Retaining maintainability throughout the scaling of software projects using automated tools

Deutscher Titel

Nikkel Mollenhauer

Universitätsbachelorarbeit
zur Erlangung des akademischen Grades

Bachelor of Science
(B. Sc.)

im Studiengang
IT-Systems Engineering

eingereicht am 30. Juni 2022 am
Fachgebiet Enterprise Platform and Integration Concepts der
Digital-Engineering-Fakultät
der Universität Potsdam

Gutachter	Dr. Michael Perscheid
Betreuer	Rainer Schlosser
	Johannes Huegle
	Alexander Kastius

Abstract

When scaling software projects, knowing when and how to start scaling maintainability is the key to success. Maintainability manifests in many different ways, be it well documented API's, a comprehensive wiki or sensible tests. Maintainability can be aided by automating many processes, as this takes strain away from developers and lowers the barrier of entry for new team members looking to contribute.

In our project, we employed many of these techniques, and this thesis aims to illustrate the goals and ideas behind the processes involved. Primary focus is the journey from a single-file project to the pip-package 'recommerce' with its automated testing pipeline, code-style checks and comprehensive documentation using automated tools such as 'Sphinx' and 'interrogate'.

Analysis of the process shows that while tests are useful in aiding developers to understand code it must always be kept in mind that while tests make certain aspects of the software development process more accessible, they are also to be maintained.

One of the biggest and most costly undertakings was the introduction of pip-packaging the project, which leads to the conclusion that while building and maintaining a clear vision of the desired state of the product as early as possible can lead to an overhead early in the process, it will also make the project grow more sustainably, leading to higher maintainability in the long run.

Problem

Background

Objective

Methodology

Results - focuses on tests while leaving out documentation and pip (which comes in conclusion?!?)

conclusion

Zusammenfassung

Acknowledgments

Contents

Abstract iii

Zusammenfassung v

Acknowledgments vii

Contents ix

I Introduction 1

1 Knowing your Users(?) 3
1.1 Feather by SAP . . . 3
1.2 Alex == Research . . . 3

2 Related Work(?) 5
2.1 Packaging discussion in the Python community . . . 5

II Challenges 7

3 Documentation 9
3.1 Wiki . . . 9
3.2 Docstrings . . . 9
3.3 Tests . . . 9

4 Tests 11
4.1 Pytest . . . 11
4.1.1 Setup & Teardown == Automation . . . 11
4.1.2 Parametrization . . . 11
4.2 What to test . . . 11
4.3 Types of tests (e.g. Unit vs. Acceptance) . . . 11
4.4 Writing maintainable tests . . . 11
4.5 Measuring Coverage . . . 11

III Automation Tools 13

5 Pre-commit 15
5.1 Flake8 15
5.2 Interrogate 15

6 Github Actions 17

7 Sphinx/Readthedocs 19

IV Packaging the Project 21

8 Why package the project? 23

9 Using Pip 25

10 Demo Chapter 27
10.1 Senseless Section 30

11 Conclusions & Outlook 33

Bibliography 35

Declaration of Authorship 37

Part I

Introduction

1 Knowing your Users(?)

1.1 Feather by SAP

1.2 Alex == Research

2 Related Work(?)

2.1 Packaging discussion in the Python community

Part II

Challenges

3 Documentation

3.1 Wiki

3.2 Docstrings

3.3 Tests

4 Tests

4.1 Pytest

4.1.1 Setup & Teardown == Automation

4.1.2 Parametrization

4.2 What to test

4.3 Types of tests (e.g. Unit vs. Acceptance)

4.4 Writing maintainable tests

4.5 Measuring Coverage

Part III

Automation Tools

5 Pre-commit

5.1 Flake8

5.2 Interrogate

6 Github Actions

7 Sphinx/Readthedocs

Part IV

Packaging the Project

8 Why package the project?

9 Using Pip

10 Demo Chapter

This is where you can write some meta information about your chapter. For example, this chapter is based on one of my publications [NC25], and I just blindly copied everything without adjusting it. Just a heads-up warning.

Sadly, if you cite your own publications, they will appear in the bibliography. Thus, make sure to cite your papers with yourself as one of the authors.

This chapter shows off some of the basic formats of this thesis. Many packages are included in order for you to be able to start immediately without having to manually add all of the important things. The features deemed most important are now presented.

Here is just some filler text.[15] The following citations use the command textcite: Name and Co-Author [NC25]; Name et al. [Nam+30]. The first reference has a short list of authors, the second one a long list.

Also note that you can include to-do notes if necessary. Delete this chapter!

We now state a theorem and restate it later on again. Have a look at the source code in order to see how the theorem is written. Many macros are used, and all of them can be used without using math mode explicitly. Note that we can refer to inequality (10.1) as an inequality through the magic of an option in its label.

▶ **Theorem 10.1 (Variable Drift).** Let $(\mathcal{F}_t)_{t\in\mathbf{N}}$ be a filtration, $(X_t)_{t\in\mathbf{N}}$ be a random process over $\mathbf{R}_0^+$ adapted to $\mathcal{F}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that

1. $X_0 \geq x_{\min}$ and that
2. there is a monotonically increasing function $h\colon D \to \mathbf{R}^+$ such that, for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid \mathcal{F}_t] \geq h(X_t)$.

Then

$$\mathrm{E}[T \mid \mathcal{F}_0] \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)}\mathrm{d}z\,. \tag{10.1}$$

◀

15 Here is a footnote with a strange number (if that floats your boat). Note how the footnote mark is *above* the period at the end of the sentence.

(a) This is the caption of the subfigure that displays the logo of the HPI.

(b) This is the caption of the subfigure that displays the logo of the UP.

Figure 10.1: These are the two logos featured on the title page. Figure 10.1 (a) belongs to the HPI, whereas Figure 10.1 (b) belongs to the UP.

Please shift your attention to Figure 10.1. This reference was created using the package cleveref, which knows in what environment the label is defined in. This way, you can easily change a theorem into a lemma, and the name of the reference will be adjusted automatically. A wrapfigure like Figure 10.2 is referenced just like a normal figure.

Of course, you can also use tables in a fancy style. See, for example, Table 10.1. This document already contains packages in order to also handle larger tables. Hence, it is possible to use tables spanning multiple pages or to rotate a page into landscape in order to fit in a wider table.

Before we continue, consider the following obvious theorem. We conjecture that it also holds for $n = 2$.

▶ **Theorem 10.2.** Let $a, b, c, n \in \mathbb{N}^+$ with $n > 2$. Then

$$a^n + b^n \neq c^n \,.$$ ◀

Since the proof is straightforward, it is omitted. Nonetheless, we present a proof in order to show off the proof environment.

Proof of Theorem 10.2. Unfortunately, there is too little space in this PDF for the proof. ■

You can have very expressive and fancy enumerations from the package enumitem. Again, we can easily reference an item like item (i).

(i) The labels of the items can be nicely chosen.

(ii) Note how the labels are left-aligned. This does not look good but should demonstrate what is easily possible.

Table 10.1: This is a nicely formatted table. Thus, the caption is *above* the content. If not, the data could not be interpreted meaningfully. As a rule of thumb, never use vertical lines[1], and use horizontal lines sparingly. If you think that a table is illegible and thus needs vertical lines, then your spacing between columns is wrong and should be increased. Always use some whitespace first before you use some additional lines.

Text	**Number**
This is some text. Thus, it is left-aligned.	**0**
Numbers are right-aligned.	**1**
The numbers are formatted in bold using the package array.	**2**

We can even interrupt this enumeration and easily resume it immediately.

(iii) We continue where we left off.

Recall that Theorem 10.1 was as follows:

▶ **Theorem 10.1 (Variable Drift).** Let $(\mathcal{F}_t)_{t\in\mathbf{N}}$ be a filtration, $(X_t)_{t\in\mathbf{N}}$ be a random process over $\mathbf{R}_0^+$ adapted to $\mathcal{F}$, $x_{\min} > 0$, and let $T = \inf\{t \mid X_t < x_{\min}\}$. Additionally, let D denote the smallest real interval that contains at least all values $x \geq x_{\min}$ that, for all $t \leq T$, any X_t can take. Furthermore, suppose that

1. $X_0 \geq x_{\min}$ and that
2. there is a monotonically increasing function $h\colon D \to \mathbf{R}^+$ such that, for all $t < T$, we have $X_t - \mathrm{E}[X_{t+1} \mid \mathcal{F}_t] \geq h(X_t)$.

Then

$$\mathrm{E}[T \mid \mathcal{F}_0] \leq \frac{x_{\min}}{h(x_{\min})} + \int_{x_{\min}}^{X_0} \frac{1}{h(z)}\mathrm{d}z\,. \tag{10.1}$$

◀

Note that the reference above still refers to the first occurrence of the theorem. However, the theorem is repeated without any noise. That is, it is identical to the other occurrence.

From the next page on, other than a warp figure and some filler text, there is not much more to see. Thank you very much for taking your time and reading so far. I hope you got an impression of what this template is capable of. Have fun using it, and create a great thesis!

1 Except you know what you are doing.

10.1 Senseless Section

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Figure 10.2: The HPI logo is sneaked in between text.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis.

At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr,

At accusam aliquyam diam diam dolore dolores duo eirmod eos erat, et nonumy sed tempor et et invidunt justo labore Stet clita ea et gubergren, kasd magna no rebum. sanctus sea sed takimata ut vero voluptua. est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat.

Consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet dom-

ing id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo.

11 Conclusions & Outlook

Bibliography

[Nam+30] My Name, First Co-Author, Second Co-Author, Third Co-Author and Fourth Co-Author. **Dear Lord! How Did This Get Accepted?** *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 42:1 (2030), 2–1024 (see page 27).

[NC25] My Name and A Co-Author. **Useless Stuff That No One Cares About**. In: *Proceedings of the coolest Annual ACM Symposium on Theory of Computing (STOC'XX)*. ACM Press, 2025, 42–1337 (see page 27).

Declaration of Authorship

I hereby declare that this thesis is my own unaided work. All direct or indirect sources used are acknowledged as references.

Potsdam, 21st March 2022

Nikkel Mollenhauer