**PŘÍLOHA č. 4 – TECHNICKÁ SPECIFIKACE PLNĚNÍ**

TECHNICKÁ SPECIFIKACE | Český rozhlas | 10. srpna 2022

Český rozhlas hledá dodavatele služby na online dodávání klíčových slov na základě analýzy textu, perexu, domicilu, titulku, analýzy dříve publikovaných textů a seznamu klíčových slov používaných na iROZHLASu.

# 1 Poskytovaná data

## 1.1 Zasílané segmenty článku k analytice:

1. Titulek
2. Domicil
3. Perex /anotace
4. Text článku
5. Pole klíčových slov, které doplnil redaktor ručně
6. Rubrika

## 1.2 Zpracování klíčových slov používaných na iROZHLAS

ČRo poskytne seznam klíčových slov,bufde-li tak vyžádáno, která se používají na zpravodajském serveru iROZHLAS tak, aby došlo k nabízení správných výrazů, které se mohou lišit od výrazů používaných na jiných zpravodajských serverech. ČRo vyžaduje upřednostňování svých forem klíčových slov.

# 2. Co dodavatel dodává

1. Dodavatel dodává API,
   1. pro zasílání podkladů (textů) k analýze, a dodavatel bude vracet návrhy klíčových slov.
   2. pro úpravy výrazů, aby API vracelo klíčová slova v požadovaném formátu (dle úzu ČRo).
2. Dodavatel dodává IPCT Media Topics.
3. Dodavatel provede analýzu práce redaktora v redakci pro zlepšení workflow a zpětné vazby
4. Dodavatel upravuje nabídku klíčových slov a témat na základě feedbacku

# 3. Předpokládané workflow redaktora

## 3.1 Workflow redaktora - způsob užití

Redaktor založí článek a vepíše titulek, domicil, anotaci, perex, text článku. Před vyplněním klíčových slov klikne na speciální tlačítko, ve kterém otevře dialogové okno s výpisem 12 vhodných klíčových slov vytvořených na základě analýzy. Výpis slov by měl být do 5 sekund. První polovina nabídky bude předvolená (zaškrtnutá), další budou nezaškrtnuté. Potvrzením se zvolená klíčová slova zavedou k článku a bude možné je doplnit ručně o další výrazy, které systém nenabídl.

## 3.2. Workflow automatického přiřazení klíčového slova

Systémy v ČRo vytvoří přepis audia, zašlou jej na API dodavatele s parametrem, kolik klíčových slov požadují zpět, a přijatou sadu klíčových slov automaticky přiřadí k přepisu.

# 4. Funkce výběru klíčových slov

## 4.1 Váha

Slova by měla být vybírána dle váhy mezi obecnými výrazy, osobami, názvy ad., kterou nadefinuje ČRo s dodavatelem.

*Např. Jména jsou důležitější než obecné výrazy, tedy spíše uvést Miloš Zeman nežli pacient.*

## 4.2 Zakázaná klíčová slova

ČRo vyžaduje online přístupný systém pro zavádění výrazů, které se nemají nabídce analytiky objevit.

*Např.: ČTK, BBC, Český rozhlas, tiskový mluvčí ad. by se neměly ve výsledku objevit.*

## 4.3 Automatická záměna / sloučení

ČRo vyžaduje online přístupný systém pro zavádění výrazů, které se mohou v textech objevit, ale v nabídce by měly být formulovány jinak.

Např. centrální banka jako Česká národní banka, Spojené státy jako USA, Volodymyr Oleksandrovyč Zelenskyj jako Volodymyr Zelenskyj.

## 4.4 Nenabízet následné výrazy po určitém slovním spojení

ČRo vyžaduje editovatelný systém (stačí ze strany dodavatele) pro skrývání jmen za konkrétními nadefinovanými výrazy.

*Např. z věty tiskový mluvčí Karel Vomáčka, tiskový mluvčí Jana Polívková systém nenabídne žádné jméno po výrazu tiskový mluvčí. ČRo tedy vytvoří seznam výrazů, po kterých obvykle následují jména a ta vyloučí z výsledků analýzy.*

## 4.5 Vystihnutí obsahu před nabízením klíčových slov

ČRo přivítá takovou analytiku, která správně vystihne téma obsahu.

*Např. Pokud osobnost řekne, že “s XY by na pivo nešla”, což se může objevit i v titulku, tak nenabízet klíčové slovo pivo, o tom text není.*

## 4.6 Domicily pro určení země

ČRo bude poskytovat domicily a vyžaduje u všech zahraničních lokalit doplnit automaticky stát.

*Např. v domicilu uvedený Norimberk znamená ve výpisu klíčových slov výraz Německo.*

## 4.7. Označení klíčových slov

Na výstupu uživatele ČRo požaduje označovat příznakem taková kličová slova, která nemá ČRo zavedená ve svém slovníku.

*Např. ČRo používá klíčové slovo Donald Trump namísto Donald F. Trump.*

## 4.8 Slučování klíčových slov

V textu může být více označení pro to samé, ČRo vyžaduje na výstupu používané klíčové slovo v ČRo

*Např. Výraz “Národní úřad pro kybernetickou a informační bezpečnost” má zkratku NÚKIB a jeho zápis může být různý, ČRo ale preferuje v nabízených klíčových slovech pouze jednu, kterou používá, tedy NÚKIB.*

# 5 Technické podmínky

## 5.1 API

Dodavatel zpřístupní zadavateli API endpoint, který bude přijímat data o článku a vracet navržená klíčová slova.

U endpointu budou nastaveny CORS hlavičky tak, aby ho bylo možné volat i přímo z javascriptu na konkrétních doménách (irozhlas.cz, rozhlas.cz, mujrozhlas.cz + vývojové, testovací a stage varianty domén).

## 5.2. Rychlost odezvy a četnost

V případě, že dodavatel dodá API s CORS, je možné požadavek zpracovat do 5 vteřin od odeslání, pokud API není možné volat přímo z JS, požadujeme zpracování do 3 vteřin.

iROZHLAS vydává 50-70 článků denně a zdaleka ne každý využívá službu.

ČRo navíc hodlá službu využívat i pro automatické transkripce. Termín spuštění není naplánován, půjde o desítky požadavků denně.

Časem by tedy mělo jít o ca 100 požadavků denně.

## 5.3. Formát předávání dat

Request i response bude data posílat ve formátu JSON.

Request bude obsahovat strukturovaně atributy článku:

* titulek (řetězec)
* domicil (řetězec)
* anotaci (řetězec)
* perex (řetězec)
* text článku (řetězec)
* již zadaná klíčová slova (pole řetězců)
* identifikátor článku na straně zadavatele (číslo)

Response bude obsahovat:

* Pole navržených štítků, u každého štítku budou následující atributy:
  + Název (řetězec)
  + Identifikátor štítku na straně dodavatele (číslo/řetězec)
  + Kategorie štítku (řetězec), (např. “Osoba”, “Země”, “Město”, “Obecný”, …)

Názvy klíčů, pod kterými budou jednotlivé atributy v request/response určí Dodavatel.