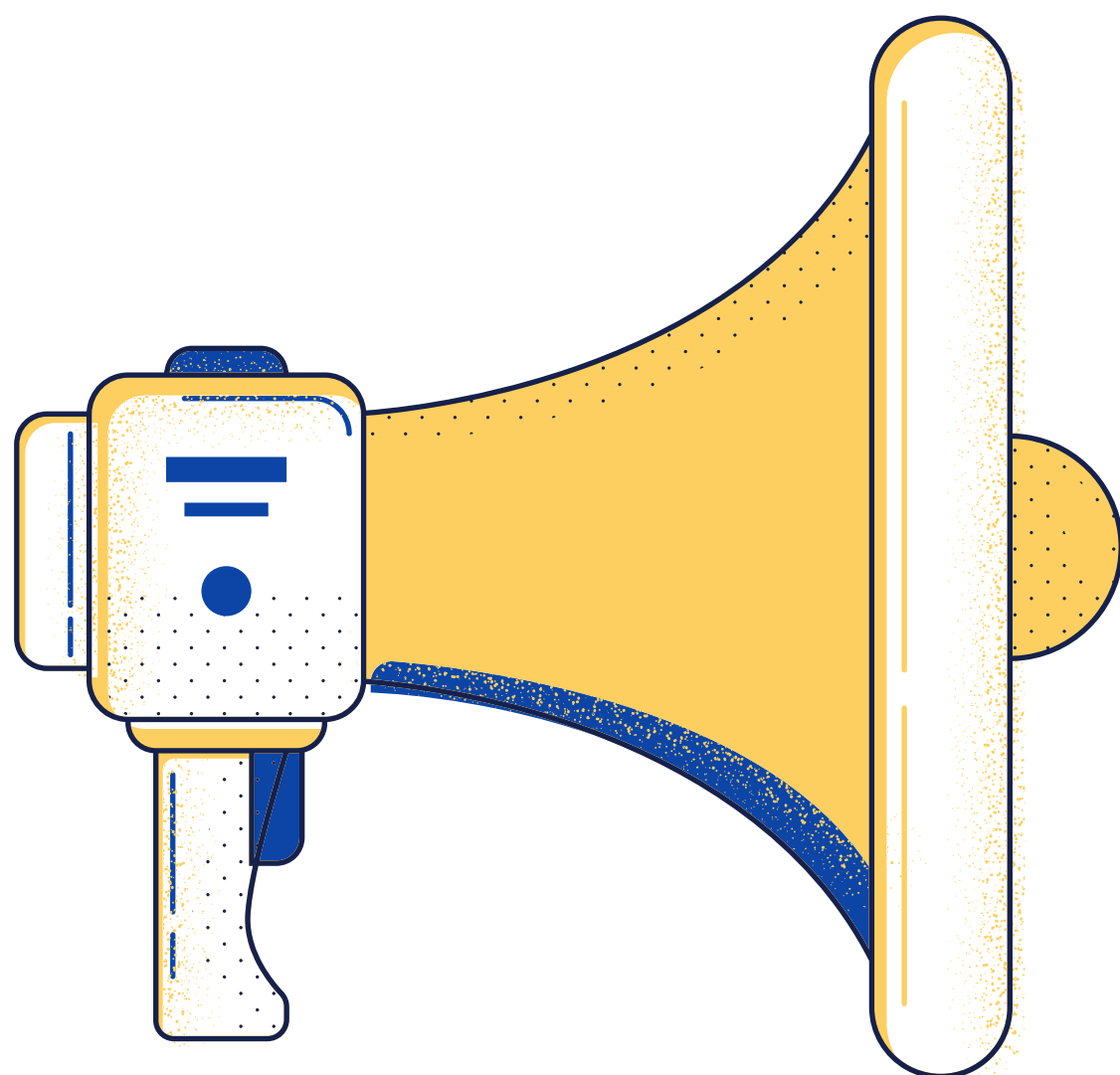


Tvorba spam filtru



Filip Kurl
Petr Kučera





- **popis technologie**
- **popis týmové práce**



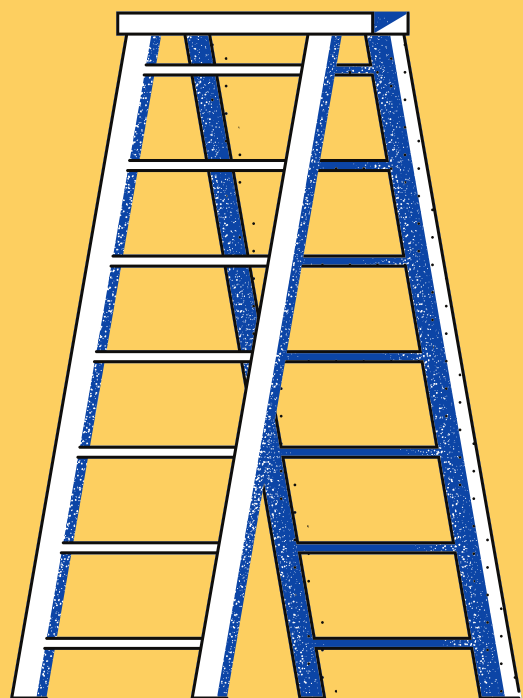
POPIS PRINCIPU ALGORITMU

- statická část
- dynamická část
- vyhodnocování



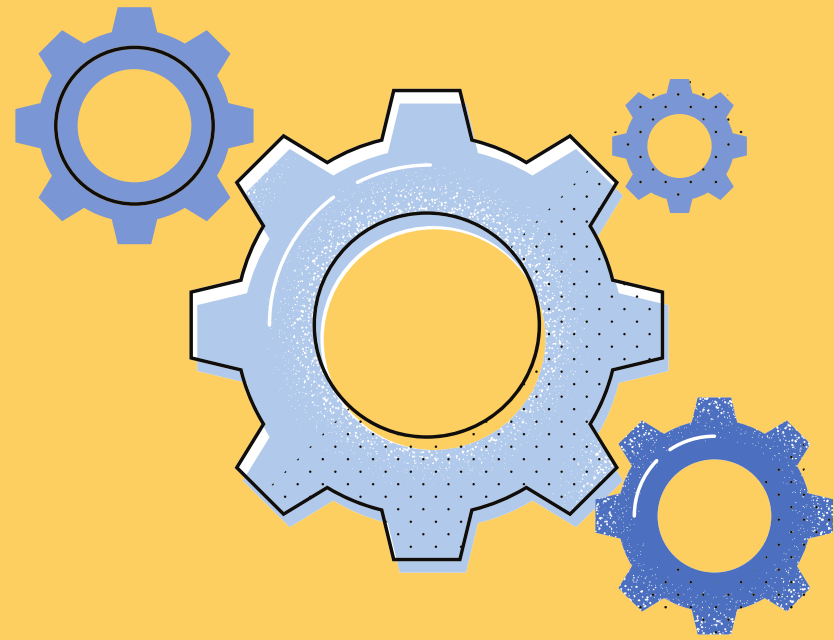
NAČÍTÁNÍ DAT

- tvorba modulu
- stopwords
- unifikování velikosti
- html tagy
- nahrazení složitých struktur



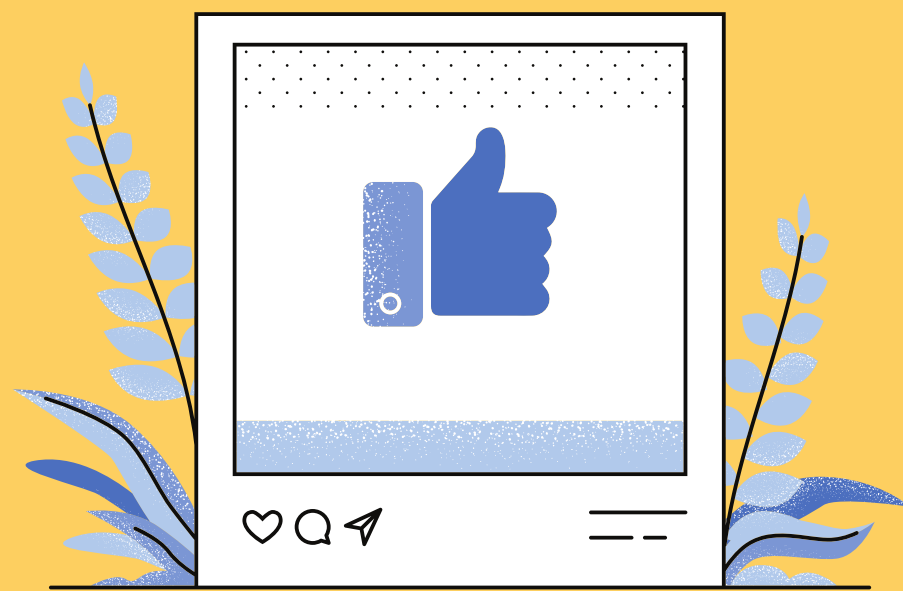
STATICKÁ ČÁST

- nadměrné užití znaků
- slova typická pro smap



DYNAMICKÁ ČÁST

- učení filtru
- Naive Bayes
- *porter_stemmer* z knihovny *nltk*



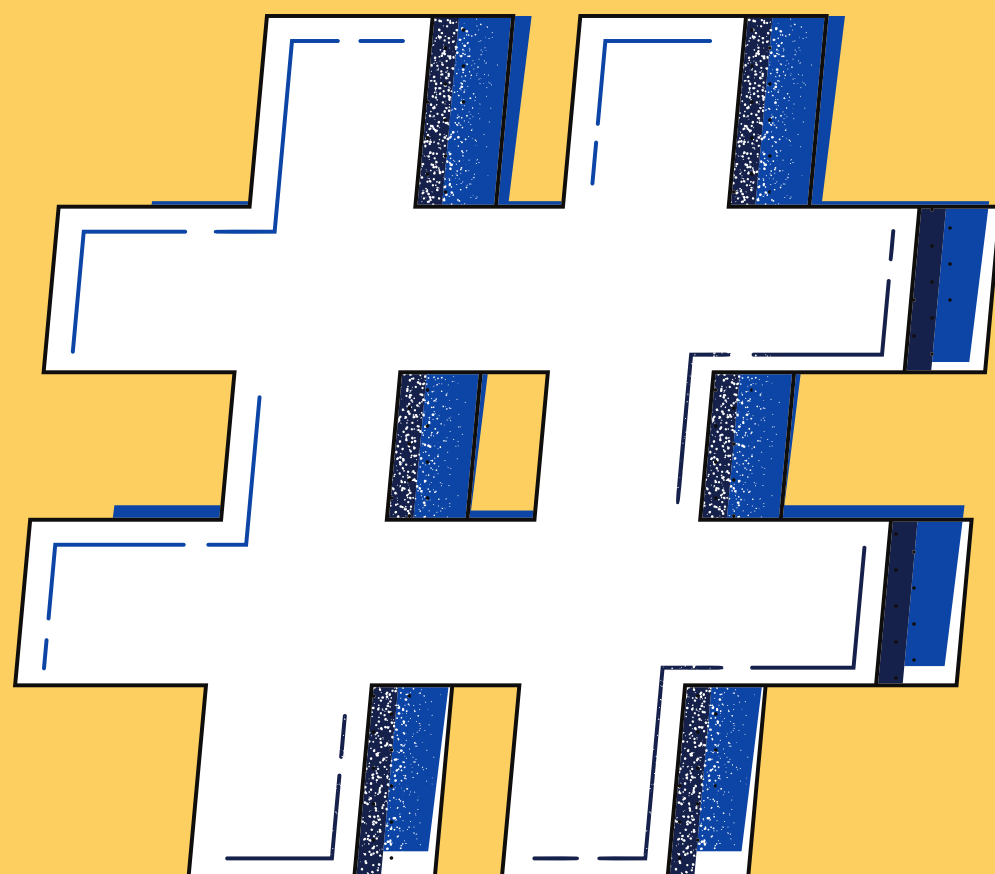
VYHODNOCENÍ STATUSU MAILU

- učení filtru
- Naive Bayes
- *porter_stemmer* z knihovny *nltk*

ROZDĚLENÍ PRÁCE

-
- plánování
 - využívané technologie
 - rozd





Filip Krul
krulfili@fel.cvut.cz

Petr Kučera
kucerp28@fel.cvut.cz



[petrkucera/spamfilter](https://github.com/petrkucera/spamfilter)