

Nützliche Metriken, um das Testen zu managen

Karol Frühauf, Dr. Ferdinand Gramsamer

Die größten Herausforderungen im Testmanagement sind neben den menschlichen Faktoren, den Testausführungsfortschritt zu verfolgen und vorherzusagen, wann ein Produkt reif ist für die Markteinführung. In unserem Beitrag „Useful Metrics for Managing Testing“ gehen wir auf einfache und praxistaugliche Metriken ein, um diesen Herausforderungen zu begegnen. Sie wurden in einem größeren Projekt mit vier Teststufen erprobt. Weniger als 1% des gesamten Testaufwandes benötigte ihre Erfassung und Auswertung. Mit den Metriken haben wir versucht, die folgenden Fragen des Projektmanagements zu beantworten:

Welche funktionalen und nicht-funktionalen Aspekte sind...
 ...auf welcher Teststufe wie oft getestet worden?
 ...sind noch nicht getestet worden?

Wie lautet der Fahrplan, um die...
 ...bereits getesteten Aspekte mittels Regressionen zu testen?
 ...noch ungetesteten Aspekte zu testen?

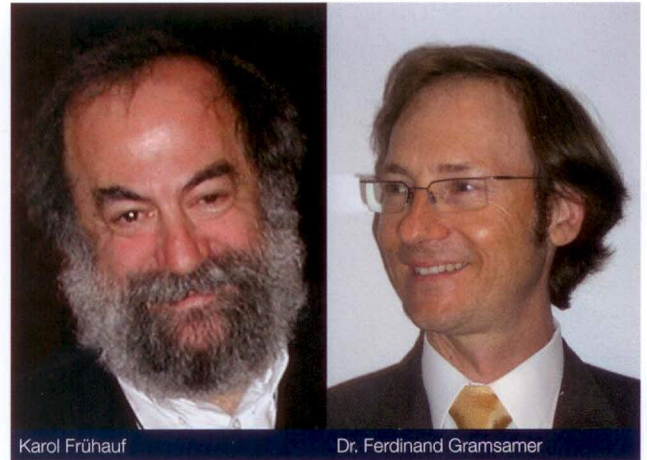
Wie viele kritische Fehler...
 ...wurden in einer Periode entdeckt?
 ...wurden in einer Periode geschlossen?
 ...sind am Ende einer Periode noch offen?

Wann können wir in Produktion gehen?

Grundlage für Testausführungs- und Testabdeckungsmetriken ist eine Strukturierung des Testumfangs auf jeder Teststufe in mehrere Testspezifikationen. Das Thema einer Testspezifikation ist das funktionale oder nicht-funktionale Verhalten eines gewissen Features oder das Verhalten einer Einheit an einer ihrer Schnittstellen. Jede Testspezifikation ist wiederum in Testbereiche aufgeteilt, die eine Anzahl Testfälle beinhalten. Jeder Testfall besteht aus Testschritten.

Der Fortschritt auf Testschritzebene ist für den Tester und seinen Testmanager interessant. Für das Projektmanagement ist der Status von Testfallebene aufwärts interessant. Ein Testfall kann „noch nicht ausgeführt“, „nicht abgeschlossen“, „abgeschlossen“ und für einen Testzyklus (ein Release) „nicht geplant“ sein. Die Zustände „noch nicht ausgeführt“ und „nicht geplant“ kann man für eine kompaktere grafische Darstellung zu „nicht ausgeführt“ zusammenfassen. Ist ein Testfall oder ganzer Testbereich „abgeschlossen“, d.h. sind alle Testschritte oder alle Testfälle ausgeführt, wird der Status „abgeschlossen“ in die Resultate „bestanden“ oder „mit Fehlern“ aufgelöst. Der Testfortschritt wird mit den Resultaten sowie den Zuständen „nicht abgeschlossen“ und „nicht ausgeführt“ dargestellt. Sie werden mit der Anzahl der spezifizierten und zur Testausführung geplanten Testfälle verglichen.

Diese Ergebnisse wurden für das Testen von einem Release wöchentlich in einem Graf eingetragen, um den Fortschritt im Testzyklus zu zeigen. Zudem wurden sie pro Testbereich und Teststufe jeweils am Ende eines Release grafisch dargestellt. Aus der Geschichte dieser Daten konnten wir die folgenden Fragen beantworten:



Sind alle Testbereiche abgedeckt?
 Ist die Testausführungsabdeckung aller Spezifikationen 100%?
 Wir oft wurde ein Testbereich ausgeführt?
 In welchem Release wurde ein Testbereich das letzte Mal fehlerfrei getestet?

Für Fehlermetriken haben wir die Anzahl der gefundenen, offenen und geschlossenen Fehler über die Zeit verfolgt. Release-bezogene Daten, z.B. wie viele Fehler in einem Release entdeckt wurden, wurden nicht erfasst, weil diese Information für die Prognose der Software-Reife irrelevant ist. Für den Entscheid, ob die Software produktiv gesetzt werden kann, wurden nur die kritischen Fehler herangezogen. Der frühe Entscheid, im Rahmen des Projekts ausschließlich solche betriebsverhindernden Fehler zu beheben, trug bei zur vergleichsweise raschen Stabilisierung des Produkts, nachdem erst jegliche Funktionalität implementiert war.

Erfasst wurden dafür die Daten über die Teststufe, auf der die Fehler entdeckt wurden. Mit der Auswertung dieser Daten konnten wir den Testprozess regelmäßig überprüfen: Wurden die meisten Fehler auf Komponentenebene entdeckt? Wieso werden im Systemtest deutlich mehr Fehler entdeckt als im Integrationstest?

Das Erfassen des Datums von Zustandsänderungen der Fehlermeldungen und die Auswertung der Zeitdifferenzen half, den Fehlerbehebungsprozess zu überwachen: Wie lang ist die Verweildauer von kritischen Fehlern von der Entdeckung bis zum Schließen nach dem erfolgreichen Nachttest? Wir berichten über keine neuen spektakulären Metriken. Das Spektakulärste im Projekt war, dass die wenigen Metriken im gesamten Projekt nicht nur konsequent gesammelt und ausgewertet, sondern auch als Entscheidungsgrundlage genutzt wurden.

Die Autoren

Karol Frühauf ist Mitgründer der Firma INFOGEM AG, in Baden, Schweiz.
Dr. Ferdinand P. Gramsamer ist Leiter von Testing Services bei der Firma bbv Software Services AG in Luzern, Schweiz.