Evidence over Opinion: An Empirical Approach to Software Engineering

Prof. Dr. Sebastian Baltes

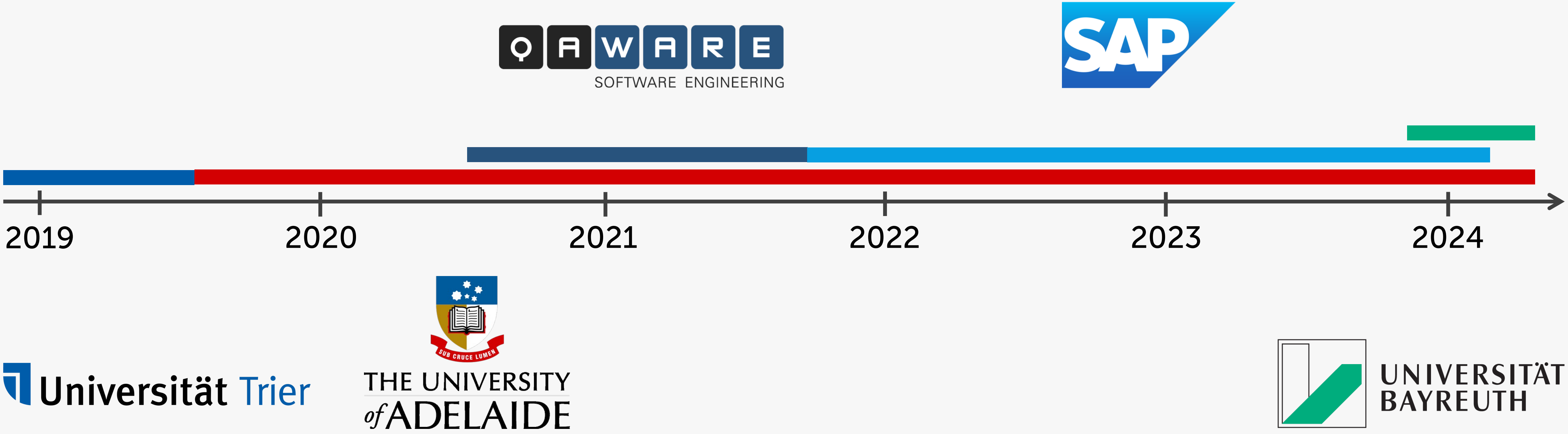
empirical-software.engineering

UNIVERSITÄT BAYREUTH

BayLDS

Lecture Series Digital Sciences at UBT
February 2024

Hi, my name is Sebastian and I’m a pracademic*



[* https://en.wikipedia.org/wiki/Pracademic](https://en.wikipedia.org/wiki/Pracademic)

A Brief History of Software Engineering

Origin of the term "Software Engineering"

Margaret Hamilton
(1965-1969)

Anthony G Oettinger
(1966)

NATO Software Engineering Conference
(1968)

The 1968 NATO SE Conference Report

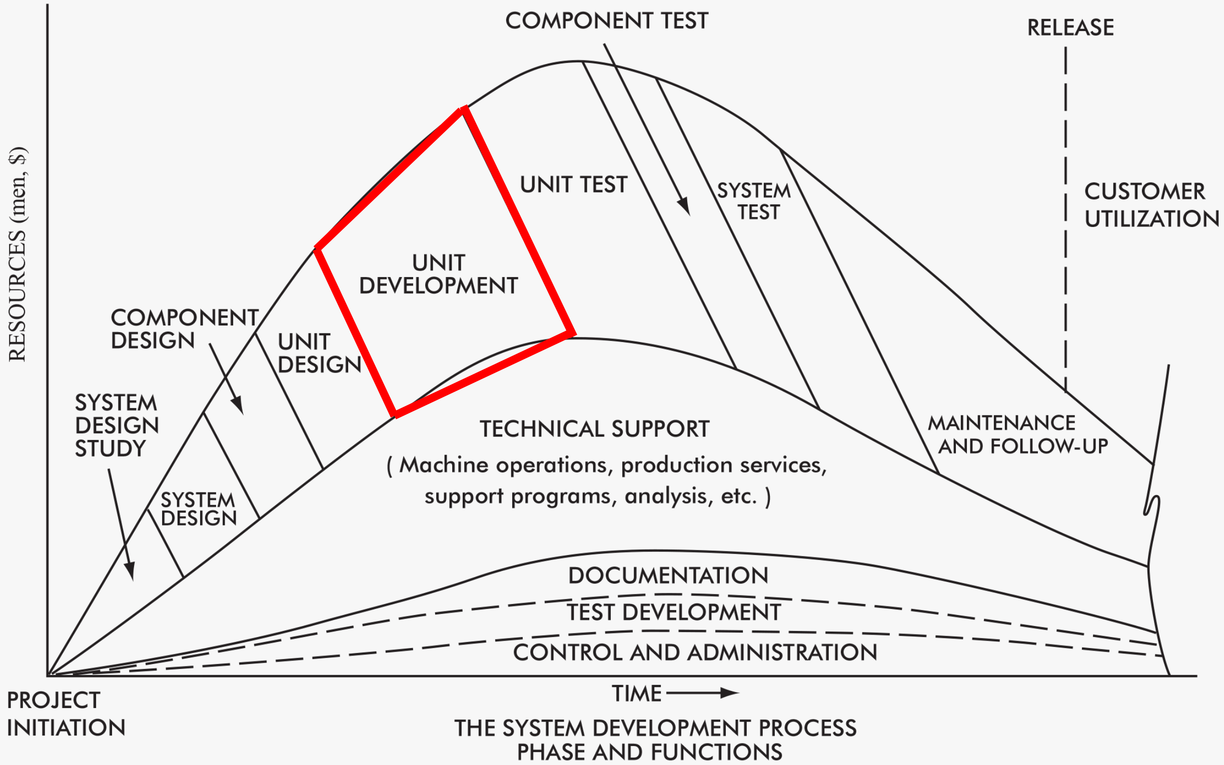
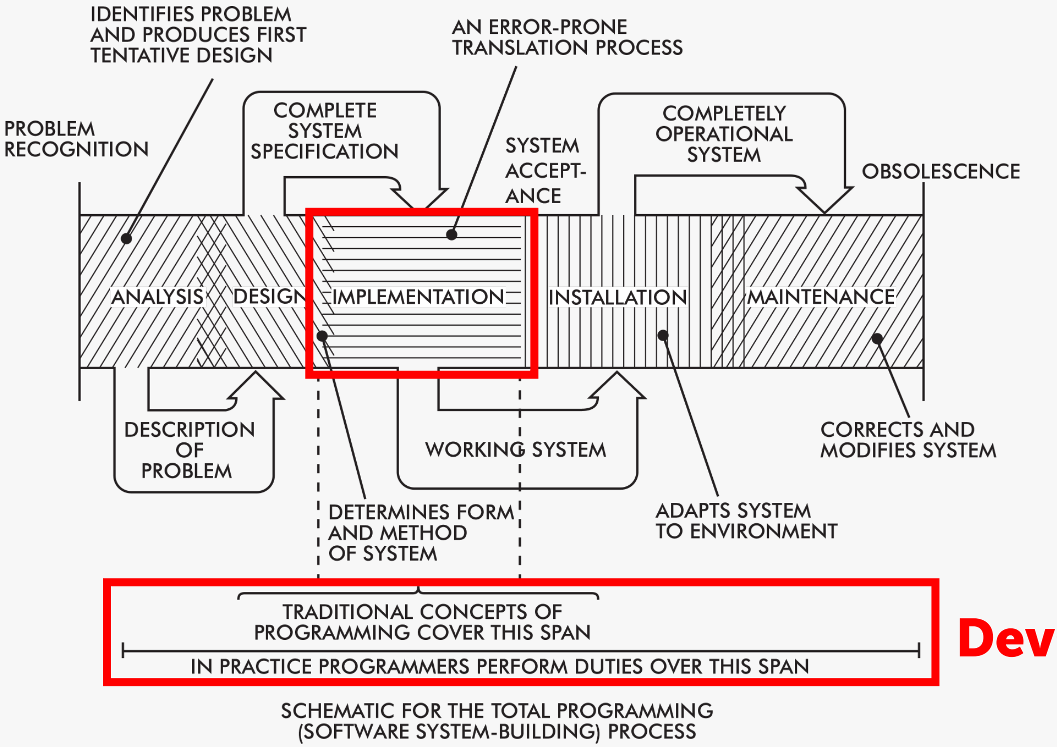


Figure 1. From Nash: Some problems in the production of large-scale software systems.

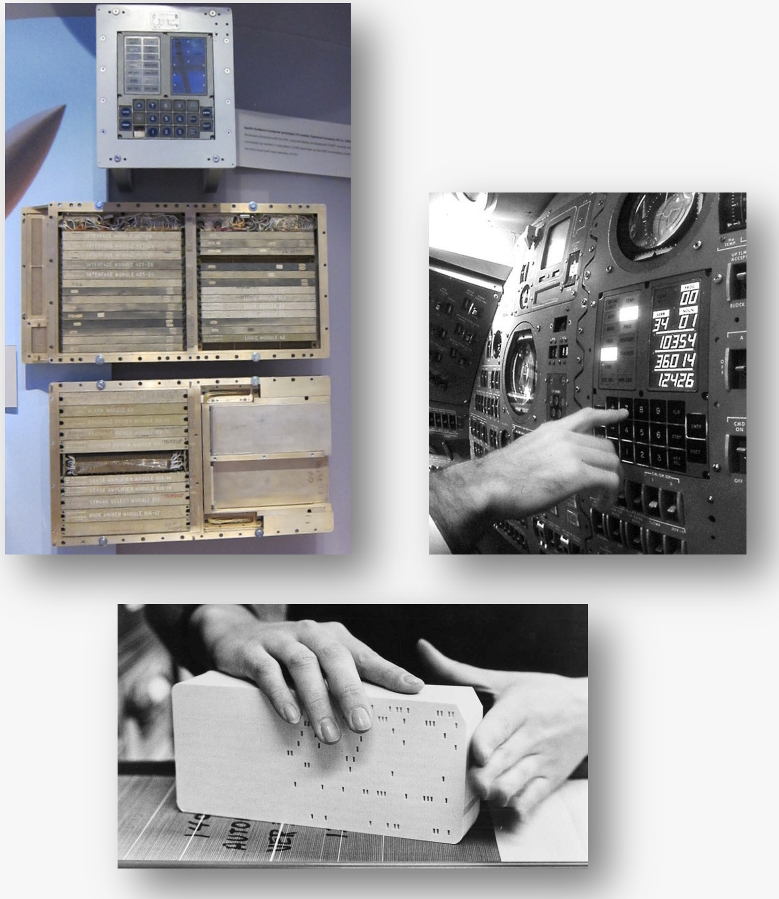

IDENTIFIES PROBLEM AND PRODUCES FIRST TENTATIVE DESIGN
AN ERROR-PRONE TRANSLATION PROCESS
PROBLEM RECOGNITION
COMPLETE SYSTEM SPECIFICATION
SYSTEM ACCEPT-ANCE
COMPLETELY OPERATIONAL SYSTEM
OBSOLESCENCE
ANALYSIS
DESIGN
IMPLEMENTATION
INSTALLATION
MAINTENANCE
DESCRIPTION OF PROBLEM
WORKING SYSTEM
CORRECTS AND MODIFIES SYSTEM
DETERMINES FORM AND METHOD OF SYSTEM
ADAPTS SYSTEM TO ENVIRONMENT
TRADITIONAL CONCEPTS OF PROGRAMMING COVER THIS SPAN
IN PRACTICE PROGRAMMERS PERFORM DUTIES OVER THIS SPAN
SCHEMATIC FOR THE TOTAL PROGRAMMING (SOFTWARE SYSTEM-BUILDING) PROCESS

DevOps

Figure 2. From Selig: Documentation for service and users. Originally due to Constantine.

http://homepages.cs.ncl.ac.uk/brian.randell/NATO/nato1968.PDF

A lot has changed since the 1960s



1966: Apollo guidance computer

AWS Lambda

Amazon Aurora

2024: Embedded systems still relevant, but way more application domains and abstractions of hardware and software.

While many of the **fundamental concepts** introduced in the first years of software engineering are **still valid**, the application domains, stakeholders, tools, infrastructure, and processes today are **more diverse** than ever before.

Disciplinary Boundaries of Software Engineering

Disciplinary Boundaries of Software Engineering

With a **traditional view** emphasizing software engineering’s **roots in computer and systems engineering** many questions of modern software development **cannot be answered**.

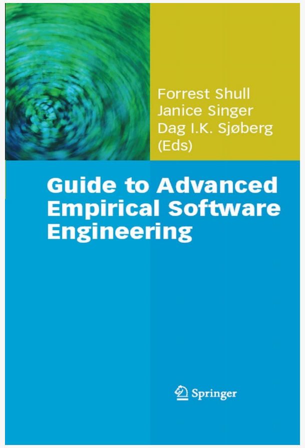
Examples:

- *How can we develop visual programming environments without knowledge of cognition?*
- *How can we study pair or mob programming without a deep understanding of verbal and non-verbal communication?*
- *How can we fully grasp the implications of AI-generated code without understanding copyright legislation and software licenses?*

Personal Observations

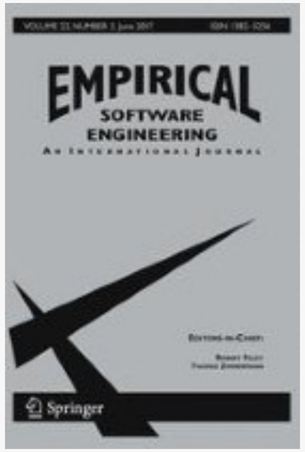
1. Many problems **relevant** in the **software industry** are rooted in software engineering but often have an **interdisciplinary** angle.
2. To impact industry, academia needs to provide **actionable recommendations** addressing **problems rooted in practitioners' actual needs**.
3. **Empirical research methods are essential** for identifying such problems (*problem space*) and corroborating recommendations/proposed solutions with empirical evidence (*solution space*).

Empirical Software Engineering



2008: “Ten years ago, it was **rare to see a conference or journal article about a software development tool or process that had empirical data to back up the claims**. Today, in contrast, it is becoming **more and more common** […].”

https://link.springer.com/book/10.1007/978-1-84800-044-5



2020: “[…] **it has become clear that empirical studies are a fundamental component of software engineering research and practice** […].”

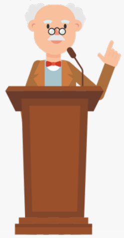
https://www.springer.com/journal/10664/aims-and-scope

Institutionalized Boundaries

Photo by David Iliff. License: CC BY-SA 3.0

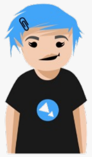
Institutionalized Boundaries

"If you were using ***MDSE**** *for building your mobile app, you'd see huge quality improvements, see papers x, y, z."*



Research

"Have you heard about things like ***time-to-market*** *and quickly responding to customer feedback? We're not building safety-critical software."*

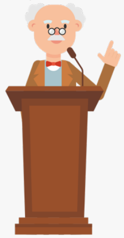


Practice

* https://softwareengineering.stackexchange.com/q/288569

Institutionalized Boundaries

*"If you were using **MDSE*** for building your mobile app, you'd see huge quality improvements, see papers x, y, z."*



*"Have you heard about things like **time-to-market** and quickly responding to customer feedback? We're not building safety-critical software."*

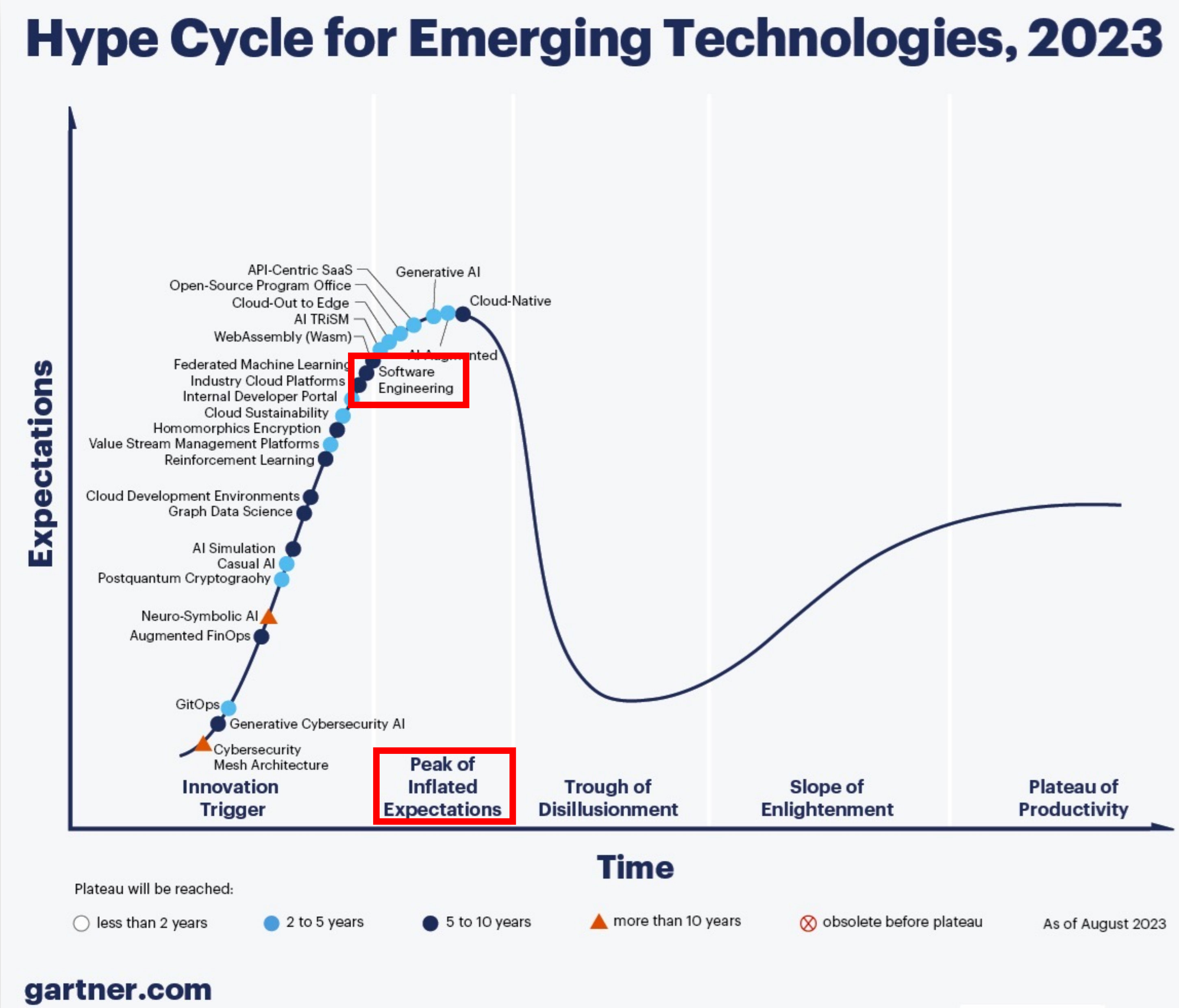
Research

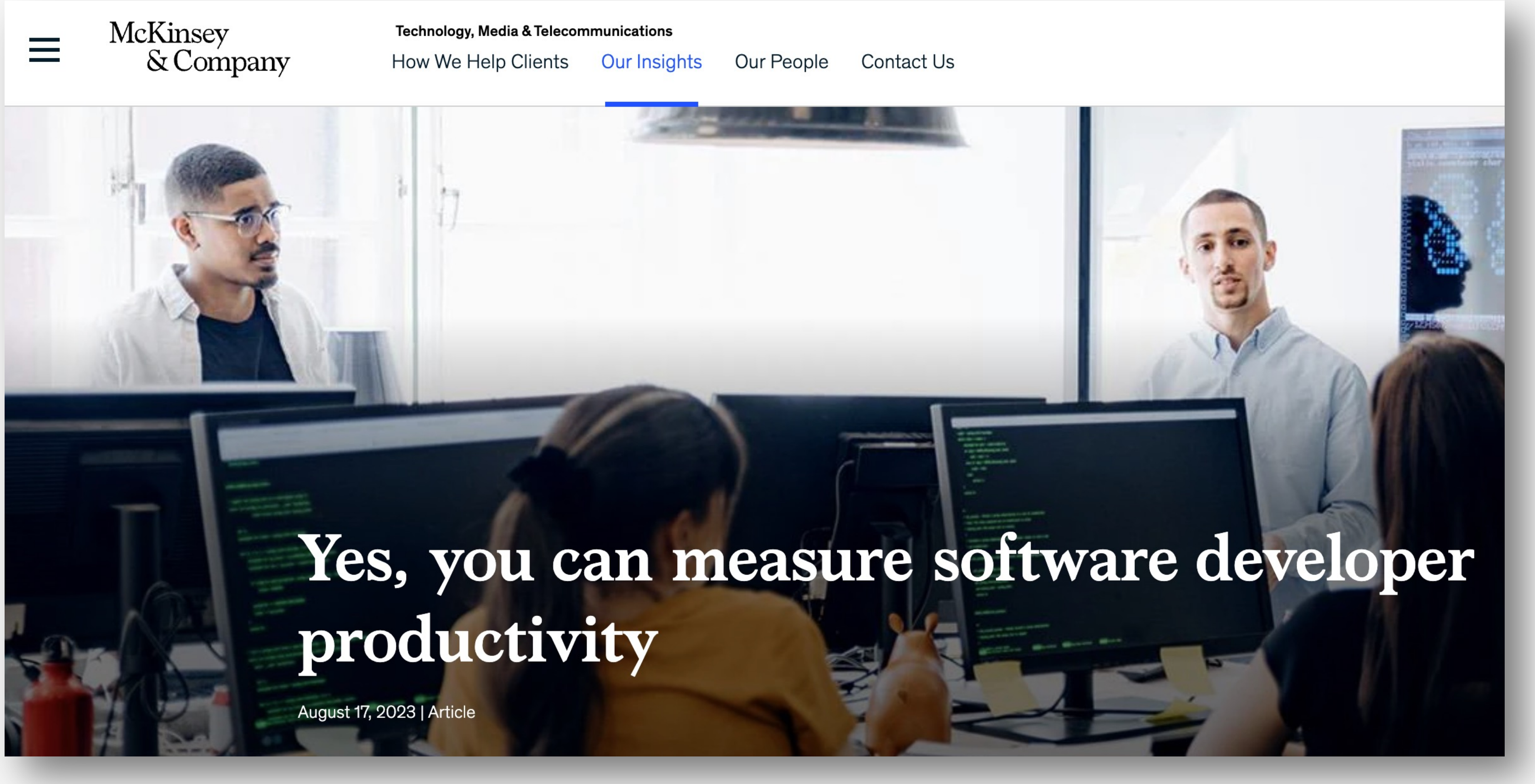
Practice

Issue in practice:

Hype-driven Software Engineering

* https://softwareengineering.stackexchange.com/q/288569

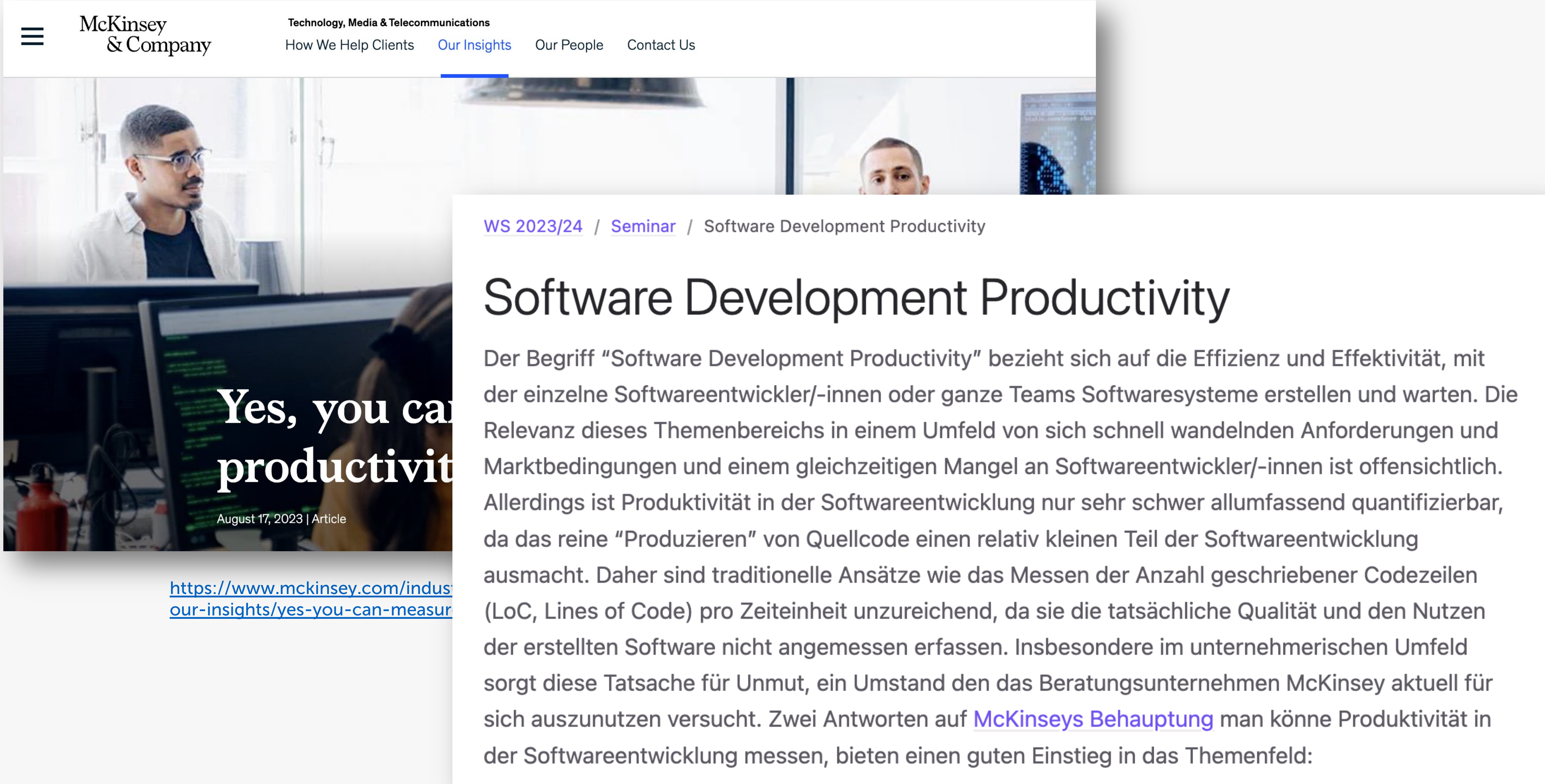

Hype Cycle for Emerging Technologies, 2023
Expectations
API-Centric SaaS
Open-Source Program Office
Cloud-Out to Edge
AI TRiSM
WebAssembly (Wasm)
Generative AI
Cloud-Native
AI-Augmented
Software Engineering
Federated Machine Learning
Industry Cloud Platforms
Internal Developer Portal
Cloud Sustainability
Homomorphics Encryption
Value Stream Management Platforms
Reinforcement Learning
Cloud Development Environments
Graph Data Science
AI Simulation
Casual AI
Postquantum Cryptograohy
Neuro-Symbolic AI
Augmented FinOps
GitOps
Generative Cybersecurity AI
Cybersecurity Mesh Architecture
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
Time
Plateau will be reached:
○ less than 2 years
● 2 to 5 years
● 5 to 10 years
▲ more than 10 years
⊗ obsolete before plateau
As of August 2023
gartner.com



Yes, you can measure software developer productivity

August 17, 2023 | Article

https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/yes-you-can-measure-software-developer-productivity



McKinsey & Company

Technology, Media & Telecommunications

How We Help Clients   Our Insights   Our People   Contact Us

Yes, you ca
productivit

August 17, 2023 | Article

https://www.mckinsey.com/indus
our-insights/yes-you-can-measur

WS 2023/24 / Seminar / Software Development Productivity

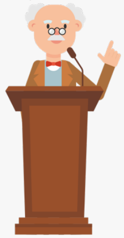
Software Development Productivity

Der Begriff “Software Development Productivity” bezieht sich auf die Effizienz und Effektivität, mit der einzelne Softwareentwickler/-innen oder ganze Teams Softwaresysteme erstellen und warten. Die Relevanz dieses Themenbereichs in einem Umfeld von sich schnell wandelnden Anforderungen und Marktbedingungen und einem gleichzeitigen Mangel an Softwareentwickler/-innen ist offensichtlich. Allerdings ist Produktivität in der Softwareentwicklung nur sehr schwer allumfassend quantifizierbar, da das reine “Produzieren” von Quellcode einen relativ kleinen Teil der Softwareentwicklung ausmacht. Daher sind traditionelle Ansätze wie das Messen der Anzahl geschriebener Codezeilen (LoC, Lines of Code) pro Zeiteinheit unzureichend, da sie die tatsächliche Qualität und den Nutzen der erstellten Software nicht angemessen erfassen. Insbesondere im unternehmerischen Umfeld sorgt diese Tatsache für Unmut, ein Umstand den das Beratungsunternehmen McKinsey aktuell für sich auszunutzen versucht. Zwei Antworten auf McKinseys Behauptung man könne Produktivität in der Softwareentwicklung messen, bieten einen guten Einstieg in das Themenfeld:

https://ubt-se.github.io/teaching/semesters/23-24_ws/seminar/productivity.html

Institutionalized Boundaries

*"If you were using **MDSE*** for building your mobile app, you'd see huge quality improvements, see papers x, y, z."*

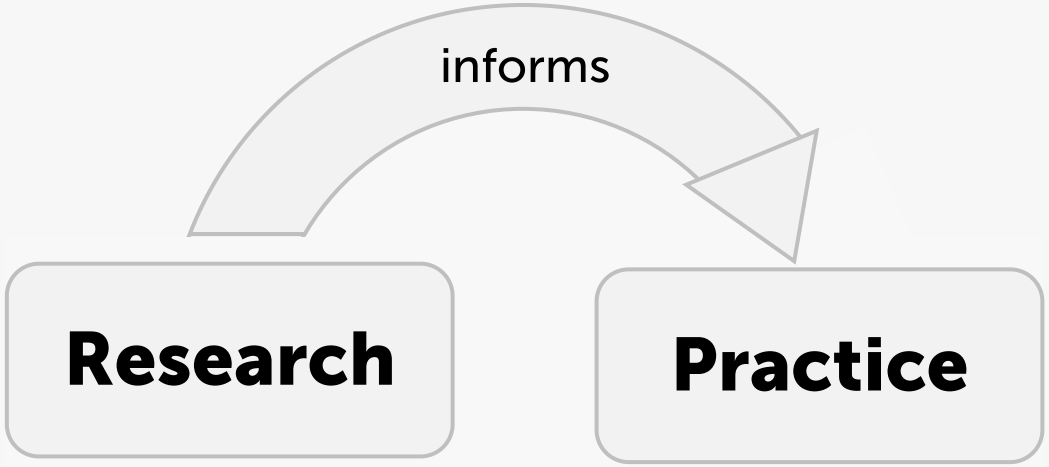


*"Have you heard about things like **time-to-market** and quickly responding to customer feedback? We're not building safety-critical software."*

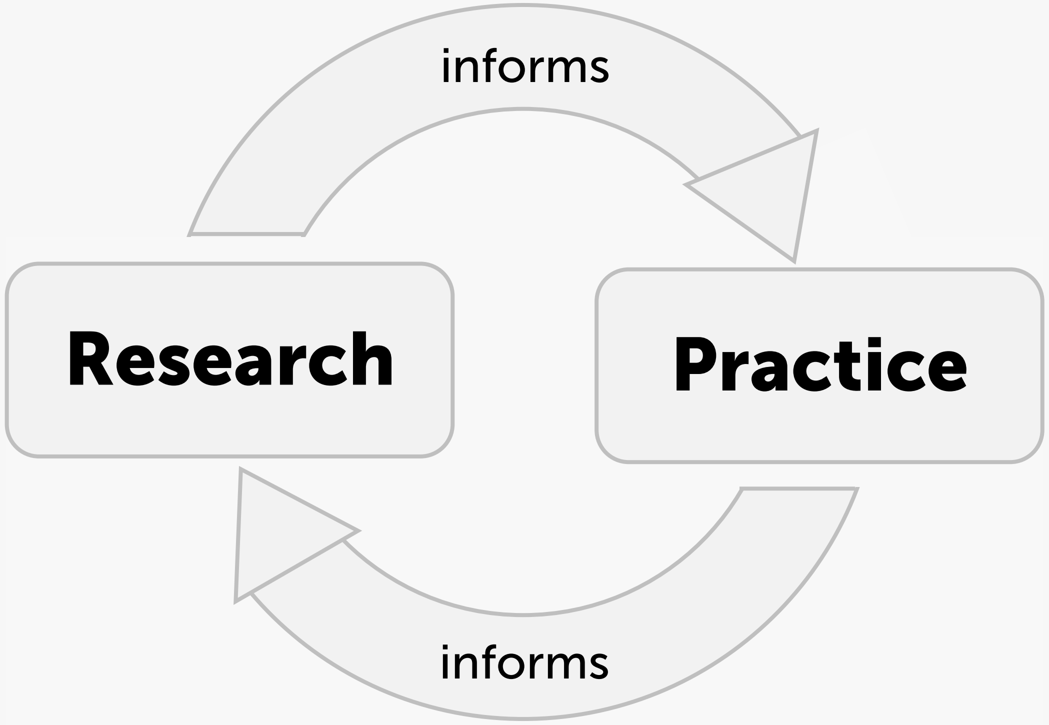
Research | **Practice**

[* https://softwareengineering.stackexchange.com/q/288569](https://softwareengineering.stackexchange.com/q/288569)

Institutionalized Boundaries



Institutionalized Boundaries



Implications for researchers:

1) Strong understanding of **state of practice** is essential.
2) To reach this understanding, we need to utilize **diverse empirical research methods** and **learn from other disciplines.**
3) To advance evidence-based practice, we need to invest effort into **communicating findings/potential solutions back** to practitioners.

An Example from Industry: Tech Stack Harmonization

Paraphrased excerpt from a status report to the SAP CTO:

- **Context:** SAP has a wide range of languages, tools, and infrastructure (“technology stacks”) being used to build, test, and deploy products and services.
- **Problem:** Maintaining too heterogenous tech stacks is expensive and inefficient, it also increases the company’s attack surface.
- **Progress:**
  - Derived a layered tech stack model based on interviews and a survey.
  - Used that model to build a data mining pipeline and dashboard for visualizing core tech stack aspects.
  - Discussed findings with product teams after a live demo.
  - Developed proposal for <CONFIDENTIAL>.

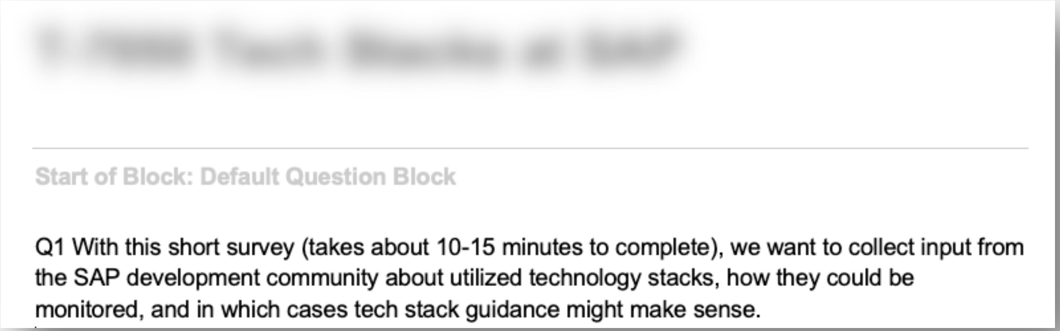
1. Strong understanding of state of practice.

2. Diverse empirical research methods.

3. Communicating findings back.

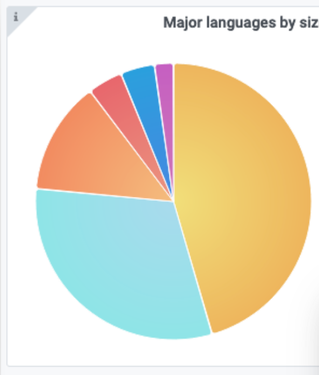
4. Solution based on research results.

An Example from Industry: Tech Stack Harmonization

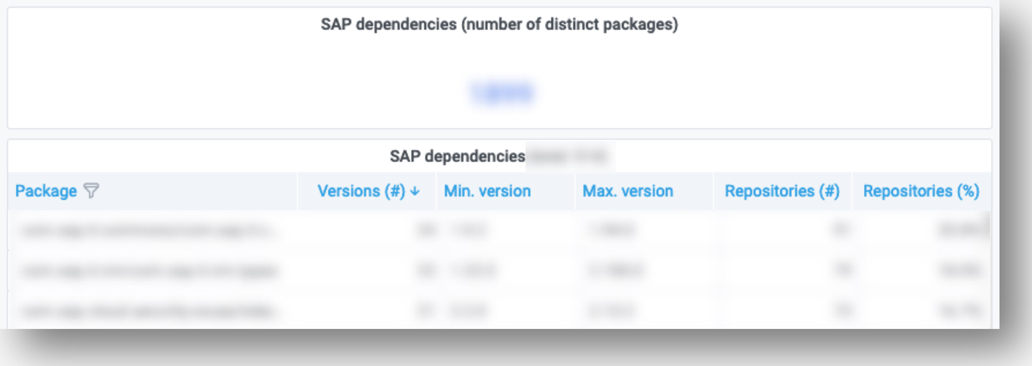


Start of Block: Default Question Block

Q1 With this short survey (takes about 10-15 minutes to complete), we want to collect input from the SAP development community about utilized technology stacks, how they could be monitored, and in which cases tech stack guidance might make sense.



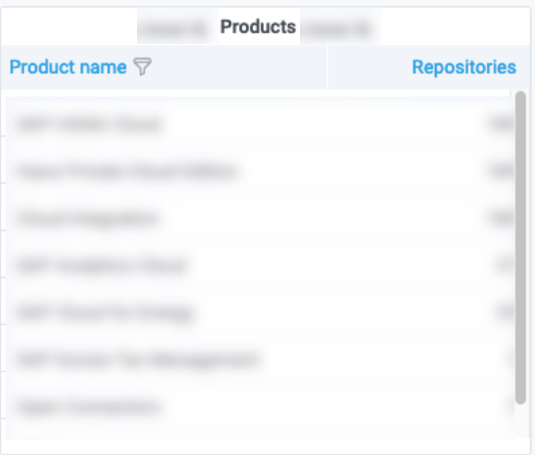
Major languages by size of source files



SAP dependencies (number of distinct packages)

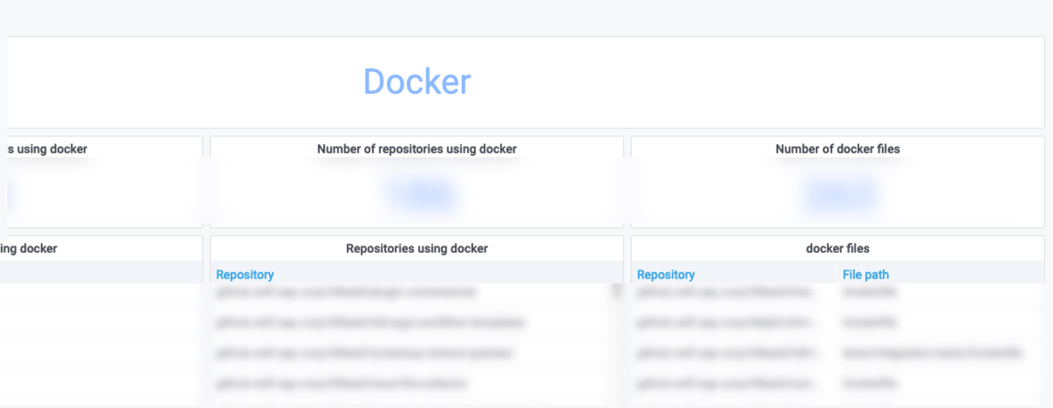
SAP dependencies

Package	Versions (#)	Min. version	Max. version	Repositories (#)	Repositories (%)



Products

Product name	Repositories



Docker

s using docker

Number of repositories using docker

Number of docker files

Products using docker

Product

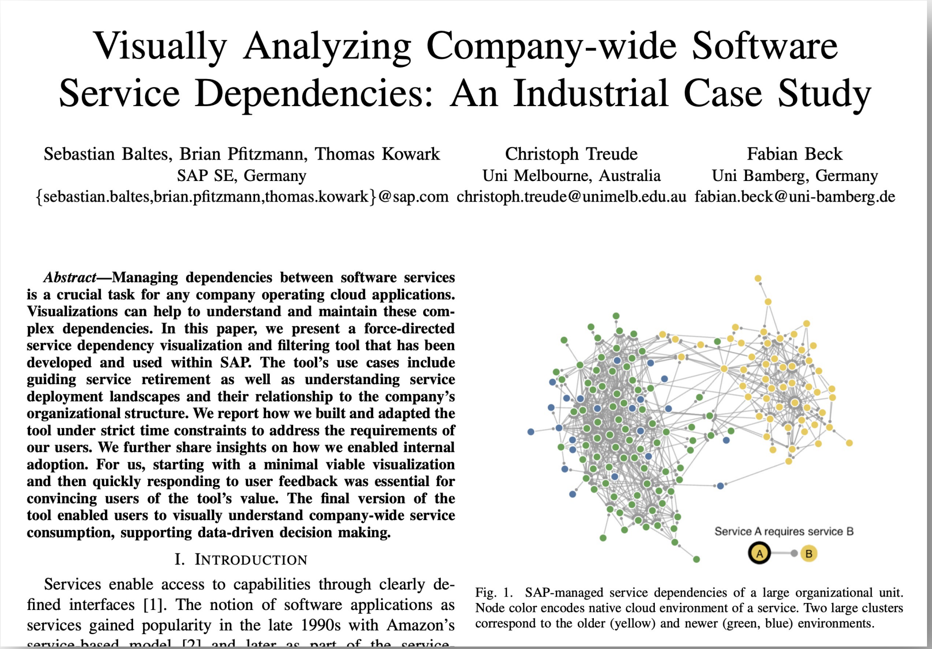
Repositories using docker

Repository

docker files

Repository	File path

An Example from Industry: Service Dependencies



Visually Analyzing Company-wide Software Service Dependencies: An Industrial Case Study

Sebastian Baltes, Brian Pfitzmann, Thomas Kowark
SAP SE, Germany
{sebastian.baltes,brian.pfitzmann,thomas.kowark}@sap.com

Christoph Treude
Uni Melbourne, Australia
christoph.treude@unimelb.edu.au

Fabian Beck
Uni Bamberg, Germany
fabian.beck@uni-bamberg.de

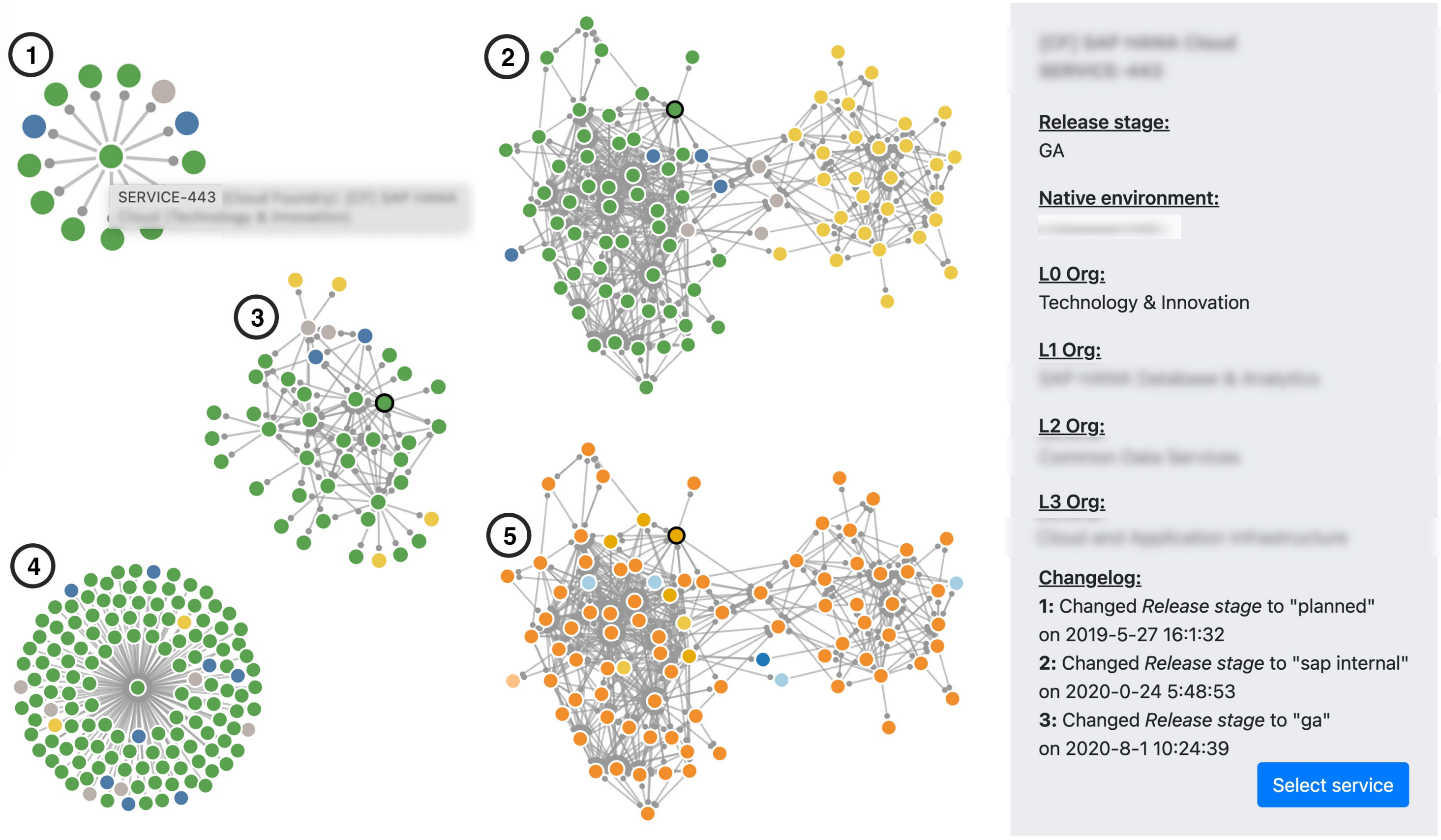
***Abstract*—Managing dependencies between software services is a crucial task for any company operating cloud applications. Visualizations can help to understand and maintain these complex dependencies. In this paper, we present a force-directed service dependency visualization and filtering tool that has been developed and used within SAP. The tool's use cases include guiding service retirement as well as understanding service deployment landscapes and their relationship to the company's organizational structure. We report how we built and adapted the tool under strict time constraints to address the requirements of our users. We further share insights on how we enabled internal adoption. For us, starting with a minimal viable visualization and then quickly responding to user feedback was essential for convincing users of the tool's value. The final version of the tool enabled users to visually understand company-wide service consumption, supporting data-driven decision making.**

I. Introduction

Services enable access to capabilities through clearly defined interfaces [1]. The notion of software applications as services gained popularity in the late 1990s with Amazon's service-based model [2] and later as part of the service-

Fig. 1. SAP-managed service dependencies of a large organizational unit. Node color encodes native cloud environment of a service. Two large clusters correspond to the older (yellow) and newer (green, blue) environments.

https://empirical-software.engineering/publications/#vissoft23-service-dependency-viz



[redacted]

[redacted]

Release stage:
GA

Native environment:
[redacted]

L0 Org:
Technology & Innovation

L1 Org:
[redacted]

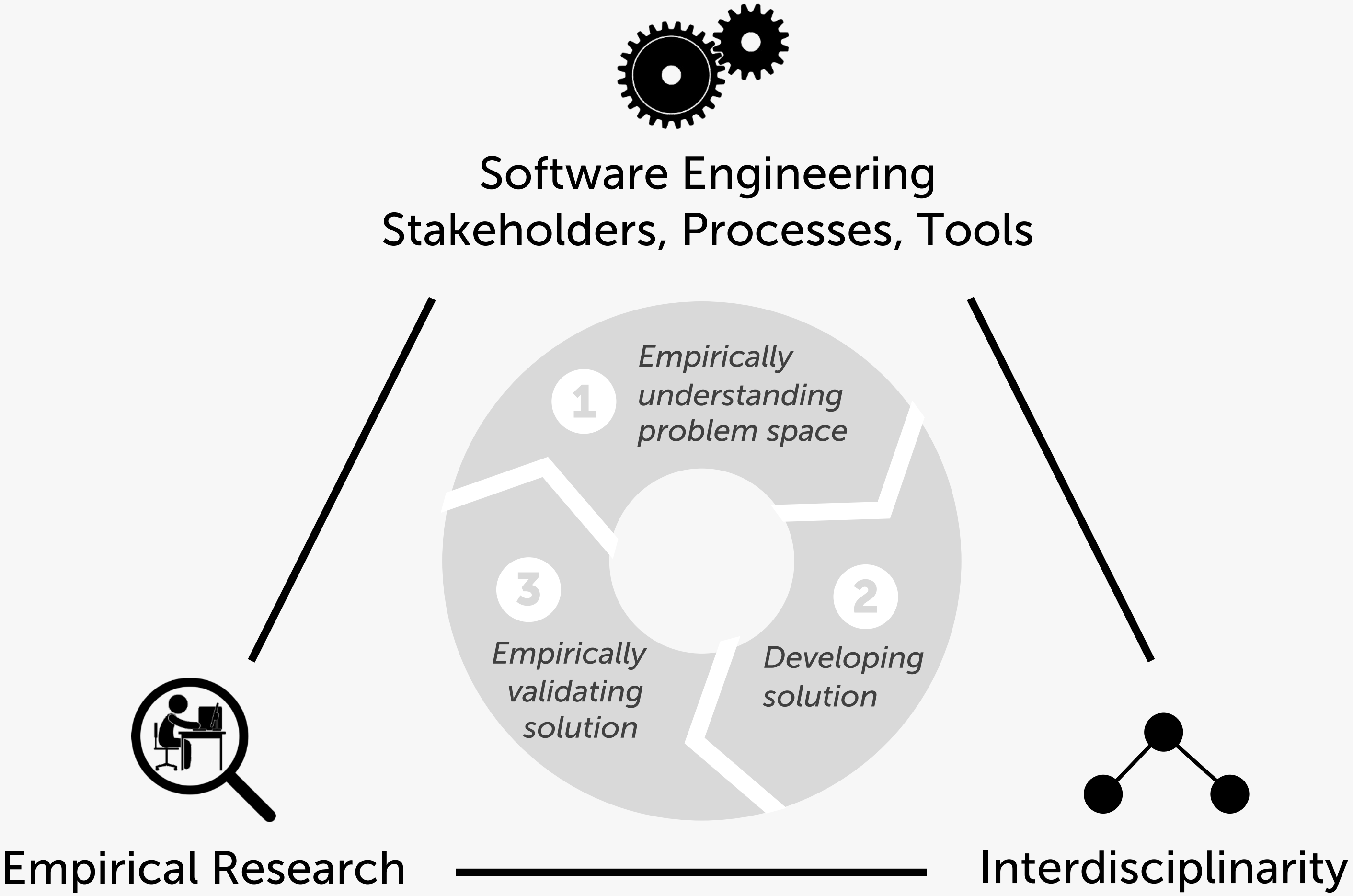
L2 Org:
[redacted]

L3 Org:
[redacted]

Changelog:
1: Changed *Release stage* to "planned" on 2019-5-27 16:1:32
2: Changed *Release stage* to "sap internal" on 2020-0-24 5:48:53
3: Changed *Release stage* to "ga" on 2020-8-1 10:24:39

My framework for Empirical Software Engineering Research

Empirical Software Engineering



Software Engineering
Stakeholders, Processes, Tools

1 *Empirically understanding problem space*

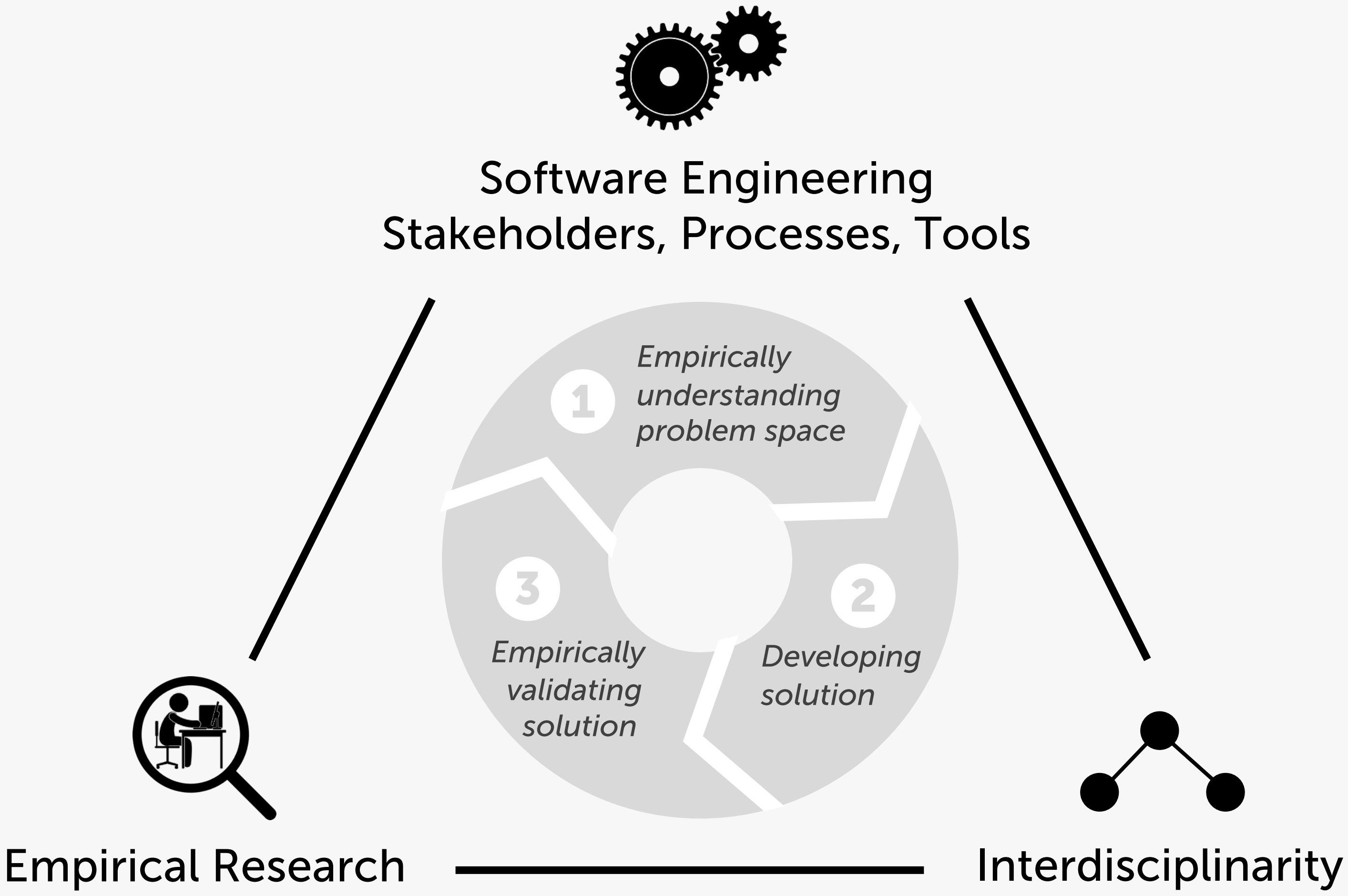
2 *Developing solution*

3 *Empirically validating solution*

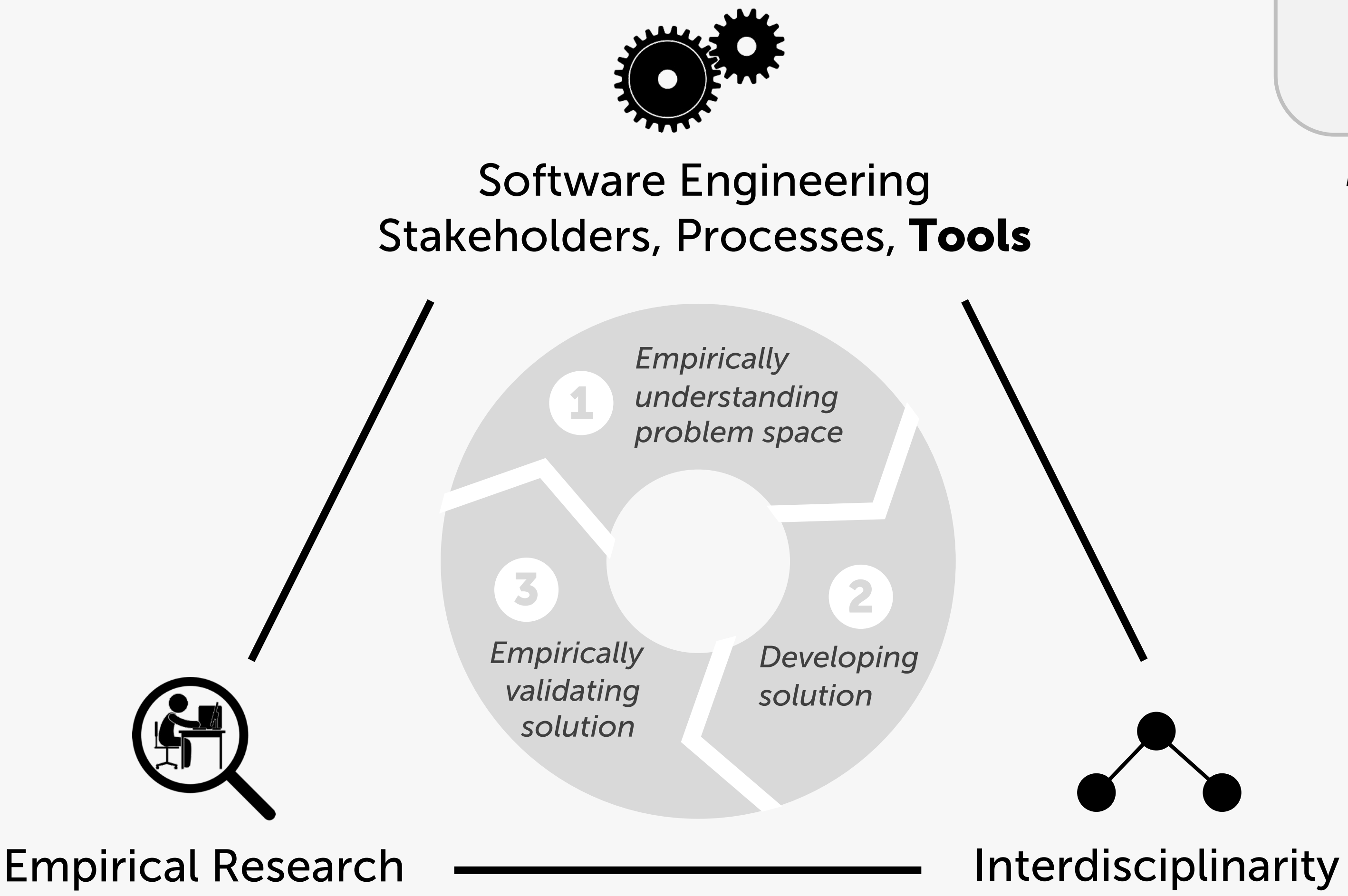
Empirical Research

Interdisciplinarity

Remainder of this talk


Software Engineering
Stakeholders, Processes, Tools
1
Empirically understanding problem space
2
Developing solution
3
Empirically validating solution
Empirical Research
Interdisciplinarity

Remainder of this talk



Software Engineering
Stakeholders, Processes, **Tools**

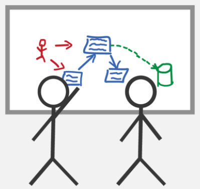
1 *Empirically understanding problem space*

2 *Developing solution*

3 *Empirically validating solution*

Empirical Research

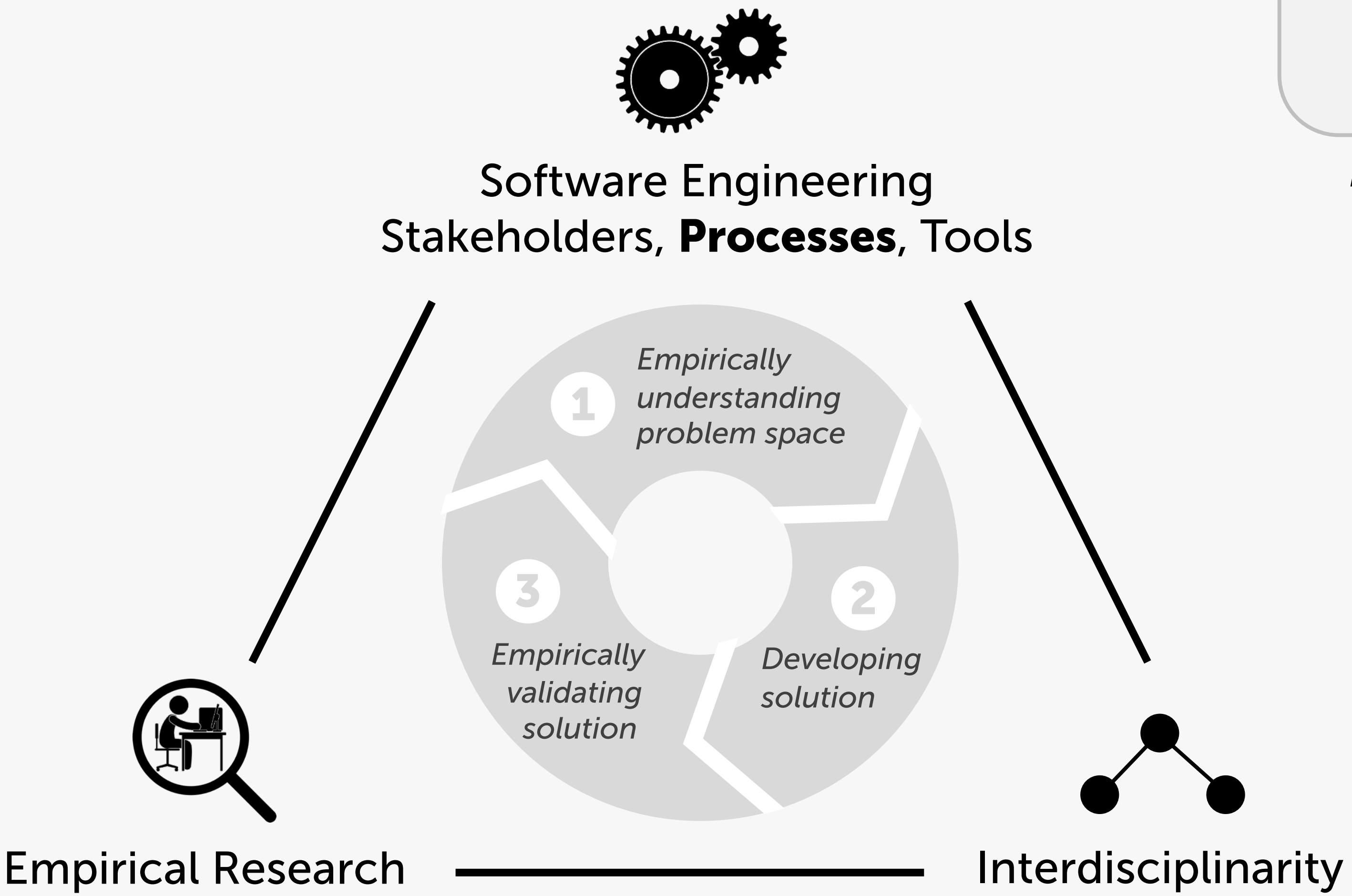
Interdisciplinarity



Sketching

FSE '14, ESEM '15,
VISSOFT '17

Remainder of this talk



Software Engineering
Stakeholders, **Processes**, Tools

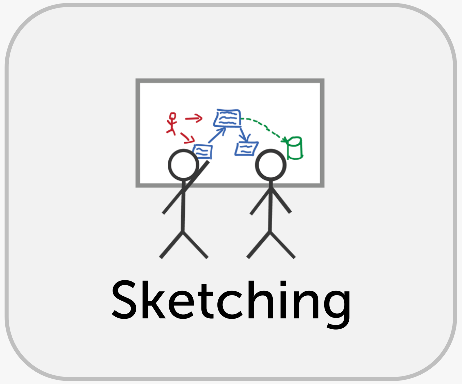
1 *Empirically understanding problem space*

2 *Developing solution*

3 *Empirically validating solution*

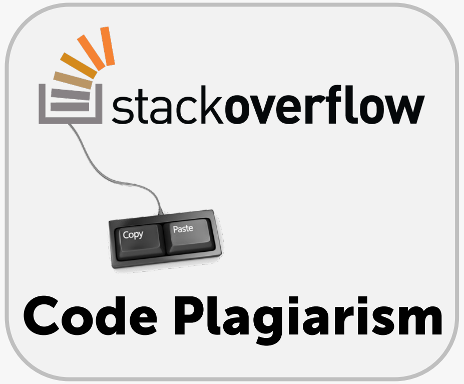
Empirical Research

Interdisciplinarity



Sketching

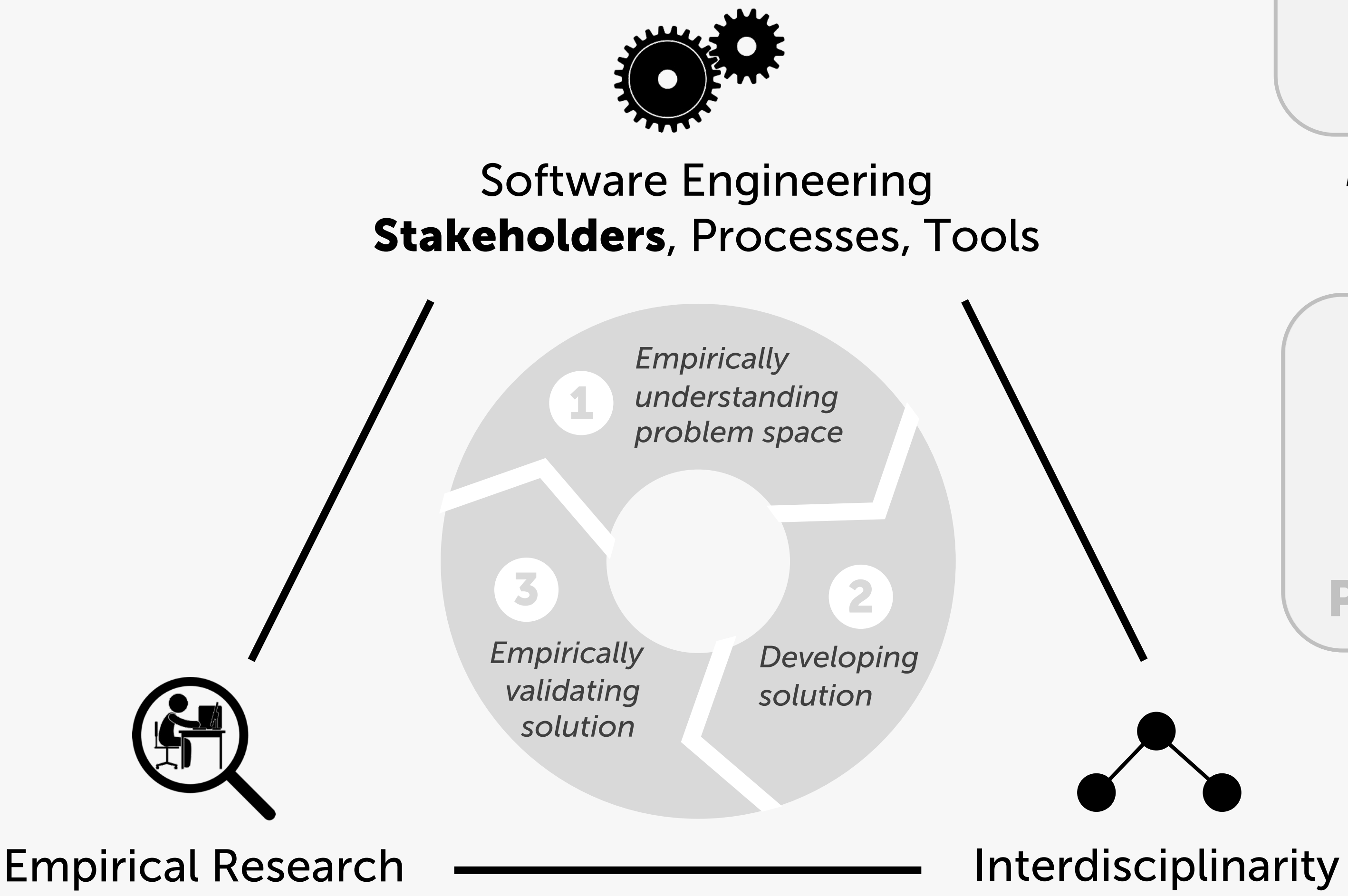
FSE '14, ESEM '15, VISSOFT '17



Code Plagiarism

EMSE '18, MSR '18, MSR '19, ICSE '20

Remainder of this talk



Software Engineering
Stakeholders, Processes, Tools

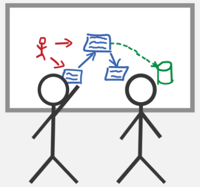
1 *Empirically understanding problem space*

2 *Developing solution*

3 *Empirically validating solution*

Empirical Research

Interdisciplinarity



Sketching

FSE '14, ESEM '15, VISSOFT '17

Code Plagiarism

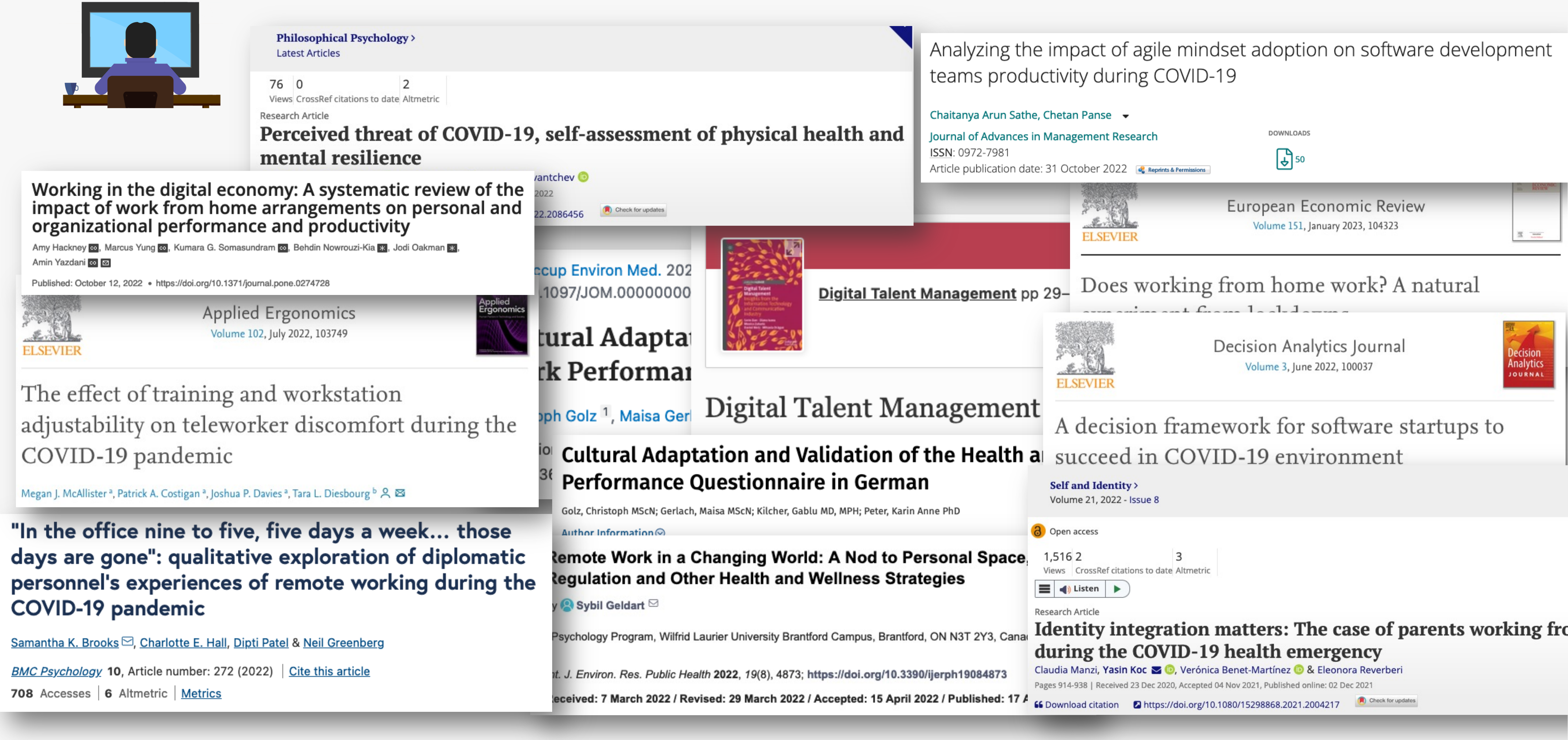
EMSE '18, MSR '18, MSR '19, ICSE '20

Pandemic Programming

EMSE '20

Pandemic Programming

Impact beyond SE



Philosophical Psychology ›
Latest Articles

76 Views | 0 CrossRef citations to date | 2 Altmetric

Research Article

Perceived threat of COVID-19, self-assessment of physical health and mental resilience

…antchev
…2022
…22.2086456 | Check for updates

Analyzing the impact of agile mindset adoption on software development teams productivity during COVID-19

Chaitanya Arun Sathe, Chetan Panse

Journal of Advances in Management Research
ISSN: 0972-7981
Article publication date: 31 October 2022 | Reprints & Permissions

DOWNLOADS 50

Working in the digital economy: A systematic review of the impact of work from home arrangements on personal and organizational performance and productivity

Amy Hackney, Marcus Yung, Kumara G. Somasundram, Behdin Nowrouzi-Kia, Jodi Oakman, Amin Yazdani

Published: October 12, 2022 • https://doi.org/10.1371/journal.pone.0274728

European Economic Review
Volume 151, January 2023, 104323

Does working from home work? A natural …

…ccup Environ Med. 202…
…1097/JOM.00000000…

…tural Adapta…
…rk Performan…

…oph Golz 1, Maisa Ger…

Digital Talent Management pp 29–…

Digital Talent Management

Applied Ergonomics
Volume 102, July 2022, 103749

The effect of training and workstation adjustability on teleworker discomfort during the COVID-19 pandemic

Megan J. McAllister a, Patrick A. Costigan a, Joshua P. Davies a, Tara L. Diesbourg b

Decision Analytics Journal
Volume 3, June 2022, 100037

A decision framework for software startups to succeed in COVID-19 environment

Cultural Adaptation and Validation of the Health a… Performance Questionnaire in German

Golz, Christoph MScN; Gerlach, Maisa MScN; Kilcher, Gablu MD, MPH; Peter, Karin Anne PhD

Author Information

Self and Identity ›
Volume 21, 2022 - Issue 8

Open access

1,516 Views | 2 CrossRef citations to date | 3 Altmetric

Listen

Research Article

Identity integration matters: The case of parents working fr… during the COVID-19 health emergency

Claudia Manzi, Yasin Koc, Verónica Benet-Martínez & Eleonora Reverberi

Pages 914-938 | Received 23 Dec 2020, Accepted 04 Nov 2021, Published online: 02 Dec 2021

Download citation | https://doi.org/10.1080/15298868.2021.2004217 | Check for updates

"In the office nine to five, five days a week… those days are gone": qualitative exploration of diplomatic personnel's experiences of remote working during the COVID-19 pandemic

Samantha K. Brooks, Charlotte E. Hall, Dipti Patel & Neil Greenberg

BMC Psychology **10**, Article number: 272 (2022) | Cite this article

708 Accesses | **6** Altmetric | Metrics

…emote Work in a Changing World: A Nod to Personal Space, …egulation and Other Health and Wellness Strategies

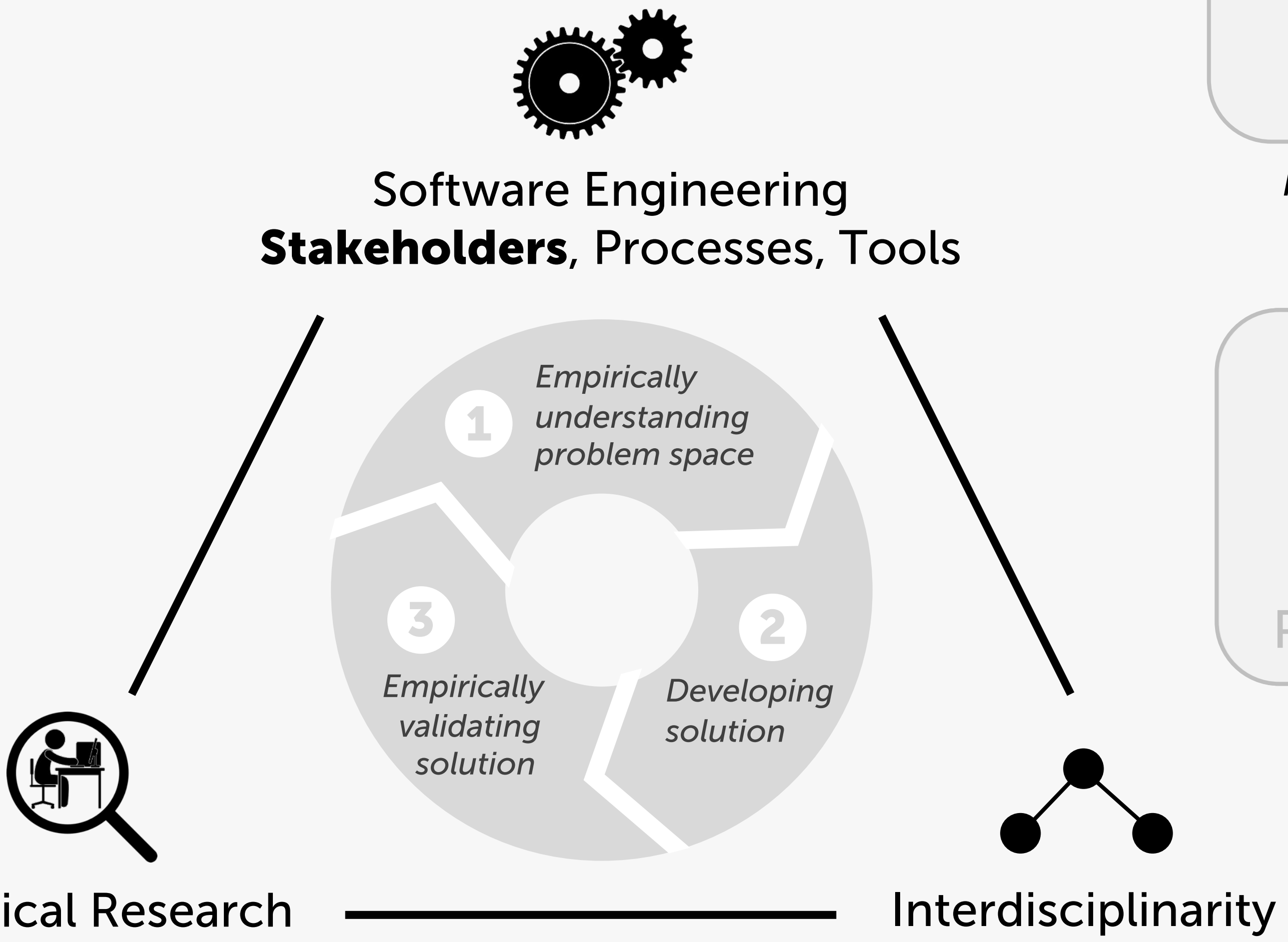
…y Sybil Geldart

…Psychology Program, Wilfrid Laurier University Brantford Campus, Brantford, ON N3T 2Y3, Cana…

…t. J. Environ. Res. Public Health **2022**, *19*(8), 4873; **https://doi.org/10.3390/ijerph19084873**

…eceived: 7 March 2022 / Revised: 29 March 2022 / Accepted: 15 April 2022 / Published: 17 A…

Remainder of this talk



Software Engineering
Stakeholders, Processes, Tools

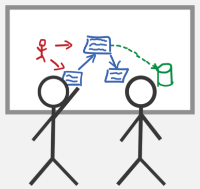
1 *Empirically understanding problem space*

2 *Developing solution*

3 *Empirically validating solution*

Empirical Research

Interdisciplinarity



Sketching

FSE '14, ESEM '15, VISSOFT '17

stackoverflow

Code Plagiarism

EMSE '18, MSR '18, MSR '19, ICSE '20

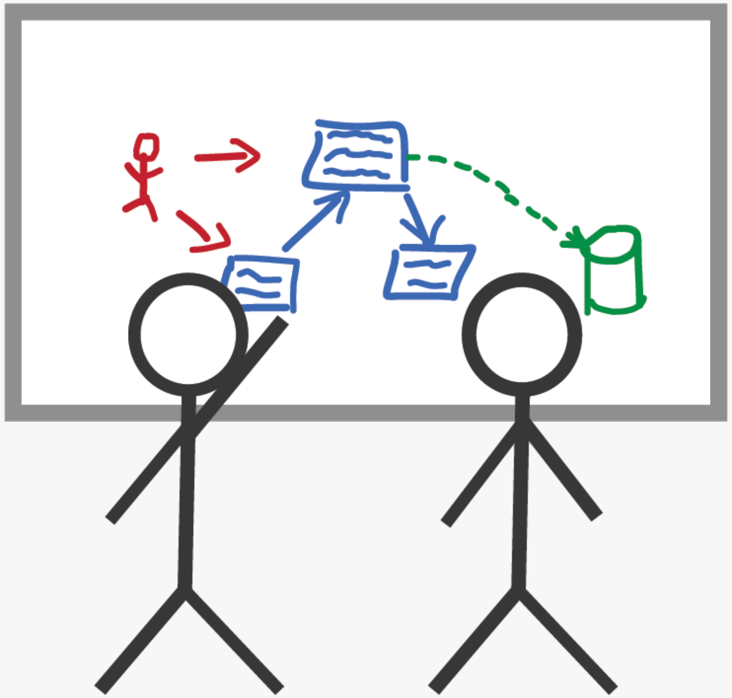
Pandemic Programming

EMSE '20

SAP HANA

Timeout Flakiness

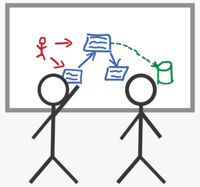
ICSE '24



Sketching

Publications:
FSE 2014, FSE 2014 Tool, ESEM 2015, VISSOFT 2017

Sketching



Research Design

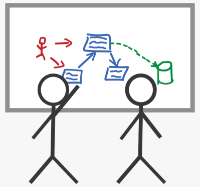
Questions:

How and **why** do software practitioners use sketches and diagrams?
How are they related to **source code**?
How can we provide better **tool support**?

Methods:

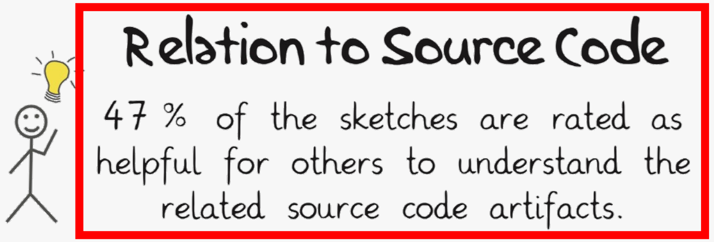
Field study, interviews, lab study, online survey, formative tool evaluations.

Sketching



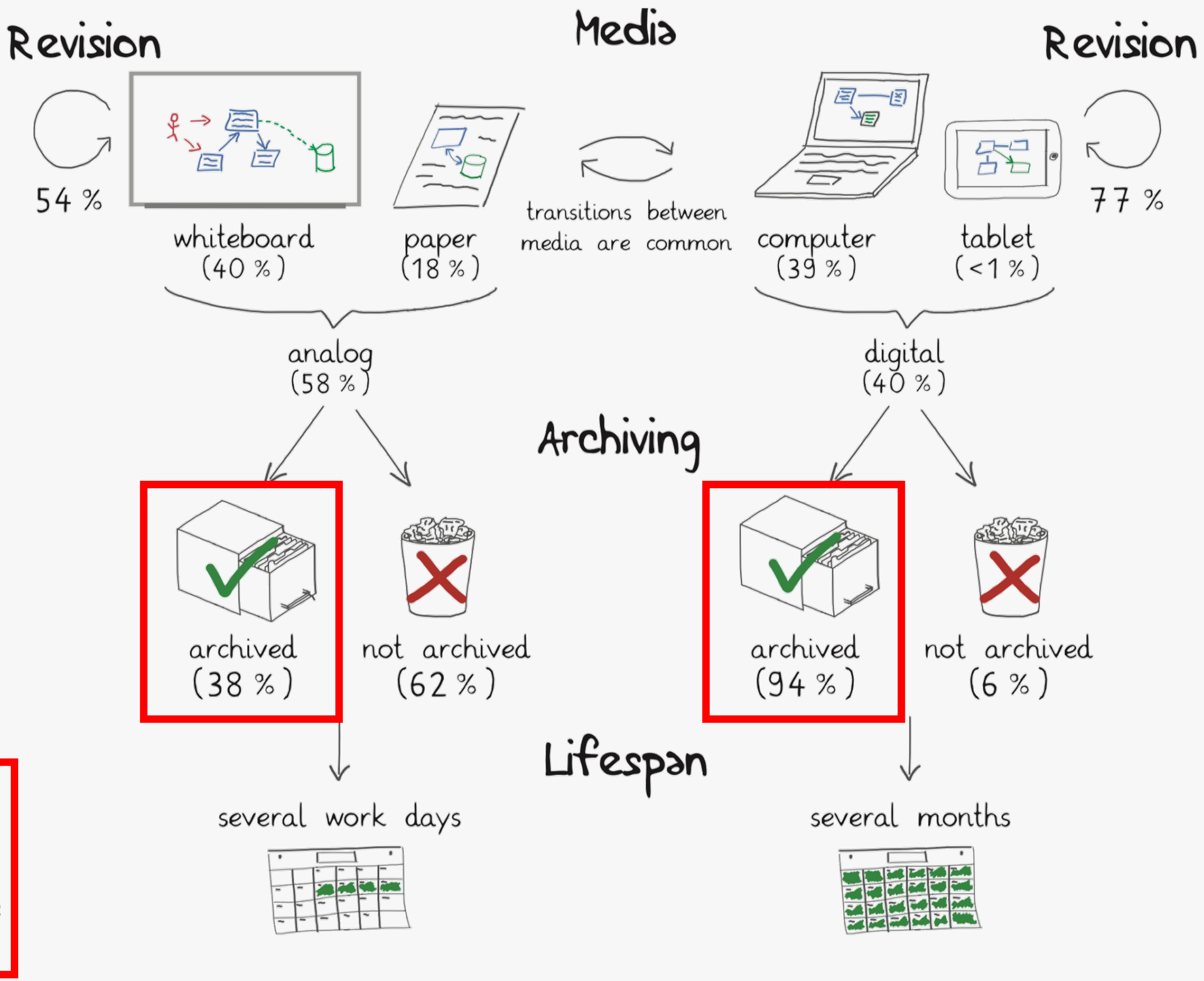
Results

Online survey with 394 participants from 32 countries, asking for last sketch/diagram created.

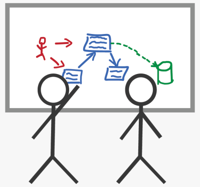


Relation to Source Code

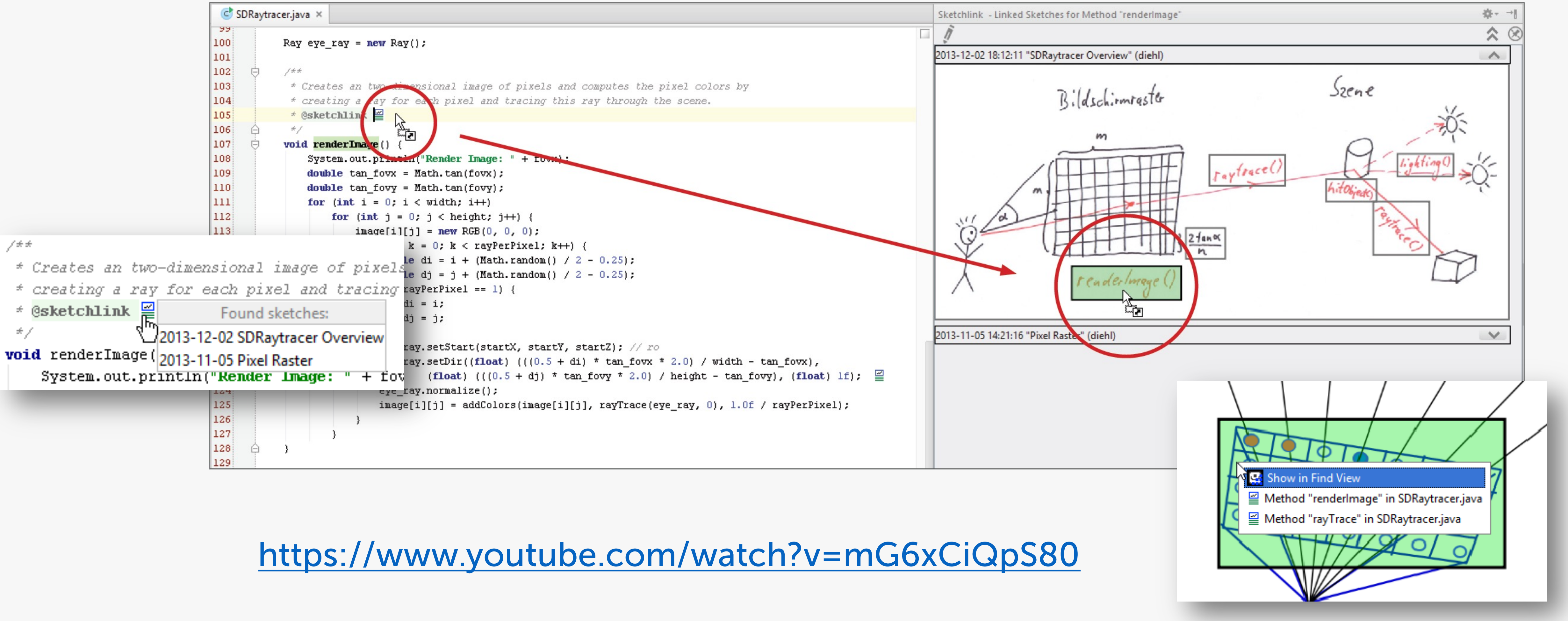
47 % of the sketches are rated as helpful for others to understand the related source code artifacts.



Sketching



SketchLink


SDRaytracer.java
Sketchlink - Linked Sketches for Method "renderImage"
2013-12-02 18:12:11 "SDRaytracer Overview" (diehl)
Bildschirmraster
Szene
m
m
α
rayTrace()
lighting()
hitObjects()
rayTrace()
2 tan α / n
renderImage()
2013-11-05 14:21:16 "Pixel Raster" (diehl)
Found sketches:
2013-12-02 SDRaytracer Overview
2013-11-05 Pixel Raster
Show in Find View
Method "renderImage" in SDRaytracer.java
Method "rayTrace" in SDRaytracer.java

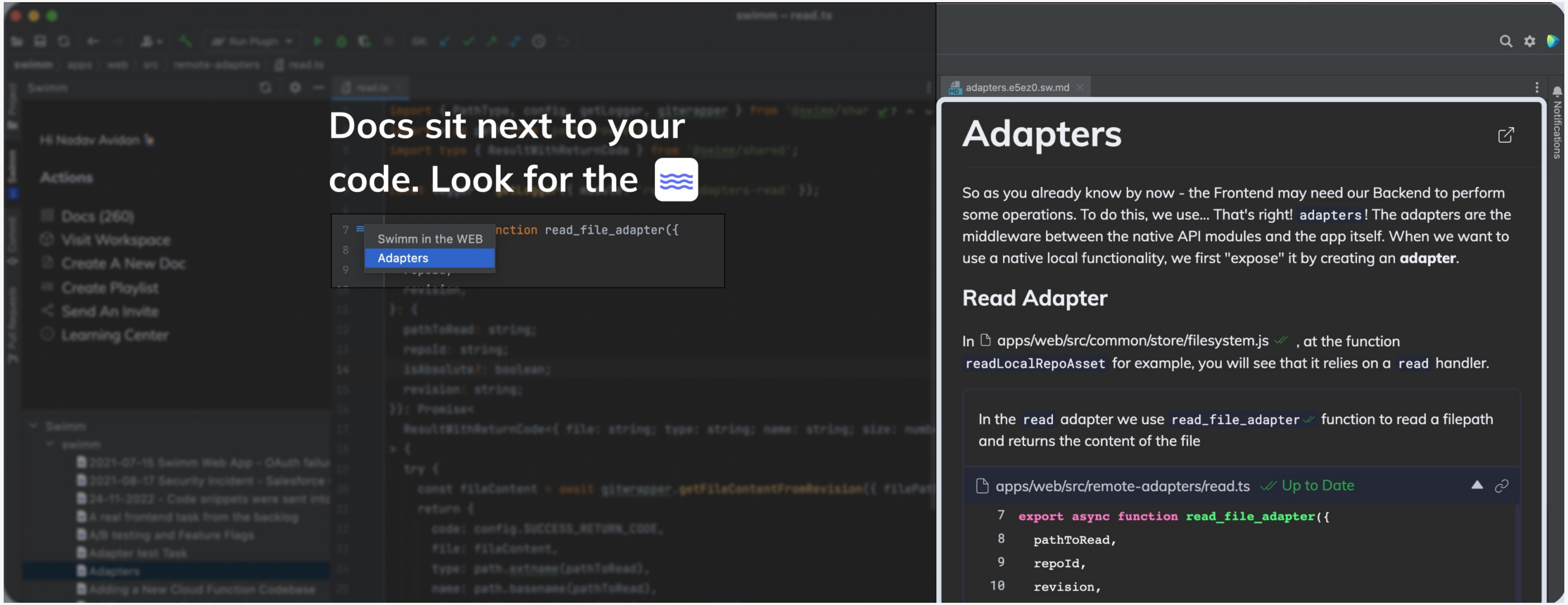
```
    Ray eye_ray = new Ray();

    /**
     * Creates an two-dimensional image of pixels and computes the pixel colors by
     * creating a ray for each pixel and tracing this ray through the scene.
     * @sketchlink
     */
    void renderImage() {
        System.out.println("Render Image: " + fovx);
        double tan_fovx = Math.tan(fovx);
        double tan_fovy = Math.tan(fovy);
        for (int i = 0; i < width; i++)
            for (int j = 0; j < height; j++) {
                image[i][j] = new RGB(0, 0, 0);
                    k = 0; k < rayPerPixel; k++) {
                    le di = i + (Math.random() / 2 - 0.25);
                    le dj = j + (Math.random() / 2 - 0.25);
                    rayPerPixel == 1) {
                    di = i;
                    dj = j;

                    ray.setStart(startX, startY, startZ); // ro
                    ray.setDir((float) (((0.5 + di) * tan_fovx * 2.0) / width - tan_fovx),
                      (float) (((0.5 + dj) * tan_fovy * 2.0) / height - tan_fovy), (float) 1f);
                    eye_ray.normalize();
                    image[i][j] = addColors(image[i][j], rayTrace(eye_ray, 0), 1.0f / rayPerPixel);
                }
            }
    }
```

```
/**
 * Creates an two-dimensional image of pixels
 * creating a ray for each pixel and tracing
 * @sketchlink
 */
void renderImage(
    System.out.println("Render Image: " + fov
```

https://www.youtube.com/watch?v=mG6xCiQpS80

A startup recently gaining traction went a step further

swimm



Docs sit next to your code. Look for the ≋

Adapters

So as you already know by now - the Frontend may need our Backend to perform some operations. To do this, we use... That's right! `adapters`! The adapters are the middleware between the native API modules and the app itself. When we want to use a native local functionality, we first "expose" it by creating an **adapter**.

Read Adapter

In apps/web/src/common/store/filesystem.js ✓ , at the function `readLocalRepoAsset` for example, you will see that it relies on a `read` handler.

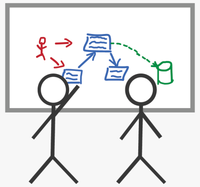
In the `read` adapter we use `read_file_adapter` ✓ function to read a filepath and returns the content of the file

apps/web/src/remote-adapters/read.ts ✓ Up to Date

```
export async function read_file_adapter({
  pathToRead,
  repoId,
  revision,
```

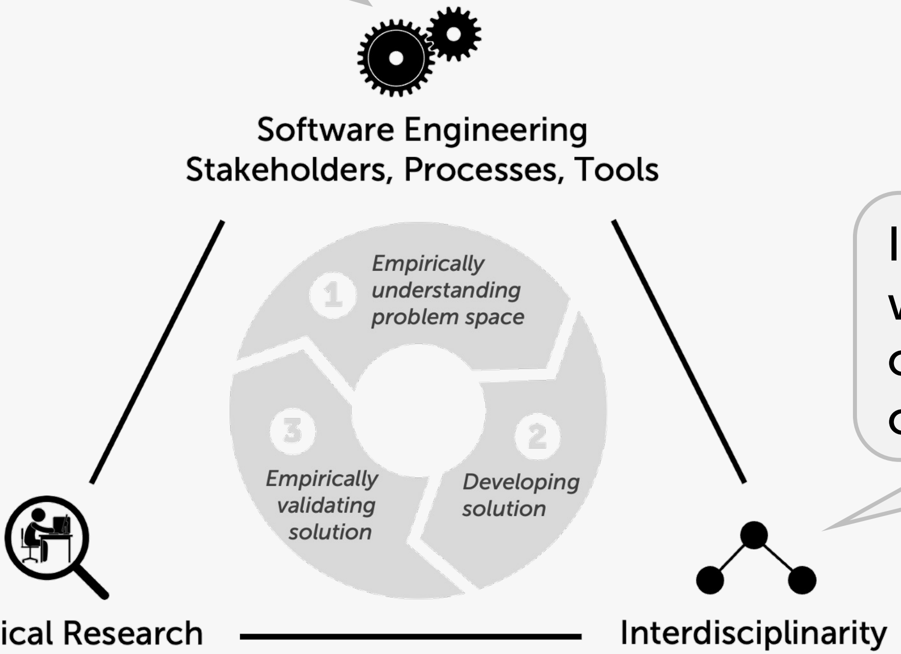
https://www.crunchbase.com/organization/swimm

Sketching



Summary

Thorough description of **state of practice**, **tool prototype** in response to identified needs.

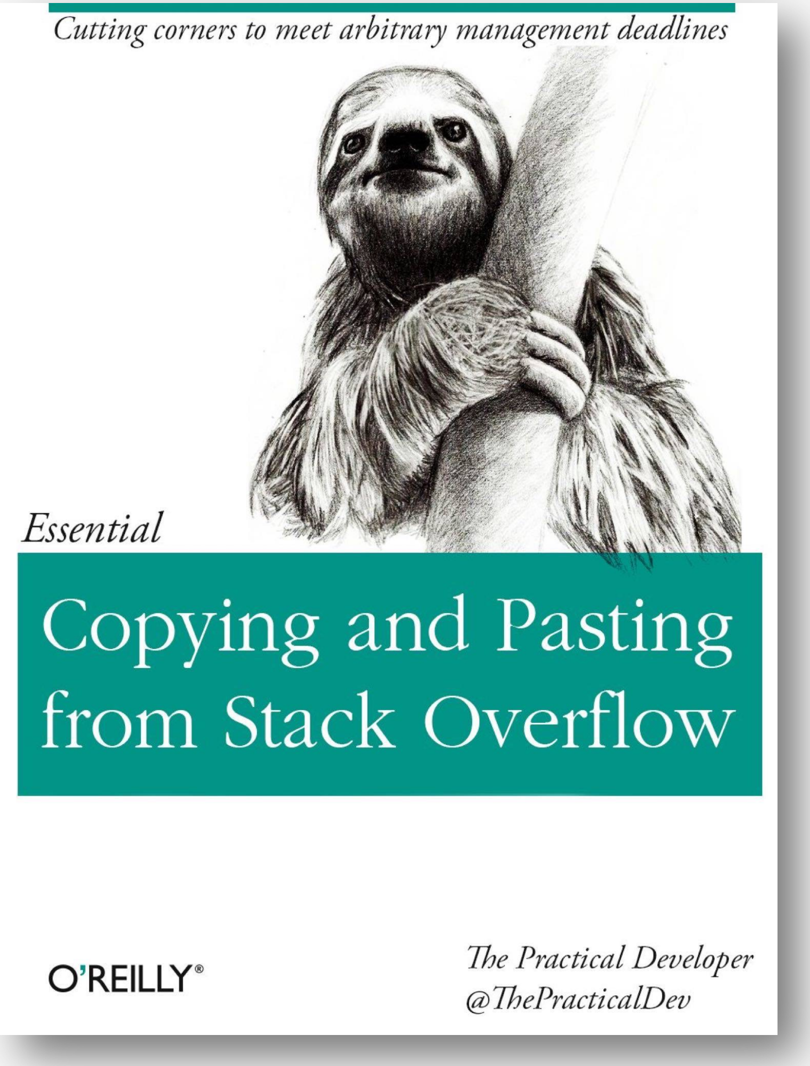

Software Engineering
Stakeholders, Processes, Tools
1
Empirically understanding problem space
2
Developing solution
3
Empirically validating solution
Empirical Research
Interdisciplinarity

Mixed-methods approach including lab and field studies, online survey, tool evaluations.

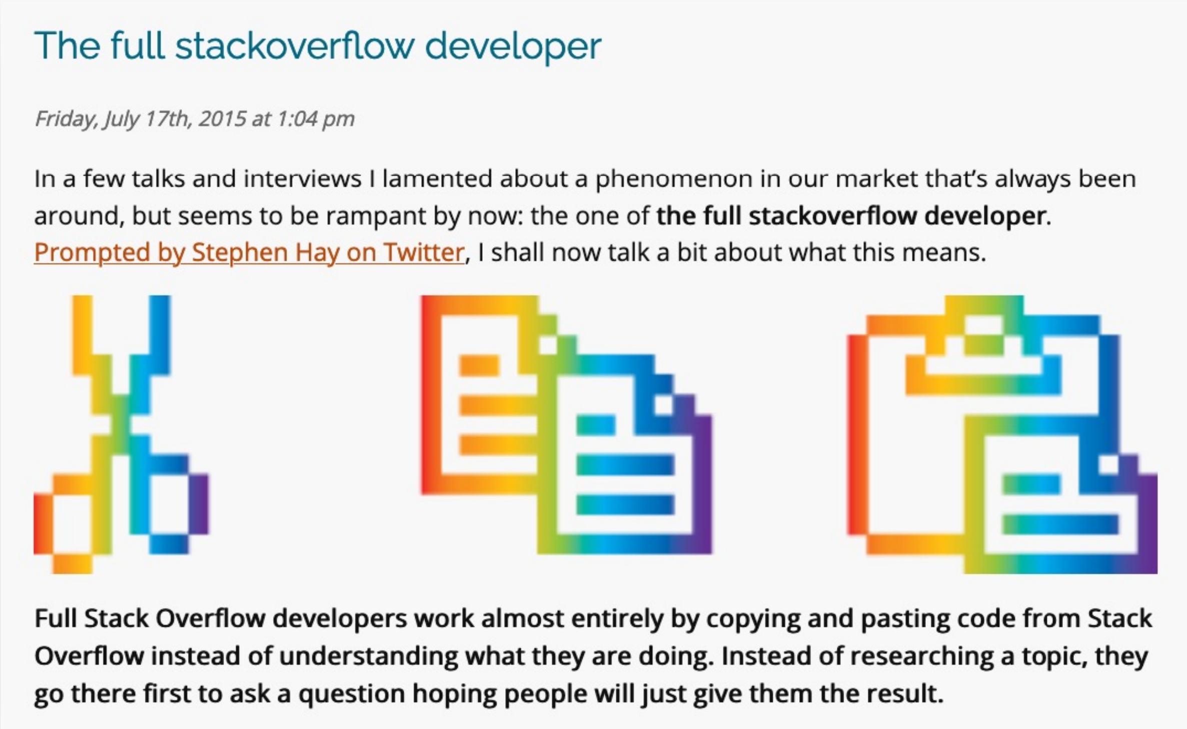
Incorporated related work from **psychology** on the positive effects of sketching.

Code Plagiarism

Publications:
EMSE 2018, MSR 2018, MSR 2019 MC, ICSE 2020 NIER


Cutting corners to meet arbitrary management deadlines
Essential
Copying and Pasting from Stack Overflow
O'REILLY®
The Practical Developer
@ThePracticalDev

https://twitter.com/ThePracticalDev/status/705825638851149824



The full stackoverflow developer

Friday, July 17th, 2015 at 1:04 pm

In a few talks and interviews I lamented about a phenomenon in our market that's always been around, but seems to be rampant by now: the one of **the full stackoverflow developer**. Prompted by Stephen Hay on Twitter, I shall now talk a bit about what this means.

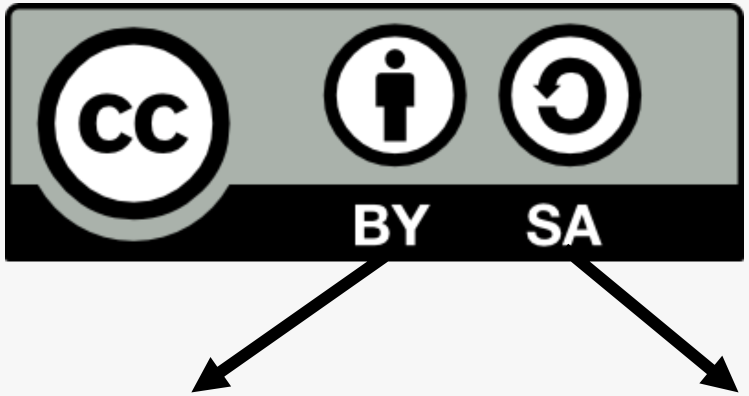
Full Stack Overflow developers work almost entirely by copying and pasting code from Stack Overflow instead of understanding what they are doing. Instead of researching a topic, they go there first to ask a question hoping people will just give them the result.

https://christianheilmann.com/2015/07/17/the-full-stackoverflow-developer/

Stack Overflow License

All content on Stack Overflow is licensed under CC BY-SA.

"You must give ***appropriate credit*** [...] and indicate if changes were made."



Attribution

Share-alike

“If you [...] ***build upon*** the material, you must ***distribute your contributions*** under the same license as the original.”

Research Design

Question:

How **frequently** is code from Stack Overflow posts used in public GitHub projects **without** the required **attribution**?

Method:

Triangulation of an estimate for the attribution ratio using three different **data mining** approaches.

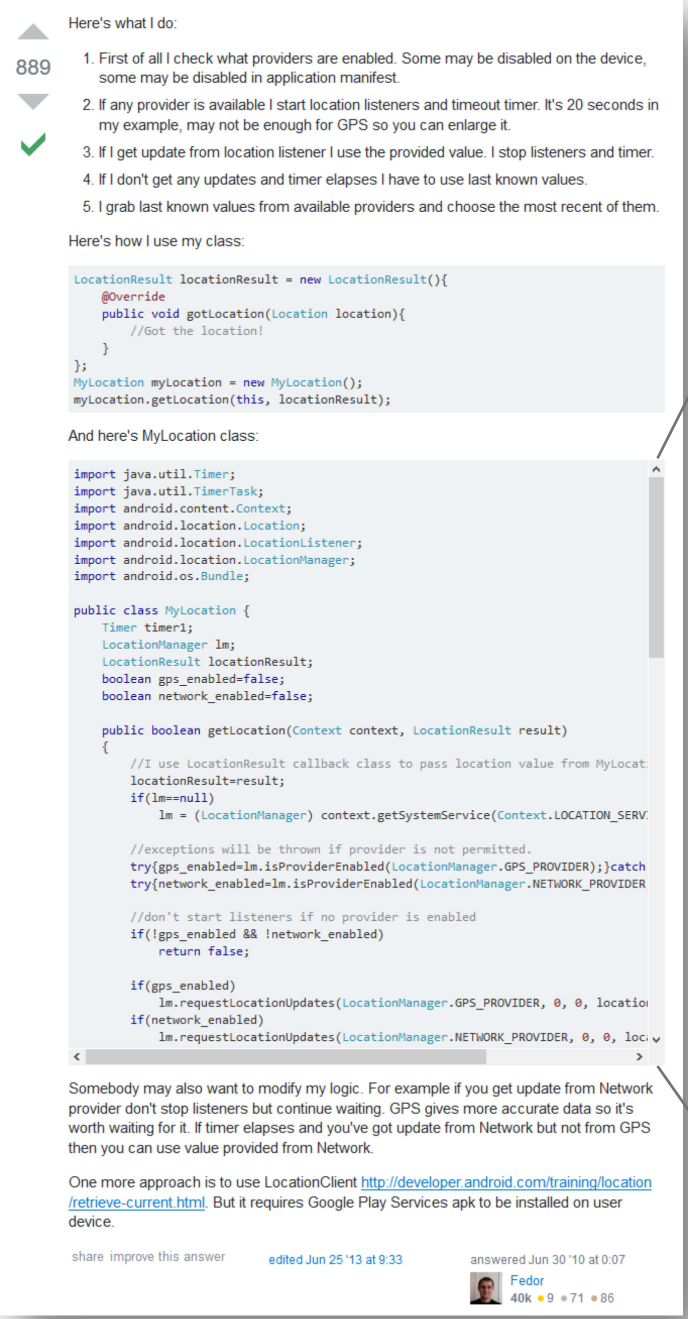
Code Plagiarism

Background

http://theconversation.com/why-universities-cant-be-expected-to-police-copyright-infringement-82677

> *"Well, but these snippets are rather trivial and not protected by copyright."*

- **Not all** snippets on Stack Overflow copyrightable, but some experts argue that the **threshold is low.** [Engelfriet 2016]
- No *"international standard for originality"*. [Creative Commons 2017b]
- CC BY-SA is a **viral copyleft license**, affecting all modifications and derived works.



889

Here's what I do:

1. First of all I check what providers are enabled. Some may be disabled on the device, some may be disabled in application manifest.
2. If any provider is available I start location listeners and timeout timer. It's 20 seconds in my example, may not be enough for GPS so you can enlarge it.
3. If I get update from location listener I use the provided value. I stop listeners and timer.
4. If I don't get any updates and timer elapses I have to use last known values.
5. I grab last known values from available providers and choose the most recent of them.

Here's how I use my class:

```
LocationResult locationResult = new LocationResult(){
    @Override
    public void gotLocation(Location location){
        //Got the location!
    }
};
MyLocation myLocation = new MyLocation();
myLocation.getLocation(this, locationResult);
```

And here's MyLocation class:

```
import java.util.Timer;
import java.util.TimerTask;
import android.content.Context;
import android.location.Location;
import android.location.LocationListener;
import android.location.LocationManager;
import android.os.Bundle;

public class MyLocation {
    Timer timer1;
    LocationManager lm;
    LocationResult locationResult;
    boolean gps_enabled=false;
    boolean network_enabled=false;

    public boolean getLocation(Context context, LocationResult result)
    {
        //I use LocationResult callback class to pass location value from MyLocat
        locationResult=result;
        if(lm==null)
            lm = (LocationManager) context.getSystemService(Context.LOCATION_SERV

        //exceptions will be thrown if provider is not permitted.
        try{gps_enabled=lm.isProviderEnabled(LocationManager.GPS_PROVIDER);}catch
        try{network_enabled=lm.isProviderEnabled(LocationManager.NETWORK_PROVIDER

        //don't start listeners if no provider is enabled
        if(!gps_enabled && !network_enabled)
            return false;

        if(gps_enabled)
            lm.requestLocationUpdates(LocationManager.GPS_PROVIDER, 0, 0, locatio
        if(network_enabled)
            lm.requestLocationUpdates(LocationManager.NETWORK_PROVIDER, 0, 0, loc
```

Somebody may also want to modify my logic. For example if you get update from Network provider don't stop listeners but continue waiting. GPS gives more accurate data so it's worth waiting for it. If timer elapses and you've got update from Network but not from GPS then you can use value provided from Network.

One more approach is to use LocationClient http://developer.android.com/training/location/retrieve-current.html. But it requires Google Play Services apk to be installed on user device.

share improve this answer

edited Jun 25 '13 at 9:33

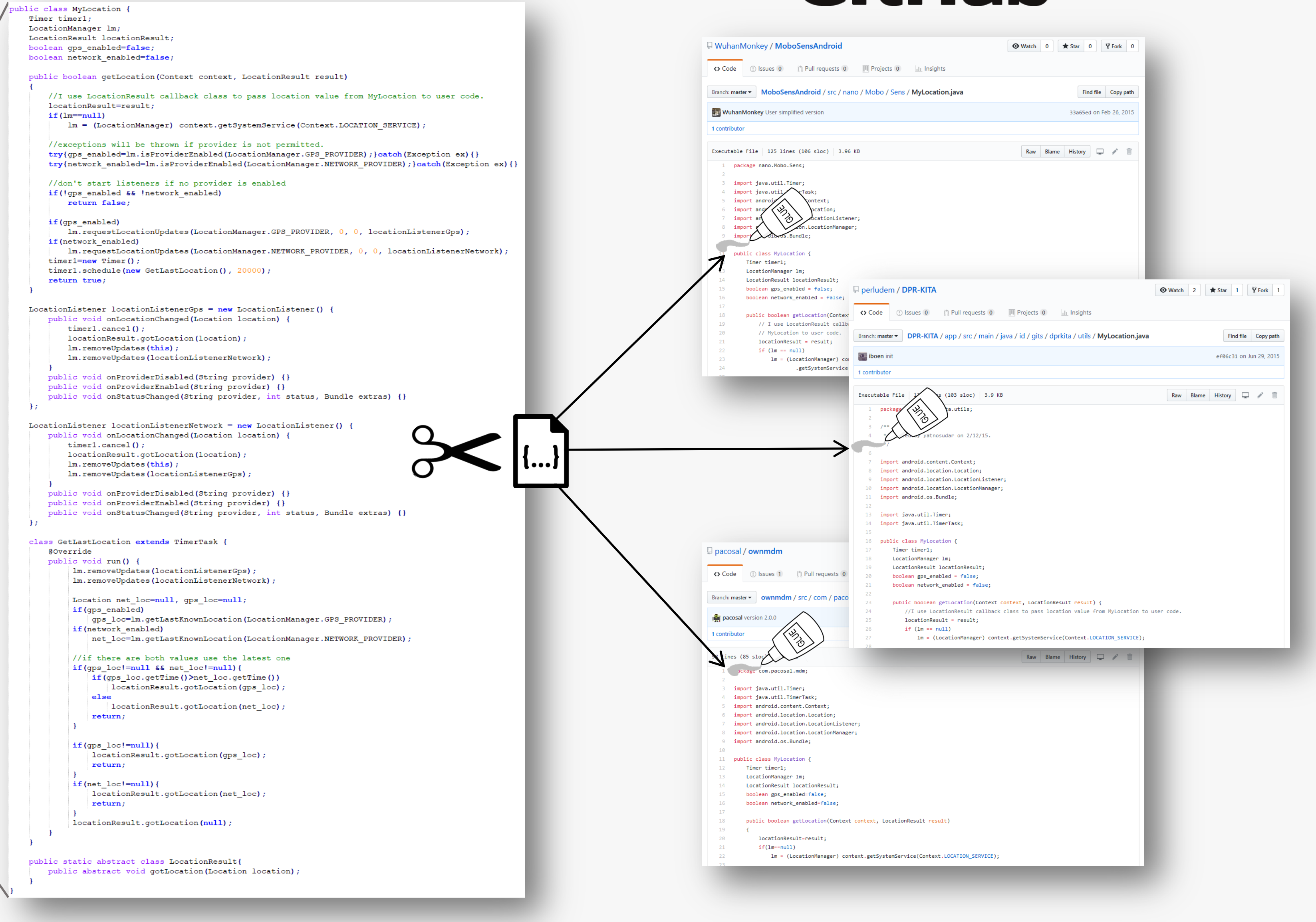
answered Jun 30 '10 at 0:07

Fedor

40k ●9 ●71 ●86

stackoverflow

GitHub



```
public class MyLocation {
    Timer timer1;
    LocationManager lm;
    LocationResult locationResult;
    boolean gps_enabled=false;
    boolean network_enabled=false;

    public boolean getLocation(Context context, LocationResult result)
    {
        //I use LocationResult callback class to pass location value from MyLocation to user code.
        locationResult=result;
        if(lm==null)
            lm = (LocationManager) context.getSystemService(Context.LOCATION_SERVICE);

        //exceptions will be thrown if provider is not permitted.
        try{gps_enabled=lm.isProviderEnabled(LocationManager.GPS_PROVIDER);}catch(Exception ex){}
        try{network_enabled=lm.isProviderEnabled(LocationManager.NETWORK_PROVIDER);}catch(Exception ex){}

        //don't start listeners if no provider is enabled
        if(!gps_enabled && !network_enabled)
            return false;

        if(gps_enabled)
            lm.requestLocationUpdates(LocationManager.GPS_PROVIDER, 0, 0, locationListenerGps);
        if(network_enabled)
            lm.requestLocationUpdates(LocationManager.NETWORK_PROVIDER, 0, 0, locationListenerNetwork);
        timer1=new Timer();
        timer1.schedule(new GetLastLocation(), 20000);
        return true;
    }

    LocationListener locationListenerGps = new LocationListener() {
        public void onLocationChanged(Location location) {
            timer1.cancel();
            locationResult.gotLocation(location);
            lm.removeUpdates(this);
            lm.removeUpdates(locationListenerNetwork);
        }
        public void onProviderDisabled(String provider) {}
        public void onProviderEnabled(String provider) {}
        public void onStatusChanged(String provider, int status, Bundle extras) {}
    };

    LocationListener locationListenerNetwork = new LocationListener() {
        public void onLocationChanged(Location location) {
            timer1.cancel();
            locationResult.gotLocation(location);
            lm.removeUpdates(this);
            lm.removeUpdates(locationListenerGps);
        }
        public void onProviderDisabled(String provider) {}
        public void onProviderEnabled(String provider) {}
        public void onStatusChanged(String provider, int status, Bundle extras) {}
    };

    class GetLastLocation extends TimerTask {
        @Override
        public void run() {
             lm.removeUpdates(locationListenerGps);
             lm.removeUpdates(locationListenerNetwork);

             Location net_loc=null, gps_loc=null;
             if(gps_enabled)
                 gps_loc=lm.getLastKnownLocation(LocationManager.GPS_PROVIDER);
             if(network_enabled)
                 net_loc=lm.getLastKnownLocation(LocationManager.NETWORK_PROVIDER);

             //if there are both values use the latest one
             if(gps_loc!=null && net_loc!=null){
                 if(gps_loc.getTime()>net_loc.getTime())
                     locationResult.gotLocation(gps_loc);
                 else
                     locationResult.gotLocation(net_loc);
                 return;
             }

             if(gps_loc!=null){
                 locationResult.gotLocation(gps_loc);
                 return;
             }
             if(net_loc!=null){
                 locationResult.gotLocation(net_loc);
                 return;
             }
             locationResult.gotLocation(null);
        }
    }

    public static abstract class LocationResult{
        public abstract void gotLocation(Location location);
    }
}
```

WuhanMonkey / MoboSensAndroid

MoboSensAndroid / src / nano / Mobo / Sens / MyLocation.java

WuhanMonkey User simplified version — 33a65ed on Feb 26, 2015

1 contributor

Executable File | 125 lines (106 sloc) | 3.96 KB

```
package nano.Mobo.Sens;

import java.util.Timer;
import java.util.  rTask;
import androi       Context;
import and          ocation;
import an           LocationListener;
import              on.LocationManager;
import         oid.os.Bundle;

public class MyLocation {
    Timer timer1;
    LocationManager lm;
    LocationResult locationResult;
    boolean gps_enabled = false;
    boolean network_enabled = false;

    public boolean getLocation(Context
        // I use LocationResult callba
        // MyLocation to user code.
        locationResult = result;
        if (lm == null)
            lm = (LocationManager) co
                    .getSystemService
```

perludem / DPR-KITA

DPR-KITA / app / src / main / java / id / gits / dprkita / utils / MyLocation.java

iboen init — ef06c31 on Jun 29, 2015

1 contributor

Executable File | 12 s (103 sloc) | 3.9 KB

```
package              ta.utils;

/**
 *     ed by yatnosudar on 2/12/15.
 */

import android.content.Context;
import android.location.Location;
import android.location.LocationListener;
import android.location.LocationManager;
import android.os.Bundle;

import java.util.Timer;
import java.util.TimerTask;

public class MyLocation {
    Timer timer1;
    LocationManager lm;
    LocationResult locationResult;
    boolean gps_enabled = false;
    boolean network_enabled = false;

    public boolean getLocation(Context context, LocationResult result) {
        //I use LocationResult callback class to pass location value from MyLocation to user code.
        locationResult = result;
        if (lm == null)
            lm = (LocationManager) context.getSystemService(Context.LOCATION_SERVICE);
```

pacosal / ownmdm

ownmdm / src / com / paco

pacosal version 2.0.0

1 contributor

 ines (85 sloc)

```
 ackage com.pacosal.mdm;

import java.util.Timer;
import java.util.TimerTask;
import android.content.Context;
import android.location.Location;
import android.location.LocationListener;
import android.location.LocationManager;
import android.os.Bundle;

public class MyLocation {
    Timer timer1;
    LocationManager lm;
    LocationResult locationResult;
    boolean gps_enabled=false;
    boolean network_enabled=false;

    public boolean getLocation(Context context, LocationResult result)
    {
        locationResult=result;
        if(lm==null)
            lm = (LocationManager) context.getSystemService(Context.LOCATION_SERVICE);
```

https://github.com/pacosal/ownmdm/blob/master/src/com/pacosal/mdm/MyLocation.java

https://stackoverflow.com/a/3145655

Triangulated Attribution Ratio

Question: **How frequently** is code from Stack Overflow posts **used** in public GitHub projects **without the required attribution**?

$$\left.\begin{array}{l}\textbf{1. Exploratory study}\\ \textbf{2. Code clone detector study}\\ \textbf{3. Exact matches study}\end{array}\right\}\ \bar{r}_{\mathrm{attr}}$$

We used **popularity** and **length** of the snippets as a **proxy for originality** and checked **external availability**.

Attribution

Attribution ratio:

- Method 1 (regular expressions): $\bar{r}_{\text{attr}} = 23\%$
- Method 2 (code clone detector): $\bar{r}_{\text{attr}} = 24\%$
- Method 3 (exact matches): $\bar{r}_{\text{attr}} = 8\%$

Conservative estimate: $\boxed{\bar{r}_{\text{attr}} \leq 25\%}$

Share-alike

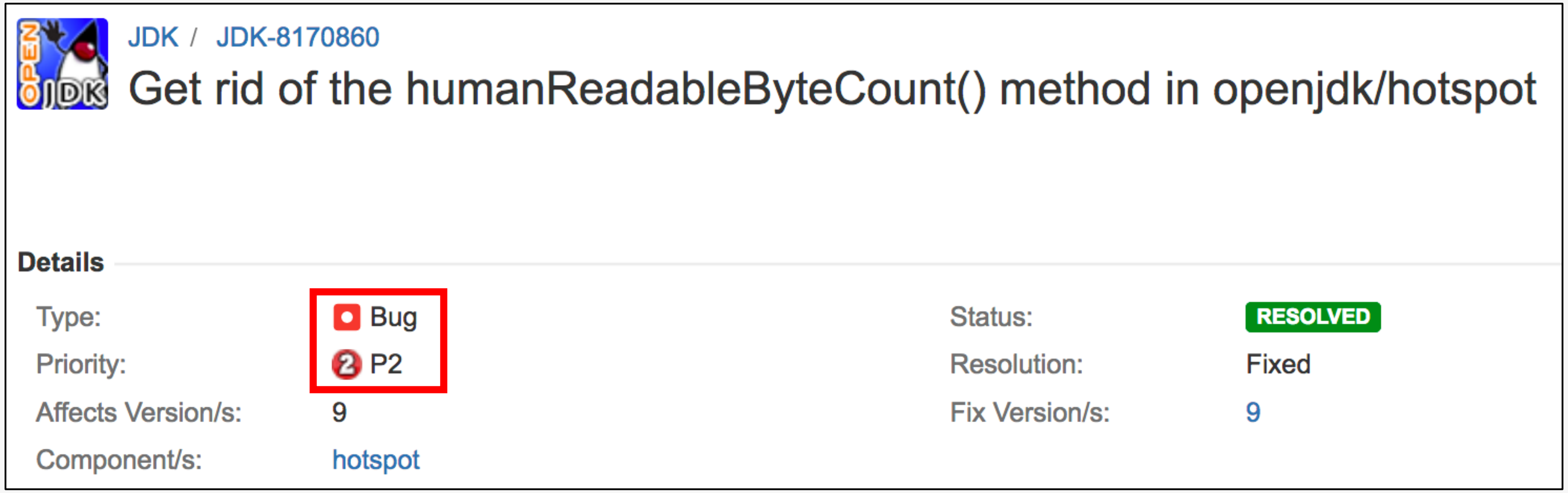
SA

Only **2%** of all analyzed repositories (methods 1-3) containing code from Stack Overflow **attributed** its source and used a **compatible license.**

Reaching out to Developers

- **Contacted owners** of GitHub repositories containing copies of Stack Overflow snippets.
- **75% not aware** of CC BY-SA licensing.
- Many thankful responses.

Stack Overflow Code in the OpenJDK

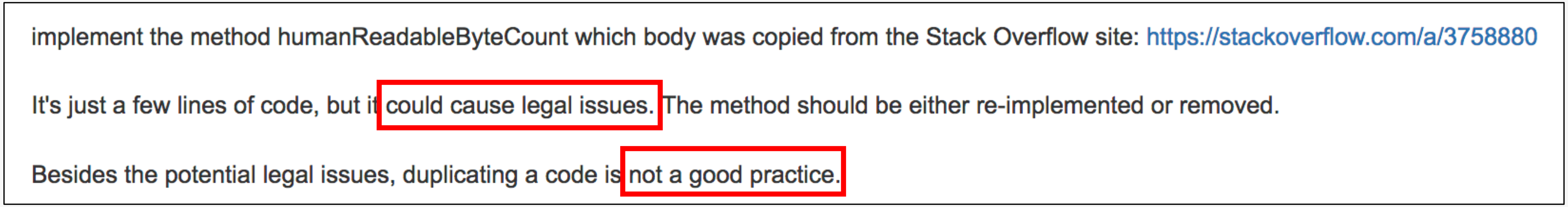


JDK / JDK-8170860

Get rid of the humanReadableByteCount() method in openjdk/hotspot

Details

Type:	Bug	Status:	RESOLVED
Priority:	P2	Resolution:	Fixed
Affects Version/s:	9	Fix Version/s:	9
Component/s:	hotspot		



implement the method humanReadableByteCount which body was copied from the Stack Overflow site: https://stackoverflow.com/a/3758880

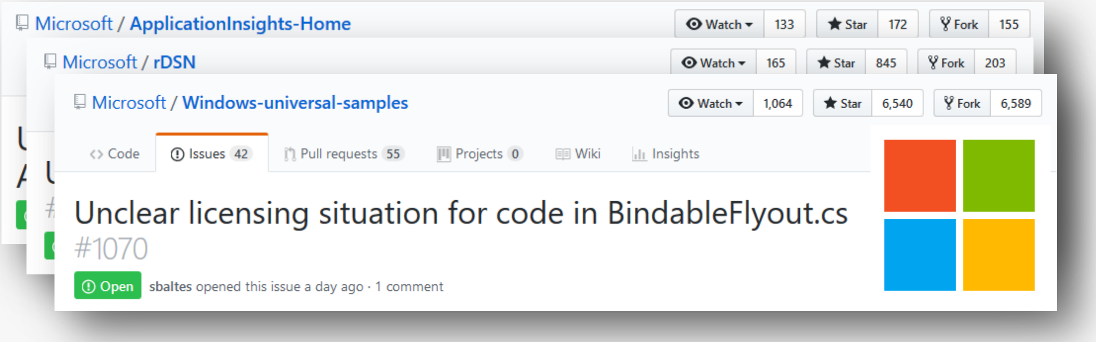
It's just a few lines of code, but it could cause legal issues. The method should be either re-implemented or removed.

Besides the potential legal issues, duplicating a code is not a good practice.

https://bugs.openjdk.java.net/browse/JDK-8170860

Code Plagiarism

Reaching out to Developers


Microsoft / ApplicationInsights-Home — Watch 133 · Star 172 · Fork 155

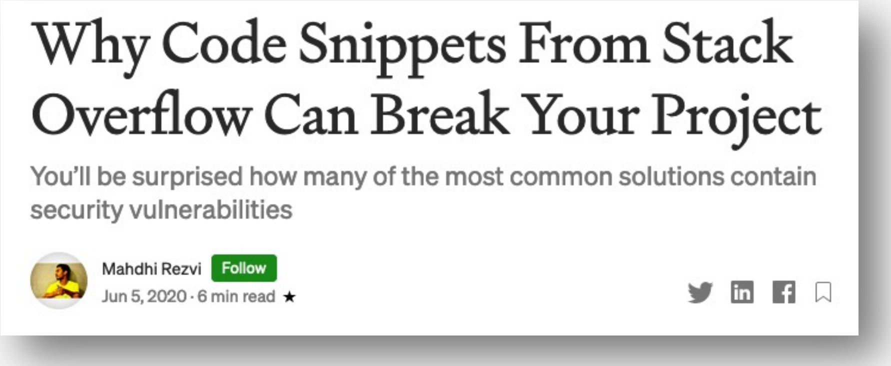
Microsoft / rDSN — Watch 165 · Star 845 · Fork 203

Microsoft / Windows-universal-samples — Watch 1,064 · Star 6,540 · Fork 6,589

Code · Issues 42 · Pull requests 55 · Projects 0 · Wiki · Insights

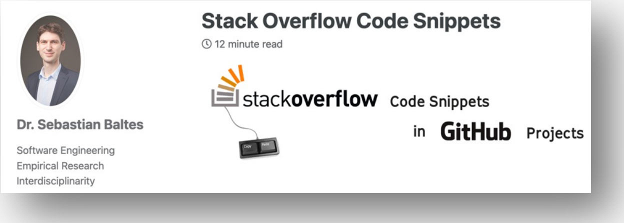
Unclear licensing situation for code in BindableFlyout.cs #1070

Open · sbaltes opened this issue a day ago · 1 comment


Why Code Snippets From Stack Overflow Can Break Your Project

You'll be surprised how many of the most common solutions contain security vulnerabilities

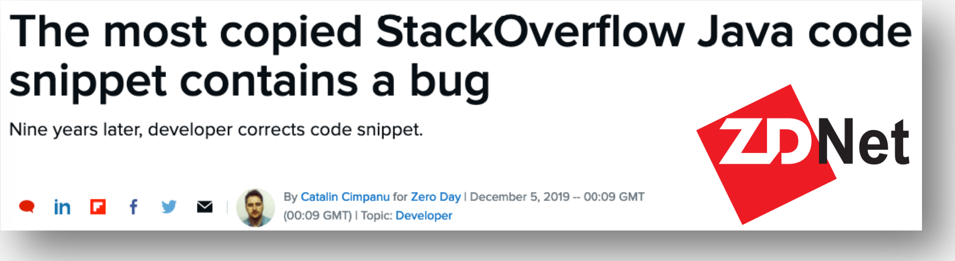
Mahdhi Rezvi Follow
Jun 5, 2020 · 6 min read ★


Dr. Sebastian Baltes

Software Engineering
Empirical Research
Interdisciplinarity

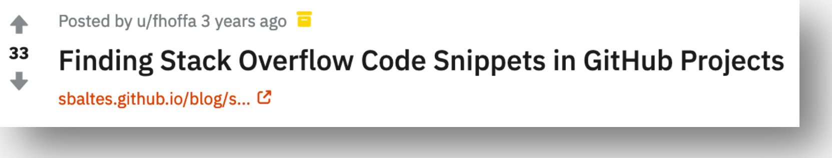
Stack Overflow Code Snippets

12 minute read


The most copied StackOverflow Java code snippet contains a bug

Nine years later, developer corrects code snippet.

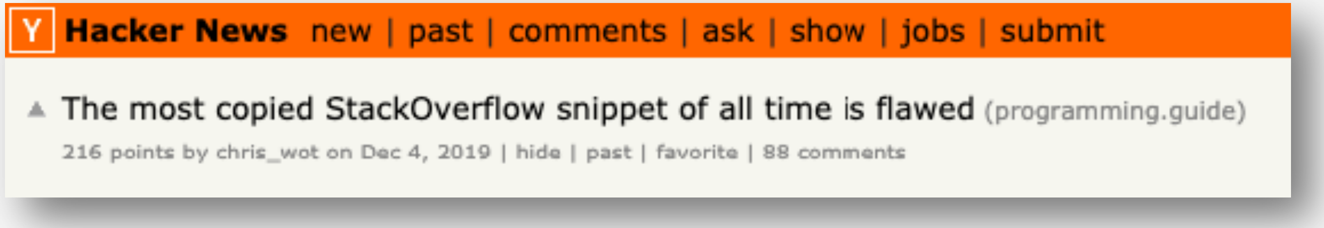
By Catalin Cimpanu for Zero Day | December 5, 2019 -- 00:09 GMT (00:09 GMT) | Topic: Developer


33

Posted by u/fhoffa 3 years ago

Finding Stack Overflow Code Snippets in GitHub Projects

sbaltes.github.io/blog/s...


Y Hacker News new | past | comments | ask | show | jobs | submit

▲ The most copied StackOverflow snippet of all time is flawed (programming.guide)

216 points by chris_wot on Dec 4, 2019 | hide | past | favorite | 88 comments

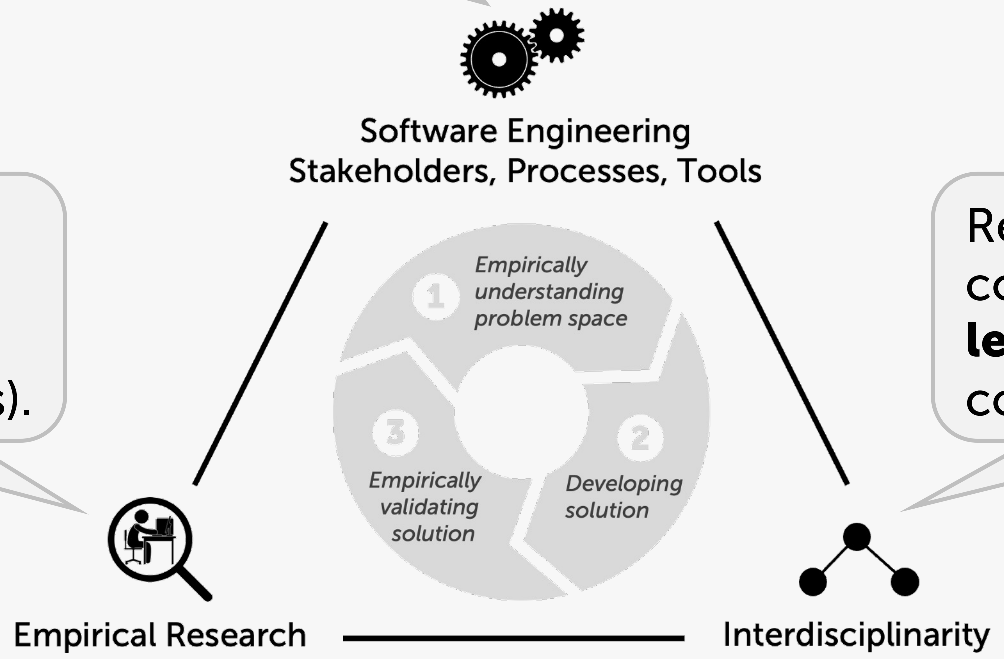
Code Plagiarism

Summary

Quantification of code plagiarism in open-source projects, **feasibility of detection**, outreach.

Triangulation using three data mining approaches, online survey, (qualit. analysis).

Research on worldwide copyright and licensing **legislation**, exemplary court cases.


Software Engineering
Stakeholders, Processes, Tools
1
Empirically understanding problem space
2
Developing solution
3
Empirically validating solution
Empirical Research
Interdisciplinarity

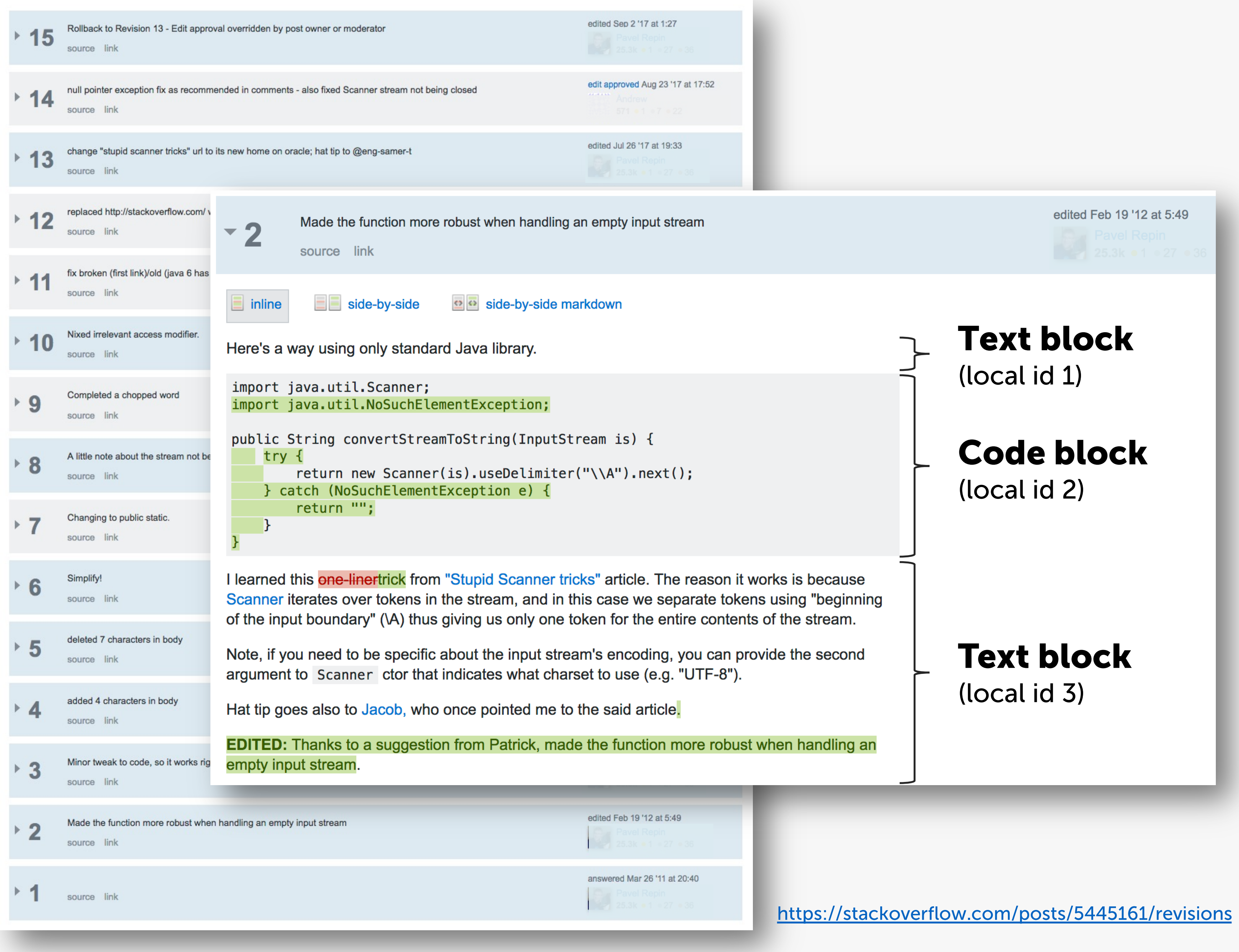
Code Plagiarism

Industry Relevance

- Initial idea: **building a tool** integrated into CI/CD.
- Nowadays: **Snippet scanning** is supported by commercial tools, but usually deactivated due to false positives.
- Especially since the US Executive Order 14028 in 2021, companies have invested in automatically creating **Software Bill of Materials** (SBOMs).
- However, SBOMs still only cover **reuse on the level of libraries** and frameworks.

One Project Became Two

Evolution



15 Rollback to Revision 13 - Edit approval overridden by post owner or moderator
source link
edited Sep 2 '17 at 1:27
Pavel Repin 25.3k 1 27 36

14 null pointer exception fix as recommended in comments - also fixed Scanner stream not being closed
source link
edit approved Aug 23 '17 at 17:52
Andrew 571 1 7 22

13 change "stupid scanner tricks" url to its new home on oracle; hat tip to @eng-samer-t
source link
edited Jul 26 '17 at 19:33
Pavel Repin 25.3k 1 27 36

12 replaced http://stackoverflow.com/ v
source link

11 fix broken (first link)/old (java 6 has
source link

10 Nixed irrelevant access modifier.
source link

9 Completed a chopped word
source link

8 A little note about the stream not be
source link

7 Changing to public static.
source link

6 Simplify!
source link

5 deleted 7 characters in body
source link

4 added 4 characters in body
source link

3 Minor tweak to code, so it works rig
source link

2 Made the function more robust when handling an empty input stream
source link
edited Feb 19 '12 at 5:49

1
source link
answered Mar 26 '11 at 20:40

2 Made the function more robust when handling an empty input stream
source link
edited Feb 19 '12 at 5:49
Pavel Repin 25.3k 1 27 36

Here's a way using only standard Java library.

```
import java.util.Scanner;
import java.util.NoSuchElementException;

public String convertStreamToString(InputStream is) {
    try {
        return new Scanner(is).useDelimiter("\\A").next();
    } catch (NoSuchElementException e) {
        return "";
    }
}
```

I learned this ~~one-liner~~trick from "Stupid Scanner tricks" article. The reason it works is because Scanner iterates over tokens in the stream, and in this case we separate tokens using "beginning of the input boundary" (\A) thus giving us only one token for the entire contents of the stream.

Note, if you need to be specific about the input stream's encoding, you can provide the second argument to `Scanner` ctor that indicates what charset to use (e.g. "UTF-8").

Hat tip goes also to Jacob, who once pointed me to the said article.

EDITED: Thanks to a suggestion from Patrick, made the function more robust when handling an empty input stream.

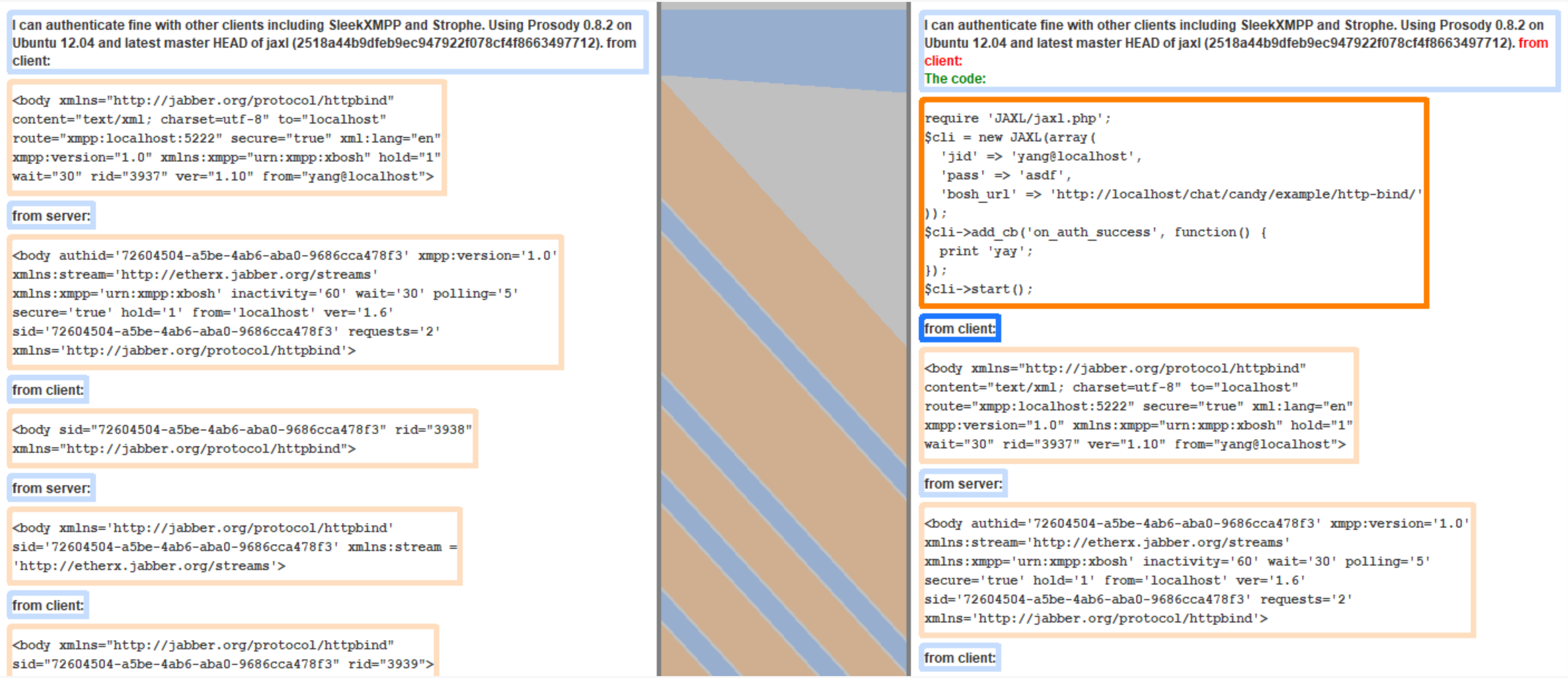
Text block
(local id 1)

Code block
(local id 2)

Text block
(local id 3)

https://stackoverflow.com/posts/5445161/revisions

Post Block Matching: Example



I can authenticate fine with other clients including SleekXMPP and Strophe. Using Prosody 0.8.2 on Ubuntu 12.04 and latest master HEAD of jaxl (2518a44b9dfeb9ec947922f078cf4f8663497712). from client:

```
<body xmlns="http://jabber.org/protocol/httpbind"
content="text/xml; charset=utf-8" to="localhost"
route="xmpp:localhost:5222" secure="true" xml:lang="en"
xmpp:version="1.0" xmlns:xmpp="urn:xmpp:xbosh" hold="1"
wait="30" rid="3937" ver="1.10" from="yang@localhost">
```

from server:

```
<body authid='72604504-a5be-4ab6-aba0-9686cca478f3' xmpp:version='1.0'
xmlns:stream='http://etherx.jabber.org/streams'
xmlns:xmpp='urn:xmpp:xbosh' inactivity='60' wait='30' polling='5'
secure='true' hold='1' from='localhost' ver='1.6'
sid='72604504-a5be-4ab6-aba0-9686cca478f3' requests='2'
xmlns='http://jabber.org/protocol/httpbind'>
```

from client:

```
<body sid="72604504-a5be-4ab6-aba0-9686cca478f3" rid="3938"
xmlns="http://jabber.org/protocol/httpbind">
```

from server:

```
<body xmlns='http://jabber.org/protocol/httpbind'
sid='72604504-a5be-4ab6-aba0-9686cca478f3' xmlns:stream =
'http://etherx.jabber.org/streams'>
```

from client:

```
<body xmlns="http://jabber.org/protocol/httpbind"
sid="72604504-a5be-4ab6-aba0-9686cca478f3" rid="3939">
```

I can authenticate fine with other clients including SleekXMPP and Strophe. Using Prosody 0.8.2 on Ubuntu 12.04 and latest master HEAD of jaxl (2518a44b9dfeb9ec947922f078cf4f8663497712). from client:
The code:

```
require 'JAXL/jaxl.php';
$cli = new JAXL(array(
  'jid' => 'yang@localhost',
  'pass' => 'asdf',
  'bosh_url' => 'http://localhost/chat/candy/example/http-bind/'
));
$cli->add_cb('on_auth_success', function() {
  print 'yay';
});
$cli->start();
```

from client:

```
<body xmlns="http://jabber.org/protocol/httpbind"
content="text/xml; charset=utf-8" to="localhost"
route="xmpp:localhost:5222" secure="true" xml:lang="en"
xmpp:version="1.0" xmlns:xmpp="urn:xmpp:xbosh" hold="1"
wait="30" rid="3937" ver="1.10" from="yang@localhost">
```

from server:

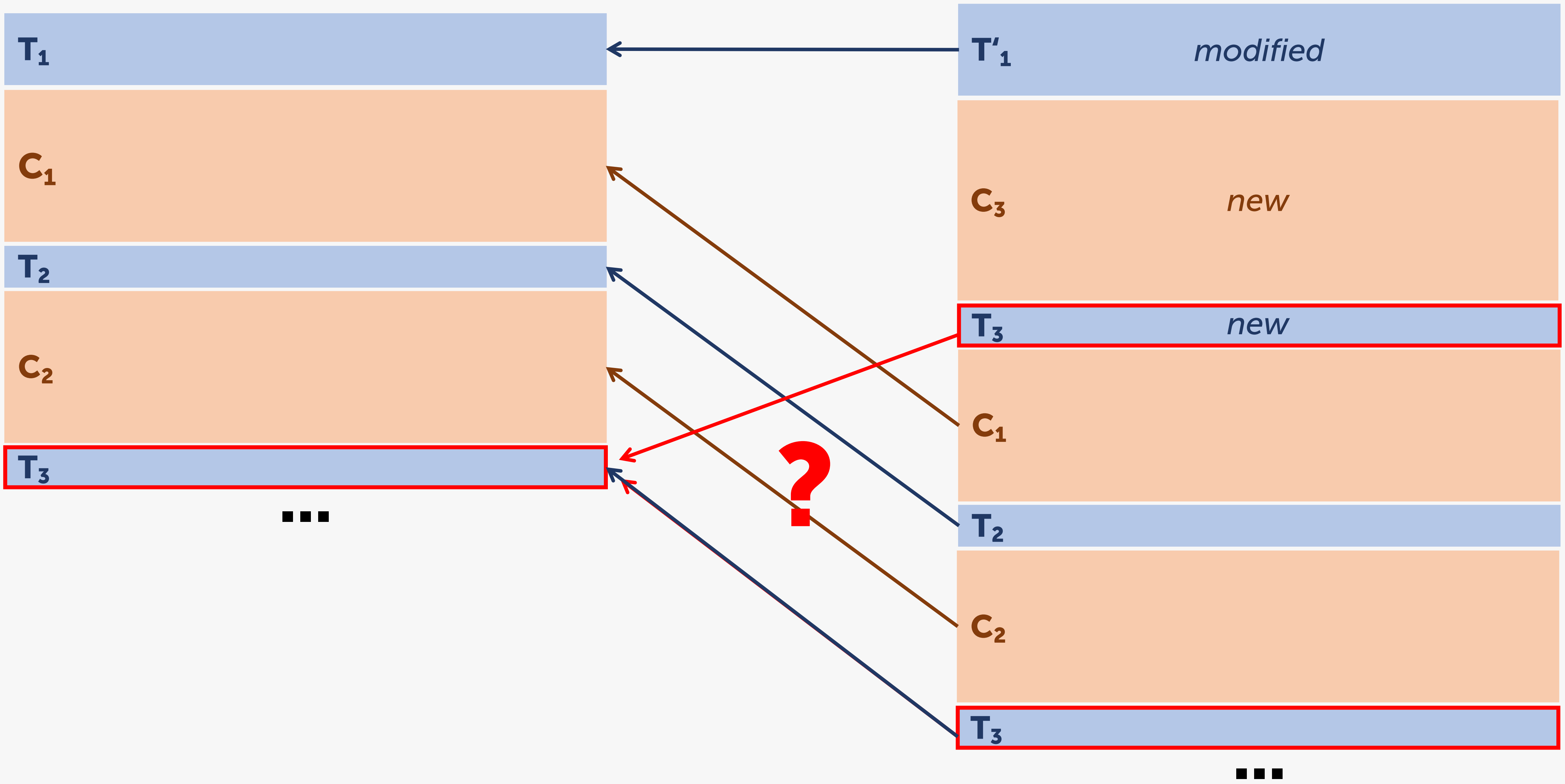
```
<body authid='72604504-a5be-4ab6-aba0-9686cca478f3' xmpp:version='1.0'
xmlns:stream='http://etherx.jabber.org/streams'
xmlns:xmpp='urn:xmpp:xbosh' inactivity='60' wait='30' polling='5'
secure='true' hold='1' from='localhost' ver='1.6'
sid='72604504-a5be-4ab6-aba0-9686cca478f3' requests='2'
xmlns='http://jabber.org/protocol/httpbind'>
```

from client:

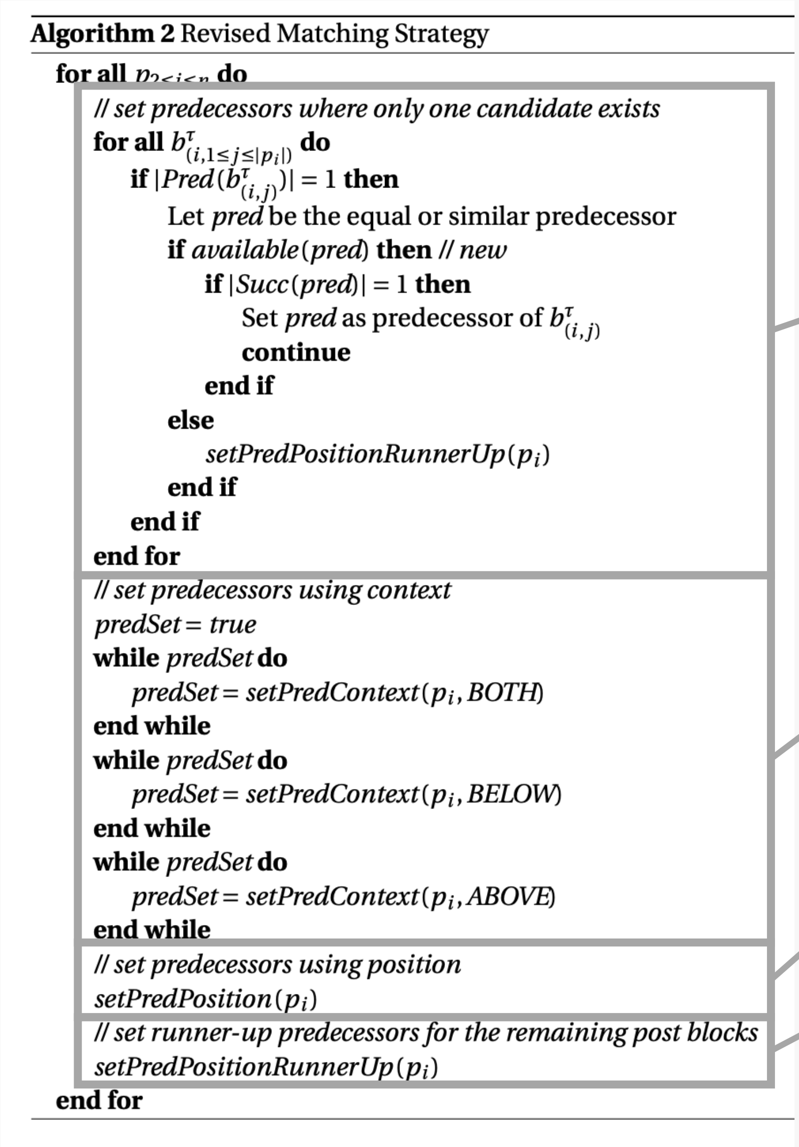
■■■

https://stackoverflow.com/q/13064858

Post Block Matching: Example



Post Block Matching: Predecessor Selection Strategy



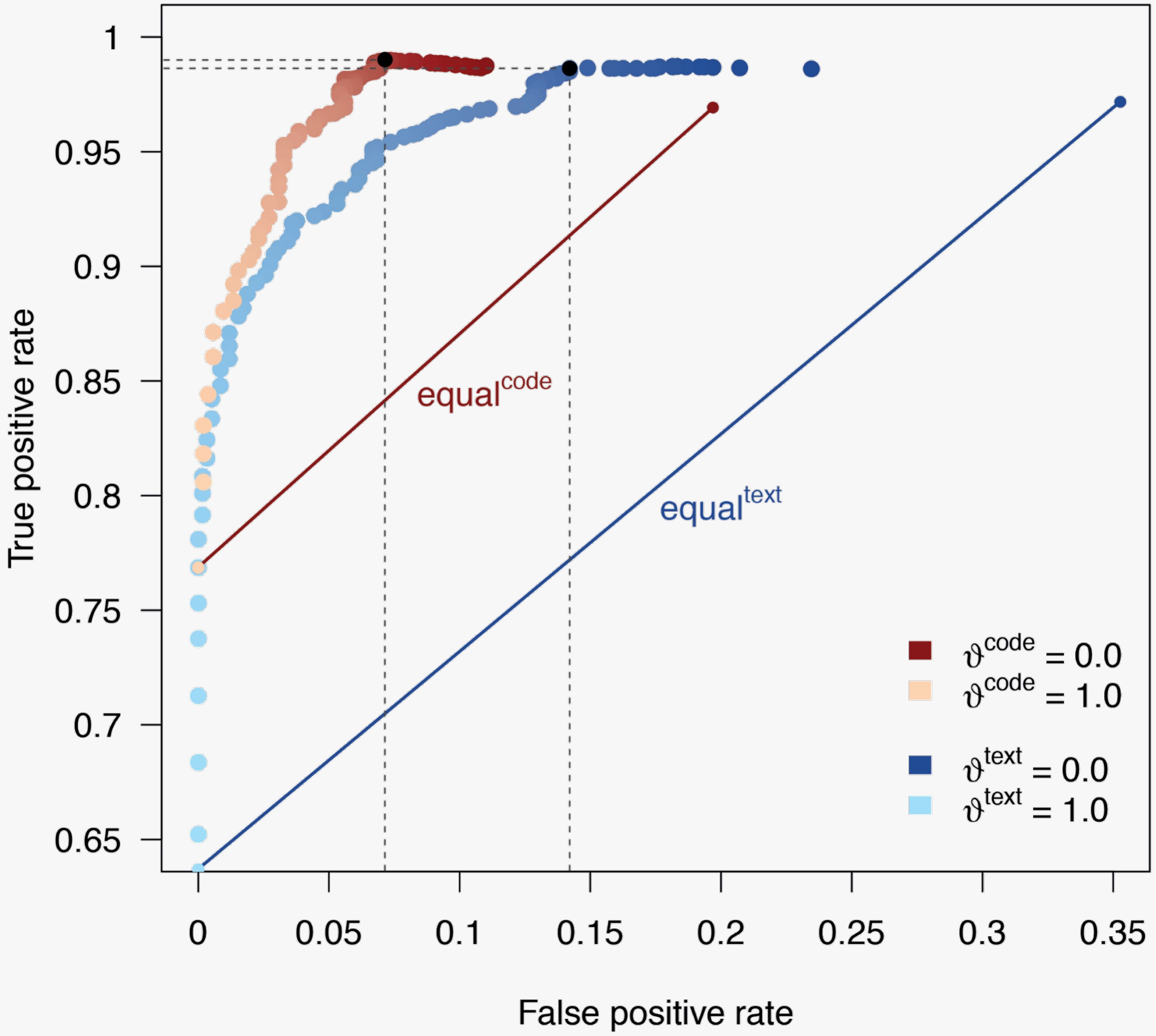
Algorithm 2 Revised Matching Strategy

```
for all $p_{2 \le i \le n}$ do
    // set predecessors where only one candidate exists
    for all $b^{\tau}_{(i,1 \le j \le |p_i|)}$ do
        if $|Pred(b^{\tau}_{(i,j)})| = 1$ then
            Let pred be the equal or similar predecessor
            if available(pred) then // new
                if $|Succ(pred)| = 1$ then
                    Set pred as predecessor of $b^{\tau}_{(i,j)}$
                    continue
                end if
            else
                $setPredPositionRunnerUp(p_i)$
            end if
        end if
    end for
    // set predecessors using context
    predSet = true
    while predSet do
        $predSet = setPredContext(p_i, BOTH)$
    end while
    while predSet do
        $predSet = setPredContext(p_i, BELOW)$
    end while
    while predSet do
        $predSet = setPredContext(p_i, ABOVE)$
    end while
    // set predecessors using position
    $setPredPosition(p_i)$
    // set runner-up predecessors for the remaining post blocks
    $setPredPositionRunnerUp(p_i)$
end for
```

- If post block has **only one possible successor** and this possible successor has **only one possible predecessor**, no strategy required
- Otherwise, consider the **context**
- Then, try to set remaining post blocks using **position** (min. local id distance)
- Sometimes exact matches are not the correct predecessor

Post Block Matching: Performance

Fork me on GitHub



Text:
manhattanFourGramNormalized
Threshold: 0.17
TPR: 0.99
FPR: 0.14
Matthews Corr.: 0.86

Code:
winnowingFourGramDiceNormalized
Threshold: 0.23
TPR: 0.99
FPR: 0.07
Matthews Corr.: 0.92

https://github.com/sotorrent/posthistory-extractor

SOTorrent

- Among other features, the dataset provides the **version history** of Stack Overflow content on the **level of individual text or code blocks**
- Was official **mining challenge** of MSR 2019

sotorrent.org

Dataset available on Zenodo

Open Data

Cloud Icon CC BY 3.0 smashicons on flaticon.com

Timeout Flakiness

SAP HANA

Publication:
ICSE 2024 SEIP

Common Flakiness Definition

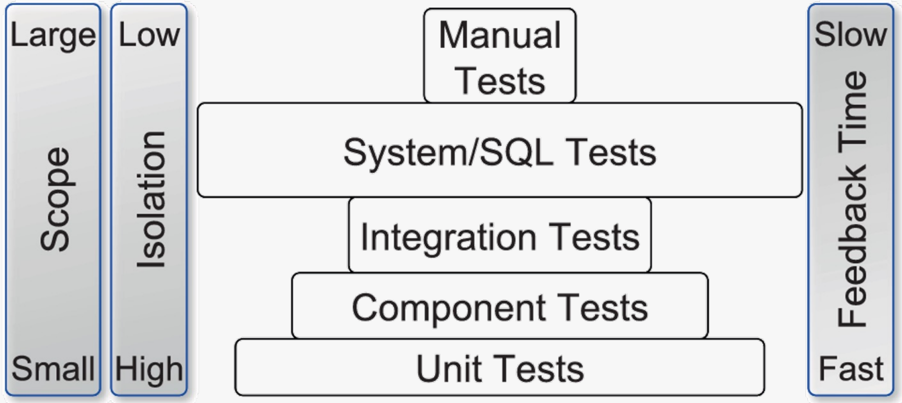
“A test can be considered flaky when it exhibits both passing and failing results for the same code.”

🟩🟥🟩🟩

SAP HANA

Context

- Flaky tests **interfere with CI/CD.**
- SAP HANA is a large industrial DBMS with **long-running system tests.**
- Flaky failures **impede automatic test run assessment** and merging.
- Standard strategy: **restart flaky tests.**
- Configurable **"max duration"** (timeout values) to prevent stuck tests from blocking resources, but this can cause flaky test executions.
- We found that **99% of CI runs are affected** by flaky failures (→ costs, delay).



SAP HANA

Research Questions

RQ1: What **level of test flakiness** do we observe in SAP HANA's system tests and what can we identify as a **major contributing factor**?

RQ2: What **impact** does **increasing timeout values** have on test flakiness in context of SAP HANA?

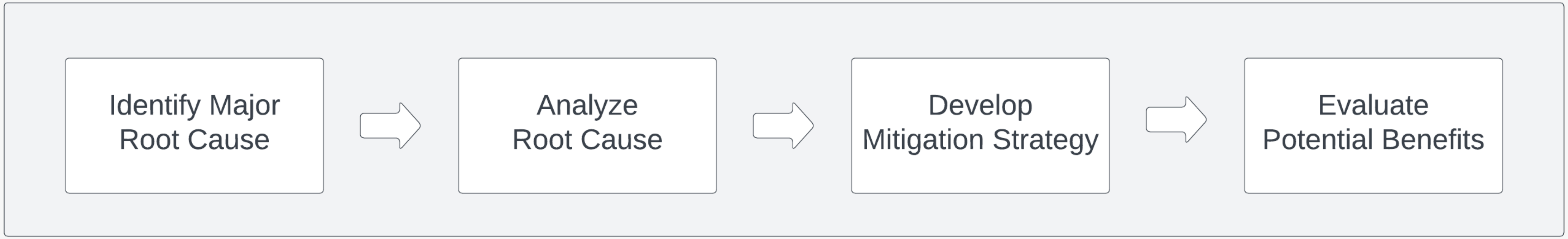
RQ3: How do **developers commonly adjust timeout values** in the context of SAP HANA?

RQ4: To what degree can we **optimize the timeout values** with respect to their average test execution costs?

Timeout Flakiness

SAP HANA

Research Approach


Identify Major Root Cause
Analyze Root Cause
Develop Mitigation Strategy
Evaluate Potential Benefits

SAP HANA

Conclusion

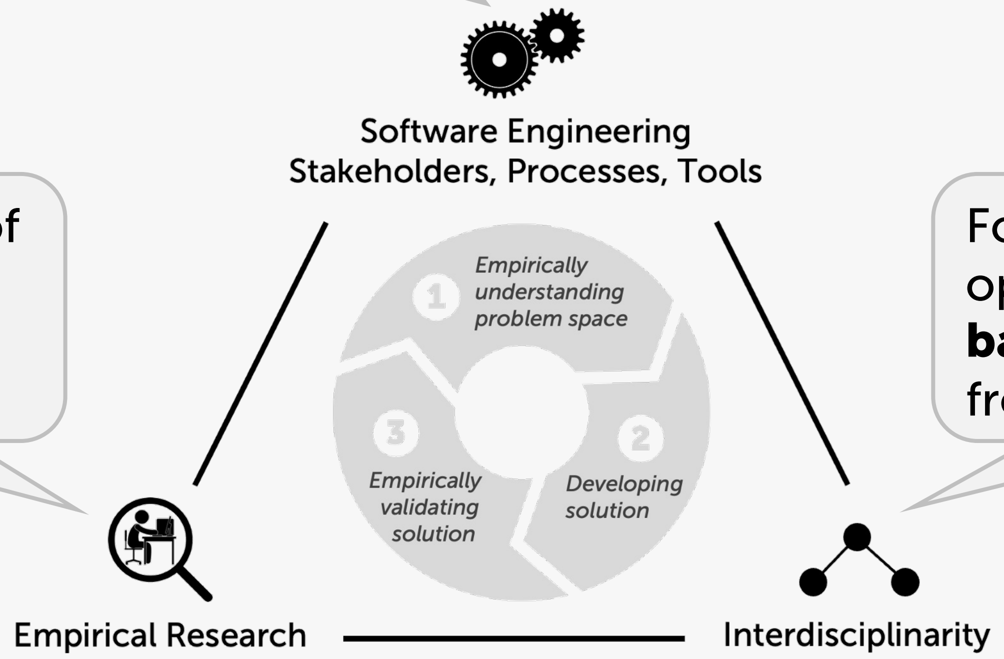
- Flakiness definition is of **little practical use**, because the test flakiness rate **converges to 1**.
- Timeout values can cause **additional costs.**
- Cost-optimal timeout values **can increase efficiency.**
- Baseline approach with a **fixed global timeout** currently being implemented at SAP.

Summary

Quantification of impact of timeouts on test flakiness, development of **cost-optimal approach**.

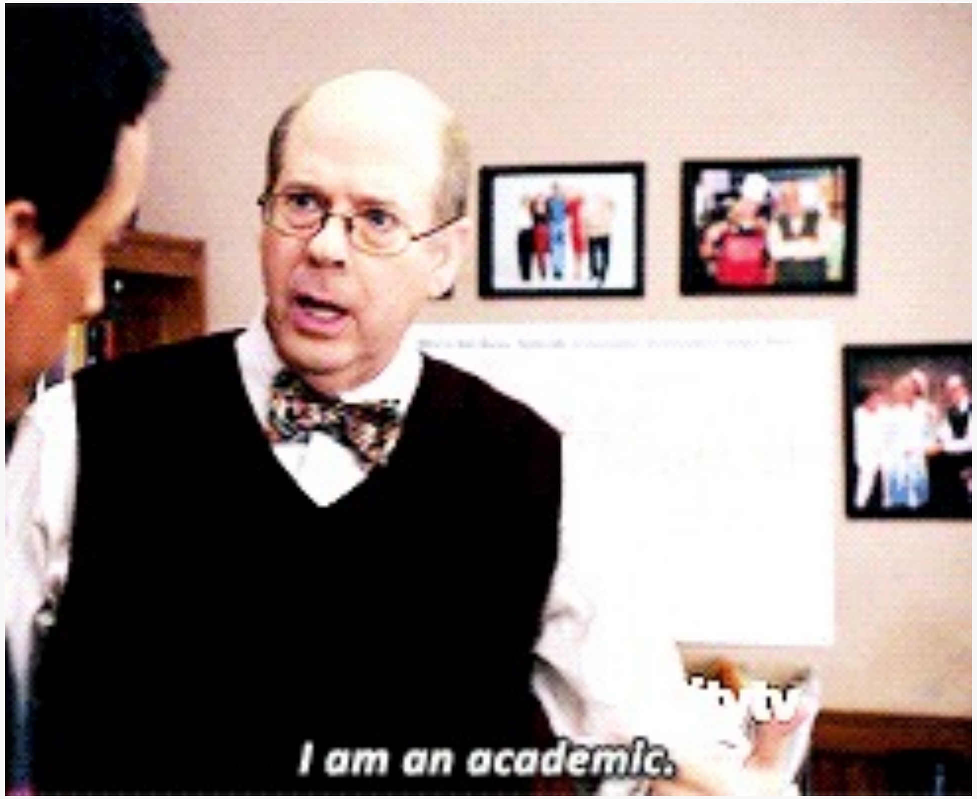
Quantitative analysis of **historical data, simulations** based on tailored dataset.

Formulation of the optimization problem **based on related work** from the AI/ML.


Software Engineering
Stakeholders, Processes, Tools
1 Empirically understanding problem space
2 Developing solution
3 Empirically validating solution
Empirical Research
Interdisciplinarity

Back to our previous topic...

Research **vs.** **Practice**



The issues of applied SE research in industry

- **Constantly changing (business) priorities**.

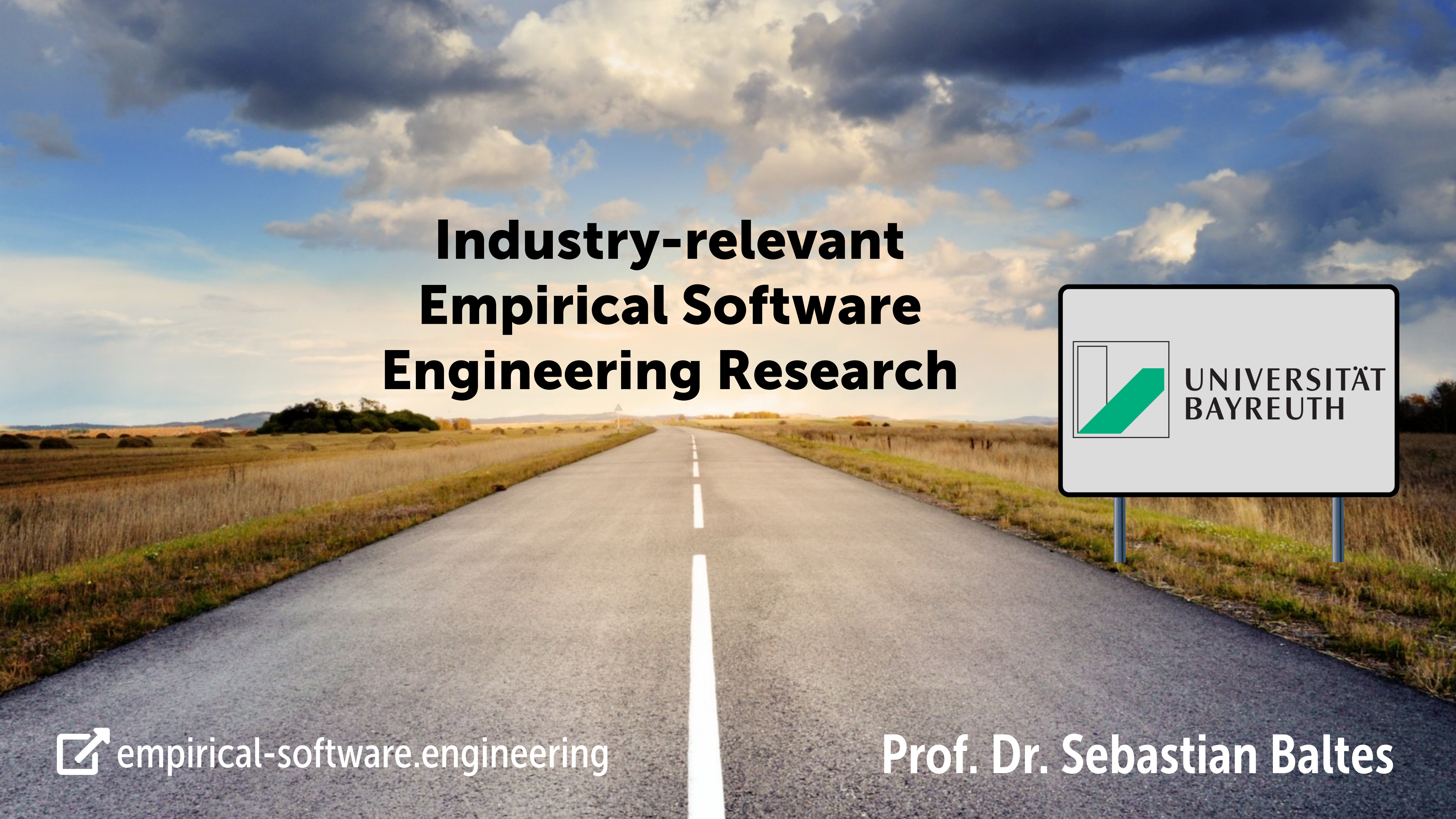
- Hard to internally “sell” **projects with potential long-term benefits** requiring short-term investment/exploration.

- Importance of (early) quantification of business impact via KPIs (key performance indicators) makes **foundational research difficult**.

…are an opportunity for academia

When doing research in **close collaboration with industry**, academic researchers can work on the topics that:

- Are **promising**, but do not (yet) have an immediate business value.
- Require considerable investment in **exploring the problem space** without an immediate quantifiable business impact.
- Require **deep understanding** of the broad spectrum of empirical software engineering methods.
- Require a **systematic screening of related (academic) work** on a topic to replicate or extend it.



Industry-relevant Empirical Software Engineering Research

UNIVERSITÄT BAYREUTH

empirical-software.engineering

Prof. Dr. Sebastian Baltes

Software Engineering @ UBT

- Lecture in Summer Semester 2024:
  - *"Software Engineering"* (Bachelor)
- New courses planned for upcoming semesters:
  - *"Advanced Software Engineering"* (Bachelor/Master)
  - *"Software Analytics"* (Master)

- **Interested in software engineering research?**
  - Bachelor/Master theses
  - Openings for PhD positions
  - Please reach out to me in case you're interested!