



Bachelorarbeit / Masterarbeit

DBMS-Independent Data Quality Metrics in Java

Im Rahmen eines Forschungsprojekts versuchen wir aktuell, Metriken zur Messung der Datenqualität relationaler Datenbanken in der Forschungsliteratur zu identifizieren und anschließend in einem Softwareprototypen zu implementieren. Dieser Prototyp soll in Zukunft die Basis für weitere Anwendungen im Bereich des sogenannten Process Minings bilden (siehe empfohlene Einstiegsliteratur).

Eine mögliche Bachelor- oder Masterarbeit in diesem Bereich beinhaltet folgende Elemente:

- Literaturarbeit: Verifikation der bestehenden Liste an Datenqualitätsmetriken sowie Identifikation weiterer Metriken in der relevanten Forschungsliteratur
- Strukturierung: Klassifikation der Metriken anhand eines ausgewählten Frameworks sowie Mapping zu Datenqualitätsdimensionen
- Implementierung: Design und Erstellung eines Softwareprototypen (voraussichtlich eine Java-Library) mit ausgewählten Metriken zur Anwendung auf beliebige relationale Datenbanken

Die Arbeit kann sowohl auf Deutsch als auch auf Englisch verfasst werden. Programmierkenntnisse (gewünscht: Java, Ersatz: Python o.ä.) werden vorausgesetzt.

Empfohlene Einstiegsliteratur:

- Richard Y. Wang & Diane M. Strong (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, *Journal of Management Information Systems*, 12:4, 5-33
- Scannapieco M., Missier P., Batini C. (2005) Data Quality at a Glance. *DatenbankSpektrum*, 14(January):6-14
- Kaiser M., Klier M., Heinrich B. (2007) How to Measure Data Quality? A Metric Based Approach. *International Conference on Information Systems*, 4801:16
- van der Aalst W. et al. (2012) Process Mining Manifesto. In: Daniel F., Barkaoui K., Dustdar S. (eds) *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing*, vol 99. Springer, Berlin, Heidelberg

Betreuer: van Dun, Christopher, M.Sc.