



Abschlussvortrag Masterarbeit Hamid Khalid

„Calculating the Information Content of Datasets in a Data Marketplace“

Data plays an important role in business decision making and the demand of high quality data is constantly increasing. Data marketplaces act as a platform where high quality data can be traded as a commodity. Since quality of the dataset varies from one data consumer to another, it is necessary that the consumer is allowed to judge the data quality before paying for the whole dataset. There is a framework which allows the data consumer to verify the dataset quality while maintaining the factors of privacy and trust by not exposing the contents of the dataset. However, It also leads to the problem of the framework abuse, allowing the data consumer to gain explicit or implicit knowledge by passing multiple requirements as an input while not providing any form of compensation to the data provider. Ideally, the data consumer must pay for the information acquired during data quality process even if he does not decide to purchase the whole dataset. In this thesis, we introduce and realize a model which calculates the amount of information referred to as an Information Content, extracted by the data consumer as a result of input requirements so it can be used to relatively charge the consumer for the acquired information during data quality check. Information Content is calculated for each requirement separately and denotes the knowledge gained by the data consumer directly or indirectly. Sum of Information Content will be directly translated to the sum of total amount which is needed to be paid by the data consumer. This model will help to improve the process of dataset trading between the data provider and consumer in the context of the data marketplaces and the quality check procedure.

Betreuer der Arbeit: Prof. Dr. Andreas Rausch, PD Dr. Christoph Knieke

Datum: Donnerstag, 09. Dezember, 10:00 Uhr

Ort: Online-Meeting über BBB

Link: <https://webconf.tu-clausthal.de/b/sim-uc9-rvy>