Software Evolution – Reader

Edition 2018/2019 – version 0.15

Paul Klint (guest lecturer) [1 5], Jurgen Vinju [1 5], Magiel Bruntink (guest lecturer) [4], Davy Landman (guest lecturer) [5], Vadim Zaytsev (guest lecturer) [6], Ana Oprescu (lab teacher) [2], and Riemer van Rozen (lecturer, lab teacher) [1 2 3]

[1]Software Analysis & Transformation, Centrum Wiskunde & Informatica
[2]Master of Software Engineering, University of Amsterdam
[3]Amsterdam University of Applied Sciences
[4]Software Improvement Group
[5]SWAT.engineering
[6]Raincode

October 28, 2018

Abstract

This is a reader to the course Software Evolution. It describes course goals, a week-by-week course schedule, obligatory assignments and grading. In a nutshell, this manual explains how to pass this course[1]. Updates are provided on Canvas.

1 Course Overview

Software Evolution is a course in the Master of Software Engineering at the University of Amsterdam of 6 ECTS. We provide descriptions of course material in Section 1.2, required course activities in Section 1.3, a detailed schedule in Section 1.5, reading in Section 1.6 and the evaluation in Section 1.4. Section 2 describe the practical assignments. This document is continuously being updated: Please check Section 3 for a log of modifications over document versions. Please read this document carefully!

[1]This reader does not explain how to get the most out of this course, that's up to you.

Software Evolution Evolution
The Software Evolution course itself has evolved over the years. Thanks and kudos for developing and maintaining this course and its assignments go to: Prof. Dr. Paul Klint, Prof. Dr. Jurgen Vinju, Dr. Magiel Bruntink and Dr. Vadim Zaytsev. Editor of the reader: ir. Riemer van Rozen.

1.1 Goals

The goals of the course are described in the study guide[2].

Exit qualification: *"The graduate masters the methods and techniques needed to analyze an existing software system and to enable it to evolve given changing requirements."*

Our objectives are three-fold:

- The first objective is to acquire an understanding and appreciation of the challenges posed by software maintenance and software evolution.
- The second objective is to learn about quality of software and source code and how it affects software maintenance and evolution.
- The final objective is to be able to select and also construct software analysis and software transformation tools to help obtain insight in software quality and to help improve software quality.

The course ties in closely with paper writing sessions where the objectives are to learn from academic literature, to develop curiosity, and to improve argumentation and writing skills.

1.2 Course Material

Slides & Papers. We provide a selection of scientific papers and lecture slides, which are available on Canvas. Additional papers can be found at the

- ACM Digital Library `http://www.acm.org/dl`
- IEEE Digital Library `http://ieeexplore.ieee.org` and `https://www.computer.org`

Rascal. For the practical lab assignments use metaprogramming language and language workbench Rascal[3]. Rascal has a built-in *Tutor* that provides explanations on concepts and interactive exercises for learning to apply language features. A non-interactive version is available online[4]. Additionally, questions can be posed on Stackoverflow[5] using the rascal tag, and issues can be reported in GitHub[6].

[2] `http://studiegids.uva.nl/xmlpages/page/2018-2019/zoek-vak/vak/62679`
[3] `http://www.rascal-mpl.org`
[4] `http://tutor.rascal-mpl.org` (may be out-dated)
[5] `http://stackoverflow.com/questions/tagged/rascal`
[6] `https://github.com/usethesource/rascal/issues`

1.3 Required Course Activities

The course consists of activities related to reading scientific papers, discussing those papers, and attending lectures and the practical lab[7].

- **Reading: Study.** Students study a selection of scientific papers each week, as well as the slides that accompany the lecture. Please check the course schedule for detailed information in Section 1.5 and weekly reading in Section 1.6.
- **Paper Sessions: Read and write.** During the paper sessions students discuss papers with the lecturer. Special attention is given to analyzing scientific papers and writing reviews. As part of this course students are asked to write an annotated bibliography.
- **Lecture Part 1 - Seminar: Introduce and discuss a topic.** During the first part of the lecture on day 1 the lecturer introduces a subject by presenting a deck of slides. The group questions the concepts, problems and solutions and discusses them.
- **Lecture Part 2 - Practical Lab: Work on assignments.** During the second part of the lecture on day 1 and on day 2 students work on practical assignments. Students work on practical assignments in teams of two students. Every practical session students are encouraged to ask for feedback with the lecturer, improving their work over several iterations before finally handing it in before the deadline, and the assignments are graded.

1.3.1 Grading

The course grade is the average of three grades, for practical lab *Series 1* and *Series 2* and an individual grade for an *Annotated Bibliography* of papers you have studied[8].

Grades are calculated as follows.

```
grade(series1, annotated_bibliography, series2) =
    (grade(series1) + grade(annotated_bibliography) + grade(series2))/3
```

1.4 Evaluation

Just like the programming students produce during the practical labs, this course is also the product of iterative improvement. We request student feedback in two ways:

- suggestions during the course, for quick on-the-fly improvements to current and upcoming lectures.
- assessment at the end of the course, for improving the course structure, content and goals for the next year.

[7] **Note:** Lecture attendance is mandatory.
[8] **Note:** Practical lab Series 0 is mandatory but not graded.

1.5 Course Schedule

Table 1 shows a week-by-week schedule of topics, lecture dates and lecturers. The columns *L* and *W* indicate the lecture week number and the calendar week number. The column *Subject* is a brief description of the subjects for that week. The column *Date* shows the date of the lecture. The column *Lecturer* shows the (guest) lecturer.

Table 2 shows reading, assignments and deadlines. The column *Reading* specifies which papers to study and columns *Practical Lab* and *Deadlines* show which practical lab to work on during that week, and which assignments are due[9].

L	W	Subject	Date	Lecturer
1	44	Introduction to Software Evolution	Oct 31st	Riemer van Rozen
2	45	Meta-programming and Rascal	Nov 7th	Paul Klint
3	46	Software Metrics, the SIG Perspective	Nov 14th	Magiel Bruntink
4	47	DSLs: SWAT.engineering perspective	Nov 21st	Davy Landman
5	48	Clone Detection and Management	Nov 28th	Riemer van Rozen
6	49	Model Evolution	Dec 5th	Riemer van Rozen
7	50	Legacy Software and Renovation	Dec 12th	Vadim Zaytsev

Table 1: Course Plan: Lecture Topics, Lecture Dates and Lecturers

L	W	Reading	Practical Lab	Deadlines
1	44	Mens [13]. Herraiz *et al.* [8].	Series 0 & 1	Series 0
2	45	Klint *et al.* [10].	Series 1	
3	46	Fenton [3]. Heitlager *et al.* [6]. Basili *et al.* [21].	Series 1	Series 1
4	47	Erdweg *et al.* [2]	Series 2	
5	48	Koschke [12]. Kapser and Godfrey [9].	Series 2	
6	49	Hermannsdörfer and Wachsmuth [7]. Alanen and Porres [1].	Series 2	
7	50	to be announced	Series 2	Series 2
8	51			Annotated bibliography

Table 2: Course Plan: Reading, Assignments and Deadlines

[9] **Note:** Deadlines are at the end of the week

1.6 Reading

This is the list of papers students read during the course, organised week-by-week the papers are covered in the lectures. Each paper on this list from weeks 1–5 is to be included in the annotated bibliography that students create during the course. The reading from weeks 6 and 7 is recommended reading only.

Lecture 1

- Tom Mens. *Software Evolution*, chapter 1. Introduction and Roadmap: History and Challenges of Software Evolution, pages 2–11. Springer, 2008
- Israel Herraiz, Daniel Rodriguez, Gregorio Robles, and Jesus M. Gonzalez-Barahona. The Evolution of the Laws of Software Evolution: A Discussion Based on a Systematic Literature Review. *ACM Comput. Surv.*, 46(2):28:1–28:28, December 2013

Lecture 2

- Paul Klint, Tijs van der Storm, and Jurgen Vinju. RASCAL: A Domain Specific Language for Source Code Analysis and Manipulation. In *Proceedings of the 2009 Ninth IEEE International Working Conference on Source Code Analysis and Manipulation*, SCAM '09, pages 168–177, Washington, DC, USA, 2009. IEEE Computer Society

Lecture 3

- N. Fenton. Software Measurement: A Necessary Scientific Basis. *IEEE Transactions on Software Engineering*, 20(3):199–206, Mar 1994
- I. Heitlager, T. Kuipers, and J. Visser. A Practical Model for Measuring Maintainability. In *Quality of Information and Communications Technology, 2007. QUATIC 2007. 6th International Conference on the*, pages 30–39, Sept 2007
- Rini van Solingen, Vic Basili, Gianluigi Caldiera, and H. Dieter Rombach. *Goal Question Metric (GQM) Approach*. John Wiley & Sons, Inc., 2002

Lecture 4

- Sebastian Erdweg, Tijs van der Storm, Markus Völter, Meinte Boersma, Remi Bosman, William R. Cook, Albert Gerritsen, Angelo Hulshout, Steven Kelly, Alex Loh, Gabriël D. P. Konat, Pedro J. Molina, Martin Palatnik, Risto Pohjonen, Eugen Schindler, Klemens Schindler, Riccardo Solmi, Vlad A. Vergu, Eelco Visser, Kevin van der Vlist, Guido H. Wachsmuth, and Jimi van der Woning. The State of the Art in Language Workbenches: Conclusions from the Language Workbench Challenge. In Martin Erwig, Richard F. Paige, and Eric Van Wyk, editors, *Software Language Engineering – Proceedings of the 6th International Conference, SLE 2013, Indianapolis, IN, USA, October 26–28, 2013*, volume 8225 of *LNCS*, pages 197–217. Springer, 2013

Lecture 5

- Rainer Koschke. *Software Evolution*, chapter 2. Identifying and Removing Software Clones, pages 15–36. Springer, 2008
- C. Kapser and M. W. Godfrey. "Cloning Considered Harmful" Considered Harmful. In *2006 13th Working Conference on Reverse Engineering*, pages 19–28, Oct 2006

Lecture 6 The following papers are recommended reading only, and are not required for the annotated bibliography assignment.

- Markus Hermansdörfer and Guido Wachsmuth. *Evolving Software Systems*, chapter 2. Coupled Evolution of Software Metamodels and Models, pages 33–63. Springer, 2014
- Marcus Alanen and Ivan Porres. Difference and Union of Models. In *International Conference on the Unified Modeling Language*, pages 2–17. Springer, 2003
- Riemer van Rozen and Tijs van der Storm. Toward Live Domain-Specific Languages: From Text Differencing to Adapting Models at Run Time. *Software & Systems Modeling*, pages 1–18, August 2017

Lecture 7 reading to be announced

Lecture 8 no additional reading

Publications related to Master Projects The following papers have resulted from student master projects related to this course. These papers serve as inspirational examples only, and are not required for the annotated bibliography assignment.

- Ammar Hamid and Vadim Zaytsev. Detecting Refactorable Clones by Slicing Program Dependence Graphs. In *SATToSE*, pages 37–48, 2014
- Riemer van Rozen and Quinten Heijn. Measuring Quality of Grammars for Procedural Level Generation. In *Proceedings of the 13th International Conference on Foundations of Digital Games, FDG 2018, as part of the 9th Workshop on Procedural Content Generation, PCG 2018, Malmö, Sweden, August 7-10, 2018*, pages 56:1–56:8. ACM, 2018

2 Assignments

Students are required to complete three obligatory practical assignment series for this course. During the first (Series 0) you work alone. This series is approved but not graded. During the second and third (Series 1 and 2) you work in the same group of two students. When you have completed the assignment you can request your lecturer to approve your work by explaining what you did. Ask your lecturer how to deliver the solutions of the assignments. Deadlines are at the end of the week. Table 3 shows how to work on assignments and Table 4 when to work on assignments and deadlines to deliver them.

Deliverable	**Type of work**
Practical Lab Series 0	Individual work
Practical Lab Series 1	Team work
Annotated Bibliography	Individual work
Practical Lab Series 2	Team work

Table 3: Assignments and how to work on them.

L	**W**	**Practical Lab**	**Writing**	**Deadline**
1	44	Series 0 and 1	Annotated Bibliography	Series 0
2	45	Series 1	Annotated Bibliography	
3	46	Series 1	Annotated Bibliography	Series 1
4	47	Grading Series 1 Series 2	Annotated Bibliography	
5	48	Series 2	Annotated Bibliography	
6	49	Series 2		
7	50	Series 2		Series 2
8	51	Grading Series 2		Annotated Bibliography

Table 4: When to work on assignments and deadlines to deliver them.

Next we describe the practical assignments, which include details on grading for each assignment series.

Annotated Bibliography

During the lectures and the paper sessions we use several papers in the field of software evolution. In this assignment you structure your own thoughts on these papers and exercise your skills at creating summaries and syntheses by writing a scientific paper.

Collaboration

You need to perform this assignment individually. You are allowed to discuss literature with other students, but you have to write the annotated bibliography alone.

Assignment

For each paper of the **first 5 weeks** on the reading list of Section 1.6.

- **Content.** Write a concise discussion (2-4 coherent paragraphs in your own words) of the major points of the paper.
- **Format.** Submissions should use the Article format, single column, standard page width, 11 point font, using the font family Times New Roman[10]. Please use the template shown in Figure 1. All submissions should be in PDF format.
- **Page Limit.** Submissions are limited to **4 pages** excluding bibliographic references. Submissions that exceed the page limit will not be graded.

A good bibliography shows that a student has applied critical thinking to the material that was read. Example questions to consider are: what do you learn from the papers, given the newly acquired knowledge what could you have done differently in past work/assignments, what are the practical implications, do you see any validity problems in the paper, etc.? The lectures provide a presentation of the material, but of course this is subject to the teacher's interpretations and preferences. In your bibliography you show your own critical opinion and the relevant argumentation. To further strengthen the bibliography, you can relate other literature you find yourself.

Hints

You are encouraged to compare some annotated bibliographies and commonly used templates for choosing how to structure your paper, e.g., Cornell's guidelines[11]. Your paper structure might contain the following elements as discussed in the reading assignments from "Preparation Master Project".

- **Introduction.** The annotated bibliography should be a self-contained article which requires an introduction. An introduction usually describes the topic, intended audience, and sketches a structure (why, what, how).
- **Annotations per paper.** The annotations per paper might be as described in Table 5, and consider omitting irrelevant sections and include only information that is noteworthy.

[10] `https://en.wikipedia.org/wiki/Times_New_Roman`
[11] `http://guides.library.cornell.edu/annotatedbibliography`

```
\documentclass[11pt]{article}
\begin{document}
\title{Title Text}
\author{Name (and student number)\\Affiliation\\Email}
\maketitle
\section{Introduction}
The text of the paper begins here.
%your sections go here
\bibliographystyle{plain}
\bibliography{papers.bib} %create a separate file containing the BibTex
\end{document}
```

Figure 1: Annotated Bibliography LaTeX Template

Section	**Description**	**approx. size**
Summary	Topic, problem statement, objectives and approach.	3 sentences
Significance	The subject addressed in this article is worthy of investigation.	1 sentence
Originality	The information presented was new.	1 sentence
Quality	Quality of technical content.	1 sentence
Relevance	Why and how is the paper relevant?	1 sentence
Readability	Is the paper well-structured and understandable?	1 sentence
Overall	Overall the conclusions were supported by the data.	1 sentence

Table 5: "Preparation Master Project" Suggested Annotations per Paper

Grading

The annotated bibliography will be graded using the following model:

Factor	**Base grade modification**
Missing name and or introduction paragraph	-0.5
Writing quality: proper spelling, grammar, and structure.	-1.0 to +1.0
Each paper on the reading list that is missing or not covered in sufficient depth in the bibliography.	-0.5 per missing paper
The bibliography clearly argues the students critical opinion on the contents of the papers.	+0.5 to +1.0
The bibliography considers literature outside of the reading list to support argumentation.	+0.5 to +1.0

Table 6: Grading Conditions and Scoring for the Annotated Bibliography

The base grade is 7. For this grade you need to produce an annotated bibliography that conforms to the assignment described above. The factors of Table 6 modify the base grade. The grade range is 1 to 10.

Deadline

The annotated bibliography should be delivered in course week 8.

Practical Lab Series 0 – Rascal Basics

In this lab you learn the basic facts about Rascal [10, 11] and quickly learn to program in it. The idea is to interact with Rascal using the RascalTutor environment and Eclipse. See `http://www.rascal-mpl.org` for download and installation instructions.

Once you have started the Rascal/Eclipse environment, you can activate the RascalTutor by selecting the menu Rascal > Start Tutor. The non-interactive tutor can also be found at `http://tutor.rascal-mpl.org`.

- Comments on the documentation can be given at the bottom of each page.
- Technical questions, related to code, should be asked on Stackoverflow.com: `http://stackoverflow.com/questions/tagged/rascal`.
- Bug reports can be submitted via `https://github.com/usethesource/rascal/issues`.

Collaboration

Please do these exercises individually. Any communication between students is allowed, but bear in mind that you should personally be capable to code in Rascal after this.

Assignment

Fully explore the Rascal course using the Tutor and teach yourself Rascal. You will be assisted in the laboratory to install the system and type your first expressions and statements. Please ask the teachers any question about Rascal or the exercises you might have. It will be hard work!

Hints

- Please do not use spaces in your Eclipse path.

Grading

This series is not graded. If you pass most of the exercises in the tutor, you will know enough Rascal to do Series 1 and 2. You are required to take the Rascal Online Test and pass it with a score of at least 50%. You may retake the test as many times as is necessary. Please post a screen shot of the results on your Git repository.

Deadline

You should finish Series 0 in the first week of the course in order to start Series 1.

Practical Lab Series 1 – Software Metrics

In Series 1 we focus on software metrics. Software metrics are used (for example) by the Software Improvement Group (`http://www.sig.eu`) to quickly gain an overview of the quality of software systems and to pinpoint problem areas that may cause low maintainability.

Some relevant questions are:

1. Which metrics are used?
2. How are these metrics computed?
3. How well do these metrics indicate what we really want to know about these systems and how can we judge that?
4. How can we improve any of the above?

In other words, you have to worry about motivation and interpretation of metrics, as well as correct implementation.

The SIG Maintainability Model provides an answer to question 1. You can read about it here:

- I. Heitlager, T. Kuipers, and J. Visser. A Practical Model for Measuring Maintainability. In *Quality of Information and Communications Technology, 2007. QUATIC 2007. 6th International Conference on the*, pages 30–39, Sept 2007.
- Additional information can be found at `https://www.sig.eu/resources/sig-models/`

The second question above ("How are these metrics computed?") is your assignment for Series 1. The third and fourth questions will be addressed during the grading session.

Collaboration

Please make groups of two students. You can work together as a pair on all aspects of this assignment. You can brainstorm with anybody else about the contents of your essay, but for this assignment you are not allowed to look at code from other groups or exchange solutions in detail with other groups.

Assignment

Using Rascal, design and build a tool that calculates the SIG Maintainability Model scores for a Java project.

Calculate at least the following metrics:

- Volume,
- Unit Size,
- Unit Complexity,
- Duplication.

For all metrics you calculate the actual metric values, for Unit Size and Unit Complexity you additionally calculate a risk profile, and finally each metric gets a score based on the SIG model ($--$, $-$, o, $+$, $++$).

Calculate scores for at least the following maintainability aspects based on the SIG model:

- Maintainability (overall),
- Analysability,
- Changeability,
- Testability.

You can earn bonus points by also implementing the Test Quality metric and a score for the Stability maintainability aspect.

Use this zip file to obtain compilable versions of two Java systems (smallsql and hsqldb): zip file[12]

- **smallsql** is a small system to use for experimentation and testing. Import as-is into Eclipse and ignore build errors.
- **hsqldb** is a larger system to demonstrate scalability. Import into Eclipse, make sure to have only `hsqldb/src` on the build path, and add the following external jars from your `eclipse/plugins/` directory: `javax.servlet_$VERSION.jar` and `org.apache.ant_$VERSION/lib/ant.jar`

Hints

- Create an Eclipse Java project with example files to test your solution on (using the Rascal test functionality).
- Create an Eclipse Java project for each of the two systems, smallsql and hsqldb too. Some few lines of code will still not compile, but commenting them out would not change the metrics too much. So commenting out just a few lines is ok in this case. It saves time!

Grading

The assignment is judged by demonstrating your results and your code to us in a small interactive session. At the end of this session you will immediately get a grade between 3 and 10.

You also have to submit a zip file with all the code and relevant documents in the Blackboard assignments. The files are checked for plagiarism automatically. If you worked in a team of two, submit your assignment twice: one for each student!

You will be graded using the following model. The base grade is 7. For this grade you need an implementation that conforms to the assignment described above. The

[12] `http://homepages.cwi.nl/~jurgenv/teaching/evolution1314/assignment1.zip`

Condition	Base grade modification
The metric value (total LOC) and/or score for Volume deviate without good motivation	-0.5 to -1.0
The metric value (%) and/or score for Duplication deviate without good motivation	-0.5 to -1.0
The risk profile and/or score for Unit Size deviate without good motivation	-0.5 to -1.0
The risk profile and/or score for Unit Complexity deviate without good motivation	-0.5 to -1.0
The scores calculated for the maintainability aspects deviate without good motivation	-0.5
Your tool produces output that allows easy verification of the correctness of the result (metric values, risk profiles, scores, etc. are neatly listed next to each other)	+0.5
You also implemented Test Quality and Stability and can argue their correctness	+0.5
Your tool produces correct output for hsqldb within the time span of the grading session (approximately 30 minutes); if clone detection is turned off you may get at most an extra half point	+0.5 to +1.0
You can demonstrate that your own code is of high maintainability and has proper automated tests	+0.5
You have found another metric in the literature that is not in the SIG Maintainability Model, and you can argument why and how it would improve the results	+0.5 to +1.0

Table 7: Grading Conditions and Scoring for Series 1

implementation consists of sensible design and code. You can explain and motivate how it actually reads in the Java code and calculates the metrics based on that. Your implementation can be run during the grading session on at least the smallsql project. For grading, import the smallsql project into Eclipse as-is and ignore the 100 or so build errors.

Table 7 shows conditions and how they modify the grade (the teachers have a reference implementation that provides the correct outputs).

Deadline

The deadline for Series 1 is before the grading sessions in week 4 **November 20th at the latest.**

Practical Lab Series 2 – Clone Detection

Code cloning is a phenomenon that is of both scientific and practical interest. In the lecture and the related papers, clone detection and management techniques were discussed, as well as the various arguments surrounding the problems that code cloning causes.

In this lab we will take a more hands-on approach by building our own clone detection and management tools. Such tools should be of help to software engineers like yourselves, so be sure that your solution will at least satisfy your own needs! Furthermore, tool building for others is a challenging activity by itself. We expect you to find some literature on this topic yourself and use it in your design. A suggested paper on visual tools for software exploration is:

- M.-A.D Storey, F.D Fracchia, and H.A Müller. Cognitive Design Elements to Support the Construction of a Mental Model during Software Exploration. *Journal of Systems and Software*, 44(3):171 – 185, 1999

Compared to Lab Series 1, this assignment will be more open. Your solution will be graded using more generic criteria, with a stronger emphasis on motivation and argumentation. You will need to use literature discussed and referenced in the lectures to find and motivate good solutions: for instance for finding an appropriate clone detection algorithm.

Collaboration

Please work in groups of two students. Complete the assignment in the same groups you did for Series 1. You can brainstorm, but for this assignment you are not allowed to look at code from other groups or exchange solutions in detail with other groups.

Assignment

The assignment consists of two main deliverables:

1. A Working prototype implementation of a clone management tool, consisting of the following elements:
   (a) A clone detector that detects at least Type 1 clones in a Java project:
      - The detected clone classes are written to file in a textual representation.
      - Clone classes that are strictly included in others are dropped from the results (subsumption).

   (b) A report of cloning statistics showing at least the % of duplicated lines, number of clones, number of clone classes, biggest clone (in lines), biggest clone class, and example clones.

   (c) At least one insightful vizualization of the cloning in a project. The lecture discusses several example vizualizations you could use.

   (d) A convincing test harness (an automated regression test framework) that ensures your clone detector works.

2. B. Design documentation that (1) describes and (2) motivates the following elements:
   (a) The 3 main requirements your tool satisfies from the perspective of a maintainer (see for instance [18]), and the related implementation choices.
   (b) The exact type of clones your tool detects. Start from Type 1, Type 2, ... but become more specific.
   (c) The core of the clone detection algorithm that you use (in pseudocode).
   (d) The vizualization(s) you implemented: how do they help a maintainer or developer?

To score a higher grade than the base grade (7), additionally do:

- Also detect Type 2 and Type 3 clone classes.
- Implement more vizualizations that provide additional insight.
- Make sure your tool works on bigger projects such as hsqlsb[13].
- Produce maintainable code that is covered by unit tests.

Implementation

Use any method you like, but go beyond what you have done in Series 1. The detection part must be in pure Rascal, the visualisation is allowed to tap into other languages. If you opt for a PDG/SDG approach, consider using a Rascal library (it is in a separate repo, clone or fork it).

Related Work on Software Clones

The following resources can help you get acquainted with clone management:

- Chanchal K. Roy, James R. Cordy, and Rainer Koschke. Comparison and Evaluation of Code Clone Detection Techniques and Tools: A Qualitative Approach. *Sci. Comput. Program.*, 74(7):470–495, May 2009
- Dhavleesh Rattan, Rajesh Bhatia, and Maninder Singh. Software Clone Detection: A Systematic Review. *Information and Software Technology*, 55(7):1165–1199, 2013.
- C. K. Roy, M. F. Zibran, and R. Koschke. The Vision of Software Clone Management: Past, Present, and Future (keynote paper). In *Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE), 2014 Software Evolution Week - IEEE Conference on*, pages 18–33, Feb 2014
- C. Kapser and M. W. Godfrey. "Cloning Considered Harmful" Considered Harmful. In *2006 13th Working Conference on Reverse Engineering*, pages 19–28, Oct 2006

[13] `http://sourceforge.net/projects/hsqldb/files/hsqldb/hsqldb_1_8_0/hsqldb_1_8_0_10.zip/download`

- Ammar Hamid and Vadim Zaytsev. Detecting Refactorable Clones by Slicing Program Dependence Graphs. In *SATToSE*, pages 37–48, 2014

The first three are overviews [17, 15, 16], the next one is a highly cited controversial piece [9], the last one is an example paper that can result from a Master thesis - it is easy to read and contains a simplified brief overview of the field [4].

Related Work on Visualization

Use any of the following resources.

Lecture slides:

- You may choose to use RASCAL's figure library. Documentation on that library can be found in Paul Klint's lecture "Towards Visual Analytics"[14]

Some inspiring papers:

- Hiroaki Murakami, Yoshiki Higo, and Shinji Kusumoto. ClonePacker: A Tool for Clone Set Visualization. In Yann-Gael Gueheneuc, Bram Adams, and Alexander Serebrenik, editors, *Proceedings of the 22nd International Conference on Software Analysis, Evolution and Reengineering*, pages 474–478. IEEE, 2015
- Lucian Voinea and Alexandru C. Telea. Visual Clone Analysis with SolidSDD. In *Proceedings of the Second IEEE Working Conference on Software Visualization*, pages 79–82. IEEE Computer Society, 2014
- Avdo Hanjalic. In *Proceedings of the First IEEE Working Conference on Software Visualization*, pages 1–4. IEEE Computer Society

Visualization Libraries:

- Salix is a new library for interactive tools and visualizations in RASCAL using a browser [15]. Several demos are available, including live programming of state machines, similar to the running example of [20]. – powerful yet experimental
- The built-in visualization modules of RASCAL are `vis::*`. Sources[16]. Tutor[17].
- External visualisation libraries are D3[18], vis[19], Vega[20] and Gephi[21].

Grading

The assignment is judged by demonstrating your results and your code to us in a small interactive session. At the end of this session you will immediately get a grade between 1 and 10.

[14] `https://blackboard.ic.uva.nl/bbcswebdav/pid-4951566-dt-content-rid-6861541_1/courses/2318N001.5364SOEV6Y.S1.1.2014/intro-visualization.pdf`
[15] `https://github.com/cwi-swat/salix`
[16] `https://github.com/usethesource/rascal/tree/master/src/org/rascalmpl/library/vis`
[17] `http://tutor.rascal-mpl.org/Rascal/Libraries/Vis/Vis.html`
[18] `https://d3js.org`
[19] `http://visjs.org`
[20] `https://vega.github.io/vega/`
[21] `https://gephi.org`

You also have to drop a zip file with all code and relevant document in the assignments. The files are checked for plagiarism automatically. If you worked in a team of two, drop your assignment twice: one for each student!

You will be graded using the following model:

The base grade is 7. To qualify for grading you first need a solution that complies to the assignment described above. During the grading session (15 minutes), your tool should complete on at least the smallsql project used in Lab Series 1.

The following conditions modify the grade:

The base grade is 7. Table 8 shows grading conditions and how they modify the grade.

Condition	**Base grade modification**
Type I clone classes are incorrectly detected.	-0.5 to -1.0
Test harness (automated test framework) is unconvincing, is limited in scope or not thorough.	-0.5 to -1.0
Cloning visualization does not give insight or does not work properly.	-1.0
The clone detection report is incomplete or incomprehensible.	-0.5 to -1.0
Type II clone classes are correctly detected and visualized	+0.5 to +1.0
Type III clone classes are correctly detected and visualized	+0.5 to +1.0
Your tool and design document include additional convincing visualizations	+0.5 to +1.0
Not able to explain the core algorithm or answer questions about it	-1.0
Documentation describes unfounded, unsupported, illogical or anecdotal motivation for visualisations	-1.0
Non-interactive static visualisation	-1.0
Does not work on hsqldb	-0.5

Table 8: Grading Conditions and Scoring for Series 2

Deadline

The deadline for Series 2 is course week 7.

3 Change Log

Table 9 shows the changes made to this document over versions.

version	date	modification	author
0.01	Oct 31st 2016	Created this document based on the work of Paul Klint, Jurgen Vinju, Magiel Bruntink and Vadim Zaytsev.	Riemer van Rozen
0.06	Nov 9th 2016	Added RASCAL online test to Series 0 and added this change log.	Riemer van Rozen
0.07	Nov 17th 2016	Added hints section to the annotated bibliography assignment. Updated the description of Series 1.	Riemer van Rozen Ana Oprescu
0.08	Nov 30th 2016	Fixed deadlines in Table 4. Added link to Cornell guidelines on *"how to prepare an annotated bibliography"*	Riemer van Rozen Ana Oprescu
0.09	Dec 22nd 2016	Minor clarifications and fixed typos.	Riemer van Rozen
0.10	Oct 29th 2017	Modified the schedule for 2017/2018.	Riemer van Rozen
0.11	Nov 5th 2017	Simplified the LaTeX template for the annotated bibliography.	Riemer van Rozen
0.12	Nov 14th 2017	Fixed a critical error that was introduced in version 0.11 of the annotated bibliography assignment description. The correct page limit is 4 pages.	Riemer van Rozen
0.13	Nov 21st 2017	Added this week's reading.	Riemer van Rozen
0.14	Sept 17th 2018	Updated course schedule and lecturers. Added link to SIG model in Series 1. Added Salix framework to Series 2.	Riemer van Rozen
0.15	Okt 28th 2018	Updated reading of lecture 4 and series 1 deadline.	Riemer van Rozen

Table 9: Change Log

References

[1] Marcus Alanen and Ivan Porres. Difference and Union of Models. In *International Conference on the Unified Modeling Language*, pages 2–17. Springer, 2003.

[2] Sebastian Erdweg, Tijs van der Storm, Markus Völter, Meinte Boersma, Remi Bosman, William R. Cook, Albert Gerritsen, Angelo Hulshout, Steven Kelly, Alex Loh, Gabriël D. P. Konat, Pedro J. Molina, Martin Palatnik, Risto Pohjonen, Eugen Schindler, Klemens Schindler, Riccardo Solmi, Vlad A. Vergu, Eelco Visser, Kevin van der Vlist, Guido H. Wachsmuth, and Jimi van der Woning. The State of the Art in Language Workbenches: Conclusions from the Language Workbench Challenge. In Martin Erwig, Richard F. Paige, and Eric Van Wyk, editors, *Software Language Engineering – Proceedings of the 6th International Conference, SLE 2013, Indianapolis, IN, USA, October 26–28, 2013*, volume 8225 of *LNCS*, pages 197–217. Springer, 2013.

[3] N. Fenton. Software Measurement: A Necessary Scientific Basis. *IEEE Transactions on Software Engineering*, 20(3):199–206, Mar 1994.

[4] Ammar Hamid and Vadim Zaytsev. Detecting Refactorable Clones by Slicing Program Dependence Graphs. In *SATToSE*, pages 37–48, 2014.

[5] Avdo Hanjalic. In *Proceedings of the First IEEE Working Conference on Software Visualization*, pages 1–4. IEEE Computer Society.

[6] I. Heitlager, T. Kuipers, and J. Visser. A Practical Model for Measuring Maintainability. In *Quality of Information and Communications Technology, 2007. QUATIC 2007. 6th International Conference on the*, pages 30–39, Sept 2007.

[7] Markus Hermansdörfer and Guido Wachsmuth. *Evolving Software Systems*, chapter 2. Coupled Evolution of Software Metamodels and Models, pages 33–63. Springer, 2014.

[8] Israel Herraiz, Daniel Rodriguez, Gregorio Robles, and Jesus M. Gonzalez-Barahona. The Evolution of the Laws of Software Evolution: A Discussion Based on a Systematic Literature Review. *ACM Comput. Surv.*, 46(2):28:1–28:28, December 2013.

[9] C. Kapser and M. W. Godfrey. "Cloning Considered Harmful" Considered Harmful. In *2006 13th Working Conference on Reverse Engineering*, pages 19–28, Oct 2006.

[10] Paul Klint, Tijs van der Storm, and Jurgen Vinju. RASCAL: A Domain Specific Language for Source Code Analysis and Manipulation. In *Proceedings of the 2009 Ninth IEEE International Working Conference on Source Code Analysis and Manipulation*, SCAM '09, pages 168–177, Washington, DC, USA, 2009. IEEE Computer Society.

[11] Paul Klint, Tijs van der Storm, and Jurgen Vinju. *EASY Meta-programming with Rascal*, pages 222–289. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[12] Rainer Koschke. *Software Evolution*, chapter 2. Identifying and Removing Software Clones, pages 15–36. Springer, 2008.

[13] Tom Mens. *Software Evolution*, chapter 1. Introduction and Roadmap: History and Challenges of Software Evolution, pages 2–11. Springer, 2008.

[14] Hiroaki Murakami, Yoshiki Higo, and Shinji Kusumoto. ClonePacker: A Tool for Clone Set Visualization. In Yann-Gael Gueheneuc, Bram Adams, and Alexander Serebrenik, editors, *Proceedings of the 22nd International Conference on Software Analysis, Evolution and Reengineering*, pages 474–478. IEEE, 2015.

[15] Dhavleesh Rattan, Rajesh Bhatia, and Maninder Singh. Software Clone Detection: A Systematic Review. *Information and Software Technology*, 55(7):1165–1199, 2013.

[16] C. K. Roy, M. F. Zibran, and R. Koschke. The Vision of Software Clone Management: Past, Present, and Future (keynote paper). In *Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE), 2014 Software Evolution Week - IEEE Conference on*, pages 18–33, Feb 2014.

[17] Chanchal K. Roy, James R. Cordy, and Rainer Koschke. Comparison and Evaluation of Code Clone Detection Techniques and Tools: A Qualitative Approach. *Sci. Comput. Program.*, 74(7):470–495, May 2009.

[18] M.-A.D Storey, F.D Fracchia, and H.A Müller. Cognitive Design Elements to Support the Construction of a Mental Model during Software Exploration. *Journal of Systems and Software*, 44(3):171 – 185, 1999.

[19] Riemer van Rozen and Quinten Heijn. Measuring Quality of Grammars for Procedural Level Generation. In *Proceedings of the 13th International Conference on Foundations of Digital Games, FDG 2018, as part of the 9th Workshop on Procedural Content Generation, PCG 2018, Malmö, Sweden, August 7-10, 2018*, pages 56:1–56:8. ACM, 2018.

[20] Riemer van Rozen and Tijs van der Storm. Toward Live Domain-Specific Languages: From Text Differencing to Adapting Models at Run Time. *Software & Systems Modeling*, pages 1–18, August 2017.

[21] Rini van Solingen, Vic Basili, Gianluigi Caldiera, and H. Dieter Rombach. *Goal Question Metric (GQM) Approach.* John Wiley & Sons, Inc., 2002.

[22] Lucian Voinea and Alexandru C. Telea. Visual Clone Analysis with SolidSDD. In *Proceedings of the Second IEEE Working Conference on Software Visualization*, pages 79–82. IEEE Computer Society, 2014.