# NLP of TV Series Scripts for Analysis of Characters and Topics

Deborah Djon and Lea Vergari, Informatik
**Supervised by: Monika Kochanowski, Informatik**

Project Description

AI can be used within NLP for giving a user a neutral insight into the content of a TV series, namely 'Stargate SG-1'. This students' work [1] focuses on analysis of the written scripts of TV series over a large number of episodes and seasons[1]. The goals are to identify and analyze the main characters and their sentiment, the effect of gender on the dialog, the evolution of topics in the series, as well as giving an overview of single episodes with various techniques [2, 3, 4].

Word Cloud, Example Topics, Sentiment and Speech Proportion
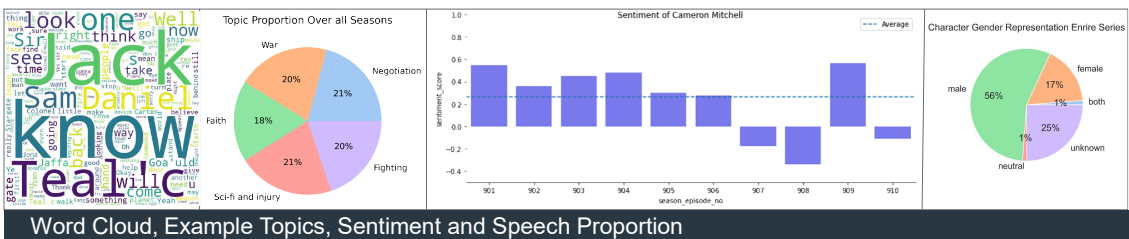
Results of the Analysis

With manual preprocessing and statistical analysis we applied AI unsupervised techniques for topic detection [4].
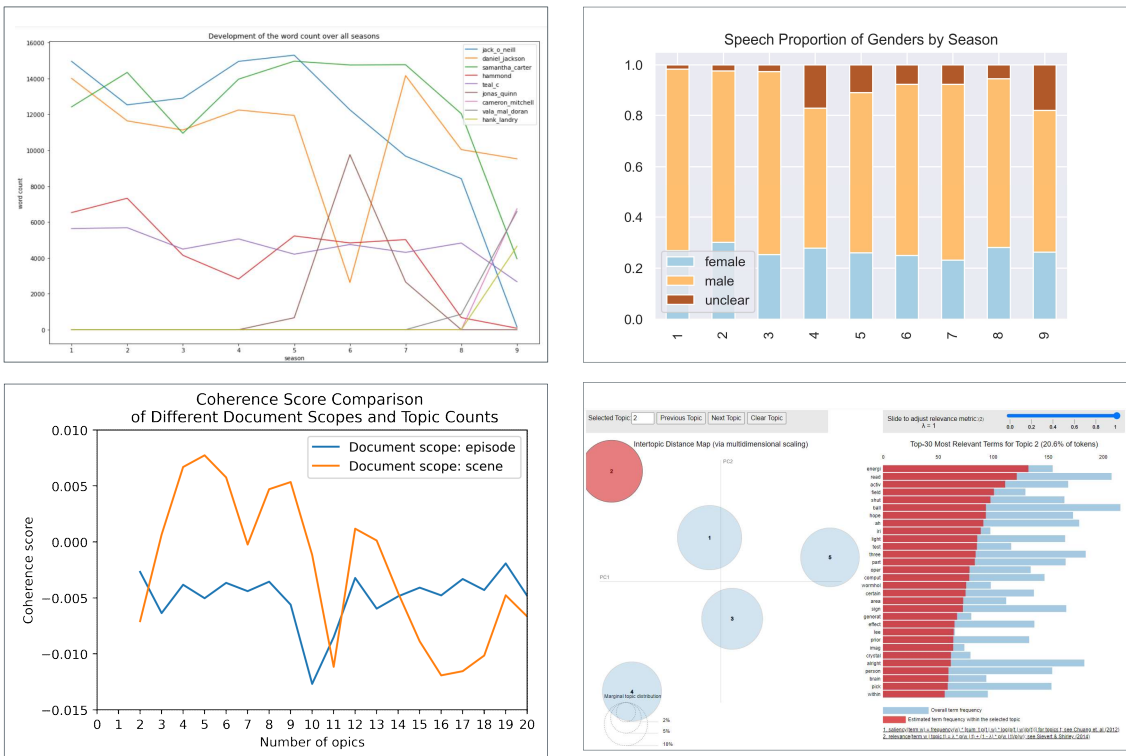
Topic Detection

Finding topics with a latent dirichlet allocation [4] shows challenging in text-only based fields, as dialogue only gives hints on the actual plot.

» Fitting and evaluating topic models is crucial

» Several metrics exist; however, they are only helpful in combination with expert knowledge for labelling

» Speech seems to differ from prose, for topic modeling this should be part of future research

Results

NLP methods are applicable for analysis of character evolution and topic detection based on textual data of series scripts.

» Identification of main characters is possible, if they speak enough, showing their participation over time

» Topic detection shows varying content in episodes and seasons

» Sentiment analysis on text only is possible

» Men and women speak about the same topics, but as more men appear they speak on the whole more.

Outlook

Identified challenges include the automatic gender detection as well as description of identified topics. Topic models can be improved as well. Additionally, it would be an area of further research to generate synopsis based on text or to include other types of non-text content like voice or video content in the analysis.

Used AI technologies
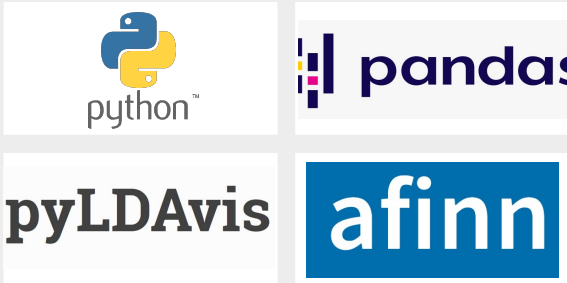
Sources

» [1] D. Deborah and L. Vergari, Natural Language Processing of Movie Scripts to Analyse the Plot Content, Characters, and Their Evolution (Student Research Paper), Stuttgart: DHBW, 2022.

» [2] S. Ghosh and D. Gunning, Natural language processing fundamentals: Build intelligent applications that can interpret the human language to deliver impactful results, Birmingham: Packt Publishing, 2019.

» [3] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, p. 82–89, 4 2013.

» [4] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, vol. 3, p. 993–1022, 2003.

[1]Data available at The Internet Movie Script Database (IMSDb) https://imsdb.com

Kontakt

Duale Hochschule Baden-Württemberg

Rotebühlplatz 41
D-70178 Stuttgart
monika.kochanowski@dhbw-stuttgart.de

Alle Informationen finden Sie unter: **www.dhbw.de**