



DIGITAL TOTAL

Beitrag ID: 70 Beitragskennung: 123

Typ: Poster

## 123 - DisKo –Das Diversitäts Korpus

DisKo steht für Diversitäts-Korpus und ist ein literaturwissenschaftliches Projekt mit Digital-Humanities-Komponente. Mit Methoden des maschinellen Lernens trainieren wir einen Machine Learning Classifiers, der nicht nur weibliche, männliche und neutrale Genderrollen in literarischen Texten aus unterschiedlichen Epochen erkennt, sondern auch weniger binäre Genderzuschreibungen. Für den überwachten Trainingsprozess bauen wir in Kooperation mit der Deutschen Nationalbibliothek Frankfurt ein Trainingskorpus auf, das Texte beinhaltet, in denen nicht-binäre Genderzuschreibungen vorkommen. Für die Korpuskonstituierung nutzen wir *Citizen Humanities*. Die manuelle Annotation der Trainingstexte bildet die Grundlage für das Machine Learning. Den projekteigenen Workflow von der Textbeschaffung bis zum Training des Classifiers und dessen Anwendung möchten wir auf der “Digital Total” vorstellen.

### Find me @ my poster

1, 2

### Keywords

Computational Literary Studies  
Gender Studies  
Citizen Humanities  
Machine Learning

### TentID

**Autoren:** Dr. SCHUMACHER, Mareike; FLÜH, Marie