

### PŘÍLOHA č. 3 – Technická specifikace

Český rozhlas hledá dodavatele služby na online analýzu textů. Většinu těchto textů budou tvořit přepisy rozhlasových pořadů, analyzované texty však mohou být i jiného charakteru (články z webu apod.).

## Popis

**Textová analýza** zpracovávaných textů by měla zahrnovat:

- **kategorizaci** textů, tj. tématické rozčlenění textů podle daných témat
- vyhledávání **entit**, zejména osob, firem, institucí a jejich představitelů, státních útvarů apod.
- určení **sentimentu**, a to jak celkového sentimentu textu jako takového, tak sentimentu vůči jednotlivým (vybraným) entitám, zejména vůči politickým stranám a jejich představitelům
- „**kdo o kom**“, možnost analyzovat text sledováním toho, kdo (typicky mluvčí či autor textu) o kom (osobě, ideálně však jakékoliv entitě) hovořil - a případně i jak
- lokaci, datum, čas a další údaje, které lze z textu získat

**Formáty textů** budou následující:

- word
- html
- prostý text

přičemž vnitřní struktura analyzovaných textů se může lišit.

## Činnost

Objednatel bude dodavateli zasílat texty ke zpracování (průběžně či dávkově) a dodavatel objednateli bude vracet zjištěné výsledky analýzy. Dodavatel bude upravovat parametry a specifika automatické analýzy textů v průběhu trvání smlouvy dle požadavků objednatele.

## Rozsah

Objednatel požaduje možnost zpracovávání až stovek textů denně, očekávaná rychlost odezvy u jednoho textu je v řádu vteřin.