

Änderungsverwaltung mit DIFF und PATCH

Jan Hendrik Dithmar

27.10.2005

Einführung

- ▶ Bei der Softwareentwicklung ist es nötig, Änderungen oder Erweiterungen nachzuvollziehen.
- ▶ *Software-Konfigurationsmanagement* (kurz SKM, engl.: *software configuration management*) hat zur Aufgabe, die Evolution von Softwaresystemen zu organisieren und zu kontrollieren.

DIFF

- ▶ DIFF vergleicht verschiedene Versionen eines Dokuments und bestimmt die Unterschiede; DIFF kann nicht nur Änderungen an Dateien, sondern auch über ganze Verzeichnisse bestimmen (→ rekursive Anwendung).
- ▶ Die Unterschiede von 2 Texten *A* und *B* bestimmt DIFF als eine Folge von Kommandos, die *A* in *B* umwandeln. Kommandos sind das Einfügen und Löschen von Zeilen.

PATCH

- ▶ PATCH wendet die Kommandos der Ausgabe von DIFF auf eine Datei oder mehrere Dateien an.
- ▶ Das Nachvollziehen und Anwenden von Änderungen lässt sich dadurch automatisieren.

Manuelle und automatische Integration

- ▶ Bei der **Zwei-Wege-Integration** wird aus zwei Dateien *A* und *B* eine neue Datei erzeugt, indem für jeden Unterschied interaktiv aus *A* oder *B* ausgewählt wird.
- ▶ Bei der **Drei-Wege-Integration** wird eine gemeinsame Urfassung hinzugezogen: Ist ein Textstück nur in *A* oder *B* verändert worden, kann die Änderung automatisch übernommen werden. Ist das Textstück in *A* und *B* geändert worden, liegt ein Konflikt vor und der Benutzer muss aktiv werden, um die Änderungen selbst zu integrieren.

Probleme bei der Integration

- ▶ Wird als Basis für die Änderung die Zeilennummer benutzt, so können Probleme bei der Ausführung der Änderungen auftreten.
- ▶ Diese Probleme lassen sich jedoch weitgehend entschärfen, indem als Ort für die Änderung nicht die Zeilennummer, sondern der Kontext betrachtet wird. Der Kontext ist ein Bereich von Zeilen, der vor und nach der jeweiligen Änderung auftreten muss.
- ▶ Kennt PATCH den Kontext einer Änderung, kann es im Text nach den Kontextzeilen suchen, um den Ort der Änderungen zu bestimmen.
- ▶ DIFF kennt ein spezielles Kontextformat, das mit einer Option aktiviert werden muss.
- ▶ Eine exakte Übereinstimmung der Zeilen ist nicht erforderlich.

Syntaktische und semantische Integration

- ▶ DIFF und PATCH gehen rein textorientiert vor, was leicht Probleme bei der Integration hervorrufen kann.
- ▶ Die Sicherheit automatischer Integration lässt sich weiter steigern, indem syntaktische und semantische Informationen des zugrunde liegenden Dokuments ausgenutzt werden.
- ▶ Die syntaktische Integration und die semantische Integration berücksichtigen zusätzlich Syntax bzw. Semantik der Dokumente, auf die die Änderungen anzuwenden sind.

Wie DIFF arbeitet

- ▶ Problem: Gegeben zwei Zeichenketten A und B . Die Bearbeitungsdistanz (engl. *edit distance*, kurz: Distanz) zweier Zeichenketten ist die Länge der kürzesten Folge von Einfüge- oder Löschkommandos, die A nach B umwandelt. Ziel ist es, eine kürzeste Kommandofolge zu finden.
- ▶ Klassische Algorithmen für dieses Problem bestimmen die längste gemeinsame Teilfolge (engl. *longest common subsequence*, LCS) von A und B ; durch den Vergleich der LCS mit A und B ergeben sich Einfüge- und Löschkommandos.
- ▶ Ein alternativer Ansatz ist der für DIFF entworfene Algorithmus von Miller und Myers. Grundlage für den Algorithmus ist eine Distanzmatrix, in der die kürzesten gefundenen Kommandofolgen für Präfixe von A und B eingetragen werden.