

Automatisierung der Rechnerbedienung durch Data Mining

Betreuer:

Artur Andrzejak: mail: andrzejak@zib.de
ZIB, Raum 3152

Zielgruppe:

Dieses Projekt richtet sich an Studenten im Hauptstudium, die Interesse an dem Thema event and sequence mining (ein Teil von Data Mining) haben, und zugleich Erfahrung in Betriebssystemprogrammierung sammeln wollen. Kenntnisse in C++ sind wünschenswert, aber nicht notwendig. Interesse an wissenschaftlichen Arbeiten und Engagement sind wünschenswert.

Motivation:

Viele Vorgänge bei der Bedienung des Computers können partiell automatisiert werden, wenn man dem Benutzer längere Zeit „über die Schulter schaut“, seine Gewohnheiten aufzeichnet und anschließend auf Gemeinsamkeiten analysiert. Beispiele für solche Gewohnheiten sind:

- bevorzugte Benutzung bestimmter Funktionen von Programmen, z.B. Aufruf von Textformatierungsdialog, regelmäßiges Speichern
- Speicherung von HTML-Seiten, Downloads, oder anderen Dokumenten an spezifischen Pfaden, die z.B. von der besuchten Seite abhängen
- Sequenzen von Programm- und Funktionsaufrufen nach dem Booten des Rechners.

Die Idee ist es, derartige Benutzeraktionen aufzuzeichnen, sie auf Gemeinsamkeiten zu analysieren, und dann Makros/Tastenzuordnungen vorzuschlagen, die die Durchführung solcher Aktionen oder Sequenzen erleichtern. Dabei sollen für die Analyse der Gemeinsamkeiten Verfahren aus dem Bereich sequence mining (ein Teil von Data Mining) und bei Bedarf andere eingesetzt werden. Die Makros/Tastenzuordnungen werden in Form von Skripten für AutoHotkey, einem frei verfügbaren Makroframework für Windows, erstellt.



Abbildung 1: Drei Phasen der Automatisierung der Rechnerbedienung

Projektziele:

Das Ziel des Projektes ist es, ein Framework mit folgenden drei Teilen zu erstellen:

- ein Logger-Modul, das die Eingaben des Benutzers (Maus und Keyboardaktivitäten) innerhalb von Windows bzw. ausgewählten Applikationen aufzeichnet
- ein Analyse-Modul, das mit Hilfe von Data Mining Verfahren die Aufzeichnungen auf häufige Ereignisse bzw. Sequenzen untersucht (ggf. von Logs verschiedener Benutzer)
- ein Generator-Modul, das die im letzten Schritt gefundenen Ereignisse/Sequenzen als Makros im AutoHotkey-Format erzeugt.

Der Fokus des Projekts sollte dabei auf folgenden einfachen Szenarios liegen:

- Vorschlagen von Pfaden in den Fileselectorboxen in Abhängigkeit von Dokumentenart, aufrufender Applikation, besuchter WWW-Seiten, Zeit, und anderen potentiell signifikanten Attributen
- Vorschlagen von „keyboard shortcuts“ für die häufigsten Funktionen in einer Applikation, in Abhängigkeit von Applikation und ggf. weiteren Attributen
- Mining von häufigen Sequenzen von Funktionsaufrufen direkt nach dem Bootvorgang (z.B. 1. Umstellen der Bildschirmauflösung, 2. Wechsel in ein bestimmtes Verzeichnis, 3. Aufruf eines Editors etc.)

Bei erfolgreichem Abschluss des Projektes wird die Veröffentlichung der Ergebnisse in einem wissenschaftlichen Artikel angestrebt.

Durchführung des Projektes:

Das Projekt soll von einer kleinen Gruppe durchgeführt werden. Start und Ende des Projektes können mit dem Betreuer flexibel vereinbart werden, genauso wie die regelmäßigen Projekttreffen.

Das Projekt beginnt mit einer kleinen Vortragsreihe, in der die Projektteilnehmer kurze Vorträge über verschiedene Grundlagen zu den Themen Selektion der Attribute, Mining von Ereignissen/Sequenzen, Ereignisprotokollierung unter Windows.