

Erstellung eines Datensatzes zur Evaluierung eines automatischen Ansatzes zum (semi-)automatischen Verfolgen von Anforderungen im Source Code

Hintergrund

Softwareprojekte werden immer komplexer und besitzen viele verschiedene Artefakte, die miteinander verbunden (Tracing) sind. Tracing von verschiedenen Artefakten kann dabei helfen die verschiedenen Artefakte miteinander zu verbinden und Änderungen besser nachzuvollziehen, z.B. eine Sicherheitslücke wurde in der CVE-Datenbank entdeckt und Änderungen am Source Code müssen dringend wahrgenommen werden. Derzeit wird Tracing jedoch eher manuell bzw. semiautomatisch genutzt. Für diesen Zweck wird derzeit in einer Masterarbeit ein Prototyp entwickelt, der einen automatischen Ansatz verfolgt. Jedoch steht dafür zum Testen nur sehr rudimentäre Datensätze zur Verfügung.

Aufgabe

In dieser Bachelorarbeit soll nun ein Datensatz erstellt werden, der es ermöglicht das Tool der Masterarbeit zu evaluieren. Dafür soll zu Beginn Literatur durchsucht werden, um ein Verständnis zum Thema „Tracing“ aufzubauen. Im nächsten Schritt soll eine geeignete Möglichkeit gefunden werden einen Datensatz (semi-) automatisch zu erstellen. Dafür gibt es verschiedene Möglichkeiten, wie zum Beispiel LLMs oder GitHub Mining. Hinsichtlich der Neueinführung von LLMs bietet sich eine Nutzung eines LLMs an. Dafür ist es möglich eine Schablone zu definieren, die automatisiert Lösungsvorschläge generiert, umso den Datensatz aufzubauen. Nachdem dieser Datensatz erstellt wurde, soll dieser im nächsten Schritt validiert werden. Dafür soll einerseits das Tool aus der Masterarbeit genutzt und auch gezeigt werden, dass dieser Datensatz funktioniert bzw. gut geeignet ist (dafür müssen Metriken aufgestellt werden). Und andererseits soll gezeigt werden, dass dieser Datensatz für ausgewählte Open Source Projekte auch funktioniert, z.B. Corona Warn App.

Bsp. für das Aussehen des Datensatzes:

Anforderung: Verschlüsselung des Passworts – Source Code: `cipher = encrypt(passkey, str1)`

1. Einarbeitung in die Literatur (Tracing, LLMs)
2. Explorieren eines geeigneten Ansatzes zum Erstellen des Datensatzes
3. Erstellen des Datensatzes
4. Validierung des Datensatzes für das Tool aus der Masterarbeit
5. Validierung des Datensatzes hinsichtlich eines Open Source Projekts

Organisatorisches

Betreuer: Alexander Specht
Prüfer: Prof. Dr. Schneider
Beginn: ab sofort möglich