

PD. Dr. Burkhard Meyer-Sickendiek
Dr. Hussein Hussein
Dr. Timo Baumann:

How to Develop Speech Annotations for Machine Learning Techniques

During the last ten years, two different ideas or aspects of "digital" working in the humanities emerged: The creation, dissemination and use of digital data or repositories of such digital content on the one hand, and the computer-assisted analysis and preparation of digital repositories using the most modern algorithmic and informal methods on the other hand. While the first was originally developed from the humanities itself and has been often referred to as "digital humanities" (DH), the second implies a stronger presence of computer-based approaches and could therefore be described as "computational humanities", as long as it is more focussed on machine learning techniques. Our workshop is dedicated to the new techniques in this field of machine learning, which is one of the most vibrant areas of "digital humanities".

In particular, we want to focus on possible applications with limited data sets and the handling of difficult problems within the scope of the Reinforcement Learnings. An example for these state-of-the-art techniques for automatic classification and recognition are the statistical probabilistic Gaussian mixture model (GMM) and Hidden markov model (HMM), Neural Networks (NNs) and convolutional neural network (CNN) as well as Machine Translation (MT). Some of these techniques are based on acoustic data, for example HMM and GMM, others are based on both acoustic and textual data like NNs.

A typical example for the use of machine learning techniques in digital humanities is speech recognition. A number of machine learning experiments on prosodic features have found that different genres like broadcast news or meetings, differ in speaking style. These prosodic features provide information about duration, pause, or intonational contours, and can be extracted from the automatic alignments of word and phone transcriptions with the speech signal. The automatic classification include feature extraction, which plays an important role in audio and image processing. The features can be classified to temporal and spectral features. The background of our workshop is a three-year project "Rhythmicalizer. A digital tool to identify free verse prosody", localized at the free university Berlin and funded by the Volkswagen Foundation in the program "Mixed Methods in the Humanities". In this project we examine "lyrikline" (<https://www.lyrikline.org/>), the most famous online portal for spoken poetry. The analysis of this acoustic corpora means to carry out the annotation work to identify different patterns in the database as well as to analyse it by identifying patterns. We are specialized on the identification of prosodic and textual features in Poems by using different tools for the following tasks: PoS-tagging, alignment, intonation, phrases and pauses, and tempo. The forced alignment toolkit is used to align audio and text. The annotation of accentuation and intonation phrase which constitute the prosody in the tone and break-index (ToBI) paradigm is used to identify the intended prosodic grouping of a poem as well as the pitch accents and boundary tones. This analysis is a first step towards an automatic classification based on machine learning techniques. Our main example in the workshop will be WEKA which is a collection of machine learning algorithms used in the for data mining area.