

Projektarbeit

Datenerhebung und Auswertung von Zugverspätungen

Betreuer: Dr. Alexander Hinneburg

hinneburg@informatik.uni-halle.de

Zugverspätungen gehören zu den leidigen Erfahrungen vieler Bahnfahrer. Oft schleicht sich das Gefühl ein, daß Verspätungen häufiger passieren und eher die Regel als die Ausnahme sind. In diesem Projekt soll untersucht werden, ob sich dieses vage Gefühl statistisch untermauern läßt. Um eine solche Untersuchung durchführen zu können, müssen regelmäßig über einen Zeitraum von z.B. einem Monat die An- und Abfahrtzeit von Zügen erfaßt werden. Ebenso ist eine Repräsentation des regulären Fahrplans notwendig. Alle diese Daten können im Internet erhoben werden.

Die Herausforderungen dieses Projektes sind der Entwurf eines Datenbankschemas zur Repräsentation der Zugverbindungen mit den entsprechenden Zeiten und ein Anwendungsprogramm, das in kurzen Zeitabständen die Bahnhofsanzeigetafeln im Internet abfragt, die Daten aus den erhaltenen HTML-Dokumenten extrahiert und in die Datenbank speichert. Zum Abschluss des Projektes soll eine einfache Auswertung der Daten erfolgen.

Ziele des Projektes

- Der Entwurf eines Datenbankschemas zur Repräsentation von Zugverbindungen (nur ICE, IC, IR und EC) und der planmäßigen und tatsächlichen An- und Abfahrtzeiten von allen Haltebahnhöfen.
- Ein Skript zur prozeß-parallelen Abfrage der relevanten Bahnhofstafeln im Internet¹ über einen zusammenhängenden Zeitraum von mindestens einem Monat.

Voraussetzungen

- Kenntnisse über Datenbankentwurf (DB I)
- Kenntnisse und praktische Erfahrung mit Perl oder Python und Unix/Linux
- Grundlagen in Graphentheorie und Statistik

Der Beginn der Arbeit ist bis November 2003 möglich.

¹www.bahn.de