

Používateľská príručka

3.5.1 Lokálne použitie konzolovej aplikácie

Na spustenie aplikácie je potrebné:

- mať nainštalovaný Python 3.7 (návod ako ho nainštalovať na danú platformu je možné nájsť [tu](#))
- nainštalovať knižnicu xlwt (v príkazovom riadku cez príkaz “python -m pip install xlwt”; návod je možné nájsť [tu](#))

Na spustenie aplikácie, treba do príkazového riadku napísať príkaz:

“<python_path> <init_path> <file_name> <file_path> <encoding> <language_path>”

<python_path> celá cesta k python súboru

<init_path> celá cesta k inicializačnému python modulu

<file_name> názov súboru, ktorý chceme spracovať

<file_path> cesta k súboru, ktorý chceme spracovať (na oddelovač priečinkov treba použiť front slash)

<encoding> kódovanie súboru, ktorý chceme spracovať (v prípade použitia UTF kódovaní treba dodať

“-sig”, teda namiesto UTF-8 zadávame UTF-8-sig)

<language_path> celá cesta konfiguračného súboru jazyka, v ktorom chceme text spracovať (na

oddelovač priečinkov treba použiť front slash)

príklad:

C:/Users/Michal/AppData/Local/Programs/Python/Python37-32/python.exe

C:/Users/Michal/Documents/slabiky/application/py_scripts/init_terminate_module.py

test.txt

C:/Users/Michal/Documents/slabiky/application/py_scripts/

UTF-8-sig

C:/Users/Michal/Documents/slabiky/application/py_scripts/configs/conf_cs_lat.json

Po spustení príkazu sa vytvoria súbory v priečinku **<file_path>**.

3.5.2 Lokálne pridanie nového jazyka

Súbory, v ktorých sú uložené pravidlá pre prácu s jazykom analyzovaného textu, sú vo formáte **json**. Pre pridanie nového jazyka je potrebné vytvoriť súbor vo formáte json, ktorého štruktúra je popísaná nižšie. Doteraz vytvorené jazyky sa nachádzajú v priečinku “py_scripts/configs”.

Štruktúra:

```
{
  "language": {
    "name": "language_name",
    "writing_system": "latin/cyrillic"
  },
  "lowercase": {
    "lowercase_letter1": {
      "phoneme-length": length or -1,
      "sign": "SONOR/CONS/VOWEL/SPEC",
      "subunit": 0/1
    },
    "lowercase_letter2": {
      "phoneme-length": length or -1,
      "sign": "SONOR/CONS/VOWEL/SPEC",
      "subunit": 0/1
    },
    ...
  },
  "clusters": {
    "letter1": {
      "letter1": {
        "except_in_subword": ["subword1", "subword2", ...],
        "only_in_subword": ["subword1", "subword2", ...],
        "length": length,
        "phonotype": "SONOR/CONS/VOWEL"
      }
    },
    ...
  },
  "phonotype_changes": {
    "letter1": {
      "preceding": "SONOR/CONS/VOWEL/NONE",
      "following": "SONOR/CONS/VOWEL/NONE",
      "becomes": "SONOR/CONS/VOWEL/NONE",
      "only_in_word": ["word1", "word2", ...]
    },
    "letter2": {
      "preceding": "SONOR/CONS/VOWEL/NONE",
      "following": "SONOR/CONS/VOWEL/NONE",
      "becomes": "SONOR/CONS/VOWEL/NONE",
      "only_in_word": ["word1", "word2", ...]
    },
    ...
  },
  "text_changes": {
    "text1": {
      "becomes": "resultText"
    },
  },
}
```

```

        "text2": {
            "becomes": "resultText"
        },
        ...
    },
    "zero-syllable_words": {
        "word1": "to_following/to_preceding",
        "word2": "to_following/to_preceding",
        ...
    },
    "special_sound_length": {
        "iu": {
            "0": [
                {
                    "preceding": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    },
                    "following": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    }
                },
                ...
            ],
            "1": [
                {
                    "preceding": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    },
                    "following": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    }
                },
                ...
            ],
            "2": [
                {
                    "preceding": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    },
                    "following": {
                        "signs": ["NONE"/"CONS"/"VOWEL"/"SPEC", ...],
                        "letters": ["letter1", "letter2", ...]
                    }
                },
                ...
            ]
        },
        ...
    }
}

```

Popis:

- *language*: obsahuje meno jazyka v angličtine, a písmo, ktoré sa v konfiguračnom súbore používa
- *lowercase*: obsahuje malé písmená jazyka, pričom pre každé písmeno uvádza zvukovú dĺžku phoneme-length (ak je uvedené -1, znamená to, že dĺžka môže byť rôzna a treba nazerať do special_sound_length), označenie fonotypu sign (môže byť SONOR, CONS, VOWEL, SPEC - pre špeciálne znaky jazyka) a informáciu o tom, či písmeno môže byť subunit (0 – nemôže, 1 – môže). Ak písmeno môže byť subunit, treba nazrieť do *clusters* a pokiaľ skúmané písmeno spĺňa podmienku tam uvedenú, viac s ním nepracovať ako so samostatným písmenom (informácie uvedené v *lowercase* sú neplatné).
- *clusters*: obsahuje písmená, ktoré môžu byť prvým písmenom zvukovo nedeliteľnej dvojice písmen. Každé uvedené písmeno obsahuje zoznam „druhých“ písmen, ktoré s ním môžu tvoriť takúto dvojicu. Ak túto dvojicu tvoria vždy okrem prípadov, kedy sa nachádzajú v určitých podslovách, zoznam týchto výnimiek je uvedený v except_in_subword. Ak túto dvojicu naopak netvoria nikdy okrem prípadov, kedy sa nachádzajú v určitých podslovách, zoznam týchto podslov je uvedený v only_in_subword. Táto dvojica je považovaná za jedno písmeno a teda je uvedená jej dĺžka aj fonotyp (jednotlivé písmená tvoriace dvojicu sa neposudzujú nikdy zvlášť). Ak je dĺžka -1, treba túto dvojicu tvoriacu jedno písmeno hľadať v special_sound_length.
- *phonotype_changes*: obsahuje písmená, ktorých fonotyp sa môže zmeniť v závislosti od okolitých písmen. Preceding spolu s following tvoria jednu podmienku, ktorá musí byť kompletne splnená, aby sa uplatnila zmena fonotypu. V preceding a following môže byť uvedený niektorý z fonotypov (SONOR, CONS, VOWEL), čo znamená, že túto časť podmienky spĺňa ktorékoľvek písmeno danej sonority, alebo NONE, čo znamená, že na danom mieste (pred/za) nesmie byť žiadne písmeno. Iná podmienka (nezávislá od predchádzajúcej) na zmenu fonotypu je, že sa písmeno nachádza v konkrétnom celom slove (neplatí pre podslová)– tieto slová sú uvedené v only_in_word. Ak je ktorákoľvek z týchto troch podmienok splnená, za fonotyp písmena sa nepovažuje to, čo bolo uvedené v *lowercase*, ale to, čo je uvedené v becomes.
- *text_changes*: obsahuje textové reťazce (v texte sa môžu nachádzať vo vnútri slova), pričom je pri každom z nich uvedené, na akú postupnosť znakov má byť reťazec prepísaný.
- *zero-syllable_words*: obsahuje slová, ktoré neobsahujú slabiku, pričom pre každé z nich je uvedené, či sa majú pri slabikovaní a analyzovaní pripojiť k slovu predchádzajúcemu (to_preceding) alebo nasledujúcemu (to_following).

- *special_sound_length*: obsahuje písmená, ktoré menia zvukovú dĺžku v závislosti od okolitých písmen. Pre každé písmeno je pre každú z dĺžok (0, 1, 2) uvedený zoznam postačujúcich podmienok, pri ktorých splnení túto dĺžku písmeno nadobúda. Každá podmienka sa skladá z dvoch „podpodmienok“ – jedna hovorí o tom, aké znaky prípadne fonotypy musia nasledovať, aby vyššia podmienka bola splnená, a druhá o tom, aké musia predchádzať skúmanému písmenu. Hlavná podmienka je potom splnená, ak predchádzajúce písmeno je buď v zozname letters „podpodmienky“ preceding alebo je jeho fonotyp v zozname signs tejto „podpodmienky“ a zároveň je nasledujúce písmeno (resp. fonotyp) v zozname letters (resp. signs) „podpodmienky“ following. Signs môže okrem SONOR, CONS, VOWEL obsahovať aj NONE, čo značí, že sa na danom mieste sa nenachádza žiadne písmeno (teda je to prvý alebo posledný znak slova).

Pre zobrazenie nového jazyka vo webovej aplikácii je potrebné pridať tento jazyk do súboru „languages.json“.

```
{
  "languages": [
    {
      "option": "Belarusian",
      "value": "conf_be_cyr"
    },
    ...
  ]
}
```

Popis:

- *option*: názov jazyka, ktorý sa zobrazí pri výbere jazykov na stránke
- *value*: názov json súboru, ku ktorému jazyk prislúcha