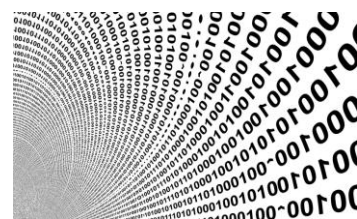


### **Warum ist das interessant?**

Die Aufwändigkeit der Berechnungen eines Algorithmus lässt sich mit Hilfe der Komplexität gut abschätzen. Bei der Realisierung des Algorithmus kann es leicht passieren, dass die tatsächliche Performanz der Implementierung von der theoretischen Abschätzung abweicht. Durch systematische Beobachtung kann man ein statistisches Performanzprofil des Algorithmus erstellen, das die Identifikation derartiger Abweichungen ermöglicht. Stehen mehrere alternative Algorithmen zur Auswahl, kann man deren Profile vergleichen, um den bestmöglichen auszuwählen. Dies ist insbesondere im Big Data Umfeld interessant, da sich die Charakteristiken der analysierten Daten zur Laufzeit ändern können, und damit eine dynamische Änderung der Algorithmen-Auswahl erforderlich machen. Die Arbeit erfolgt im Kontext des EU Projekts QualiMaster, in dem eine derartige adaptive Big Data Analyse Infrastruktur entwickelt wird.



### **Was soll ich tun?**

In dieser Arbeit sollen Sie verschiedene existierenden Ansätze zur Ermittlung von Performanzprofilen (z.B., [1, 2, 3]) auf ihre Anwendbarkeit im oben beschriebenen Kontext überprüfen und charakterisieren. Dann analysieren Sie ausgewählte Ansätze indem Sie Prototypen (existierende oder eigene erstellt bspw. mithilfe des Statistik-Werkzeuges R [4]) zur Performance-Messungen existierender Algorithmen einsetzen. Abschließend realisieren Sie den aussichtsreichsten Kandidaten als Java-Version (basierend auf Statistik-Bibliotheken) und integrieren diese in die QualiMaster Infrastruktur.

### **Welche Ergebnisse werden erwartet?**

Ihre Arbeit diskutiert die konzeptuelle und analytische Bewertung der Ansätze, sowie das Design, die Implementierung und die Validierung des Java-Prototyps. Weiterhin sollte die Arbeit die Integration in die QualiMaster Infrastruktur sowie Verbesserungsvorschläge der Ansätze im Anwendungskontext diskutieren und evaluieren.

### **Was bringt mir das?**

Neben den Kreditpunkten und der Note für den oben genannten Typ dieser Arbeit:

- Einblicke in die Performanz-Analyse von Software-Komponenten
- Einblick in aktuelle Forschung im Bereich Big Data
- Kenntnisse der Statistik-Software R
- Vertiefung ihrer Kenntnisse der Programmiersprache Java

[1] I. Shafer, K. Ren, V. N. Boddeti, Y. Abe, G. R. Granger, C. Faloutsos, RainMon: An Integrated Approach to Mining Bursty Timeseries Monitoring Data, KDD'12, 1158-1166

[2] S. Spinner, G. Casale, X. Zhu, S. Kounev, LibReDE: A Library for Resource Demand Estimation, ICPE'14, 227-228

[3] T. Zheng, M. Litoiu, M. Woodside, Integrated Estimation and Tracking of Performance Model Parameters with Autoregressive Trends, ICPE'11, 157-166

[4] The R-Project for Statistical Computing, <http://www.r-project.org>

### **Kontakt**

Dr. Holger Eichelberger  
eichelberger@sse.uni-hildesheim.de

### **Aufteilung der Arbeit**

Theorie	Implementierung	Literatur
30%	50%	20%