

Identifikation von Test-Klonen

Bachelorarbeit, betreut von Mathias Landhäußer

Johann Böhler | 19. April 2013

IPD TICHY



Test-Klone

- Semantisch und syntaktisch ähnliche Testfälle

@Test

```
public void testAcceleration() {  
    Car car = new Trabbi();  
    car.accelerate();  
    assertTrue(car.getSpeed() > 5);  
}
```

@Test

```
public void testAcceleration() {  
    Car car = new Trabbi(-) Ferrari();  
    car.accelerate();  
    assertTrue(car.getSpeed() > 560);  
}
```

@Test

```
public void testIncrement() {  
    Counter counter;  
    counter.increment();  
    assertTrue(counter.getCount() > x);  
}
```

- 100 Testfälle \approx 5.000 manuelle Vergleiche
 - Wiederholte Muster?
 - Rückschluss auf Funktionalität?
- \Rightarrow Automatisierung!

Test-Klone

- Semantisch und syntaktisch (?) ähnliche Testfälle

Kategorie: [Computer](#) > [Computer-Zubehör](#) > [Eingabe- & Lesegeräte](#) > [Tastaturen](#)



Tastatur mit ergonomischem "Gutenberg Tastaturlayout"

Artikelzustand: **Als Ersatzteil / defekt**

Restzeit: **1 Tag 20 Stunden** (23. Feb. 2011 16:10:01 MEZ)

Aktuelles Gebot: **EUR 15,50** [18 Gebote]

Geben Sie mindestens EUR 16,00 ein

Versand: **EUR 3,90** - Standardversand [Weitere Versandarten](#) [Rabatte anzeigen](#) |

[Alle Details anzeigen](#)

Artikelstandort: NRW, Deutschland

Versand nach: Deutschland

Lieferung: **Voraussichtlich innerhalb von 2-4 Werktagen**

Zahlungen: [Überweisung](#) | [Siehe Zahlungsinformationen](#)



25 € Startguthaben sichern

- **Mit der neuen eBay-Kreditkarte.**

Rücknahmen: **Keine Rücknahme.** Dies ist ein Privatverkauf.

Beschreibung

Versand und Zahlungsmethoden

Der Verkäufer ist für dieses Angebot verantwortlich.

Auftreten und Entstehung

- Ähnliche, zu testende Funktionalität
- Kopieren und Einfügen (*Copy & Paste*)

Plagiatssuche

- **Textbasierte Analyse**
 - z.B. *diff*, Zeichen- oder zeilenweiser Vergleich
- **Metrische Analyse**
 - Maßzahlen (z.B. Anweisungen, Schleifen, ...)
- **Lexikalische Analyse**
 - Zerlegung in Tokens → LCS
- **Syntaktische Analyse**
 - Vergleiche abstrakte Syntaxbäume
- **Semantische Analyse**
 - Vergleiche Datenfluss-Graphen

Grundidee

- Erstelle abstrakten Syntaxbaum
- Vergleiche Anweisungen \rightarrow LCS
- Ähnlichkeit $:= \frac{LCS}{Anweisungen}$



■ Vorfilter

- @Test, void test. . . ()
- Mindestlänge

■ Wandler

- Normalisiere Zusicherungen (*JUnit*, *TestNG*)

■ Metriken

- Paket- oder Klassenzugehörigkeit
- Anzahl und Typen der Zusicherungen
- @Test(expected=?)

■ Analyse



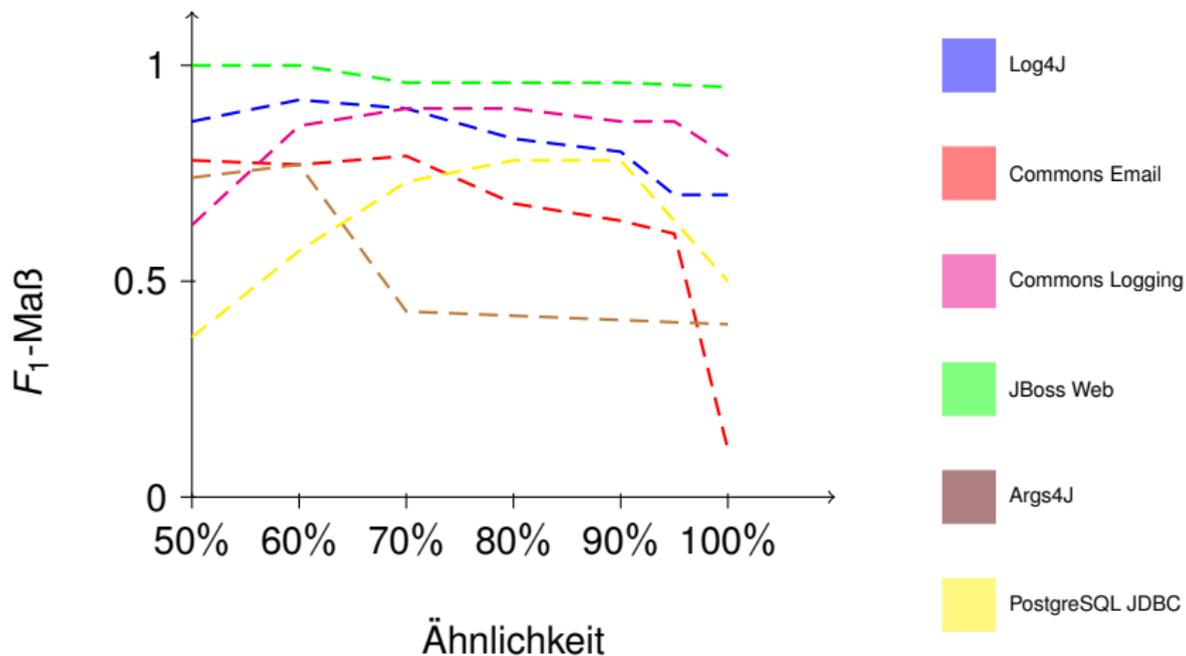
The logo features a green square on the left containing a white DNA double helix. To the right of the square is a white speech bubble tail pointing downwards and to the right. To the right of the square is the text 'QJUTCI' in a bold, green, sans-serif font.

Vorgehen

- ~ 10.000 Paare manuell untersucht → *Jutci*
- Maßzahlen
 - Präzision: „*richtig*“ erkannt
 - Ausbeute: „*wie viele*“ erkannt

Untersuchungsziele

- 6 quelloffene Bibliotheken
- ~ je 100 Testfälle, 50% Klone
- Bei 60%iger Ähnlichkeit
 - \emptyset Präzision: 87%
 - \emptyset Ausbeute: 85%



Untersuchung Industrieprojekte

- ~ 13.500 Testfälle ≈ 90 Mio. Paare
- 45% – 88% Klone

- Reihenfolge der Befehlsanweisungen
- Kurze Testfälle
 - Ungewollte Überschneidungen (*projektübergreifend*)
 - Leichte Abweichungen

- Identifikation von Test-Klonen
 - 4-stufiges Lösungsverfahren
 - Gute Ergebnisse → leicht erweiterbar

- Automatische Testfall-Generierung
- Synchrone Testfall-Adaptierung
- ...

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	60	47	13	1	78,33%	97,92%	.87
60%	50	45	5	3	90%	93,75%	.92
70%	43	41	2	7	95,35%	85,42%	.9
80%	36	35	1	13	97,22%	72,92%	.83
90%	32	32	0	16	100%	66,67%	.8
95%	26	26	0	22	100%	54,17%	.7
100%	24	24	0	24	100%	50%	.67

Tabelle: Zusammenfassung der Ergebnisse für Log4J

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	189	120	69	0	63,49%	100%	.78
60%	126	95	31	25	75,4%	79,17%	.77
70%	108	90	18	30	83,33%	75%	.79
80%	66	63	3	57	95,45%	52,5%	.68
90%	58	57	1	63	98,28%	47,5%	.64
95%	53	53	0	67	100%	44,17%	.61
100%	7	7	0	113	100%	5,83%	.11

Tabelle: Zusammenfassung der Ergebnisse für Commons Email

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	142	66	76	0	46,48%	100%	.63
60%	67	57	10	9	85,07%	86,36%	.86
70%	59	56	3	10	94,92%	84,85%	.9
80%	56	55	1	11	98,21%	83,33%	.9
90%	53	52	1	14	98,11%	78,79%	.87
95%	51	51	0	15	100%	77,27%	.87
100%	43	43	0	23	100%	65,15%	.79

Tabelle: Zusammenfassung der Ergebnisse für Commons Logging

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	984	982	2	0	99,8%	100%	1
60%	983	982	1	0	99,9%	100%	1
70%	911	911	0	71	100%	92,77%	.96
80%	910	910	0	72	100%	92,67%	.96
90%	909	909	0	73	100%	92,57%	.96
95%	907	907	0	75	100%	92,36%	.96
100%	894	894	0	88	100%	91,04%	.95

Tabelle: Zusammenfassung der Ergebnisse für JBoss Web

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	748	168	580	0	22,46%	100%	.37
60%	357	150	207	18	42,02%	89,29%	.57
70%	224	144	80	24	64,29%	85,71%	.73
80%	135	118	17	50	87,41%	70,24%	.78
90%	110	108	2	60	98,18%	64,29%	.78
95%	79	79	0	89	100%	47,02%	.64
100%	56	56	0	112	100%	33,33%	.5

Tabelle: Zusammenfassung der Ergebnisse für PostgreSQL JDBC

Schwelle	K_a	r_+	f_+	f_-	p	a	F_1
50%	60	47	13	1	78,33%	97,92%	.87
60%	50	45	5	3	90%	93,75%	.92
70%	43	41	2	7	95,35%	85,42%	.9
80%	36	35	1	13	97,22%	72,92%	.83
90%	32	32	0	16	100%	66,67%	.8
95%	26	26	0	22	100%	54,17%	.7
100%	24	24	0	24	100%	50%	.67

Tabelle: Zusammenfassung der Ergebnisse für Log4J

Ä-S.	K_a	r_+	f_+	f_-	p	a	F_1
50%	19.121	8.400	10.721	0	43,93%	100%	.61
100%	6.105	3.400	2.705	5.000	55,69%	40,48%	.46

Tabelle: Zusammenfassung der Ergebnisse für Derby

Ä-S.	K_a	r_+	f_+	f_-	p	a	F_1
50%	63.471	30.800	32.671	0	48,52%	100%	.65
100%	9.088	7.538	1.550	23.262	82,29%	24,18%	.37

Tabelle: Zusammenfassung der Ergebnisse für Guava

- 10.135 Testfälle
- 446 Test-Klone
- Kurz, ohne semantischen Zusammenhang

- < 1000 Testfälle \rightarrow wenige Sekunden
- $= 5000$ Testfälle \rightarrow 1 min.
- > 10.000 Testfälle \rightarrow mehr als 5 min.

Vorteile der **syntaktischen Analyse** ...

- Objektorientierung
- Verwendung AST in allen Schritten

Nachteile der **syntaktischen Analyse** . . .

- Hohes Speichervolumen
- Verwendung gebunden an konkrete Implementierung

- MVC
- Abstraktion von AST
- Entwurfsmuster
- org.reflections