

Identifikation semantischer Clone in Informationssystemen

- Bachelorthesis
 Masterthesis

Aufgabenstellung

In nahezu jedem größeren Unternehmen bilden betriebliche Informationssysteme (BIS) geschäftskritische Prozesse ab. In ihre individuelle Erstellung und Weiterentwicklung werden jedes Jahr Milliardenbeträge investiert und viele Systeme werden über Jahrzehnte weiterentwickelt und gewartet. Ein wichtiges Problem hierbei sind Redundanzen in der Implementierung, die zu Mehraufwänden und Inkonsistenzen führen. In vielen Fällen entstehen diese durch „Copy & Paste“-Programmierung. Diese Form von (syntaktisch ähnlichen) Clonen ist gut untersucht und kann durch diverse Werkzeuge zur Clone-Erkennung identifiziert werden kann.

Schwieriger ist es jedoch bei Redundanz, die entsteht, weil ein Entwickler sich nicht bewusst ist, dass eine Funktionalität bereits im System existiert, und diese erneut umsetzt. Gerade bei unklarer Projektstruktur, Fluktuation im Team und langjähriger Weiterentwicklung kommt dies häufig vor. Da sich die beiden Implementierungen i.d.R. syntaktisch unterscheiden werden sie von üblichen Clone-Detektoren nicht gefunden. Hierfür sind andere Algorithmen und Werkzeuge erforderlich, die auch bereits erforscht werden („semantic cloning“). Allerdings liefern diese Werkzeuge häufig zu viele „false positives“ für den Praxiseinsatz. Für die vorliegende Arbeit soll daher keine allgemeine Lösung entwickelt werden, sondern fokussiert werden auf den Einsatz in der Wartung von Informationssystemen, in der Erwartung dass damit eine präzisere Aussage möglich ist.

Zielsetzung der Arbeit

Ziel der Arbeit ist die Konzeption, prototypische Umsetzung und Erprobung eines Werkzeugs zur Identifikation semantischer Clone in Informationssystemen, ggf. aufbauend auf existierenden Ansätzen. Das Werkzeug soll Entwicklern helfen zu erkennen, dass die Funktionalität, die sie gerade umsetzen, anderswo bereits vorhanden ist. Zudem soll es für ein existierendes System vorhandene (wahrscheinliche) Clone auflisten.

Hierfür sind vsl. erforderlich:

- Literaturrecherche zum Stand der Forschung
- Analyse realer Informationssysteme zur Identifikation relevanter Beispiele und ihrer Charakteristika
- Konzeption und prototypische Umsetzung eines Werkzeugs (ggf. auch auf Basis existierender Arbeiten)
- Erprobung anhand realer Systeme, insb. im Hinblick auf in der Praxis akzeptable Trefferquoten

Mögliche Charakteristika semantischer Clone und damit Ansätze zur Erkennung könnten sein:

- Verwendung ähnlicher Identifier oder Literale
- Nutzung ähnlicher oder identischer Bibliotheken
- Ähnlichkeit von Kommentaren
- Ähnlichkeit von Strukturen wie Schleifen

Unternehmensprofil

Diese Arbeit wird in Kooperation mit der itestra GmbH (www.itestra.de) angeboten. Die itestra GmbH ist ein innovativer, unabhängiger und international tätiger Software-Dienstleister im Bereich unternehmenskritischer Prozesse, Systeme und Anwendungen. Zum Portfolio gehören die Analyse von Geschäftsprozessen, der Entwurf und die Realisierung leistungsfähiger Lösungen in Projekten sowie das strategische Management bereits vorhandener Softwaresysteme.

Betreuung (itestra GmbH)

Jonathan Streit (streit@itestra.de)