

Strojové učení v Seznam.cz

Vladimír Kadlec

Seznam.cz a.s.



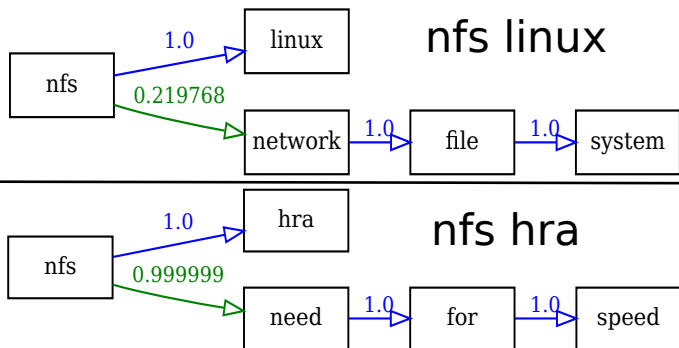
Email.cz, Firmy.cz, Kupi.cz, Mapy.cz, Obrázky.cz, Počasí.cz, Sauto.cz, Sbazar.cz, Sdovolená.cz, Seznam.cz (domovská stránka), Sklik.cz, Sport.cz, Sreality.cz, Stream.cz, Seznam Zprávy, Vyhledávání, Zboží.cz, ...

Každá služba Seznam.cz používá strojové učení.



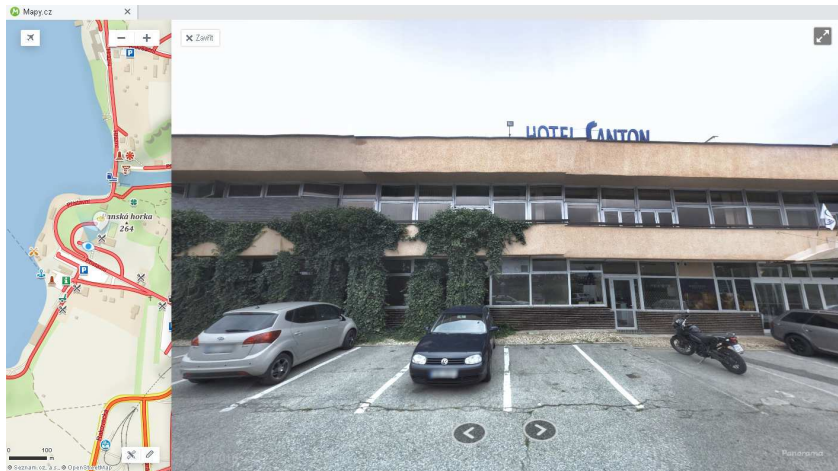
... používá strojové učení pro:

- Prioritizace návštěv webových stránek.
- Download/Parse/Index, pouze “užitečné” stránky.
- Klasifikace stránek bez znalosti obsahu (video, spam, pr0n).
- Detekce jazyka.



- Oprava překlepů.
- Přidání diakritiky, lemmatizace.
- Expanze dotazu – synonyma, zkratky, ohýbání slov, ...
- Detekce (pojmenovaných) entit – lokality, lidé, firmy, zboží, ...
- Klasifikace jazyka dotazu.

Mapy.cz, anonymizace obličejů a registračních značek



Mapy.cz, anonymizace obličejů a registračních značek

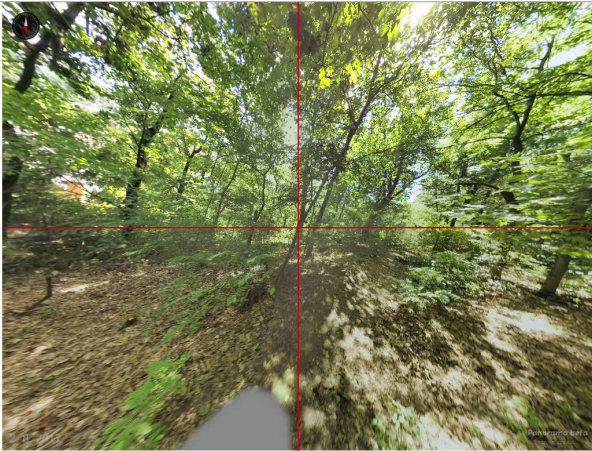




Mapy.cz, krosna, natočení



- Korekce natočení snímků oproti skutečnosti.



Panorama

244936

Autor

jmeno.prijmeni

Omega: 12.5

Phi: -38.3

Heading: -147.6

Pitch: -31.8

Roll: 25.6

Přiblížení

+ -

Nastavení

Nastavit sever

Nastavit jh

Nastavit západ

Nastavit východ

Špatná fotka

Uložit

- Korekce natočení snímků oproti skutečnosti.



Identita podle obličeje

- na použitelné úrovni vyřešená úloha.
- lze částečně použít i k autorizaci.
- technologie se vejde do telefonu.

Cílená reklama

- je všude ...

Strojové učení

- Random Forest (interní RC-Rank, XGBoost).
- Lineární/Logistická regrese.

Zpracování obrazu

- Transfer learning.
- Caffe, ResNet50, InceptionV3, ...

Nástroje

- Python, Numpy, scikit-learn.
- word2vec, FastText.
- Tensor Flow, Keras.

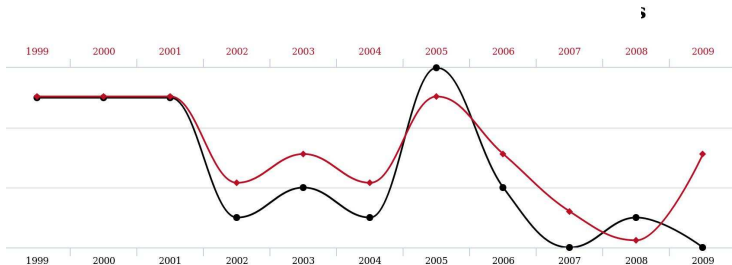
Data, data, data



- 90% úspěchu jsou dobrá, čistá, velká, reprezentativní data.
- 10% jsou algoritmy, strojové učení, optimalizace.

Můžete mít kvalitní strojové učení bez sofistikovaných algoritmů, ale nikoliv bez dobrých dat. (Daniel Tunkelang, Quora)

- Do matematických modelů často není vidět.
- Strojové učení nemá znalosti o reálném světě.



- Do matematických modelů často není vidět.
- Strojové učení nemá znalosti o reálném světě.

