# Daten- und Fehlerbeschreibung

Trainingsdaten:

Bei den Trainingsdaten handelt es sich um den LUBM-Benchmark <http://swat.cse.lehigh.edu/projects/lubm/>

Der LUBM-Datensatz beschreibt eine oder mehrere Universitäten mit verschiedenen Einrichtungen, Personen und weiteren Merkmalen. Der Datensatz kann in (fast) beliebiger Größe automatisch erstellt werden und ist vollständig und fehlerfrei.

Anschließend wurde jeder Person im Datensatz noch ein Alter hinzugefügt.

LUBM-Datensatz + Eigenschaft “age” (Integer) für jede Person

Festgelegte Altersbereiche für Personen:

- UndergraduateStudents: range(17 - 30)

- GraduateStudents: range(20 - 35)

- AssociateProfessor: range(30 - 45)

- AssistantProfessor: range(32 - 45)

- FullProfessor: range(35 - 70)

## Daten:

* 74k\_w\_age: Datensatz ohne Fehler
* SetX: Datensatz mit Fehlern
* SetX\_errors: Enthält nur die fehlerhaften Datensätze
* mergedGraph257: enthält Datensätze 2, 5 und 7

### Set2:

Ca. 1 % der UndergraduateStudents wurde ein Fehler hinzugefügt:

Format des Alters auf String geändert.

### Set5:

Ca. 1 % der UndergraduateStudents wurde ein Fehler hinzugefügt:

Altersbereich range(55 - 76)

### Set7:

Alle Personen haben eine Telefonnummer mit Muster (xxx-xxx-xxx).

Ca. 1 % der UndergraduateStudents wurde ein Fehler hinzugefügt:

Muster: xxx-xx-x