

Uživatel:ParcervenychBot/Pískoviště

parcervenychBot je stránka a uživatelské rozhraní pythonovského botu [1] (<https://en.wikipedia.org/wiki/Wikipedia:Bots>) uživatele Pavla Duška.

Úkoly, které bot provádí

- vyhledávání článků, které postrádají odkaz na portálu

Úkoly, které bota ještě čekají (TODO)

- odstranění duplicitních nadpisů (článek začíná stejným nadpisem, jako je název stránky)
- řízení redakčního zpracování článků (vkládání šablony redakce na všechny diskusní stránky hlavního jmenného prostoru)
- vyhledávání souvisejících článků na jiné wiki nebo bráně MEFANET
- opravy html značek (nepárové `<tag>` → `<tag />`; uzavření neuzavřených párových)
- pravidelné uhrabávání pískoviště

Odkazy

- dodělat dokumentaci na Fórum:Dokumentace
- Wikipedia:Bots (<https://en.wikipedia.org/wiki/Wikipedia:Bots>)
- Fórum:Obecné#Boti
- Wikiskripta:Boti
- Nekategorizované stránky

Dokumentace

Úvodní nastavení

Pro správnou funkci bota je nutné ve zdrojovém kódu nastavit hodnotu

- `user_name` na přihlašovací jméno uživatele wikiskript a
- `password` na jeho přihlašovací heslo.

Dále by měla být nastavena hodnota

- `url` na URL API serveru wikiskript (<https://www.wikiskripta.eu/api.php>),
- `upload_server` na URL komunikujícího serveru pro nahrávání výstupu bota (vytvořil jsem takový skript na <http://myform.xf.cz/wiki/upload.php>) a
- `noportal_whitelist_link` na URL textového souboru, v němž je uložen seznam názvů článků, které mají být při hledání článku bez odkazu na portálu ignorovány. V tomto textovém souboru by měl být každý název článku na samostatném řádku.

V případě, že nemá být hlášení vypisováno na obrazovku (třeba pokud běží program na pozadí), ale do souboru, je třeba nastavit hodnotu

- `logfile_name` na název souboru, kam mají být hlášení ukládána.

Opakované pouštění

- Pokud má být bot pouštěn opakovaně, ale nebude to provádět vnější skript či operační systém, je třeba nastavit `time_period` počet vteřin, za jak dlouho se má bot znovu spustit, a on se již o opakované spouštění postará sám.
- Nemá-li být opakovaně spouštěn, musí být hodnota `time_period` nastavena na prázdný řetězec ('').
- Má-li být pouštěn opakovaně, ale postará se o to vnější skript či operační systém, `time_period` musí být nastavena na prázdný řetězec a počet vteřin do dalšího spuštění se programu předává pomocí argumentu příkazové řádky.

Hlavní program

Program postupuje přibližně takto:

- nejprve se přihlásí k API serveru wikiskript pomocí funkce `login()`
- získá seznam všech článků v hlavním jmenném prostoru pomocí funkce `list_all_pages()`
- získá seznam článků, které budou při hledání ignorovány (funkce `get_whitelist()`)
- získá **seznam všech portálů** pomocí funkce `list_all_pages()` (jmenný prostor 100)

- pro každý článek zjistí stránky, které na něho odkazují, pomocí funkce `list_noportaly()`; pokud se mezi nimi nenachází žádný portál, zařadí jej do **seznamu článků bez odkazu na portálu'**. ***Pokud se mezi nimi nachází jen portály zkouškových otázek, zařadí jej do seznamu článků s odkazem jen na portálu zkouškových otázek;*** ignorují se články získané funkcí `get_whitelist()`
- od článků bez odkazu na portálu zjistí kategorie pomocí funkce `list_pages_categories()`
- ze seznamu článků bez odkazu na portálu vyčlení **články postgraduálního studia**, články, které jsou **podstránky**, a články v kategorii **hesla**
- vytvoří **textový výstup** seznamů ve správném formátu pomocí funkce `make_output()`
- připojí k textovému výstupu hlavičku a odešle komunikujícímu serveru provádějícímu upload na internet – funkce `make_upload()`
- odhlásí se od API serveru pomocí funkce `logout()`

Průběh své aktivity popisuje pomocí funkce `logging()`, která v případě, že je nastavena hodnota `logfile_name`, ukládá svá hlášení do souboru, jinak je vypisuje na obrazovku (standardní výstup).

Dokumentace funkcí