

Fórum:ParcervenychBot.py (dokumentace)

Tato stránka je dokumentace bota ParcervenychBot, jehož zdrojový kód je umístěn na MediaWiki:ParcervenychBot.py. Tento bot v žádném případě neaspiruje na to, aby byl profesionálně a dokonale naprogramován.

Editace, které bot provádí

Bot provádí automatické úpravy textu článků v hlavním jmenném prostoru, ve jmenném prostoru *Fórum*, *WikiSkripta* a *Nápověda* podle přesně formulovaných kritérií. Jedná se o

- úpravu obsahu článků — funkce `check_page_content(pageTitle)` —, zaměňují se nesprávně zapsaný název projekt „WikiSkripta“ (absence camelcase), špatně použité znaky pro spojovník pomlčkou, tři tečky za znak pro výpustku a anglické uvozovky českými,
- přidávání redakční šablony — funkce `add_redaction_template(pageTitle)` — diskusním stránkám článků z hlavního prostoru, v nichž chybí.

Pro každý z těchto úkolů je vyhrazen editační limit maximálního počtu editací při jednom spuštění bota (proměnná `editLimit`).

Pokud text prošlý níže popsány procesy není shodný s textem původním, dojde k editaci. V opačném případě k editaci nedochází.

„Wikiskripta“ → „WikiSkripta“, „wikiskripta“ → „WikiSkripta“

Nejprve dojde k záměně všech sekvencí obsahujících „wikiskripta“ nebo „Wikiskripta“ za „WikiSkripta“, posléze dojde k záměně „wikiskripta.eu“ za „wikiskripta.eu“.

Spojovník → pomlčka

Pokud

- se na stránce nachází sekvence „ - “ (mezera — spojovník — mezera),
- není zde tag `<gallery>`,
- není zde tag `<pre>` a
- není zde tag `<source>`,

dojde k nahrazení „ - “ za „ - “ a pak se v původním textu naleznou všechny

- odkazy (text uvozený „[[“ a „]]“),
- matematické vzorce (text uvozený „$“ a „$“),
- texty uvozené tagy „<nowiki>“ a „</nowiki>“ a
- texty uvozené tagy „<code>“ a „</code>“,

u nichž dojde k ekvivalentní záměně „ - “ za „ - “ a pak zpětnému nahrazení těchto fragmentů v textu stránky.

„...“ → „...“

Pokud

- se na stránce nachází sekvence „...“ (tři tečky),
- není zde tag `<gallery>`,
- není zde tag `<pre>` a
- není zde tag `<source>`,

dojde k nahrazení „...“ za „...“ a pak se v původním textu naleznou všechny

- odkazy (text uvozený „[[“ a „]]“),
- matematické vzorce (text uvozený „$“ a „$“),
- texty uvozené tagy „<nowiki>“ a „</nowiki>“ a
- texty uvozené tagy „<code>“ a „</code>“,

u nichž dojde k ekvivalentní záměně „...“ za „...“ a pak zpětnému nahrazení těchto fragmentů v textu stránky.

"*" → „*“

Pokud

- se na stránce nachází anglické uvozovky,
- nejsou zde nikde na stránce české (pravé české uvozovky využívají stejný znak jako anglické levé, což by při záměnách dělalo neplechu),
- není zde tag `<gallery>`,
- není zde tag `<pre>` a
- není zde tag `<source>`,

je každý text uvozený anglickými uvozovkami nahrazen stejným textem uvozeným českými uvozovkami — **v případě, že tyto uvozovky těsně nepředchází znak pro rovnítko**. (Zabrání se tak nahrazování uvozeného textu ve zdrojovém kódu HTML, CSS a JavaScriptu.)

Zároveň je zde připojen cyklus, který prochází všechny možné části uvozeného textu rozděleného dle případných dalších uvozevek a provádí další případné náhrady, jelikož regulární výrazy v Pythonu neprovádí nejkratší možný match, nýbrž nejdelší.

Redakční šablona

Pokud

- vlastní článek neobsahuje text „{{Pracuje se“,
- vlastní článek neobsahuje text „{{Rozcestník“,
- vlastní článek obsahuje více než 100 znaků,
- diskusní stránka neobsahuje text „{{Redakce“,
- diskusní stránka neobsahuje text „{{Zkontrolovat“,

- diskusní stránka neobsahuje text „#PŘESMĚRUJ“ (diskusní stránka by pak přesměrovala na jinou diskusní stránku),

před vlastní text diskusní stránky je vložen text „{{Redakce\\n|citace\\n|členění\\n|kategorie\\n|licence\\n|nadpisy\\n|synonyma\\n|obrázky\\n|prolinkování|\\nzdroje\\n|portály\\n|pravopis\\n|typografie}}“.(Sekvence „\\n“ je Pythonem interpretována jako nový řádek.)

Seznamy, které bot generuje

Bot generuje seznamy článků se specifickým obsahem:

- WikiSkripta:Check WikiSkripta/Wikifikace — články, které potřebují z různého důvodu wikifikaci
- WikiSkripta:Check WikiSkripta/Portály — články, na něž neodkazuje žádný portál
- WikiSkripta:Check WikiSkripta/Redirects — návrhy přesměrování

Wikifikace

- Jedním z nežádoucích jevů na wiki je používání html tagů, které se dají opsat wikikódem. Do seznamu **Články, které potřebují wikifikaci** je článek zařazen, pokud se v něm nachází jeden z těchto řetězců
 - „
“, „
“, „
“,
 - „<hr>“, „<hr />“, „<hr/>“,
 - „<h1>“, „<h2>“, „<h3>“, „<h1 />“, „<h2 />“, „<h3 />“, „<h1/>“, „<h2/>“, „<h3/>“, „</h1>“, „</h2>“, „</h3>“,
 - „“, „“, „“, „“,
 - „“, „“, „“, „“,
 - „<p>“, „<p />“, „<p/>“, „</p>“,
 - „<i>“, „<i />“, „<i/>“, „</i>“,
 - „<u>“, „<u />“, „<u/>“, „</u>“,
 - „<a>“, „<a />“, „<a/>“ nebo „<a>“.
- Do seznamu **Stránky, které špatně používají CSS styly** je článek zařazen, pokud se v něm nachází řetězec „style=““ nebo „style=““ (styly by měly být definovány pomocí parametru *class*).
- Do seznamu **Stránky s neuzavřenými závorkami** je článek zařazen, pokud
 - se v jeho název nezačíná řetězcem „Fórum:“ (na stránkách na fóru se velmi často vyskytují emotikony); a zároveň
 - není v článku shodný počet znaků „[“ a „]“; nebo
 - není v článku shodný počet znaků „{“ a „}“; nebo
 - není v článku shodný počet znaků „(“ a „)“; nebo
 - není sudý počet apostrofů.
- Do seznamu **„Opravy obsahu“** je článek zařazen, pokud se u něj bot pokusil o automatickou opravu obsahu („Wikiskripta“ → „WikiSkripta“, „wikiskripta“ → „WikiSkripta“, spojovník → pomlčka, „…“ → „…“, „*“ → „*“).

Portály

Článek je zařazen do seznamu **„Stránky bez odkazu na portálu“** v případě, že Aplication programming interface (API) WikiSkript nevrací žádné *backlinks* (stránky, jež na článek odkazují) ve jmenném prostoru *Portál*, a do seznamu **„Stránky s odkazem jen na portálu zkouškových otázek“** v případě, že jsou API vráceny pouze portály s řetězcem „Otázky“ nebo „otázky“ v názvu, a zároveň pokud

- není ve whitelistu
- speciální stránka *Odkazuje sem* pro daný článek v sobě nemá řetězec „Portál“ (API občas špatně vyhodnotí přesměrovávání – pokud portál odkazuje na stránku A, která přesměrovává stránku B, API může chybně vracet backlinks pro stránku B)
- nemá kategorii *Kategorie:Rozcestník* (rozcestníky nepotřebují odkaz na portálu, ten potřebují cílové stránky)
- nemá v názvu „/PGS“ — v daném případě je zařazen do seznamu **„Články postgraduálního studia bez odkazu na portálu“**
- nemá v názvu „/“ — v daném případě, pokud nemá v názvu „/PGS“, je zařazen do seznamu **„Podstránky bez odkazu na portálu“**
- nemá kategorii *Kategorie:Heslo* — v daném případě je zařazen do seznamu **„Hesla bez odkazu na portálu“**

Seznam navrhovaných redirectů

Seznam navrhovaných redirectů se tvoří tak, že v seznamu všech stránek (kromě whitelistu) se oboustranně nahradí dvojice řetězců:

- „asa“ a „áza“,
- „asy“ a „ázy“,
- „osa“ a „óza“,
- „asie“ a „ázie“,
- „asia“ a „ázia“,
- „itis“ a „itida“,
- „choroba“ a „nemoc“,
- „Choroba“ a „Nemoc“,
- „emia“ a „émie“,
- „emie“ a „émie“,
- „poesa“ a „poéza“,
- „poeza“ a „poéza“,
- „ismus“ a „izmus“,
- „pathie“ a „patie“,
- „Chromosom“ a „Chromozom“,
- „chromosom“ a „chromozom“.

Z tohoto seznamu se dále vyřadí položky obsahující řetězce „ochoroba“, „Dignosa“, „diagosa“, „žlas“, „chémie“ a „vasan“. Pokud nově vytvořený název stránky není v seznamu všech stránek, zařadí se do seznamu navrhovaných redirectů.

Přehled úvodních nastavení

Proměnná	Standardní hodnota	Význam proměnné
user_name	'parcervenyhBot'	Přihlašovací jméno (login) účtu, na něhož se bot přihlašuje
password	heslo k účtu ParcervenyhBot	Heslo k účtu uvedeném v proměnné user_name
testurl	'http://www.test-wiki.lf1.cuni.cz/api.php'	URI API serveru testovací wiki; pokud chcete spouštět bota na testovací wiki, je nutné tuto hodnotu přiřadit do proměnné url: url = testurl
url	'https://www.wikiskripta.eu/api.php'	URI API serveru WikiSkript
urlWikiHome	'https://www.wikiskripta.eu/index.php'	URI WikiSkript, slouží k vyhledávání nenapřímených odkazů na portálu pomocí speciální stránky, tedy vyhledávání mimo API
upload_server	'http://myform.xf.cz/wiki/upload.php'	URI serveru, který provádí nahrávání výstupu bota na internet; pokud bot edituje a nahrává tak výstup přímo do WikiSkript (Check WikiSkript), zůstává nevyužito
logfile_name	'parcervenyhbot.log'	Název textového souboru, kam má bot ukládat výstup na příkazovou řádku, pokud nenastaveno (""), tiskne pouze do standardního výstupu
noportal_whitelist_link	'http://myform.xf.cz/wiki/NPWhiteList.txt'	URI stránky, v níž jsou uvedeny tituly článků, které má bot při vyhledávání článků postrádajících odkaz na portálu ignorovat; každý titul musí být na samostatném řádku
redirect_whitelist_link	'http://myform.xf.cz/wiki/RedirectWhiteList.txt'	stejně jako noportal_whitelist_link, jen pro navrhování redirectů (aby byl návrh ignorován, lze zadat jak stávající název článku, tak návrh redirectu)
cj	cookieLib.CookieJar()	Objekt, který se stará o Cookies při komunikaci se serverem (API to potřebuje pro identifikaci); zkratka z CookieJar
editCount	0	Proměnná, která udává počet provedených editací, standardně se na začátek určuje jako 0, stejně tak před editací Check WikiSkripta
editLimit	60	Limit editací na jedno spuštění bota