

API modul pre CiteSeerX

Nasledujúci dokument popisujú funkcie, ktoré zapúzdrujú operácie pre prácu s CiteSeerX. 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 výsledok a pod.. Jednotlivé funkcie ošetrujú aj nevalidné operácie. V prípade, že nastane chyba funkcia vráti príslušnú výnimku inak vráti všeobecnú výnimku, ktorú musí ďalej užívateľ ošetriť. **Vyžaduje prítomnosť knižníc: sys, BeautifulSoup, time, urllib2, re, codecs.**

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 funkcia zabezpečujú vyhľadávanie prostredníctvom služby CiteSeerX. 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 2 parameter:

- **Kľúčové slovo**(myslí sa vyhľadávací refazec, ktorý sa zadá službe CiteSeerX)
- **IncludeCitation**(má 2 hodnoty:False nezahrnutie citácií pri vyhľadávaní, True zahrnutie citácií do vyhľadávania,konkrétne počtu citácií), parameter je povinný.
- **Sort** – 1-5: 1-triedenie podľa citácií, 2 – triedenie podľa dôveryhodnosti, 3 - trieda podľa rokov klesajúco,4-triedenie podľa rokov stúpajúco,5-triedenie podľa novosti.

Táto funkcia predstavuje základnú možnosť vyhľadávania. Funkcia poskytuje ako výstup asociatívne pole, ktoré vyzerá nasledovne:

1:['názov_publicácie','meno/mená autorov ','zanru publikacie', 'roku_vydania','krátky_obsah','citácia z diela','abstrakt', 'odkaz_do_kniznice','odkazy_na_citácie','pocet_citacii'],

2:['názov_publicácie','meno/mená autorov ','zanru publikacie', 'roku_vydania','krátky_obsah','citácia z diela','0', 'odkaz_do_kniznice','odkazy_na_citácie','pocet_citacii'],

3:['názov_publicácie','meno/mená autorov ','zanru publikacie', 'roku_vydania','krátky_obsah','citácia z diela','abstrakt','odkaz_do_kniznice','odkazy_na_citácie','0'],...}.

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

/**

* @param: keyword – vyhľadacia fráza, ktorá môže obsahovať aj viac kľúčových slov, ktoré predstavuje kľúč na vyhľadávanie, IncludeCitation - *zahrnutie citácií

*/

asociativne_pole basicSearch(keyword,IncludeCitation);

Rozšírené vyhľadávanie

Funkcia rozšíreného vyhľadávania umožňuje špecifickejšie určiť vyhľadávané články. Funkcia má 12 parametrov:

- **Text** - kľúčové slovo/slová podľa ktorého/ktorých sa bude vyhľadávať(hodnota **False** znamená nepoužitie parametru),
- **Title** – titulok článku(hodnota **False** znamená nepoužitie parametru),
- **Author** – meno autora podľa, ktorého sa vyhľadáva(hodnota **False** znamená nepoužitie parametru),
- **AuthorAffi** – príslušnosť autora (hodnota **False** znamená nepoužitie parametru),
- **PublicVenue** – miesto publikácie (hodnota **False** znamená nepoužitie parametru),
- **Keywords** – vyhľadávacia fráza(v prípade zadania parametra Text sa vyhľadáva podľa oboch kľúčových slov) (hodnota **False** znamená nepoužitie parametru),
- **Abstract** – vyhľadávanie podľa abstraktu(hodnota **False** znamená nepoužitie parametru) ,
- **Citations(hodnoty False/True)** – nezahrnutie citácií/zahrnutie citácií do vyhľadávania,
- **MinNumCitations** – minimálny počet citácií(hodnota **False** znamená nepoužitie tejto podmienky),
- **Year** – parameter, ktorý obmedzuje vyhľadávanie podľa roku autora(hodnota **False** znamená nepoužitie tohto parametra, hodnota 1 použitie).
- **YearArg** – zoznam(list), ktorý má 2 položky v prípade, že sa zadá hodnota parametru Year na 1 a to je buď 1 alebo 2 roky, 1. Položka v zozname – vyhľadávanie od roku, 2. Položka – vyhľadávanie do istého roku,
- **SortBy** – 1-5: 1-triedenie podľa citácií, 2 – triedenie podľa dôveryhodnosti, 3 - trieda podľa rokov klesajúco, 4-triedenie podľa rokov stúpajúco, 5-triedenie podľa novosti.

Výsledný spôsob organizácie výsledkov:

{1:['názov_publicácie','meno/mená autorov ','zanru publikácie', 'roku_vydania','krátky_obsah','citácia z diela','abstrakt',

'odkaz_do_kniznice','odkazy_na_citácie','pocet_citacii'],

2:['názov_publicácie','meno/mená autorov ','zanru publikácie', 'roku_vydania','krátky_obsah','citácia z diela','0',

'odkaz_do_kniznice','odkazy_na_citácie','pocet_citacii'],

3:['názov_publicácie','meno/mená autorov ','zanru publikácie', 'roku_vydania','krátky_obsah','citácia z diela','abstrakt',

'odkaz_do_kniznice','odkazy_na_citácie','0'],...}.

V prípade, že výsledku vyhľadávania nebude hodnota zahrnutá, napr. nebude vo výsledku zobrazený rok na jeho miesto v zoznamu sa doplní 0. Poradie jednotlivých položiek je záväzné. Spracovanie jednotlivých nájdení je

možné podľa kľúčov zoznamu, ktoré majú hodnotu od 1 až do n.

/**

*[@param](#): **Text** – kľúčové slovo na vyhľadávanie, **Title** – titulok, **Author** – meno autora, **AuthorAffi** – príslušnosť autora, **PublicVenue** – miesto publikácie, **Keywords** – kľúčové slová, **Abstract** – abstract, **Year** – vyhľadávanie podľa roku, **YearArg** – zoznam rokov, **MinCitations** – minimálny počet citácií, **Citations** – zahrnutie citácií, **SortBy** – triedenie výsledkov podľa možností.

*/

Asociatívne_pole extendedSearch(Text ;Title; AuthorAffi ; PublicVenue;Keywords ;Abstract ;Year ; YearArg ; MinCitations ;Citations ; SortBy)