

# Uživatelská dokumentace pro klicova\_slova

Lukáš Štupák

June 13, 2023

## 1 Použití

```
perl klicova_slova.pl [moznosti]
```

Extrahuje a seřadí klíčová slova z textu poskytnutého na standardním vstupu.

## 2 Možnosti

- `--stopwords DIR`: Cesta ke slovníku stop slov (výchozí: stopwords).
- `--lang LANG`: Jazyk pro stemování a stop slova (výchozí: cs).
- `--num_keywords NUM`: Počet výstupních klíčových slov (výchozí: 10).
- `--output_format FMT`: Formát výstupu (text, json, csv) (výchozí: text).
- `--help`: Zobrazí tuto nápovědu.

## 3 Popis

Skript `script.pl` extrahuje a seřadí klíčová slova z vstupního textu. Načte stop slova ze souboru a použije algoritmus TextRank k určení důležitosti slov. Výsledná klíčová slova se vypíší na standardní výstup ve formátu specifikovaném možností `--output_format`.

## 4 Příklady použití

1. Extrahovat a vypsat 10 klíčových slov z vstupního textu:

```
perl klicova_slova.pl
```

2. Extrahovat a vypsat 5 klíčových slov z vstupního textu ve formátu JSON:

```
perl klicova_slova.pl --num_keywords 5 --output_format json
```

3. Extrahovat a vypsat 15 klíčových slov z vstupního textu do CSV souboru:

```
perl klicova_slova.pl --num_keywords 15 --output_format csv >  
↪ keywords.csv
```

## 5 Poznámky

- Skript předpokládá, že balíčky `Lingua::Stem::Snowball` a `Getopt::Long` jsou nainstalovány. Pokud nejsou, můžete je nainstalovat pomocí následujících příkazů:

```
cpan Lingua::Stem::Snowball
cpan Getopt::Long
```

- Skript očekává, že stop slova pro zvolený jazyk jsou k dispozici ve souboru s názvem `<lang>.txt` v adresáři `stopwords`, kde `<lang>` je kód jazyka (např. "cs" pro češtinu). Pokud takový soubor neexistuje, skript bude pokračovat bez použití stop slov.
- Vstupní text je očekáván na standardním vstupu. Můžete jej zadat ručně nebo upravit zdrojový kód skriptu tak, aby četl z jiného zdroje.