Evaluation of customer reviews using Natural Language Processing

Graduate Candidates

Tobias Rothlin

Giovanni Triulzi

Definition of Task: The World Wide Web provides a tremendous amount of text data, which contains a lot of information. Especially hotel reviews contain extensive information about customer satisfaction with the hotel and its services. For a hotel employee, it is unattainable to analyze and evaluate this vast amount of data in a reasonable amount of time. In this bachelor thesis, a system will be developed which analyzes hotel reviews in both English and German. It classifies each sentence by its sentiment and category using Natural Language Processing. Statistical metrics are then collected and visualized in a dashboard. The dashboard design will be based on a dashboard concept developed by the team in St.Gallen. This team of business IT specialists will create a dashboard mock-up and define the relevant topics to classify. The team will conduct a study with two hotels and the WTT to analyze the system's requirements.

Approach: After researching the topic in more detail, different approaches were considered. The first step was to find a suitable hotel review dataset to train the machine learning models. After transforming the dataset and labelling, statistical machine learning models, e.g. random forest and support vector machines, were trained and evaluated. After achieving good results with statistical machine learning, the state of the art approach, based on a BERT model, was implemented and tested. BERT transforms the information in a sentence into a vector. Together with a classification head, the BERT model was then trained on the hotel review dataset.

Result: This thesis shows that it is possible to automatically extract information from text data using Natural Language Processing. In addition to developing the machine learning models to classify German and English reviews, an end-to-end system was designed to receive a review, split it into sentences, classify the sentences and save the results in a database. From this database, a dashboard loads the data and visualizes the results.
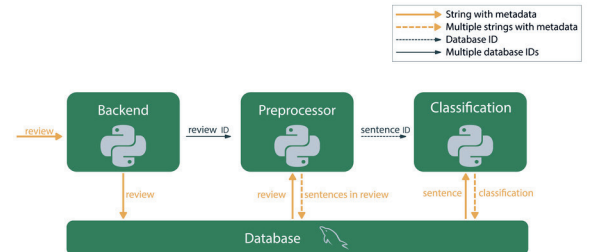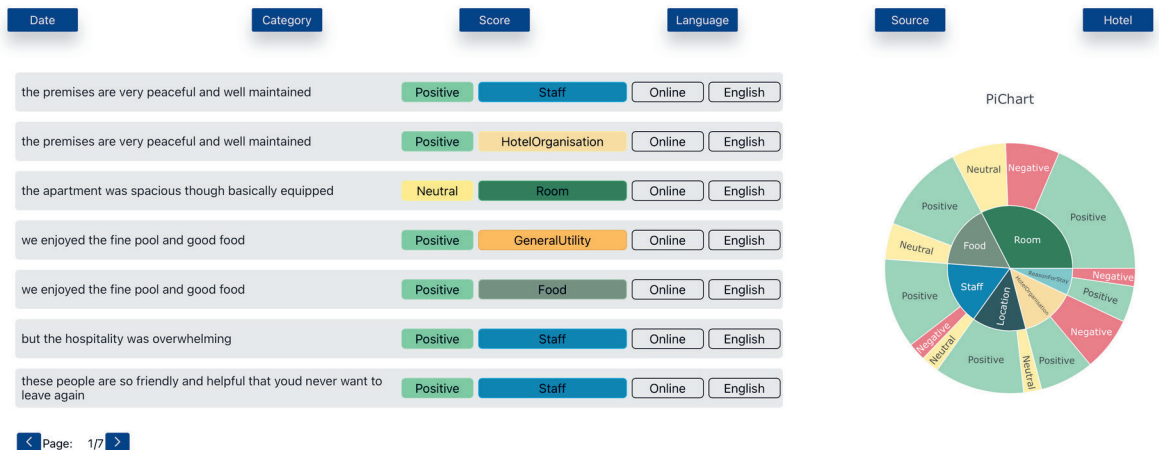
NLP Tree
Own presentment
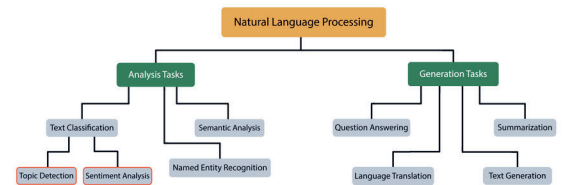
Classification Pipeline
Own presentment

Dashboard
Own presentment

Advisor
Hannes Badertscher

Co-Examiner
Gabriel Sidler, Teamup Solutions AG, Zürich, ZH

Subject Area
Artificial Intelligence

Project Partner
Wissenstransferstelle Wirtschaft (WTT) der OST, St. Gallen, SG