



Big Data Scenarios

Task 1

Laurenz Brahner, Jan Landmann



Table of contents

1. Case studies:
 - a. **Summarization** of news articles
 - b. **NER** with sentences from wikipedia article
 - c. **Sentiment analysis** of yelp reviews
 - d. **Translation** of text with **language detection**
 - e. **Zero-shot classification** of recipes and corresponding cuisines
2. Limitations
3. Conclusion



Summarization of news articles

- With this code we demonstrate the ability of the transformer models to summarize texts.
- The user can select an article from a data set of different news articles, as well as the desired result length.
- The desired article is then summarized and output by the program.
- We used the „facebook/bart-large-cnn“



Summarization of news articles

Limitation:

- The maximum input length is limited. If only truncated, the missing context may influence the result negatively.

Summarization *Interface*

News article summarizer

Choose an article you would like to have summarized

Choose your desired summarization length

☐

Short

☐

Medium

☐

Long

Clear

Submit

Summarized Text

Flag

Summarization *Interface*

News article summarizer

Choose an article you would like to have summarized

Seven demonstrators were arrested for tryin_i ▼

Choose your desired summarization length

☒ Short

☐ Medium

☐ Long

Clear

Submit

Summarized Text

More than 50 people gathered outside the New York Public Library on Fifth Avenue to protest the grand jury decision not to indict Officer Darren Wilson in the August killing of unarmed black teenager Michael Brown. Protesters marched down the street carrying banners that read '

Flag

Summarization *Interface*

Choose an article you would like to have summarized

Seven demonstrators were arrested for tryin ▼

Choose your desired summarization length

☐ Short ☐ Medium ☒ Long

Clear

Submit

Summarized Text

More than 50 people gathered outside the New York Public Library on Fifth Avenue to protest the grand jury decision not to indict Officer Darren Wilson in the August killing of unarmed black teenager Michael Brown. Protesters marched down the street carrying banners that read 'Black Lives Matter' and indirectly calling police officers 'pigs' Police say protesters pushed against NYPD barricades and one of them flipped over a trash can. All the people arrested are facing disorderly conduct charges. NYPD Commissioner William Bratton had said in no uncertain terms that he would not let anyone disrupt the Thanksgiving celebration.

Flag



NER with sentences from wikipedia article

- For this purpose, a sentence can be selected from an already loaded Wikipedia article for which the NER is then performed.
- You can either filter for specific entity types or output all entity types.
- As a result, the program outputs a table showing the words sorted by frequency. The same information is also clearly displayed in a scatterplot.
- We used the „dbmdz/bert-large-cased-finetuned-conll03-english“ model



NER with sentences from wikipedia article

Limitations:

- The model in its current state sometimes struggles to categorize entities correctly.
- The maximum input sequence length is limited.

NER Interface

Named Entity Recognition

Choose a sentence from the wikipedia page

Choose the entities you want to see

If you select nothing, every entity will be selected

☐ Location

☐ Miscellaneous

☐ Person

☐ Organisation

Clear

Submit

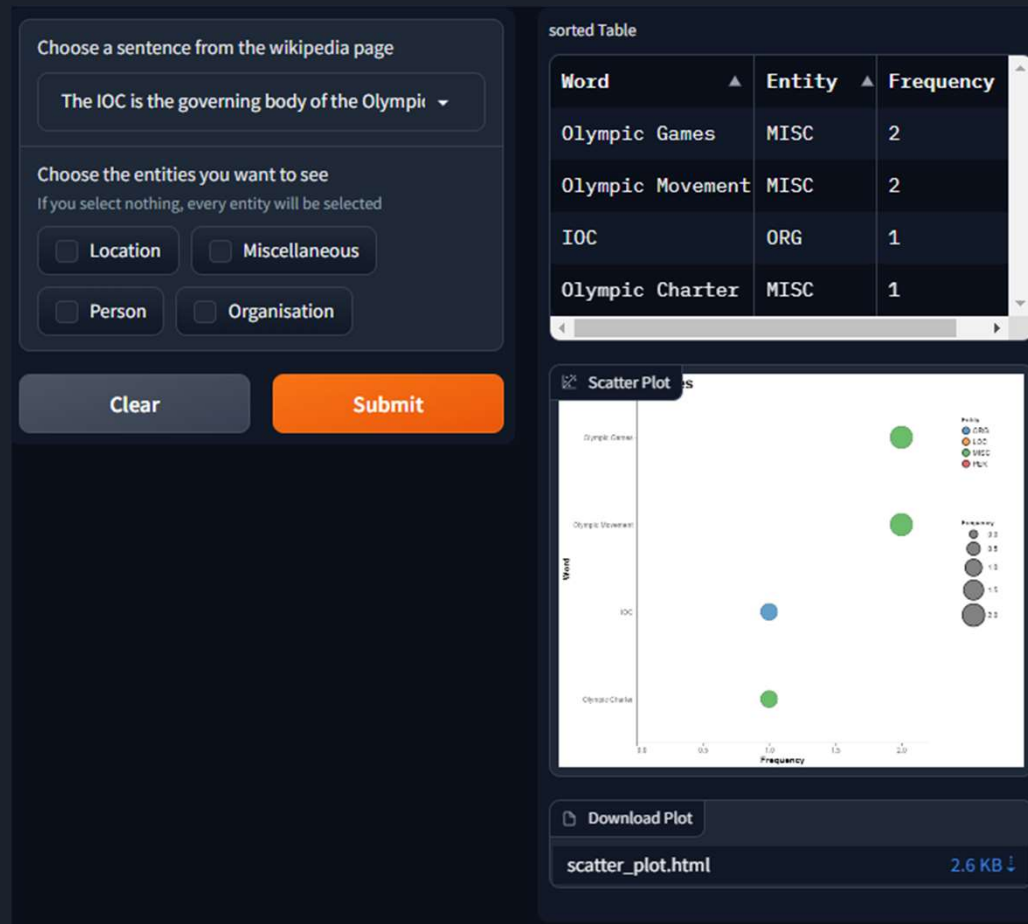
sorted Table

| 1 | 2 | 3 |
|---|---|---|
| | | |

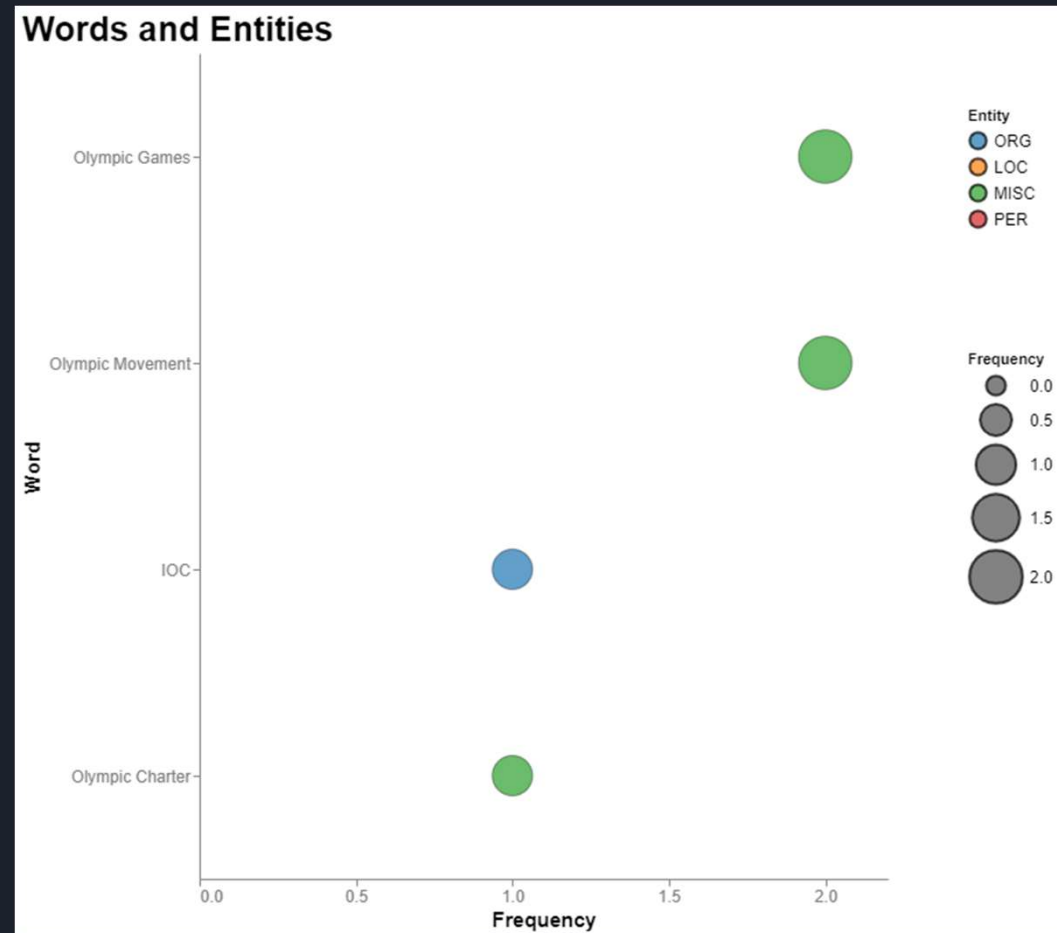
Scatter Plot

Download Plot

NER Interface



NER Interface



NER Interface

Choose a sentence from the wikipedia page

The IOC is the governing body of the Olympic ▼

Choose the entities you want to see
If you select nothing, every entity will be selected

☐ Location ☒ Miscellaneous

☐ Person ☐ Organisation

Clear Submit

sorted Table

| Word | Entity | Frequency |
|------------------|--------|-----------|
| Olympic Games | MISC | 2 |
| Olympic Movement | MISC | 2 |
| Olympic Charter | MISC | 1 |

Scatter Plot

Download Plot

scatter_plot.html 2.5 KB



Sentiment analysis of yelp reviews

- Here we demonstrate the ability of the transformer models to perform sentiment analysis.
- We tried different models for example the „*avichr/heBERT_sentiment_analysis*“ which also can classify as neutral.
- But the defaulted model: “distilbert-base-uncased-finetuned-sst-2-english” was more reliable
- You can either enter a text yourself or select a review from a dataset from yelp reviews.
- The program then determines whether the input is rated as positive or negative. The corresponding score is also displayed in numerical form.



Sentiment analysis of yelp reviews

Limitations:

- If the given input is neither positive nor negative the sentiment analysis is not able to determine whether the input is neutral.
- It also struggles to decode colloquial language or more complex inputs.

| | |
|--|--|
| <p>Review</p> <p>This place sucks, but the waitress looks good</p> | <p>Is the review postive?</p> <p>The supplied text is being labeled as: POSITIVE with a score of: 0.997.</p> |
|--|--|

Sentiment analysis *Interface*

Sentiment analysis of Yelp reviews.

Review

Enter Review manually

Choose a review from the dataset



Clear

Submit

Is the review postive?

Flag

Sentiment analysis *Interface*

Sentiment analysis of Yelp reviews.

Review

Enter Review manually

Choose a review from the dataset

Is the review postive?

Flag

eating at el-Charro's - note the empnasiz on the proper rorming or the Socarrat\n\nkegardless or all this. Common courtesy is something that I expect from any restaurant. This place comes up short on just about every thing.

Went in on a sunday afternoon. Place was dead. Wasn't greeted, sat at a table and was not helped. Waited 10 minutes, still no service. Walked out and still heard nothing from staff. \n\nlf time and service are important to you stay away from Hob Nob.

I have lived here in Phoenix for 9 months now and have had a tough time finding a hair salon as I have thick, curly hair. I have been buying groupons for many places trying them out and none have worked. I bought the groupon for The Green Room to try it out and am absolutely in love with them! \n\nSteph was my stylist and she was amazing! I came in with a specific color I wanted as I will be leaving for two out of town weddings this month and she consulted the owner to make sure I got the right color. They would not let me leave until my color was perfect (my cut was amazing as well)! Eric (the owner) even took time to blow dry my hair after Stephanie's next client came.\n\nI will definitely be going back in the future for cuts and colors. The place was cute and cozy and the staff was super friendly. They definitely made me feel like they cared about me and what I wanted where the other salons just put a color on and let me go. So glad I found them and have found my new salon!

Sentiment analysis *Interface*

Sentiment analysis of Yelp reviews.

Review

Enter Review manually

Choose a review from the dataset

Totally gone down the tube. Way overpriced ▾


Clear

Submit

Is the review postive?

The supplied text is being labeled as: **NEGATIVE**
with a score of: 0.998.

Flag



Translation of text with language detection

- Here we demonstrate the ability of the transformer models to translate texts with speech recognition.
- The user can enter any text in the input field and select the desired output language from a list of five languages.
- During input, the entered language is automatically recognized and translated into the selected language.
- For Language-Detection we used „papluca/xlm-roberta-base-language-detection“ for translation „Helsinki-NLP/opus-mt-{}-{} “



Translation of text with language detection

Limitations:

- Long sentences take a long time to translate



Translation *Interface*

Translator

Write your own text in any language you want

Choose an exapmle text from the dataset

Choose your target Language

Please select a target language, the input Language will be determined automatically

☐ German

☐ Italian

☐ Spanish

☐ French

☐ English

Clear

Submit

Translated text

Flag

Translation *Interface*

Translator

Write your own text in any language you want

Je suis sophie

Choose an exapmle text from the dataset

Choose your target Language

Please select a target language, the input Language will be determined automatically

☒ German

☐ Italian

☐ Spanish

☐ French

☐ English

Clear


Submit

Translated text

Ich bin Sophie.

Flag

Translation *Interface*



Write your own text in any language you want

Choose an example text from the dataset

alexla turn off the lamp in the living room ▼

Choose your target Language

Please select a target language, the input Language will be determined automatically

☐ German

☐ Italian

☐ Spanish

☒ French

☐ English

Clear

Submit

Translated text

alexla éteindre la lampe dans le salon

Flag



Zero-shot classification of recipes and corresponding cuisines

- Here we demonstrate the ability of the transformer models to perform a zero shot classification.
- You have the option of making an entry yourself, selecting from a pool of existing ingredient lists or having an entry selected automatically from the pool.
- In this example, the national cuisines serve as labels for the zero shot classification. The result is a table that is sorted in descending order by match. The results are also displayed graphically in a bar chart
- We used the „facebook/bart-large-mnli“ model



Zero-shot classification of recipes and corresponding cuisines

Limitations:

The model sometimes struggles to choose between different labels if the input is associated with multiple labels.

Zero-shot classification *Interface*

Zero-Shot Recipe Classification

Ingredients

Enter ingredients manually

Choose a recipe from the dataset

☐ Use Random Recipe

Choose countries

If nothing is selected, all options will be selected automatically.

☐ Chinese☐ Indian☐ Italian☐ Mexican☐ Southern Usa☐ Cajun Creole

☐ French☐ Thai

Clear

Submit

Used Recipe

This is the recipe you used

Table sorted score

| 1 | 2 | 3 |
|---|---|---|
| | | |

Barplot

Download plot here

Zero-shot classification *Interface*

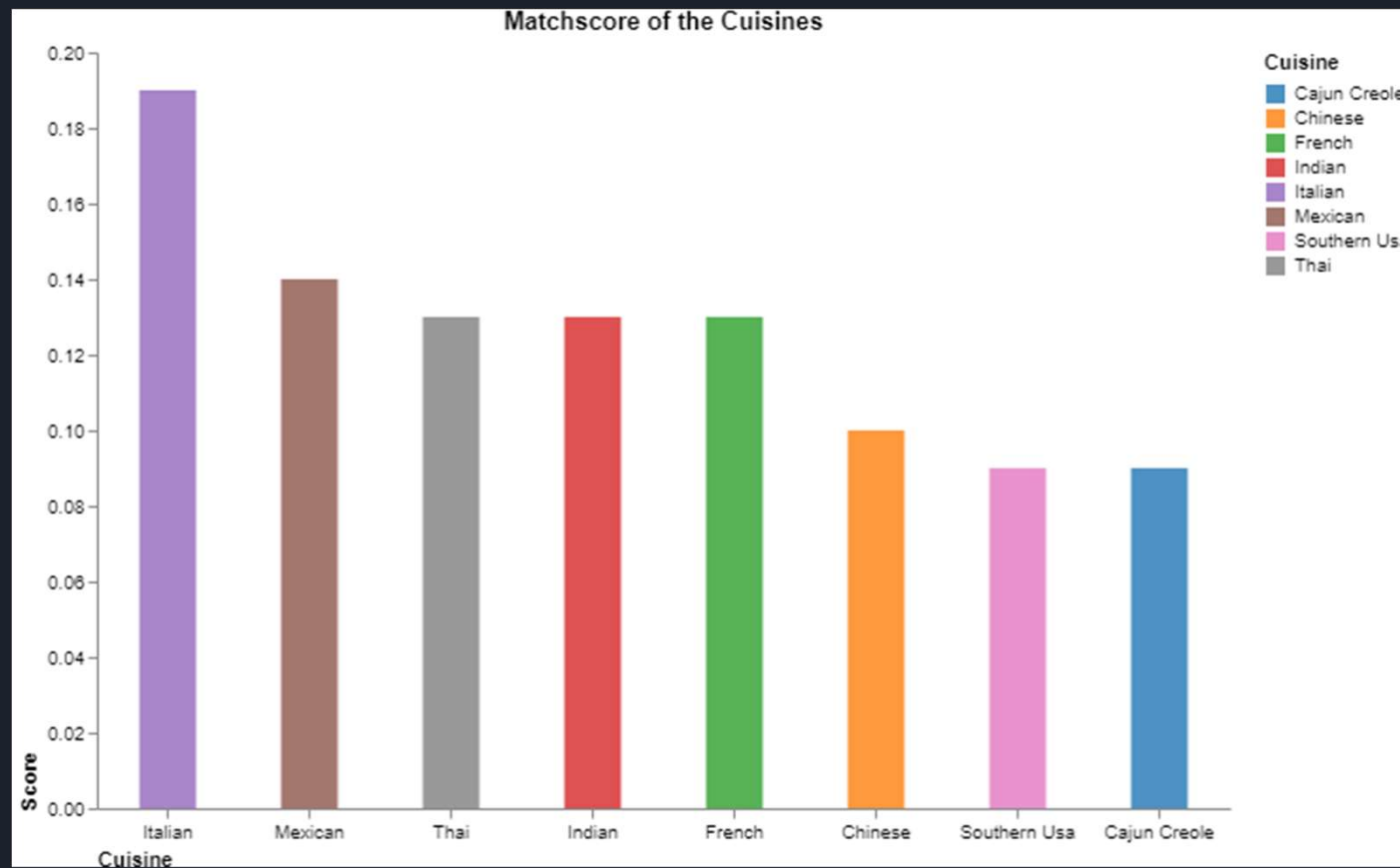
Used Recipe

ground black pepper, yellow corn meal, salt, olive oil, green tomatoes

Table sorted score

| Cuisine ▲ | Score ▲ |
|--------------|---------|
| Italian | 0.19 |
| Mexican | 0.14 |
| Thai | 0.13 |
| Indian | 0.13 |
| French | 0.13 |
| Chinese | 0.1 |
| Southern Usa | 0.09 |
| Cajun Creole | 0.09 |

Zero-shot classification *Interface*





Conclusion

- It works already quite well
 - The defaulted models worked better for us
 - but especially the Text Classification and the sentiment analysis had problems with our data
- So we have to train our models with our data to get better results!