

Digital humanities

Cvičení: Analýza sentimentu pomocí velkého jazykového modelu

Jindřich Marek

Cvičení 1: Analýza sentimentu

- technika zpracování přirozeného jazyka (NLP)
 - identifikace, extrakce a kvantifikace subjektivních informací a emočního náboje v textu
- cílem je určit postoj, názor nebo emocionální reakci autora k určitému tématu, produktu, osobě nebo události
- *kvantitativní analýza kvalitativních dat*
- *vhodné pro ukotvení subjektivních názorů na objektivní škále*

Přehled

- seznámíme se s analýzou sentimentu – určováním emocionálního tónu textu (kladný, záporný, neutrální)
- pracujeme s autentickými dopisy z 1. světové války a používáme velký jazykový model (LLM) pro automatickou analýzu
- chceme znát obsah a citové zabarvení dopisů

Cíle cvičení

- po dokončení tohoto cvičení budete schopni:
 - pochopit principy analýzy sentimentu
 - používat OpenAI API pro analýzu sentimentu
 - strukturovat výstup pomocí Pydantic
 - vytvářet reporty z analýzy sentimentu

Požadované nástroje

- JupyterLab notebook (VS Code)
 - pro psaní a spouštění kódu
 - dokumentace: <https://code.visualstudio.com/docs/setup/setup-overview>, <https://code.visualstudio.com/docs/datascience/jupyter-notebooks>
- LM Studio
 - pro spouštění lokálních modelů (grafické prostředí)
 - dokumentace: <https://lmstudio.ai/docs/>

Dataset

- několik osobních dopisů z doby první světové války
 - zdroj: web Imperial War Museum (IWM),
<https://www.iwm.org.uk/history/letters-to-loved-ones>
 - formát: TXT, každý dopis uložen v samostatném textovém souboru
 - jazyk: angličtina
 - téma: osobní zkušenosti, obavy, naděje, smutek

Letters To Loved Ones

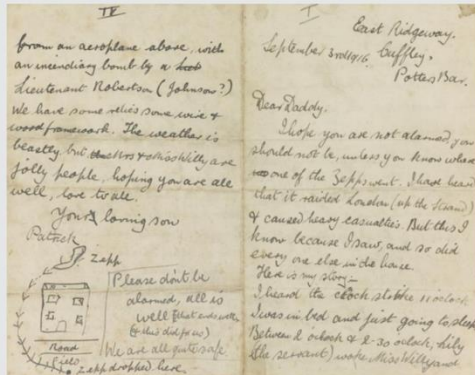
During the **First World War**, letter writing was the main form of communication between soldiers and their loved ones, helping to ease the pain of separation.

The British Army Postal Service delivered around 2 billion letters during the war. In 1917 alone, over 19,000 mailbags crossed the English Channel each day, transporting letters and parcels to British troops on the Western Front.

Soldiers wrote letters in spare moments, sometimes from front line trenches or in the calmer surroundings behind the lines. Censorship dictated what servicemen were permitted to disclose in their letters. However, in practice, men often found ways to impart information, and their letters offer a powerful and highly personal insight into the experience of war.

Receiving letters from family and friends was also vital to morale, keeping men and women connected to the homes they had left behind. Letters written on the home front to family and friends are today a fascinating source of information about everyday life in wartime Britain.

We hold around 7,500 collections of personal letters from the First World War in our archives. Here are just a few of them.



PRIVATE PAPERS

Schoolboy Patrick Blundstone to his father

Letter writing was a popular form of communication amongst all ages. Letters written by children during the First World War can be just as valuable a source of information as those written by servicemen. This letter from schoolboy Patrick Blundstone to his father contains a fascinating eye-witness account of the destruction in September 1916 of a Zeppelin airship near Cuffley in Hertfordshire by William Leefe-Robinson VC.

[See object record >](#)

Read the letter in full below.

'Dear Daddy, I hope you are not alarmed, you should not be, unless you know where one of the Zepps went. I have heard that it raided London (up the Strand) and caused heavy casualties. But this I know because I saw, and so did everyone else in the house.

Here is my story: I heard the clock strike 11 o'clock. I was in bed and just going to sleep. Between 2 o'clock and 2.30 o'clock, Lily (the servant) woke Miss Willy and told her she could hear the guns. Miss Willy woke Poolman and told him to wake me. He did so. Miss Willy helped Mrs Willy downstairs. We were all awake by

'Dear Daddy, I hope you are not alarmed, you should not be, unless you know where one of the Zepps went. I have heard that it raided London (up the Strand) and caused heavy casualties. But this I know because I saw, and so did everyone else in the house.

Here is my story: I heard the clock strike 11 o'clock. I was in bed and just going to sleep. Between 2 o'clock and 2.30 o'clock, Lily (the servant) woke Miss Willy and told her she could hear the guns. Miss Willy woke Poolman and told him to wake me. He did so. Miss Willy helped Mrs Willy downstairs. We were all awake by now, we had a Miss Blair staying with us for the weekend. We saw flashes and then heard "Bangs" and "Pops".

Suddenly a bright yellow light appeared and died down again. "Oh! It's alright" said Poolman. "It's only a star shell". That light appeared again and we Miss Blair, Poolman and I rushed to the window and looked out and there right above us was the Zepp! It had broken in half, and was like this: it was in flames, roaring, and crackling. It went slightly to the right, and crashed down into a field!! It was about a 100 yards away from the house and directly opposite us!!! It nearly burnt itself out, when it was finished by the Cheshunt Fire Brigade.

I would rather not describe the condition of the crew, of course they were dead - burnt to death. They were roasted, there is absolutely no other word for it. They were brown, like the outside of Roast Beef. One had his legs off at the knees, and you could see the joint!

The Zepp was bombed from an aeroplane above, with an incendiary bomb by a Lieutenant Robertson (Johnson?). We have some relics some wire and wood framework.

The weather is beastly but Mrs and Miss Willy are jolly people, hoping you are all well, love to all. Your loving son Patrick.

Postup

- [stažení modelu](#) v LM Studio
- [spuštění serveru](#) v LM Studio
- [spuštění kódu](#) v notebooku
 - instalace potřebných balíčků
 - import balíčků
 - načtení všech dopisů a předložení LLM
 - zjištění obsahu a kontextu
 - analýza sentimentu

Analýza sentimentu: úkol

- navrhnete prompt, který zjistí obsah a citové zabarvení dopisů
 - nejlépe v angličtině, protože v ní jsou zdrojové texty



<https://flinga.fi/s/FDBL752>

Prompt

Act as a social historian and answer the following questions:

- Who wrote the following letter to whom?
- What is the main theme?
- What is the sentiment of the text: Positive, negative, or neutral?

Jak vybrat správný lokální model

- velmi dobré pro tento úkol jsou „větší střední“ modely
 - využijí asi 14 GB RAM (ve výchozí kvantizaci)
 - např. Mistral Small 3.2 (24B)
- fungují i mnohem menší modely
 - využijí asi 3-6 GB RAM (ve výchozí kvantizaci)
 - ne tak přesné, více pravděpodobné faktické chyby: pro slabá PC
 - např. Ministral 3 (3B), Gemma 3 (4B), Gemma 4 E2B (2.3B/5.1B), Gemma 4 E4B
- zvolený slabší model: Gemma 4 E4B (4.5B/8B)
 - nový (duben 2026)
 - je možné spustit i v systému s 16 GB RAM (bez samostatné grafické karty)

analyza_sentimentu.ipynb ×

⚙️ 📄 📄 ...

cviceni > cviceni_08 > 01_analyza_sentimentu > analiza_sentimentu.ipynb > M↓ Analýza sentimentu historických dopisů pomocí LLM > M↓ 10. E

+ Code + Markdown | ▶ Run All ↺ Restart ☰ Clear All Outputs | 📄 View data ...

📄 jupyter (Python 3.12.12)

Analýza sentimentu historických dopisů pomocí LLM

Tento notebook ukazuje, jak využít velké jazykové modely (LLM) k analýze historických textů.

Zdroj dopisů: <https://www.iwm.org.uk/history/letters-to-loved-ones>

1. Instalace závislostí

Spusťte následující buňku pro instalaci potřebných knihoven.

▶ ▼

```
1 %pip install -Uqq openai pandas matplotlib ipywidgets
2 %pip install -Uqq wordcloud
```

[19]

Python

...

Note: you may need to restart the kernel to use updated packages.
Note: you may need to restart the kernel to use updated packages.

+ Code

+ Markdown

2. Import knihoven a nastavení

```
1 import json
2 import re
3 from pathlib import Path
4 from typing import List, Optional
5
6 import pandas as pd
7 import matplotlib.pyplot as plt
8 from openai import OpenAI
9 from pydantic import BaseModel, ValidationError
10
11 plt.rcParams["figure.figsize"] = (10, 6)
12 plt.rcParams["font.size"] = 12
```

[32]

Python

3. Konfigurace modelu

Porovnejte výstup z lokálního modelu vs. odpověď nástroje Copilot.

- zaměřte se na kvalitu a strukturovanost výstupu