EvaP

Jennifer Stamm, Stefan Neubert

Winter Term 2015/16

Contents

1 Introduction

1.1 EvaP – An Evaluation System

The online platform EvaP is used for evaluation of courses at the Hasso Plattner Institue (HPI). Its development started in 2011 when the student representatives decided to redevelop the former system EvaJ. Now, it is an Open Source project hosted on GitHub, even though the main developers are still part of the student representative team.

EvaP's time to shine is at the end of each semester. Each university course is evaluated with specific questionaires chosen by the lecturer. Students are allowed to give anonymous feedback to different aspects of the lecture, the lecturer itself and additional tutors through a grading system and comments. EvaP encourages evaluation by rewarding participation with points. These reward points can be redeemed for currency to use at HPI events. EvaP's latest feature is the distribution of course grades through the platform.

1.2 Motivation – Why should EvaP be Analyzed, Verified and Tested?

EvaP is an important source for feedback at HPI. The platform offers students a way to express their critique anonymously. It documents the feedback for lecturers. Thus, they do not need to collect and save the feedback themselves. Additionally, they can take their time to evaluate the student's feedback. Furthermore, the evalutation of all courses is saved centrally. This allows to gain an overview over the quality of HPI courses and compare feedback over time as well as with other courses. Therefore, EvaP is an important tool at the HPI.

Since EvaP is developed by students, the responsible persons and main developers change regularly as older students graduate and new ones enroll. The change of responsible persons shifts the view of which features, programming paradigms and quality assurance are most important. Consequently, the requirements change on a regular basis as well. A change of main developers comes with an inevitable loss of knowledge about the existing code. Besides, the Open Source aspect allows developers without inside knowledge to contribute code as well. Even though all code is reviewed and checked by the main developers, they may not grasp it as well as self-written code. Events like Hackdays or Hacking Hours are used to promote EvaP's development. While these practices ensure the advancement of EvaP, they may endanger its quality.

Because of EvaP's importance at the HPI, its quality should be ensured. Therefore, we will analyze, verify and test the software within the scope of this lecture.

2 Milestone 1

As suggested, the duration of milestone 1 is from the beginning of the project until the end of December. Milestone 1 includes the set up of the software project which led to

the discovery of the first bug. It includes the first steps taken to gather information, get comfortable with the project and planning of the project. Lastly, it includes the phase of testing the project regarding graph coverage.

2.1 Set Up

Instead of running natively on the developer's machine, EvaP is wrapped in a Vagrant development environment[1]. This allows developers to develop features with their preferred tools on their favorite operating system out of Mac OS X, Windows, Debian and Centos, without having to go through the complicated process of collecting dependencies for their specific platform. The complete project set up normally consists of installing git, a virtual machine provider for Vagrant and Vagrant itself, cloning the repository and running the shell command `vagrant up`.

As all core developers of EvaP use unix-based platforms, a bug in an external sub module used by EvaP remained unnoticed and undealt with: When we tried to set up the project on windows machines that use a line feed and a carriage return (`LF CR`) to end lines in text files — as opposed to unix systems that only use a single line feed (`LF`) — the set up failed. The bug was fixed quickly: firstly only for EvaP, subsequently in the external module the line ending policy had to enforce `LF` only.

Now, the set up therefore works on all major operating systems, allowing an easy start for all developers who want to contribute to EvaP, and more insights into the system for us.

2.2 First Steps

Incidentally, this is the first semester for the student representatives to host *EvaP Hacking Hours* biweekly. This is an event to give students a space to develop and work on EvaP. We will use it as a possibility to stay in contact with the main developers. As testers of EvaP it is a valuable resource to be able to talk directly with developers. This way we can gather current information easily. We are able to verify the content of old artifacts with them as well as our future findings.

2.2.1 Application Survey

The current main developer of new features is Johannes Wolf. He is supported by Johannes Linke who is mainly responsible for a good coding style, including code review and refactoring. We interview them about information regarding the first steps of our project.

Development Paradigm and Development Languages There are no development paradigms explicitly determined. But as stated there is a main developer responsible for code review. As it is, all newly developed code is reviewed before it is accepted.

[1] `https://www.vagrantup.com/`

Everyone is able to review code and discuss it with the author and other reviewers. This practice shall ensure readable code and distribute knowledge about code changes. Additionally, it is implicitely assumed that paradigms of the used development languages are followed. The development languages are:

- Python 3 through the Django framework
- HTML
- Javascript
- CSS

As an example for paradigms given by the languages we checked the document known as *PEP 0008*[2]. This is the style guide for Python code written by Guido van Rossum, the author of Python, and followed by most open source projects in Python.

Requirements / Specification / Documentation / Artefacts Requirements were elaborated at the start of EvaP's predecessor EvaJ several years ago. No original artifacts were stored even though most requirements still hold. Examples are:

- Evaluation should be anonymous
- Written feedback is only readable by a small, strongly specified selection of people
- Written feedback is reviewed before it is available to the lecturer

Symptomatic for an open source project managed by students there is no formal specification of EvaP. Though there are different information gathered in the project wiki hosted on Github[3]. The responsibility to specify new features is on the main developer, Johannes Wolf. He collaborates directly with the users of the feature.

Current testing status / Bug repositories The project is hosted open source on GitHub. Different tools allow to include badges on the overview page to display the status of the latest build (Figure 1). The following tools are already used:

- *Travis CI*: A continuous integration service to build and test projects hosted on GitHub
- *Gemnasium*: An automated service for monitoring project dependencies for possible updates
- *Landscape*: A service checking the code for errors, code smells and deviations from stylistic conventions
- *Coveralls*: A service relying on Travis CI that tracks the code coverage

[2] `https://www.python.org/dev/peps/pep-0008/`
[3] `https://github.com/fsr-itse/EvaP/wiki`

- Another feature of Github is used to track bugs: the label *[T] Bug* for *issues*.

Since this is already an extensive list, we will not to spend much time on researching further tools.

Figure 1: Badges of testing tools on the GitHub overview page (01.12.2015)

Personal involvement Our first contact with EvaP was as users. As HPI bachelor students we evaluated courses of the first semester. As a tutor for a course Stefan used EvaP to read feedback. As a member of the student representatives Jennifer reviewed comments before publishing the evaluation. During the *Evap Hackday 2014* and *Evap Hackday 2015* Jennifer joined the team developers. She familiarized herself with the development practice to fork, code and create pull requests for review and solved several small issues.

2.2.2 Initial Test Plan

We developed an initial plan based on our findings in the application survey. As suggested we followed the questions discussed during the lecture. The test plan includes evaluating and enhancing the existing artifacts, cross-checking the results found by the tools used and eventually improving the testing status.

Five V&V Questions We started with evaluating EvaP regarding the five basic verification and validation questions:

1. When do verification and validation start? When are they complete?

   Verification and validation started along the start of the project and will be an important part during the duration of the project.

2. What particular techniques should be applied during development?

   As EvaP is a software too big to be wholly tested by us within the scope of the seminar project, we will try to apply as many techniques as possible on a small subpart of the project. We will evaluate the techniques and make recommendations for further testing to the main developers.

3. How can we assess the readiness of a product?

   Since EvaP is already in use it is certainly ready. Our testing will not influence the readiness.

4. How can we control the quality of successive releases?

   Additionally to continous integrationa and automated tests, a code review before merge is already implemented in the development process to control the quality of successive releases. We hope to enable developers to gain more knowledge about the system by enhancing the available artifacts and documentation.

5. How can the development process itself be improved?

   The main weakness of the development process is the sparse distribution of knowledge and the change of developers over time. Since only students are interested in developing EvaP we do not see a possibility to improve this part of the process.

Test Classifications and Approaches Considering the following test approaches we came to the conlusion described below:

1. Validation vs. Defect Testing

   Since there is no requirements document or even a list of features, it is hard to implement validation testing. As we will not draw up a corresponding document, we will not apply validation testing. However, if possible, we will try to detect defects in the software by finding inputs leading to incorrect behavior. Thus, we will apply defect testing.

2. Development, Release, User Testing

   Because of the applied continuous integration process, development testing is already in use. Release and acceptance testing is not possible; there are no specified releases. User testing is applied in a way as most users are familiar with the developing process and report encountered bugs directly.

3. Unit/Component, Integration, System Testing

   EvaP is a web application, so there are no hardware components directly involved. While there are software dependencies, this aspect is already covered by the tool Gemnasium. For a web application it would be important to work in the most popular web browsers (desktop and mobile). Since there are no complaints known about EvaP not working in a certain web browser and no plans exist about developments that would use web browser specific features, we will not focus on integration testing. Similar to validation testing, without a specified requirements we can not execute system testing. Instead we will focus on unit and component testing.

In conclusion, we will use tests to find unnoticed defects and we will apply different testing techniques on units or components of EvaP. These approaches align with our goals to help EvaP's developers and to learn about different testing techniques.

Available Artifacts The most thorough artifact, the GitHub wiki, offers two artifacts with potential regarding coverage-based testing:

- A finite state machine describing the states of courses in the evaluation process (Figure 2)[4]
- Description of a few use cases with UML use case diagrams[5]

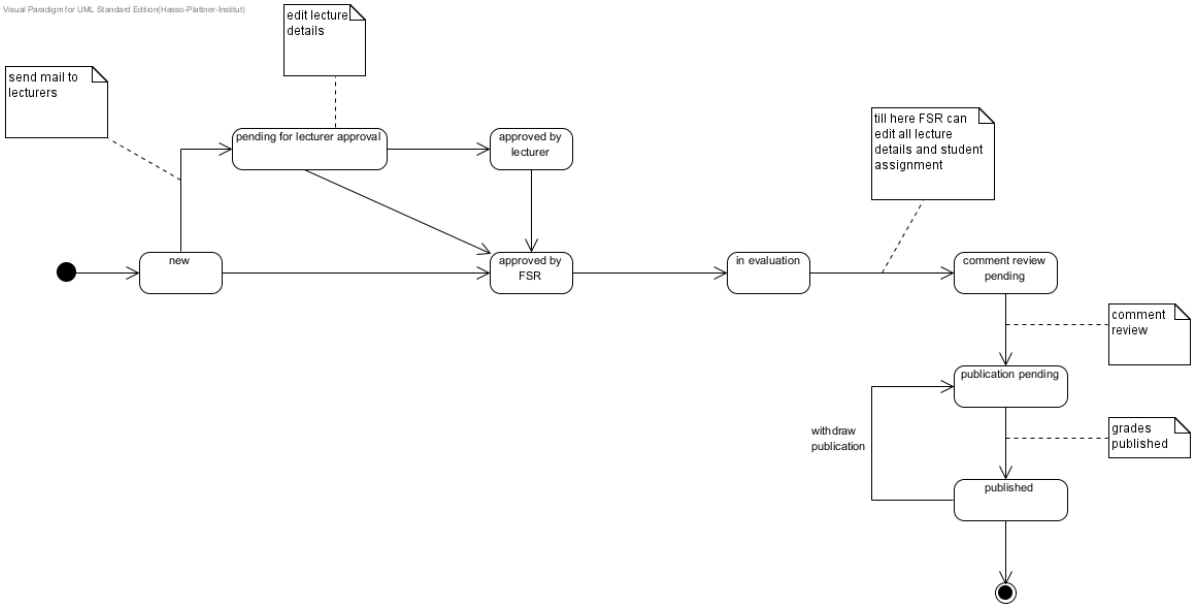

Figure 2: Original FSM: Possible states of a course

Initial Test Plan Based on our findings, the initial test plan is as follows:

- Create a control flow graph of a function, apply coverage criteria to define test sets and implement tests
- Check and if necessary update the FSM of evaluation states
- Use or create a UML use case diagram including its elaboration to develop an activity diagram
- Investigate coverage found by the tool *Coveralls*

[4] `https://github.com/fsr-itse/EvaP/wiki/Evaluation-States`
[5] `https://github.com/fsr-itse/EvaP/wiki/Use-Cases`

2.2.3 Test Automation

The project is already covered by several tools. Our research showed the used tools are popular in the Python community if the project is hosted on GitHub. Because of this and since the developers are comfortable with their choices, we will observe the functionality of the tools during the project. We will focus this section on describing our experience with the set up of running tests locally.

Content of this part: Stefan has fun with PyCharm

2.3 Graph Coverage

According to our test plan we chose parts of EvaP to test based on graph coverage. We will test the function `send_publish_notifications` by creating a control flow graph and test cases based on path coverage criteria. We will investigate the documented finite state machine (FSM). We will create an activity diagram based on the documented use-case Finally, we will investigate if our added test cases changed the line coverage, since this is the given measurement by the used tool COVERALLS.

2.3.1 Selected Control Flow Graph

We chose the function, because

During our research we found two promising tools for control flow graph creation. The first tool is hosted open source on GitHub (`https://github.com/danielrandall/python-control-flow-graph`). Unfortunately, it is not documented and its execution lead to errors. After a few tries, we estimated that fixing the tool would take more time than creating the control flow graph by hand. The second tool found was even more promising as it was a report of a master graduate from the university of Texas titled: "Control flow graph visualisation and its application to coverage and fault localization in Python" by Jackson Lee Salling. His tool even visualized edge-pair and prime path coverage in control flow graphs. But unfortunately he did not respond to our request if we could use the tool for our project or if he had tips for published tools.

Therefore, we created the control flow graph by hand. We experienced some difficulties with Python's code style:

- `def send_publish_notifications( grade_document_courses=None, evaluation_results_co`
- `grade_document_courses = grade_document_courses or []`

  This assignment of either the `grade_document_courses` or an empty list is dependent on the evaluation of `grade_document_courses`. Here are several statements hidden in one line.

-

```
def send_publish_notifications(grade_document_courses=None,
    evaluation_results_courses=None):
    grade_document_courses = grade_document_courses or []
    evaluation_results_courses = evaluation_results_courses or []

    publish_notifications = defaultdict(lambda: CourseLists(set(), set()))

    for course in evaluation_results_courses:
        # for published courses all contributors and participants get a
            notification
        if course.can_publish_grades:
            for participant in course.participants.all():
                publish_notifications[participant].evaluation_results_courses.add
                    (course)
            for contribution in course.contributions.all():
                if contribution.contributor:
                    publish_notifications[contribution.contributor].
                        evaluation_results_courses.add(course)
        # if a course was not published notifications are only sent for
            contributors who can see comments
        elif len(course.textanswer_set) > 0:
            for textanswer in course.textanswer_set:
                if textanswer.contribution.contributor:
                    publish_notifications[textanswer.contribution.contributor].
                        evaluation_results_courses.add(course)
            publish_notifications[course.responsible_contributor].
                evaluation_results_courses.add(course)
    for course in grade_document_courses:
        # all participants who can download grades get a notification
        for participant in course.participants.all():
            if participant.can_download_grades:
                publish_notifications[participant].grade_document_courses.add(
                    course)

    for user, course_lists in publish_notifications.items():
        EmailTemplate.send_publish_notifications_to_user(
            user,
            grade_document_courses=list(course_lists.grade_document_courses),
            evaluation_results_courses=list(course_lists.
                evaluation_results_courses)
        )
```

Listing 1: The control flow graph is based on this function.

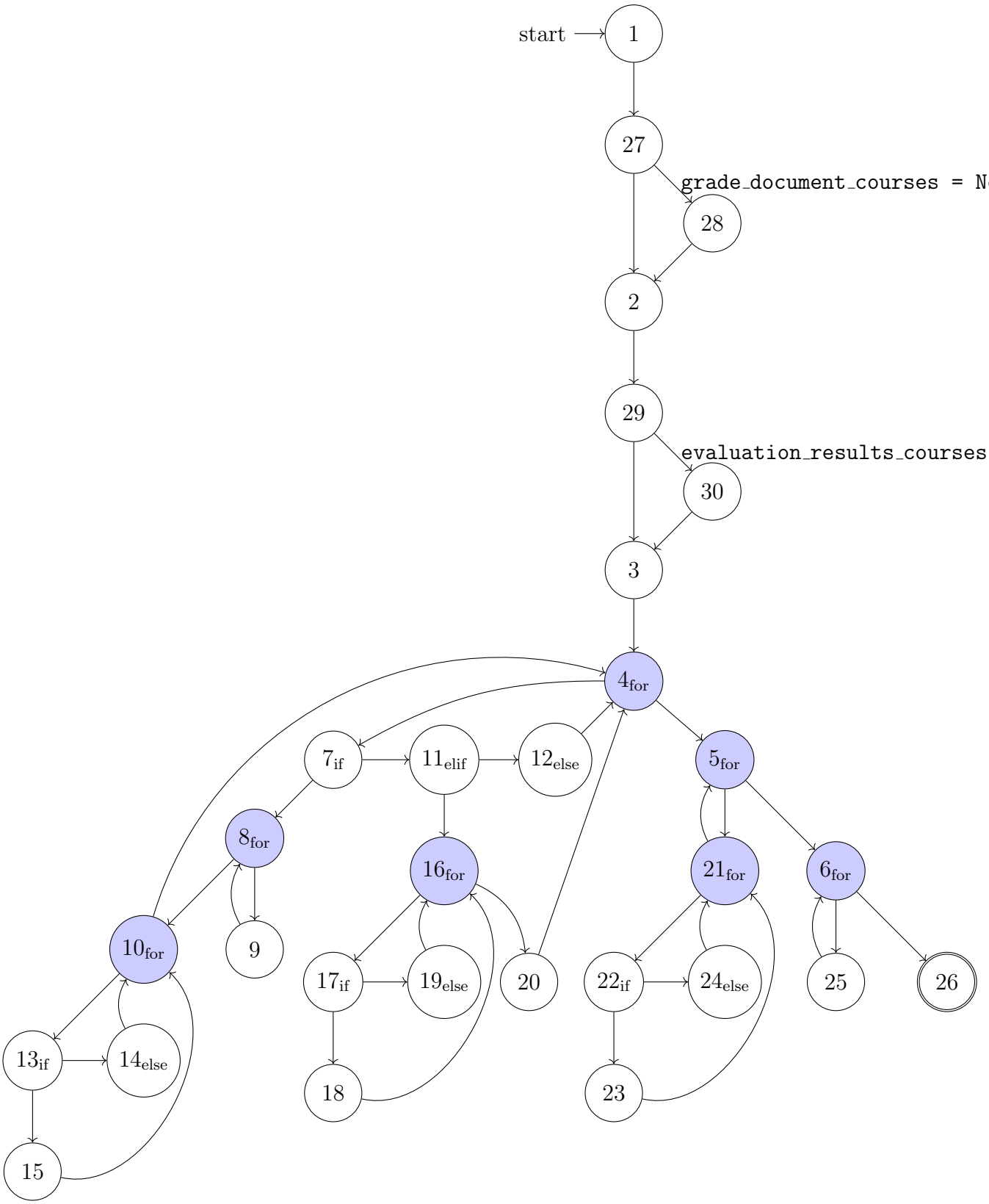

Figure 3: Control Flow Graph of the Function `send_publish_notifications` in `evap/evaluation/tools.py`

Landscape