

Slovenská technická univerzita
Fakulta informatiky a informačných technológií
Ilkovičova 2, 842 16 Bratislava

Vyhľadávanie informácií - Konzultácia 4

Katarína Juhásová

Cvičiaci: Mgr. Martin Šeleng, PhD.

Študijný odbor: Inteligentné softvérové systémy

Ročník: 1. ročník, Ing. štúdium

Akademický rok: 2020/2021

Implementácia zgrupovania autorov podľa mena

Autori sú zgrupovaní v prípade, že články týchto autorov majú aspoň 1 spoločný študijný odbor a že mená autorov spĺňajú aspoň jednu z nasledovných podmienok:

1. mená autorov sú zhodné
2. mená autorov sú zhodné podľa knihnice whoswho
(<https://github.com/rlieb/whoswho>)
3. mená autorov po odstránení interpunkcie a konvertovaní na lowercase sú zhodné
4. jedna z permutácií mien prvého autora je zhodná z menom druhého autora

Priebežné výstupy

Príklady zgrupovaných autorov a ich článkov:

```
"2460278700": {
  "otherId": [
    "2469317235"
  ],
  "name": "David Bell",
  "otherNames": [
    "David A. Bell"
  ],
  "paperCount": "1",
  "otherPaperCount": [
    "1"
  ],
  "citationCount": "0",
  "otherCitationCount": [
    "0"
  ],
  "papers": [
    {
      "fieldsOfStudy": [
        "Electronic circuit",
        "Electronic engineering",
        "Computer science"
      ],
      "title": "Electric Circuits",
      "publicationDate": "1998-01-01",
      "referenceCount": "0",
      "citationCount": "0",
      "estimatedCitationCount": "0"
    },
    {
      "fieldsOfStudy": [
        "Electronic circuit",
        "Electronic engineering",
        "Computer science"
      ],
      "title": "Electric Circuits",
      "publicationDate": "1998-01-01",
      "referenceCount": "0",
      "citationCount": "0",
      "estimatedCitationCount": "0"
    }
  ]
}
```

```

"2868772321": {
  "otherId": [
    "2882530662"
  ],
  "name": "Li Bo",
  "otherNames": [
    "Li Bo"
  ],
  "paperCount": "11",
  "otherPaperCount": [
    "14"
  ],
  "citationCount": "2",
  "otherCitationCount": [
    "3"
  ],
  "papers": [
    {
      "fieldsOfStudy": [
        "Copper",
        "Soldering",
        "Materials science",
        "Composite material"
      ],
      "title": "Soldering quality control method of large-section copper strand wire of turbonator",
      "publicationDate": "2013-03-20",
      "referenceCount": "5",
      "citationCount": "0",
      "estimatedCitationCount": "0"
    },
    {
      "fieldsOfStudy": [
        "Liquid-crystal display",
        "Electronic engineering",
        "Materials science"
      ],
      "title": "Liquid crystal display module for power monitoring and control devices",
      "publicationDate": "2015-06-10",
      "referenceCount": "0",
      "citationCount": "0",
      "estimatedCitationCount": "0"
    }
  ]
}
},

```

Plán d'álšej práce

Na základe pripomienok z konzultácií a postrehov z doterajšej implementácie budú do projektu zahrnuté nasledujúce vylepšenia:

1. uloženie dát vo formáte jsonlines namiesto json
2. odstránenie duplicitných záznamov o článkoch
3. zgrupovanie autorov nie len na základe podobnosti mena a spoločných študijných odborov ale aj podľa spoločných spoluautorov podieľajúcich sa na článkoch