

User Documentation for klicova_slova.pl

Lukáš Štupák

June 13, 2023

1 Usage

```
perl klicova_slova.pl [options]
```

Extracts and ranks keywords from text provided on standard input.

2 Options

- `--stopwords DIR`: Path to stopwords directory (default: stopwords).
- `--lang LANG`: Language for stemming and stop words(opt: cz, sk, en) (default: en).
- `--num_keywords NUM`: Number of keywords to output (default: 10).
- `--output_format FMT`: Output format (text, json, csv) (default: text).
- `--help`: Show this help message.

3 Description

The `script.pl` script extracts and ranks keywords from the input text. It loads stop words from the corresponding file and uses the TextRank algorithm to determine the importance of words. The resulting keywords are printed to the standard output in the format specified by the `--output_format` option.

4 Examples

1. Extract and print 10 keywords from the input text:

```
perl klicova_slova.pl
```

2. Extract and print 5 keywords from the input text in JSON format:

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

3. Extract and print 15 keywords from the input text to a CSV file:

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

5 Notes

- The script assumes that the `Lingua::Stem::Snowball` and `Getopt::Long` packages are installed. If not, you can install them using the following commands:

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

- The script expects that the stop words for the specified language are available in a file named `<lang>.txt` in the `stopwords` directory, where `<lang>` is the language code (e.g., "en" for English). If such a file does not exist, the script will continue without using stop words.
- The input text is expected to be provided on the standard input. You can enter it manually or modify the script's source code to read from a different source.