

Z Wikipedie, otevřené encyklopedie

Delexání z textu (anglicky *text mining*) je vědecká disciplína na pomezí lingvistiky, informatiky, strojového učení a počítačové lingvistiky. Zpracovává textové data a snaží se z nich vytáhnout nějakou informaci. Je to tedy proces, který umožňuje automaticky analyzovat velké množství textů a najít v nich nějaké vzory nebo tendence. Delexání z textu se používá například v marketingu, v lékařství, v právu, v historii a v mnoha dalších oblastech. Je to velmi široké pole, které zahrnuje mnoho různých technik a nástrojů. Delexání z textu je tedy proces, který umožňuje automaticky analyzovat velké množství textů a najít v nich nějakou informaci. Je to tedy proces, který umožňuje automaticky analyzovat velké množství textů a najít v nich nějakou informaci.

 CC BY-SA

generování souvislého textu dosud není na takové úrovni, aby překonala výsledky prvně zmíněného přístupu.

Analýza sentimentu

Podrobnější informace naleznete v článku [Analýza sentimentu](#).

Analýza sentimentu, anglicky *sentiment analysis*. Na základě výskytu [citově zabarvených slov](#) lze usuzovat na autorův pozitivní či negativní postoj k předmětu zprávy. To může být užitečná informace obzvláště ve spojení s tematicky zaměřenými diskusními fóry.

Extrakce konceptů; rozpoznání pojmenovaných entit

V angličtině *concept extraction* či konkrétněji *named-entity recognition* (NRE), jde o určení entit, které jsou v textu zmíněny. Nástroj pro NRE by je měl identifikovat a zároveň klasifikovat do předem definovaných kategorií.^[3] Například v článku o V. Klausovi by tedy výrazy „Václav Klaus“ a „prezident“ měly být přiřazeny stejné entitě. Problém úzce souvisí s [desambiguací slovních významů](#) a tudíž patří k těm základním problémům zpracování přirozeného jazyka.

Určení vztahu mezi entitami

Dokážeme-li v textu určit pojmenované entity, můžeme na základě analýzy vět (např. pomocí rámce – [FrameNet](#)) určit jejich vztahy (např. z výrazu "Sarkozy se oženil s Bruinovou" je možné získat vztah, že Bruinová je manželkou Sarkozyho).

Odkazy

Reference

- [↑] Unstructured Data and the 80 Percent Rule. *www.clarabridge.com* [online]. [cit. 2010-06-10]. [Dostupné v archivu](#) pořízeném dne 2010-07-02.
- [↑] Wayback Machine. *web.archive.org* [online]. [cit. 2023-07-08]. [Dostupné v archivu](#) pořízeném z [originálu](#) dne 2023-07-08.

3. [↑](#) ROZPOZNÁVÁNÍ POJMENOVANÝCH ENTIT | Nový encyklopedický slovník češtiny. www.czechency.org [online]. [cit. 2022-04-10]. [Dostupné online](#).

Externí odkazy

- Obrázky, zvuky či videa k tématu [dolování z textu](#) na Wikimedia Commons

Citováno z „https://cs.wikipedia.org/w/index.php?title=Dolování_z_textu&oldid=25375429“

Kategorie:

- [Počítačová lingvistika](#)
- [Informatika](#)
- [Zpracování přirozeného jazyka](#)

Skryté kategorie:

- [Monitoring:Články s identifikátorem NKC](#)
- [Monitoring:Články s identifikátorem LNB](#)
- [Monitoring:Články s identifikátorem NDL](#)
- [Monitoring:Články s identifikátorem NLI](#)

Hledání

Speciální:Hledání

Hledat

Dolování z textu

33 jazyků

[Přidat téma](#)