

Proposal NLP Project - Review Writing Helper

Dataset:

<https://www.kaggle.com/rusiano/madrid-airbnb-data>

The Kaggle dataset contains AirBnB data. Including reviews from customers.

Special characteristic: Not all entries are in english language.

Abstract:

In recent years, NLP algorithms have become increasingly more powerful. Goal of this project will be to clean the data, as well as create a visual overview of the frequency of certain words and constellations.

Subsequently, word predictions and sentence completions are predicted.

This is done by:

- Stupid Backoff Algorithm
- Markov Chains

Afterwards, users can select other semantically similar words.

Problem to solve:

Language barriers can make it difficult for users to find the right words.

This problem should be solved.

Experiment(s) done:

The experiment will be an interactive console application in which the user can type sentences and then get suggested words and word groups for completion.

Analysis of results:

- Visualisation of the dataset → (Pots & Wordcloud)
- Calculation of right predicted words (SBO)
- Calculation of similarity of predictions (SBO and Markov)