

Entwicklung einer generischen Testplattform zur Anwendung von Data Quality Profiling auf Stammdaten von Informationssystemen.

Im heutigen Informationszeitalter sind Daten allgegenwärtig. Egal ob am Barcode-Scanner an der Supermarktkasse, beim Besuch von Facebook oder im heimischen Fußballverein, überall werden fleißig Daten gesammelt. Diese Sammelleidenschaft hat zur Folge, dass mittlerweile weltweit schätzungsweise 8 Zettabyte (das sind 8,6 Milliarden Terabyte) an Daten existieren und diese Zahl sich in schätzungsweise 2 Jahren verdoppelt haben wird. Dieses unvorstellbare Wachstum macht den Unternehmen, welche die Daten verwalten zu schaffen. Ein großes Problem ist dabei mangelnde Datenqualität, die aus veralteten, falschen und unvollständigen Daten im Bestand resultiert. Die Pflege der Datenbanken verschlingt Unmengen an Geld, Zeit und Ressourcen, da die Fehler von unterschiedlichsten Quellen stammen können und meist nur im Kontext mit weiteren Daten als Fehler erkennbar sind.

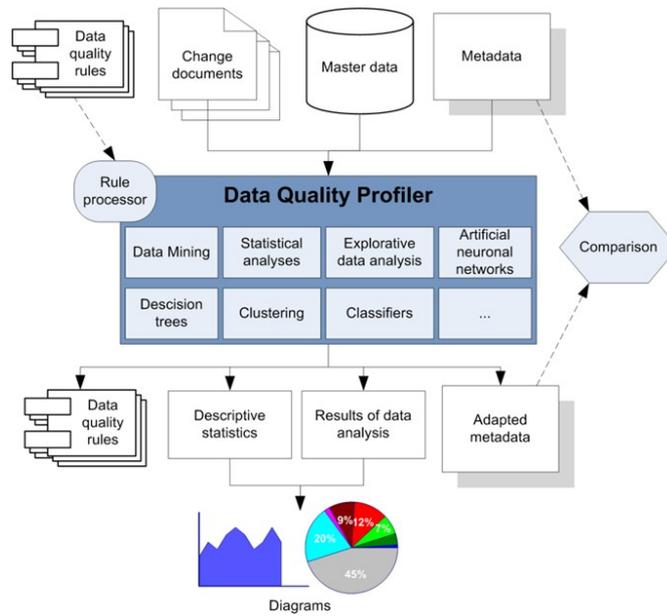
Daher befasst sich die Masterthesis von Achim Guldner mit dem Thema, wie man Anomalien in Daten aus Informationssystemen erkennen und visualisieren kann. Hierzu wurden Testdaten eines Europäischen Einzelhändlers verwendet, die mittels verschiedener Verfahren aus den Bereichen Statistik, künstliche Intelligenz und Data-Mining analysiert wurden. Die Ergebnisse der Analysen werden von der entwickelten Testplattform für den Benutzer verständlich aufbereitet und meist in Form von Graphiken und Ergebnislisten für den Benutzer präsentiert.

So können einfache Häufigkeitsberechnungen, Zeitreihen und komplexe Data-Mining Methoden wie Assoziations- und Sequenzanalyse oder Regelinduktionen verwendet werden, um den Datenbestand zu evaluieren und den Benutzer aus verschiedenen Visualisierungstechniken die Verständlichsten auswählen zu lassen.

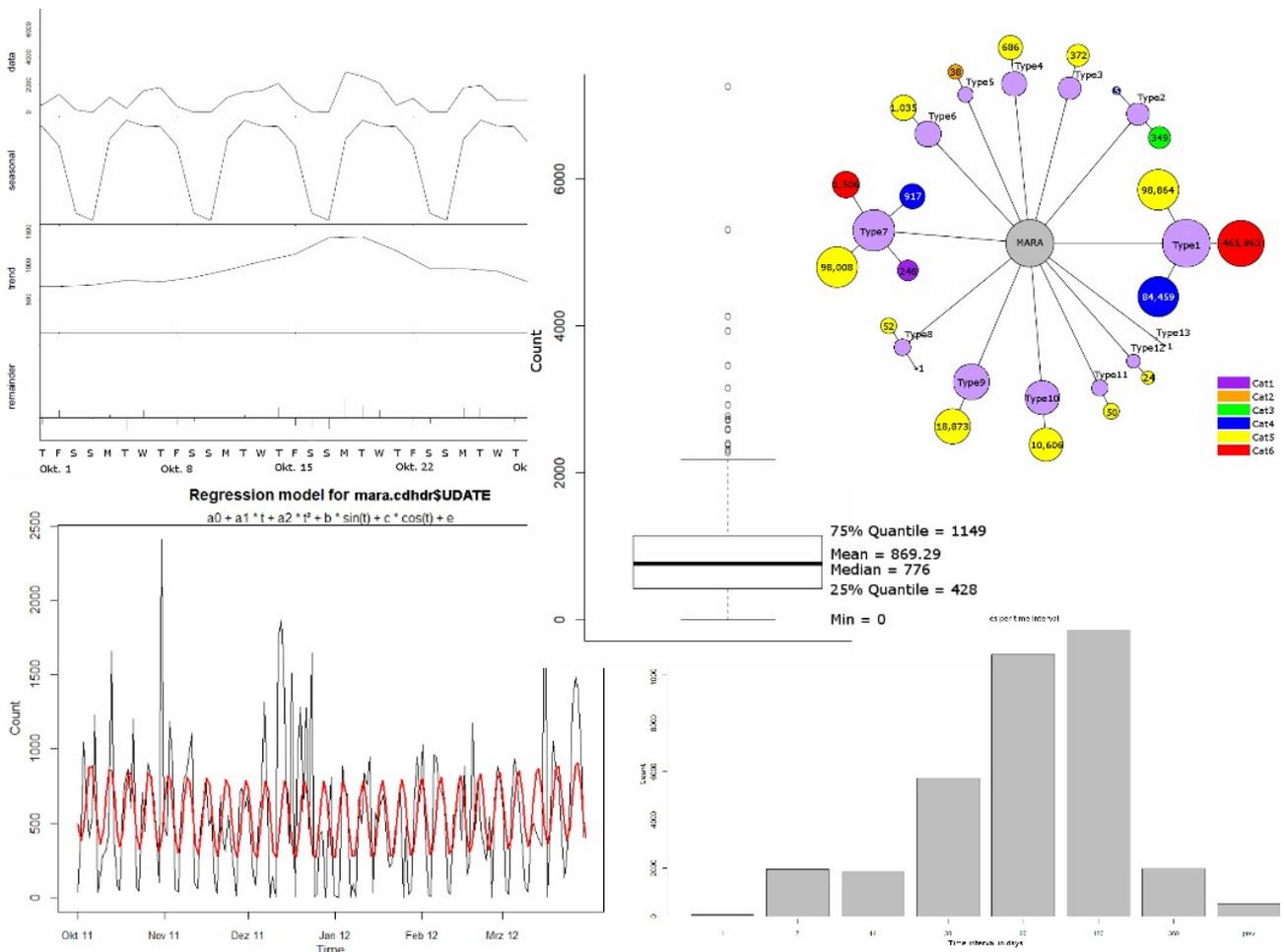
Die Thesis entstand am Institut für Softwaresysteme im Bereich Wirtschaftsinformatik und wurde von Prof. Dr. Rolf Krieger betreut. Die Inhalte und Ergebnisse sollen nun in einem Drittmittelprojekt weiter untersucht und vertieft werden, um aus der Testplattform ein Server-basiertes Tool zu entwickeln, mit dem die Analyse- und Visualisierungstechniken sowie die Benutzerschnittstelle noch weiter ausgebaut werden können.



Achim Guldner



Schematische Darstellung der Testplattform



Beispielhafte Ergebnispräsentation