



Bachelor-, Master- und Doktorandenseminar  
des Instituts für Informatik

## Data analysis pipelines for graph data with R

Jean Pascal Njiki Kouajie, B.Sc. , TU Clausthal

Die Pipeline ist eine weit verbreitete Technik zur Datenanalyse sowohl in Unternehmen, als auch im alltäglichen Leben. Eine große Menge an Daten kann in „Massenspeichern“ wie etwa graphische Daten gespeichert werden. Es wird essentiell zu wissen, wie die Technik der Pipelines genutzt werden kann um eine aussagekräftige Interpretation der graphischen Daten durchführen zu können um die Produktivität der Unternehmen verbessern zu können. Die R Programmiersprache ist dabei ein starkes Werkzeug, das für diesen Zweck genutzt wurde. In dieser Arbeit wurden zwei Arten von Pipelines konstruiert. Die „Direkte Pipeline“, bestehend aus vier analytischen Levels (primär, sekundär, tertiär und quartär) und die „Indirekte Pipeline“, bestehend aus fünf analytischen Levels (primär, sekundär, tertiär, quartär und quintär). Die graphischen Daten wurden erfolgreich in dem R programmierenden Milieu rekonstruiert.

Beide, die direkte und indirekte Pipeline, ergaben identische Graphen und bewiesen so ihre Skalierbarkeit. Der Vergleich der beiden Modelle zeigte, dass die indirekte Pipeline günstiger ist, da es weniger Systemquellen (CPU, Memory) benötigt.

Donnerstag, den 30.06.2016, 11 Uhr s.t. im  
Seminarraum 210, Ifl, Am Regenbogen 15