

Bachelor project 2017 / 2018

## Measuring and predicting software usage with intelligent monitoring

System Analysis and Modeling Research Group of Prof. Dr. Giese

### Project background

Software usage measurement helps to understand how software is actually used. We want to investigate how to deliver usage measurement results fast and with a small environmental footprint in the sense of green computing. Therefore we want to investigate how to make data collection more flexible and focused by selectively obtaining and processing relevant data.

Usually software already contains many measurement points, like log statements or spots for instrumentation. However they are mostly used to collect data during the software development cycle, for troubleshooting or to ensure service availability while operating the software. Thus the value of those points for usage measurement and in particular for capturing chains of related events is often unclear.

However, the introduction of new measurement points causes efforts for placing them properly and to implement them e.g. by adding dedicated log statements or calls of a usage measurement API (a). The corresponding delays can be significant e.g. due to software release cycles and deferred update deployments. In comparison reusing already existing measurement points would allow adapting fast to changing usage measurement objectives and starting data collection quickly (b).

Also continuous data collection over a large number of measurement points is not desirable, especially for performance and resource consumption reasons. Focusing instead only on current measurement objectives and selectively collecting data would result in a positive impact on the environmental footprint and performance as well as lower cost for software operation (c). Thus we are looking for an efficient way to identify those existing measurement points, which are relevant for a particular usage measurement objective.

We want to investigate the reuse of measurement points and the selective data collection in the domain of massive open online courses (MOOCs). Here users create interesting chains of related events by navigating through a course. Those chains can be captured through usage measurement, analyzed and used for predictions. But which events and consequently which corresponding measurement points are relevant for those predictions?

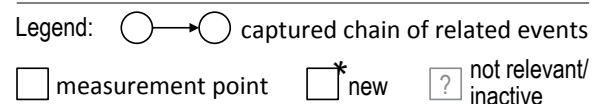
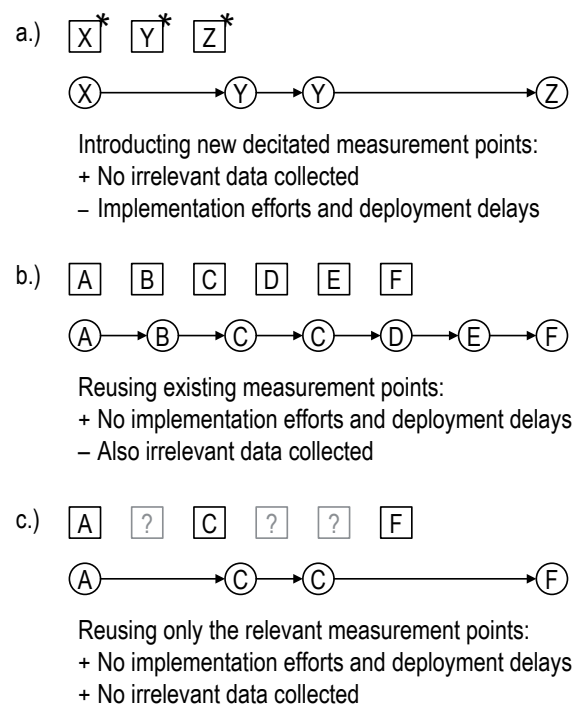


Figure: Usage measurement approaches

For MOOC providers the question whether participants complete a course or will dropout is major concern. Thus, as our first use case for software usage measurement we plan to predict likely dropouts from occurring event chains. The results can serve the MOOC providers as a starting point, e.g. for allocating their mentoring efforts effectively.

For our project partners, openSAP and openHPI, we want to find an efficient way to identify those existing measurement points within their microservices-based MOOC platform, which yield the relevant event data for such predictions. Thereby we want to minimize the efforts and delays for introducing new measurement points as well as the amount of the collected data.

It is worth mentioning that we can build on already existing scientific work regarding dropouts. We plan to use this well-researched topic as an exchangeable use case since we rather want to concentrate our study on software usage measurement approaches. However, there are already a number of ideas for possible additional use cases, which are related.

## Project objectives

The overall goal is to investigate how to make software usage measurement more elegant while increasing flexibility and capping resource consumption. For this purpose the following objectives shall be achieved during the project:

- > Understand the monitoring capabilities of the openSAP / openHPI microservices-based MOOC platform and assess them regarding their feasibility for usage measurement.
- > Design algorithms to determine the relevant measurement points for a given software usage measurement objective. For this purpose captured event chains shall be analyzed and data mining mechanisms like feature selection shall be investigated.
- > Collect data from relevant measurement points and apply machine learning algorithms to predict future software usage and to cluster certain usage types.
- > Provide a prototype for at least one use case, which can be integrated into the existing microservices-based MOOC platform of openSAP and openHPI. This includes configuring the monitoring, executing the data analysis and presenting the results.
- > Data collection for usage measurement can be adapted by changing the monitoring configuration manually. Computer-aided configuration however would be especially helpful when dealing with a large number of measurement points, their fine grained selection and multiple, potentially overlapping measurement objectives. Thus challenges and opportunities for automating the process to derive and adapt the monitoring configuration shall be explored.

## Main takeaways for students

In this project students can gain practical project experiences like coordinating multiple stakeholders, planning, work-sharing and collaborating. The project team needs to apply different aspects of software engineering like requirements analysis, solution design and quality assurance. While working towards the project objectives the students will further acquire insights and experiences with:

- > A real world microservices system architecture (openSAP / openHPI platform)
- > Data collection and monitoring for software usage measurement
- > Data mining and machine learning algorithms on chains of related events, which are actually simple graph constructs

## Project contacts

This bachelor project is offered by the System Analysis and Modeling chair of Professor Giese. From the chair Thomas Brand and Christian Adriano will coach the students and be their main contacts throughout the entire project (room: A-2.27, e-mail: [thomas.brand@hpi.de](mailto:thomas.brand@hpi.de), [christian.adriano@hpi.de](mailto:christian.adriano@hpi.de)). Our external project partner is the openSAP team around Clemens Link and Florian Schwerer. They will mainly be available for interviews and are interested in the software usage measurement results for the courses they provide. With the openHPI team we also have an internal project partner. Jan Renz and Tobias Rohloff will be our contacts. Besides providing MOOCs the openHPI team also develops and operates the microservices-based MOOC platform and monitoring infrastructure on which also openSAP is based.