

## Masterarbeit

# Effizienz und Effektivität mehrstufiger Lernverfahren zur Qualitätsbewertung von Texten

### Motivation

Von Internet-Nutzern generierte Online-Enzyklopädien, insbesondere Wikipedia, stellen heutzutage für viele Menschen die wichtigste Wissensquelle dar. Der enorme Umfang solcher Enzyklopädien erweckt den Bedarf, die Qualität der enthaltenen Texte automatisch bewerten lassen zu können. Algorithmische Verfahren für dieses Problem analysieren einen Text schrittweise bezüglich einer Reihe von Mängeln, etwa mit dem Ziel, diejenigen Texte zu identifizieren, welche mindestens einen Qualitätsmangel aufweisen. Dabei können sowohl die Reihenfolge als auch die Tiefe der Analyse-schritte signifikanten Einfluss auf die Effizienz und Effektivität des Bewertungsprozesses haben.



### Aufgabenstellung

Ziel der Masterarbeit ist einerseits, bereits existierende Klassifikationsalgorithmen der Bauhaus-Universität Weimar zur Qualitätsbewertung von Texten zu verfeinern und für verschiedene Tiefen der Analyse anzupassen. Andererseits soll eines der am s-lab entwickelten Verfahren zum effizienten Scheduling mehrstufiger Textanalyse-Prozesse geeignet ausgewählt und auf die Qualitätsbewertung übertragen werden, um die Reihenfolge ihrer Analyseschritte für die Erkennung von Texten mit Qualitätsmängeln zu optimieren.

Auf Basis einer zu implementierenden Umsetzung des Scheduling-Verfahrens für die Klassifikationsalgorithmen, sollen dann Effizienz und Effektivität verschiedener Analysetiefen der Qualitätsbewertung anhand eines großen Wikipedia-Datensatzes evaluiert werden, welcher im Vorhinein bezüglich der für die Aufgabe relevanten Kriterien zu analysieren ist.

### Voraussetzungen

Sicherer Umgang mit den Grundlagen des maschinellen Lernens

### Bearbeiter

Daniel Mex

### Wissenschaftliche Leitung

Prof. Dr. Gregor Engels

### Ansprechpartner

s-lab – Universität Paderborn  
Henning Wachsmuth  
Raum ZM1.03-11, Zukunftsmeile 1  
Tel: +49 (0) 5251 / 60-3359  
Email: hwachsmuth@s-lab.upb.de