................................................. 1
2. Overall Objectives .................................................. 2
3. Research Program .................................................... 2
    3.1. Research method 2
    3.2. Software analysis 3
    3.3. Refactoring and Transformation 4
    3.4. The Rascal Meta-programming language 4
    3.5. Domain-specific Languages 5
4. Highlights of the Year ............................................... 6
5. New Software and Platforms ....................................... 6
    5.1. MicroMachinations 6
    5.2. OSSMETER 6
    5.3. Rascal 7
    5.4. Meerkat 7
    5.5. Iguana 7
    5.6. Capsule 8
6. New Results ........................................................... 8
    6.1. Faster Immutable Data Structures for the JVM 8
    6.2. Automated Measurement and Analysis of Open Source Software 8
    6.3. Modular Interpreters for the Masses 8
    6.4. One Parser to Rule Them All 9
    6.5. A Pattern-Based Game Mechanics Design Assistant 9
7. Bilateral Contracts and Grants with Industry .................... 9
8. Partnerships and Cooperations .................................... 9
    8.1. National Initiatives ........................................... 9
    8.1.1. Master Software Engineering 9
    8.1.2. Early Quality Assurance in Software Production 10
    8.1.3. Next Generation Auditing: Data-assurance as a service 10
    8.1.4. Domain-Specific Languages: A Big Future for Small Programs 10
    8.2. European Initiatives .......................................... 10
    8.2.1. FP7 & H2020 Projects ....................................... 10
    8.2.2. Collaborations with Major European Organizations 10
    8.3. International Initiatives ....................................... 10
        8.3.1.1. Informal International Partners 10
        8.3.1.2. Research stays abroad .................................. 11
9. Dissemination .......................................................... 11
    9.1. Promoting Scientific Activities .............................. 11
    9.1.1. Scientific events organisation 11
    9.1.1.1. General chair, scientific chair 11
    9.1.1.2. Member of the organizing committees 11
    9.1.2. Scientific events selection ................................ 11
    9.1.2.1. Member of the conference program committees 11
    9.1.2.2. Reviewer .................................................. 11
    9.1.3. Invited talks ................................................ 11
    9.1.4. Leadership within the scientific community ............... 12
    9.1.4.1. Member of steering committees 12
    9.1.4.2. Member of other groups .................................. 12
    9.2. Teaching - Supervision - Juries ............................ 12
    9.2.1. Teaching .................................................... 12
9.2.2. Supervision 12
9.2.3. Juries 13
9.3. Popularization 13
10. Bibliography 13
Project-Team ATEAMS

Creation of the Project-Team: 2009 July 01, end of the Project-Team: 2015 December 31

Keywords:

Computer Science and Digital Science:
1. - Architectures, systems and networks
2. - Software
2.1.1. - Semantics of programming languages
2.1.10. - Domain-specific languages
2.1.2. - Object-oriented programming
2.1.3. - Functional programming
2.2.1. - Static analysis
2.5. - Software engineering

Other Research Topics and Application Domains:
6.1. - Software industry
6.1.1. - Software engineering
6.1.2. - Software evolution, maintenance
6.6. - Embedded systems

1. Members

Research Scientists
Jurgen Vinju [Team leader, Centrum Wiskunde & Informatica, Professor]
Tijs Van Der Storm [Centrum Wiskunde & Informatica]
Jan Van Eijck [Centrum Wiskunde & Informatica]

Faculty Member
Paul Klint [Centrum Wiskunde & Informatica, Professor]

Engineers
Maarten Dijkema [Centrum Wiskunde & Informatica]
Bert Lisser [Centrum Wiskunde & Informatica]

PhD Students
Ali Afroozeh [Centrum Wiskunde & Informatica]
Pablo Inostroza Valdera [Centrum Wiskunde & Informatica]
Anastasia Izmaylova [Centrum Wiskunde & Informatica]
Davy Landman [Centrum Wiskunde & Informatica]
Michael Steindorfer [Centrum Wiskunde & Informatica]
Riemer Van Rozen [Centrum Wiskunde & Informatica]
Jouke Stoel [Centrum Wiskunde & Informatica]

Administrative Assistant
Sandrine Meilen [Inria]
2. Overall Objectives

2.1. Presentation

Software is very complex, and it seems to become more complex every year. Over the last decades, computer science has delivered various insights on how to organize software better. Via structured programming, modules, objects, components, and agents, these days software systems are more and more evolving into “systems of systems” that provide services to each other. Each system is large, uses incompatible — new, outdated or non-standard — technology and above all, exhibits failures.

It is becoming more and more urgent to analyze the properties of these complicated, heterogeneous and very large software systems and to refactor and transform them to make them simpler and to keep them up-to-date. With the plethora of different languages and technology platforms it is becoming very difficult and very expensive to construct tools to achieve this.

The main challenge of ATEAMS is to address this combination of a need for all kinds of novel analysis and transformation tools and the existence of the diversity of programming environments. We do this by investing in a virtual laboratory called “Rascal”. It is a domain specific programming language for source code analysis, transformation, and generation. Rascal is a programming language parametric, such that it can be used to analyze, transform, or generate source code in any language. By combining concepts from both program analysis and transformation into this language we can efficiently experiment with all kinds of tools and algorithms.

We now focus on three sub-problems. Firstly, we study software analysis: to extract information from existing software systems and to analyze it. The extracted information is vital to construct sound abstract models that can be used in further analysis. We apply these extraction techniques now to analyze (large bodies of) source code: finding bugs and finding the causes of software complexity.

Secondly, we study refactoring: to semi-automatically improve the quality of a software system without changing its behavior. Refactoring tools are a combination of analysis and transformations. Implementations of refactoring tools are complex and often broken. We study better ways of designing refactorings and we study ways to enable new (more advanced and useful) refactorings. We apply these refactorings now to isolate design choices in large software systems and compare systems that are equal except a single design choice.

Finally, we study code generation from domain specific languages (DSLs). Here we also find a combination of analysis and transformation. Designing, implementing and, very importantly, maintaining DSLs is costly. We focus on application areas such as Computational Auditing, Game Economies, and Core Banking to experiment with this subject. In Computational Auditing we are focusing on modeling interactive questionnaires. The Game economies domain involves modeling and verifying the dynamic behavior of game play. Core banking requires the formal modeling of financial services and products.

3. Research Program

3.1. Research method

We are inspired by formal methods and logic to construct new tools for software analysis, transformation and generation. We try and proof the correctness of new algorithms using any means necessary.

Nevertheless we mainly focus on the study of existing (large) software artifacts to validate the effectiveness of new tools. We apply the scientific method. To (in)validate our hypothesis we often use detailed manual source code analysis, or we use software metrics, and we have started to use more human subjects (programmers).

Note that we maintain ties with the CWI spinoff “Software Improvement Group” which services most of the Dutch software industry and government and many European companies as well. This provides access to software systems and information about software systems that is valuable in our research.
3.2. Software analysis

This research focuses on source code; to analyze it, transform it and generate it. Each analysis or transformation begins with fact extraction. After that we may analyze specific software systems or large bodies of software systems. Our goal is to improve software systems by understanding and resolving the causes of software complexity. The approach is captured in the EASY acronym: Extract Analyze SYnthesize. The first step is to extract facts from source code. These facts are then enriched and refined in an analysis phase. Finally the result is synthesized in the form of transformed or generated source code, a metrics report, a visualization or some other output artifact.

The mother and father of fact extraction techniques are probably Lex, a scanner generator, and AWK, a language intended for fact extraction from textual records and report generation. Lex is intended to read a file character-by-character and produce output when certain regular expressions (for identifiers, floating point constants, keywords) are recognized. AWK reads its input line-by-line and regular expression matches are applied to each line to extract facts. User-defined actions (in particular print statements) can be associated with each successful match. This approach based on regular expressions is in wide use for solving many problems such as data collection, data mining, fact extraction, consistency checking, and system administration. This same approach is used in languages like Perl, Python, and Ruby. Murphy and Notkin have specialized the AWK-approach for the domain of fact extraction from source code. The key idea is to extend the expressivity of regular expressions by adding context information, in such a way that, for instance, the begin and end of a procedure declaration can be recognized. This approach has, for instance, been used for call graph extraction but becomes cumbersome when more complex context information has to be taken into account such as scope information, variable qualification, or nested language constructs. This suggests using grammar-based approaches as will be pursued in the proposed project. Another line of research is the explicit instrumentation of existing compilers with fact extraction capabilities. Examples are: the GNU C compiler GCC, the CPPX C++ compiler, and the Columbus C/C++ analysis framework. The Rigi system provides several fixed fact extractors for a number of languages. The extracted facts are represented as tuples (see below). The CodeSurfer source code analysis tool extracts a standard collection of facts that can be further analyzed with built-in tools or user-defined programs written in Scheme. In all these cases the programming language as well as the set of extracted facts are fixed thus limiting the range of problems that can be solved.

The approach we are exploring is the use of syntax-related program patterns for fact extraction. An early proposal for such a pattern-based approach consisted of extending a fixed base language (either C or PL/1 variant) with pattern matching primitives. In our own previous work on RScript we have already proposed a query algebra to express direct queries on the syntax tree. It also allows the querying of information that is attached to the syntax tree via annotations. A unifying view is to consider the syntax tree itself as “facts” and to represent it as a relation. This idea is already quite old. For instance, Linton proposes to represent all syntactic as well as semantic aspects of a program as relations and to use SQL to query them. Due to the lack of expressiveness of SQL (notably the lack of transitive closure) and the performance problems encountered, this approach has not seen wider use.

Parsing is a fundamental tool for fact extraction for source code. Our group has longstanding contributions in the field of Generalized LR parsing and Scannerless parsing. Such generalized parsing techniques enable generation of parsers for a wide range of existing (legacy) programming languages, which is highly relevant for experimental research and validation.

Extracted facts are often refined, enriched and queried in the analysis phase. We propose to use a relational formalization of the facts. That is, facts are represented as sets of tuples, which can then be queried using relational algebra operators (e.g., domain, transitive closure, projection, composition etc.). This relational representation facilitates dealing with graphs, which are commonly needed during program analysis, for instance when processing control-flow or data-flow graphs. The Rascal language integrates a relational sub-language by providing comprehensions over different kinds of data types, in combination with powerful pattern matching and built-in primitives for computing (transitive/reflexive) closures and fixpoint computations (equation solving).
3.2.1. Goals
The main goal is to replace labour-intensive manual programming of fact extractors by automatic generation based on concise and formal specification. There is a wide open scientific challenge here: to create a uniform and generic framework for fact extraction that is superior to current more ad-hoc approaches, yet flexible enough to be customized to the analysis case at hand. We expect to develop new ideas and techniques for generic (language-parametric) fact extraction from source code and other software artifacts.

Given the advances made in fact extraction we are starting to apply our techniques to observe source code and analyze it in detail.

3.3. Refactoring and Transformation
The second goal, to be able to safely refactor or transform source code can be realized in strong collaboration with extraction and analysis.

Software refactoring is usually understood as changing software with the purpose of increasing its readability and maintainability rather than changing its external behavior. Refactoring is an essential tool in all agile software engineering methodologies. Refactoring is usually supported by an interactive refactoring tool and consists of the following steps:

- Select a code fragment to refactor.
- Select a refactoring to apply to it.
- Optionally, provide extra parameter needed by the refactoring (e.g., a new name in a renaming).

The refactoring tool will now test whether the preconditions for the refactoring are satisfied. Note that this requires fact extraction from the source code. If this fails the user is informed. The refactoring tool shows the effects of the refactoring before effectuating them. This gives the user the opportunity to disable the refactoring in specific cases. The refactoring tool applies the refactoring for all enabled cases. Note that this implies a transformation of the source code. Some refactorings can be applied to any programming language (e.g., rename) and others are language specific (e.g., Pull Up Method). At http://www.refactoring.com an extensive list of refactorings can be found.

There is hardly any general and pragmatic theory for refactoring, since each refactoring requires different static analysis techniques to be able to check the preconditions. Full blown semantic specification of programming languages have turned out to be infeasible, let alone easily adaptable to small changes in language semantics. On the other hand, each refactoring is an instance of the extract, analyze and transform paradigm. Software transformation regards more general changes such as adding functionality and improving non-functional properties like performance and reliability. It also includes transformation from/to the same language (source-to-source translation) and transformation between different languages (conversion, translation). The underlying techniques for refactoring and transformation are mostly the same. We base our source code transformation techniques on the classical concept of term rewriting, or aspects thereof. It offers simple but powerful pattern matching and pattern construction features (list matching, AC Matching), and type-safe heterogeneous data-structure traversal methods that are certainly applicable for source code transformation.

3.3.1. Goals
Our goal is to integrate the techniques from program transformation completely with relational queries. Refactoring and transformation form the Achilles Heel of any effort to change and improve software. Our innovation is in the strict language-parametric approach that may yield a library of generic analyses and transformations that can be reused across a wide range of programming and application languages. The challenge is to make this approach scale to large bodies of source code and rapid response times for precondition checking.

3.4. The Rascal Meta-programming language
The Rascal Domain-Specific Language for Source code analysis and Transformation is developed by ATeams. It is a language specifically designed for any kind of meta programming.
Meta programming is a large and diverse area both conceptually and technologically. There are plentiful libraries, tools and languages available but integrated facilities that combine both source code analysis and source code transformation are scarce. Both domains depend on a wide range of concepts such as grammars and parsing, abstract syntax trees, pattern matching, generalized tree traversal, constraint solving, type inference, high fidelity transformations, slicing, abstract interpretation, model checking, and abstract state machines. Examples of tools that implement some of these concepts are ANTLR, ASF+SDF, CodeSurfer, Crocopat, DMS, Grok, Stratego, TOM and TXL. These tools either specialize in analysis or in transformation, but not in both. As a result, combinations of analysis and transformation tools are used to get the job done. For instance, ASF+SDF relies on RScript for querying and TXL interfaces with databases or query tools. In other approaches, analysis and transformation are implemented from scratch, as done in the Eclipse JDT. The TOM tool adds transformation primitives to Java, such that libraries for analysis can be used directly. In either approach, the job of integrating analysis with transformation has to be done over and over again for each application and this requires a significant investment.

We propose a more radical solution by completely merging the set of concepts for analysis and transformation of source code into a single language called Rascal. This language covers the range of applications from pure analyses to pure transformations and everything in between. Our contribution does not consist of new concepts or language features per se, but rather the careful collaboration, integration and cross-fertilization of existing concepts and language features.

3.4.1. Goals

The goals of Rascal are: (a) to remove the cognitive and computational overhead of integrating analysis and transformation tools, (b) to provide a safe and interactive environment for constructing and experimenting with large and complicated source code analyses and transformations such as, for instance, needed for refactorings, and (c) to be easily understandable by a large group of computer programming experts. Rascal is not limited to one particular object programming language, but is generically applicable. Reusable, language specific, functionality is realized as libraries. As an end-result we envision Rascal to be a one-stop shop for source code analysis, transformation, generation and visualization.

3.5. Domain-specific Languages

Our final goal is centered around Domain-specific languages (DSLs), which are software languages tailored to a specific problem domain. DSLs can provide orders of magnitude improvement in terms of software quality and productivity. However, the implementation of DSLs is challenging and requires not only thorough knowledge of the problem domain (e.g., finance, digital forensics, insurance, auditing etc.), but also knowledge of language implementation (e.g., parsing, compilation, type checking etc.). Tools for language implementation have been around since the archetypical parser generator YACC. However, many of such tools are characterized by high learning curves, lack of integration of language implementation facets, and lead to implementations that are hard to maintain. This line of research focuses on two topics: improve the practice and experience of DSL implementation, and evaluate the success of DSLs in industrial practice.

Language workbenches [5] are integrated environments to facilitate the development of all aspects of DSLs. This includes IDE support (e.g., syntax coloring, outlining, reference resolving etc.) for the defined languages. Rascal can be seen as a language workbench that focuses on flexibility, programmability and modularity. DSL implementation is, in essence, an instance of source code analysis and transformation. As a result, Rascal’s features for fact extraction, analysis, tree traversal and synthesis are an excellent fit for this area. An important aspect in this line of research is bringing the IDE closer to the source code. This will involve investigation of heterogeneous representations of source code, by integrating graphical, tabular or forms-based user interface elements. As a result, we propose Rascal as a feature-rich workbench for model-driven software development.

The second component of this research is concerned with evaluating DSLs in industrial contexts. This means that DSLs constructed using Rascal will be applied in real-life environments so that expected improvements in quality, performance, or productivity can be observed. We already have experience with this in the domain of digital forensics, computational auditing and games.
3.5.1. Goals

The goal of this research topic is to improve the practice of DSL-based software development through language design and tool support. A primary focus is to extend the IDE support provided by Rascal, and to facilitate incremental, and iterative design of DSLs. The latter is supported by new (meta-)language constructs for extending existing language implementations. This will require research into extensible programming and composition of compilers, interpreters and type checkers. Finally, a DSL is never an island: it will have to integrate with (third-party) source code, such as host language, libraries, runtime systems etc. This leads to the vision of multi-lingual programming environments [15].

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

Prof.dr. Paul Klint won the IEEE TCSE Software Engineering Distinguished Service Award 2015. This award is presented “annually for outstanding and/or sustained contributions and service to the software engineering community”.

5. New Software and Platforms

5.1. MicroMachinations

**FUNCTIONAL DESCRIPTION**

Objective: To create an integrated, live environment for modelling and evolving game economies. This will allow game designers to experiment with different strategies to realise game mechanics. The environment integrates with the SPIN model checker to prove properties (reachability, liveness). A runtime system for executing game economies allows MicroMachinations models to be embedded in actual games.

Impact: One of the important problems in game software development is the distance between game design and implementation in software. MicroMachinations has the potential to bridge this gap by providing live design tools that directly modify or create the desired software behaviours.

- Participants: Paul Klint and Riemer Van Rozen
- Contact: Riemer Van Rozen
- URL: https://github.com/vrozen/MM-Lib

5.2. OSSMETER

**KEYWORDS:** Software Quality, Metrics, Open-source

**SCIENTIFIC DESCRIPTION:** OSSMETER meets the challenge of software project quality assessment via fact-based business intelligence. The goal of the project was to design and evaluate a platform for incremental analysis of long lasting open-source projects to support decision making on the corporate level. **FUNCTIONAL DESCRIPTION:** OSSMETER is a platform which integrates metrics of open-source projects: their source code quality, the contents of their social interactions and their activity in issue tracking systems. It includes a fully programmable user-defined quality model utility and configurable dash-board user-interface. The basic metrics of the platform and their aggregation to the project level are carefully considered and rationalised.

- Participants: Paul Klint, Jurgen Vinju, Tijs Van Der Storm, Ashim Shahi, Bas Basten.
- Contact: Jurgen Vinju
- URL: http://www.ossmeter.org/
5.3. Rascal

**KEYWORDS:** Metaprogramming - Language

**SCIENTIFIC DESCRIPTION**

Rascal primitives include immutable data, context-free grammars and algebraic data-types, relations, relational calculus operators, advanced patterns matching, generic type-safe traversal, comprehensions, concrete syntax for objects, lexically scoped backtracking, and string templates for code generation. It has libraries for integrating language front-ends, for reusing analysis algorithms, for getting typed meta-data out of version management systems, for interactive visualization, etc.

**FUNCTIONAL DESCRIPTION**

Rascal is a programming language, such that meta programs can be created by, understood by, and debugged by programmers.

You want to use the best tool for the job when analyzing, transforming or generating source code, so normally you will end up with many different tools, possibly even written in different languages. Now the problem is to integrate these tools again. Rascal solves this problem by integrating source code analysis, transformation, and generation primitives on the language level. Use it for any kind of metaprogramming task: to construct parsers for programming languages, to analyze and transform source code, or to define new DSLs with full IDE support.

- Participants: Paul Klint, Jurgen Vinju, Tijs Van Der Storm, Davy Landman, Bert Lisser, Atze Van Der Ploeg, Vadim Zaytsev, Anastasia Izmaylova, Michael Steindorfer, Jouke Stoel, Ali Afroozeh and Ashim Shahi
- Contact: Paul Klint
- URL: http://www.rascal-mpl.org/

5.4. Meerkat

**FUNCTIONAL DESCRIPTION**

Objective: To enable fully context-free general parsing using a parser combinator library (including allowing left recursion and arbitrary context-sensitive disambiguation).

Impact: Meerkat explores algorithmic advances in context-free general parsing (based on the GLL parsing algorithm and memoized continuations) in the context of a scala parsing combinator library. This library uniquely combines the worst-case execution time guarantees of GLL with the flexibility of parsing combinators. [47]

- Participants: Anastasia Izmaylova, Ali Afroozeh and Tijs van der Storm.
- Contact: Anastasia Izmaylova, Ali Afroozeh
- URL: http://meerkat-parser.github.io/

5.5. Iguana

**FUNCTIONAL DESCRIPTION**

Objective: To provide a data-dependent context-free general parsing infra-structure for parsing programming languages and other formal data, program and modeling notations.

Impact: Iguana is a fast implementation of data-dependent grammars based on the GLL context-free parsing algorithm with data-dependent non-terminals and constraints on top. It comes with a number of high-level disambiguation constructs which are translated to the intermediate layer of data-dependent (E)BNF before being loaded into an object-oriented implementation of GLL based on abstract transition network. Using Iguana parsers for languages which are considered to be hard to parse (such as Haskell and OCAML) are within reach of being generated from simple declarative specifications [25].

- Participants: Anastasia Izmaylova, Ali Afroozeh.
- Contact: Anastasia Izmaylova, Ali Afroozeh
- URL: http://iguana-parser.github.io/
5.6. Capsule

**FUNCTIONAL DESCRIPTION**

Objective: A generic and highly optimised product-family of immutable collection data-structures.

Impact: Capsule is a library for immutable sets, maps and tables. The code is generated using high-level descriptions of the requirements and internal trade-offs of hash-trie map based implementations. We are using this code generator to experiment with the fastest and leanest representations of these persistent data-types to satisfy the requirements of Rascal meta-programming applications in static analysis, empirical research in software engineering and software analytics [37].

- Participants: Michael Steindorfer, Jurgen Vinju
- Contact: Michael Steindorfer
- URL: http://usethesource.io/projects/capsule/

6. New Results

6.1. Faster Immutable Data Structures for the JVM

Immutable data structures involve copying when updating. Efficient implementations use persistent data-structures, so that most of the unchanged data is shared between the copies. Existing libraries for such data structures in the context of the Java virtual machine (JVM), such as the data structures in Clojure and Scala, are based on Hash Array-Mapped Tries (HAMTs), which provide efficient insertion and concatenation operations for persistent maps and sets. In [37] Steindorfer and Vinju presented additional optimisation which allow such operations to be up to 28 times faster than in the Clojure and Scala libraries. Furthermore, the cost of equality checking of such data structures is lower as well. All this, without incurring additional memory.

6.2. Automated Measurement and Analysis of Open Source Software

Deciding whether an open source software (OSS) meets the required standards for adoption in terms of quality, maturity, activity of development and user support is not a straightforward process. It involves analysing various sources of information, including the project’s source code repositories, communication channels, and bug tracking systems. OSSMETER extends state-of-the-art techniques in the field of automated analysis and measurement of open-source software (OSS), and develops a platform that supports decision makers in the process of discovering, comparing, assessing and monitoring the health, quality, impact and activity of opensource software. To achieve this, OSSMETER computes trustworthy quality indicators by performing advanced analysis and integration of information from diverse sources including the project metadata, source code repositories, communication channels and bug tracking systems of OSS projects [29], [26].

This result comes from intensive collaboration in the FP7 STREP project “OSSMETER”. The ATEAMS contribution is focused around source code metrics and activity analysis for Java and PHP.

6.3. Modular Interpreters for the Masses

Object Algebras [46] are new design pattern for increased modularity and extensibility of tree based, abstract data types. By modelling the abstract syntax of a language as a generic factory interface, implementations of this interface provide multiple semantics of the data. For instance, one can define evaluation, type checking and pretty printing of the abstract syntax fully modularly. Additionally, the pattern allows syntax extension: adding a new constructor to the datatype, and modularly extending any existing interpretations to deal with the construct. The same interpretation of different constructs, however, might involve different kinds of context information. For instance, evaluation of arithmetic expressions does not require any context information, but...
evaluation of variables and binders requires and environment. In [34], Inostroza and Van der Storm introduce a simple, modular, and type safe technique to allow such interpretations to be composed anyway. It is based on lifting one interpreter to implicitly propagate the context information it does not require, so that the signatures of the interpreters become compatible. As a result, semantic definitions of language modules do not have to anticipate all kinds of context information that might be required by other modules with which it might be composed. The technique is simple, does not sacrifice separate compilation, is easy automate, and works in mainstream languages. It provides a first step towards a foundation for defining language by assembling modular building blocks.

6.4. One Parser to Rule Them All

Parsing realistic languages requires much more than just a parsing algorithm. Different kinds of language require advanced disambiguation, operator priorities, off-side rule checking, whitespace dependence or data dependence. In [25], Afroozeh and Izmaylova showed how most of these concerns are actually instances of data dependent parsing: the parsing process depends on the value of previously parsed input. They provided an encoding of indentation sensitive parsing, operator precedence and parsing in the presence of preprocessor directives, to a simple, data dependent core language which is executed using the general parsing algorithm GLL. By exposing the data dependent machinery at the level of the grammar formalism, this opens up a range of possibilities for custom parsing aspects, and provides a clear semantics for existing concerns like disambiguation.

6.5. A Pattern-Based Game Mechanics Design Assistant

Video game designers iteratively improve player experience by play testing game software and adjusting its design. Deciding how to improve gameplay, however, is difficult and time-consuming: designers lack an effective means for exploring decision alternatives and modifying a game’s mechanics. In [35], Van Rozen presented the Mechanics Pattern Language (MPL) for encoding common game economy structures and design intent, and a Mechanics Design Assistant (MeDeA) for analyzing, explaining, understanding existing mechanics, and generating, filtering, exploring and applying design alternatives for modifying mechanics. As a result, game designers’ productivity and game quality is increased by providing feedback and design alternatives early in the development cycle. Furthermore, the game economy modifications are applied at runtime using the MicroMachinations library, so that the effect of changes can be immediately experienced.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants with Industry

With the ING bank we are running a four-year project on advising and research in functional and non-functional properties of a part of the ING IT-infrastructure. The project involves modelling a large part of the product portfolio and using state-of-the-art MDE technology to simulate, verify and generate part of its IT infrastructure. The funding of this project is approximately 50% industry, 50% grants from CWI & NWO.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. Master Software Engineering

ATEAMS is a core partner in the Master Software Engineering at Universiteit van Amsterdam. This master is a collaboration between SWAT/AITEMS, Universiteit van Amsterdam, Vrije Universiteit and Hogeschool van Amsterdam.
8.1.2. Early Quality Assurance in Software Production

The EQUA project is a collaboration among Hogeschool van Amsterdam (main partner) Centrum Wiskunde & Informatica (CWI), Technisch Universiteit Delft, Laboratory for Quality of Software (LaQuSo), Info Support, Software Improvement Group (SIG), and Fontys Hogeschool Eindhoven.

8.1.3. Next Generation Auditing: Data-assurance as a service

This project is a collaboration between Centrum Wiskunde & Informatica (CWI) PriceWaterhouseCoopers (PWC), Belastingdienst (National Tax Office), and Computational Auditing, is to enable research in the field of computational auditing.

8.1.4. Domain-Specific Languages: A Big Future for Small Programs

Software and programming have a brilliant past that has brought us the automation of many expected and unexpected human and societal activities ranging from banking and consumer electronics to mobile networking, search engines and social networks. The present of software is overwhelming: many software systems have sizes in the range of 10–100 million lines of source code and contain tens of thousands of errors that are yet to be discovered. We claim that software will only have a big future if software itself becomes smaller. Smaller software leads to higher software productivity (we have to write less) and higher software quality (quality guarantees become part of the language and not of the program).

This project is funded by NWO (the Dutch national science foundation).

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

- FP7 STREP “OSSMETER — Automated Measurement and Analysis of Open Source Software” (ended in 2015)

8.2.2. Collaborations with Major European Organizations

Centrum Wiskunde & Informatica (CWI): Software Analysis & Transformation (Netherlands)

CWI SWAT is the research team associated directly with ATEAMS.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

ATEAMS collaborates with the following research teams:

- Eindhoven Technical University - SET (Eindhoven, The Netherlands)
- Universiteit van Amsterdam - Systems and Network Engineering (Amsterdam, The Netherlands)
- Royal Holloway University of London - Dept. of Computer Science
- The University of Hong Kong (China) - Computer Science
- Delft Technical University (The Netherlands)
- University of Texas at Austin (USA)
- TU Darmstadt (Germany)
8.3.1.2. Research stays abroad
- Michael Steindorfer stayed for 3 months at Oracle Labs in Austria to study efficient data-structures and data-structure optimisations on the JVM.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair
- Tijs van der Storm: co-organizer of the 3rd Workshop on Domain-Specific Language Design and Implementation (DSLDe’15).

9.1.1.2. Member of the organizing committees
- Tijs van der Storm: Publicity co-chair SPLASH and track organizer SPLASH-I
- Jurgen Vinju: co-organizer Bits & Chips Software Event (Legacy Software Track), Eindhoven, The Netherlands

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees
- Paul Klint: International Conference on Software Language Engineering (SLE’15), 14th edition of the BElgian-NEtherlands software eVOLution seminar (BENEVOL 2015), 22nd IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER’15)

9.1.2.2. Reviewer

9.1.2.2.1. Journal
- Member of the editorial boards:
  - Paul Klint: editor of Science of Computer Programming (Elsevier SCP)
- Reviewer - Reviewing activities:
  - Tijs van der Storm: Science of Computer Programming, Computer Languages and Systems (COMLAN)

9.1.3. Invited talks
- Textual and projectional language workbenches. Tijs van der Storm, Dagstuhl on “Domain-Specific Languages”.
• “Software Engineering: The War Against Complexity”, Jurgen Vinju. CHA-Q Open Tool Demonstrations Event (keynote), February 24, Antwerpen, Belgium.
• “Program Analysis and Transformation with Rascal” tutorial at the 36th annual ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI) in Portland, USA. Mark Hills, Jurgen Vinju, Paul Klint.

9.1.4. Leadership within the scientific community

9.1.4.1. Member of steering committees

• Jurgen Vinju: International Conference on Software Language Engineering (ACM SLE) (chair)
• Jurgen Vinju: International Working Conference on Source Analysis and Manipulation (IEEE SCAM)

9.1.4.2. Member of other groups

• Jurgen Vinju: Member of IFIP WG 2.3 — Working Group on Software Implementation Technology
• Tijs van der Storm: Member of IFIP WG 2.16 — Working Group on Language Design
• Tijs van der Storm: Member of the board of European Association for Programming Languages and Systems (EAPLS)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Tijs van der Storm, “Software Construction”, Masters, Universiteit van Amsterdam
Jurgen J. Vinju, “Software Evolution”, Masters, Universiteit van Amsterdam

9.2.2. Supervision

PhD : Atze van der Ploeg, “Efficient Abstractions for Visualization and Interaction”, Universiteit van Amsterdam, April, 8th, 2015 [48] (advisors: Tijs van der Storm and Paul Klint)
PhD in progress : Pablo Inostroza Valdera, “Modular Language Implementation with Object Algebras”, 2018 (advisors: Tijs van der Storm and Paul Klint)
PhD in progress : Anastasia Izmaylova, “Parser Combinators Revisited”, 2017 (advisors: Jurgen Vinju and Paul Klint)
PhD in progress : Riemen van Rozen, “Domain-Specific Languages for Game Development”, 2017 (advisors: Paul Klint and Tijs van der Storm)
PhD in progress : Michael Steindorfer, “Fast and Lean Immutable Data Structures”, 2017 (advisors: Jurgen Vinju and Paul Klint)
PhD in progress : Jouke Stoel, “Solving the Bank”, 2019 (advisors: Jurgen Vinju and Tijs van der Storm)
Msc : Zisimopolous, P.: “A Grid Scheduling infrastructure for SmartConnect’s performance monitoring calculations”, Universiteit van Amsterdam, The Netherlands
Msc : Zhelyazkov, A.T.: “Form controls in WebGL. A stepping stone to a WebGL library for developing commercial interactive 3D websites”, Universiteit van Amsterdam, The Netherlands
Msc: Harezlak, H.: “Geographically-aware scaling for real-time persistent WebSocket applications”, Universiteit van Amsterdam, The Netherlands
Msc: Iwan Flameling: “An automatic CSRF protection tool”, Universiteit van Amsterdam, The Netherlands
Msc: Omar Pakker: “Graph-Based Querying On top of the Entity Framework”, Universiteit van Amsterdam, The Netherlands

9.2.3. Juries

- Jurgen Vinju:
  - Phd Reza Yazdanshenas, A. — Universitet i Oslo, Norway
  - Phd Hafeez Osman, M. — Universiteit Leiden, The Netherlands

9.3. Popularization

- Paul Klint: “Paul Klint in RTL Nieuws over het nieuwe betalingssysteem van het pgb” (radio interview).
- Paul Klint: “De evolutie van codetaal”, media appearance, NRC Handelsblad
- Paul Klint: “Première internetfilm over Nederlands informatica erfgoed”, media appearance, Automatisering Gids.
- Jurgen Vinju: “Public/Private Collaboration {in,for,with} Software Engineering”, 21st Annual Conference of the European Association of Research Managers and Administrators, Leiden, The Netherlands
- Jurgen Vinju: OSSMETER Pitch EU Concertation Meeting - Turning cloud research into innovative software & services, March 25th 2015, Brussels, Belgium.

10. Bibliography

Major publications by the team in recent years


[8] M. HILLS, P. KLINT, T. VAN DER STORM, J. VINIU. *A One-Stop Shop For Software Evolution Tool Construction*, in "ERCIM News", 2012, n0 88, pp. 11 - 12, http://hal.inria.fr/hal-00756876


Publications of the year

Articles in International Peer-Reviewed Journals


International Conferences with Proceedings


Scientific Books (or Scientific Book chapters)

[40] T. V. D. STORM, S. ERDWEG (editors). Proceedings Of The 3rd Workshop On Domain-Specific Language Design And Implementation (DSLDI’15), August 2015, https://hal.inria.fr/hal-01261478


References in notes

