

IAB-DiskAB

Einladung

Text Mining mit R am Beispiel von Stellenangeboten

Christian Koch (IT-Systemhaus, SEP6)

Das Thema Data Mining gewinnt in der Wirtschaft und in der Forschung zunehmend an Bedeutung. Predictive Modelling und Machine Learning prägen moderne Datenarchitekturen. Ein häufiger Anwendungsfall in den Sozialwissenschaften ist die strukturierte Analyse von Text. Ursprünglich entwickelt in den Digital Humanities, umfasst Text Mining heute verschiedenste Methoden. Sie reichen von der Bereinigung über die Visualisierung und Klassifikation bis hin zur Interpretation von Text. Doch wie kann man die Methoden konkret einsetzen? Welche Technologien sind gut für Text Mining geeignet? Welche Probleme stellen sich bei der Programmierung und wie kann man sie lösen? Mit diesen Fragen beschäftigt sich der Vortrag.

Die Präsentation gibt einen Überblick über grundlegende Methoden des Text Mining mit der Programmiersprache R. Nach einer theoretischen Einführung werden verschiedene Techniken am Beispiel von Stellenangeboten vorgestellt. Hierbei wird der Data Mining Zyklus einmal vollständig durchlaufen. Die Präsentation gibt einen Überblick über Beschaffung, Visualisierung, Bereinigung und Modellierung von Texten.

Montag, 23. April 2018

10:00 – 11:00 Uhr

Raum Re100 E10