

API modul pre Google Scholar

Nasledujúci dokument popisuje funkcie, ktoré zapúzdrujú operácie pre prácu s Google Scholar. Na základe predchádzajúcej analýzy (https://merlin.fit.vutbr.cz/nlp-wiki/index.php/Rrs_cite_comparsion) tejto služby. Funkcie popisujú elementárne operácie, ktoré možno nad touto službou vykonávať. Funkcie sú logicky delené do celkov, napr. vyhľadávacie, spracovanie výsledkov a pod.. Jednotlivé funkcie ošetrujú aj nevalidné operácie. V prípade, že nastane chyba funkcia vráti príslušnú výnimku pokiaľ je schopná ju určiť inak vráti všeobecnú výnimku. Pre funkčnosť API sú nutné nasledovné knižnice: **sys, BeautifulSoup, time, re, urllib2**.

Prehľad výnimiek

ArithmeticError	Aritmetická chyba(pretečenie,delenie nulou)
EnvironmentError	Chyba prostredia(vstupno vystupna chyba,OS chyba)
ValueError	Hodnotova chyba(Chyba pre dekodovani unicodu)
Exception	Všeobecná výnimka, ktorá pokrýva vyššie neuvedené príklady

Vyhľadávacie funkcie

Tieto funkcie zabezpečujú vyhľadávanie prostredníctvom služby GoogleScholar. Funkcie sú rozdelené na základné vyhľadávanie a rozšírené vyhľadávanie. Každá funkcia má špecifické parametre, ktoré ovplyvňujú spôsob vyhľadávania. V prípade, že nastane pri vykonávaní funkcie chyba, funkcia vyvolá výnimku, ktorú si musí následne užívateľ odchytiť.

Základná vyhľadávacia funkcia

Funkcia vyhľadá informácie v zozname podľa daného kľúčového slova. Funkcia má 1 parameter:

- **Kľúčové slovo:**(myslí sa vyhľadávací reťazec).

Táto funkcia predstavuje základnú možnosť vyhľadávania. Funkcia vytvorí prázdny zoznam, do ktorého sa v prípade nájdenia uloží položka. Funkcia poskytuje ako výstup asociatívne pole, ktoré vyzerá nasledovne:

```
{1:['nazov_publicacie','meno_autoravratene_priezviska','typ_publicacie','datum_vydania',  
'Abstrakt','odkaz_na_citacie','pocet_citacii','odkaz_do_kniznice'],  
2:['nazov_publicacie','0','0','datum_vydania','Abstrakt','odkaz_na_citacie','pocet_citacii','0'],  
3:["windows",0,0,"2012",'0','20','0','0'], ...}.
```

V prípade, že výsledku vyhľadávania nebude hodnota zahrnutá, napr. nebude vo výsledku zobrazený rok na jeho miesto v zozname sa doplní 0. Poradie jednotlivých položiek je záväzné. Spracovanie jednotlivých nájdených výsledkov je možné podľa kľúčov zoznamu, ktoré majú hodnotu od 1 až do n podľa počet nájdených položiek.

/**

* @param: keyword – kľúčové slovo/slová, ktoré predstavuje kľúč na vyhľadávanie

*/

asociativne_pole basicSearch(keyword);

Rozšírená vyhľadávacia funkcia

Funkcia rozšíreného vyhľadávania umožňuje špecifickejšie určiť vyhľadávané články.

Funkcia má **9 parametrov**:

- **AllWord** - Zhoda so všetkými slovami, ktoré sa zahrnú tomuto argumentu
V prípade nepoužitia tohto parametru použitie zadanie **False**.
- **WithCorrectPhrase** – Zahrnúť do vyhľadávania frázy. V prípade nepoužitia tohto parametru treba zadať hodnotu **False**.
- **LeastOneWord** – vyhľadávanie v názve článku/vyhľadávanie v celom článku,
- **WithoutWords** – parameter, ktorý vyhľadáva podľa konkrétneho autora(hodnota 0 značí nepoužitie parametra), **PublicPlace** – miesto publikácie článku(hodnota 0 znamená nepoužitie tejto podmienky),
- **Occurence** – parameter, ktorý obmedzuje vyhľadávanie podľa miesta výskytu kľúčových slov. Hodnota 1 znamená, že dané kľúčové slovo sa môže nachádzať kdekoľvek v článku. Hodnota 2 znamená, že dané kľúčové slová sa môžu vyskytovať v titulku. Zadanie hodnoty 1 alebo 2 je povinné.
- **Author** – Názov autora podľa, ktorého sa má vyhľadávať, v prípade nepoužitia parametru treba zadať hodnotu **False**.
- **Venue** – Miesto/miestá publikácie článku. V prípade nepoužitia tohto parametru zadať hodnotu **False**.
- **StartYear** – Začiatkový rok, od ktorého majú byť zahrnuté výsledky. V prípade, nepoužitia zadať hodnotu **False**.
- **EndYear** – Konečný rok, do ktorého majú byť zahrnuté výsledky. V prípade, nepoužitia zadať hodnotu **False**.

V prípade, že nastane nejaká chyba vyvolá sa výnimka. Funkcia vracia asociatívne pole:

```
{1:['nazov_publicacie','meno_autoravratene_priezviska','typ_publicacie','datum_vydania',  
'Abstrakt','odkaz_na_citacie','pocet_citacii','odkaz_do_kniznice'],  
2:['nazov_publicacie','0','0','datum_vydania','Abstrakt','odkaz_na_citacie','pocet_citacii','0'],  
3:["windows",0,0,"2012","0","20","0","0"], ...}.
```

V prípade, že výsledku vyhľadávania nebude hodnota zahrnutá, napr. nebude vo výsledku zobrazený rok na jeho miesto v zoznamu sa doplí 0. Poradie jednotlivých položiek je záväzné. Spracovanie jednotlivých nájdení je možné podľa kľúčov zoznamu, ktoré majú hodnotu od 1 až do n.

/**

*@param: AllWord, WithCorrectPhrase, LeastOneWord, WithoutWords, Occurence, Author, Venue, * StartYear, EndYear.

*/

Asociatívne_pole extendedSearch(AllWord, WithCorrectPhrase, LeastOneWord, WithoutWords, Occurence, Author, Venue, StartYear, EndYear)