Informatikseminar

Projects

Prof. Dr. Mascha Kurpicz-Briki

FS 2021

Project 1 – Training Word Embeddings
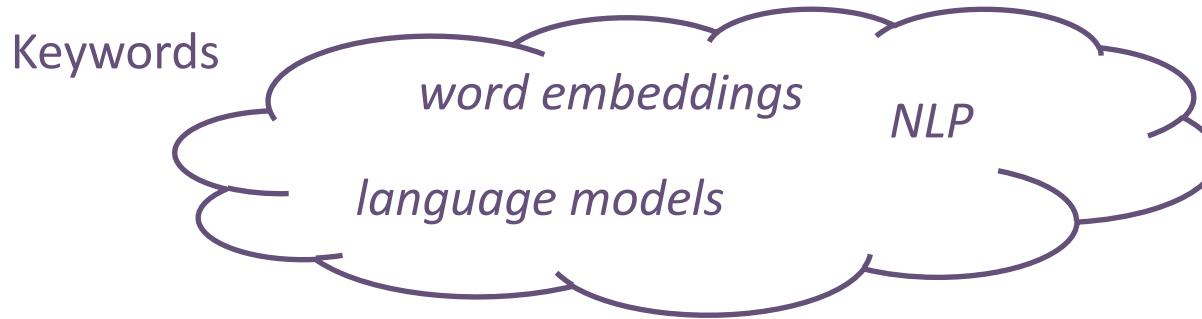
Keywords

word embeddings

NLP

*language models*

Word Embeddings are language models that represent each word as a vector. Due to this, we can imply meaning of the words using mathematical vector operations. To train such word embeddings, different algorithms can be used. In this project, we study the different algorithms and datasets used to train publicly available word embeddings, examine how they differ and summarize the advantages and disadvantages of each.

For further information, look for word2vec, GloVe, fasttext

Project 2 – Mitigation of Bias in Word Embeddings
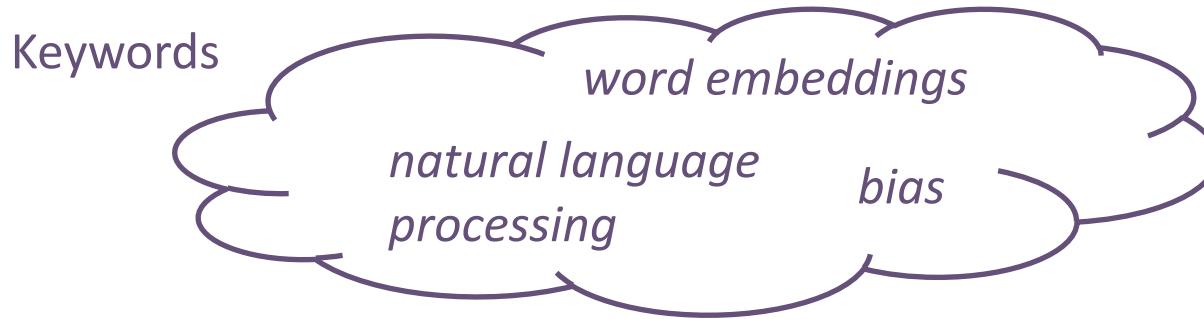
Keywords

word embeddings

natural language processing

bias

Word Embeddings are language models that represent each word as a vector. Due to this, we can imply meaning of the words using mathematical vector operations. However, these word embeddings also reflect stereotypes of the society included in the training data to build them. For example, «*man is to doctor, as woman is to nurse*». In this project, we investigate based on research literature, how such bias can be mitigated.

Additional information about the topic:

https://www.societybyte.swiss/2020/08/19/gesellschaftliche-stereotypen-in-vortrainierten-sprachmodellen/

Project 3 – Ethical Considerations in ML Projects

Keywords

Machine Learning

Digital Ethics

NLP

Project Management

Many examples from literature and practice show that software based on machine learning or natural language processing is often biased which leads to unfair decisions. For example, people of specific groups can be excluded from jobs or not get a loan. What can companies do to avoid their software behaving in such unfair manner? What best practices have been established to support their project management for such projects?

https://www.witty.works/post/voreingenommenheit-in-ki

Project 4 – Contact Tracing Applications

Keywords

SwissCovid   *COVID-19*

contact tracing

During the COVID-19 pandemic situation, so-called contact tracing apps are used to identify people that were in contact with positive-tested persons. In Switzerland, the SwissCovid app is used. In this project we investigate how the application works, and how it compares to other countries' solutions. Which solutions were more successful and what are the reasons for that?

https://www.bag.admin.ch/bag/de/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/swisscovid-app-und-contact-tracing.html

Project 5 – API Access to Online Platforms

Keywords

social media   *data*

API access
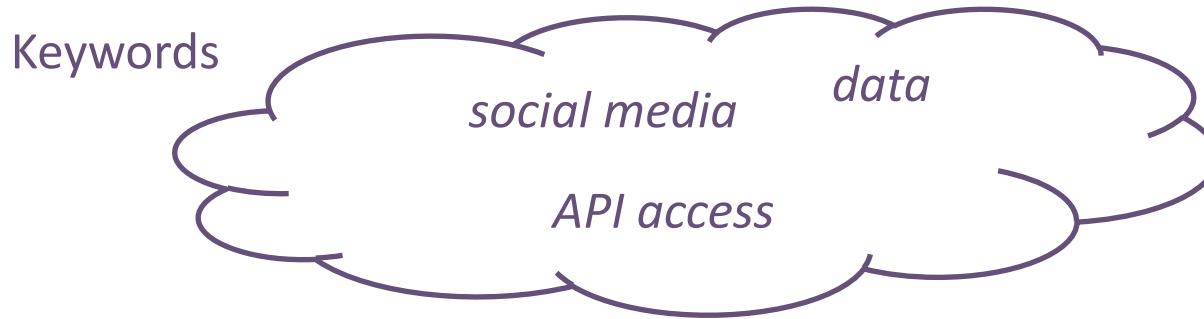
In internet research, often social media data such as Twitter or Facebook and search engine data for example from Google are used. However, due to privacy discussions and due to the value of such data, some providers have become more restrictive and do not provide access to everybody. In this project, we investigate what options developers currently have to access the data from social media and other online platforms.

https://developer.twitter.com/en/docs

Project 6 – TikTok – the future of social media?

Keywords

social media *TikTok*

API

The application TikTok has become very popular in the last years. Instead of text or image-based posts as on other social media, users post videos. In this project we investigate the following questions: what makes TikTok so successful? Who are the users? How does advertising work in TikTok? How can contents be accessed via API, and what data is available next to the videos itself?