

# Synergieeffekte bei Systembeobachtung und -analyse durch Crowdsourcing in sozialen Netzwerken

Promotionsvorhaben  
AG Systemsoftware und Verteilte Systeme, Universität Trier

Dipl.-Inf. David Christ

17. März 2013

Computersysteme können auf unterschiedliche Arten und Weisen versagen. Daher sind Methoden der Systembeobachtung und Fehleranalyse seit jeher elementarer Bestandteil der Softwareentwicklung.

Große Firmen wie Apple, Canonical (Ubuntu), Microsoft und Mozilla haben Mechanismen entwickelt, um Berichte über Softwarefehler auf den Computern der Endanwender zu erstellen und an die Hersteller zurück zu senden. Zum Umgang mit dieser Datenflut werden die Berichte dort in großen Datenbanken gesammelt und vorverarbeitet.

Für den weiteren Verlauf des Textes ist wichtig, dass unterschiedliche Fehlerberichte der gleichen Ursache mit hinreichender Genauigkeit zu einer Gruppe zusammengefasst werden können.

Das im Rahmen des Frühjahrstreffens vorzustellende Projekt möchte der existierenden und bewährten Methodik zwei Elemente hinzufügen.

- Zugänglichkeit für die Allgemeinheit in Form eines sozialen Netzwerks.
- Ausfallsicherheit durch Dezentralität.

Im Rahmen des sozialen Netzwerks sollen alle Teilnehmer an einer Software – interne und externe Entwickler, Support-Mitarbeiter, Anwender – auf einfache und intuitive Art in den Prozess der Softwareanalyse einbezogen werden können. Gruppen von Fehlerberichten gleicher Ursache fungieren dann (wie Gruppen in konventionellen sozialen Netzwerken) als Treffpunkt für bestimmte Interessen (hier dem Lösen des entsprechenden Fehlers). Durch die offene Gestaltung können Berichte durch Crowdsourcing – dem Nutzen der Weisheit der Masse – mit nutzergenerierten semantischen Informationen erweitert werden:

- Ähnliche Fehlerberichtsgruppen können verknüpft werden, wodurch Inferenzen gezogen werden können.
- Durch diese Ähnlichkeitsbeziehungen können Cluster ähnlicher Fehlerberichtsgruppen als Themengebiete aufgefasst werden.
- Erfolgreich in bestimmten Clustern tätige Mitglieder können als Experten für diese gekennzeichnet werden.

Das soziale Netzwerk liefert auch eine virtuelle Kollokation im Sinne der Computer-supported Cooperative Work (CSCW).

Durch die große Anzahl an Teilnehmern wird das System die Möglichkeit bieten müssen, mehrere Zweige von Lösungswegen parallel zu verfolgen und etwaige erkenntnisbringende Ergebnisse aus diesen wieder in einen Hauptzweig zurückzuführen, vergleichbar mit Git.

Die größte Herausforderung, welche aus der offenen Gestaltung heraus erwächst, ist die des Datenschutzes. Einige Elemente einer Fehlerberichtsdatenbank sollten nicht an die Öffentlichkeit geraten – hierzu zählen Arbeitsspeicher-Abbilder, welche sensible Nutzerdaten enthalten können oder Auszüge von Softwarequellcode in Closed-Source-Projekten, welche in internen Systemen einfach eingetragen werden könnten.

Als Folge muss ein Mechanismus für die dezentrale (und gegebenenfalls redundante) Verwaltung der anfallenden Daten gefunden werden, welcher einerseits einer etwaigen zuständigen Instanz die nötige Kontrolle an die Hand legt, andererseits aber auch so viel Information wie möglich der breiten Öffentlichkeit zugänglich macht.

Die Lösung liegt in einem Verbund aus Servern – gleich ob in Firmenbesitz oder in der Cloud – welche jeder für sich eine separate Verwaltung der Zugriffsrechte erlauben, jedoch über eine definierte Infrastruktur deren Daten untereinander und mit menschlichen Anwendern kommunizieren können. Jeder Fehlerbericht einer bestimmten Software wandert so zu „seinem“ entsprechenden Server und wird von dort aus, möglicherweise in reduzierter Form, öffentlich gemacht.

Für Open-Source-Projekte könnten öffentliche Server bereitgestellt werden, an denen sich beliebige Hersteller freier Software registrieren können, wie etwa jedes GitHub-Projekt einen eigenen „Issues“-Bereich besitzt.

Nutzerkonten werden, vergleichbar mit XMPP, auf einem jeweiligen Server angelegt, wodurch diesem die Möglichkeit gegeben wird, zwischen internen und externen Nutzern zu unterscheiden und entsprechend unterschiedliche Zugriffsrechte zu vergeben. Ein von allen Servern geteilter Namensdienst enthält die Menge aller im System registrierten Server.

Zur Zeit wird ein Prototyp entwickelt, an welchem die oben erörterten Theorien und Möglichkeiten erprobt werden können. Hierzu wird gerade die Kooperation mit einem Firmenpartner aufgebaut. So sollen auch weiteres Potential aufgezeigt und mögliche Synergieeffekte verdeutlicht werden.