Opis przetwarzania przykładowego zbioru danych przez skrypt

Weźmy przykładowy rekord:

@article{ WOS:000670074400004,

Author = {Ji, J. C. and Luo, Quantian and Ye, Kan},

Title = {{Vibration control based metamaterials and origami structures: A

state-of-the-art review}},

Journal = {{MECHANICAL SYSTEMS AND SIGNAL PROCESSING}},

Year = {{2021}},

Volume = {{161}},

Month = {{DEC}},

DOI = {{10.1016/j.ymssp.2021.107945}},

Article-Number = {{107945}},

ISSN = {{0888-3270}},

EISSN = {{1096-1216}},

ResearcherID-Numbers = {{Luo, Quantian/I-2802-2018}},

ORCID-Numbers = {{Luo, Quantian/0000-0001-5727-9290}},

Unique-ID = {{WOS:000670074400004}},

}

Skrypt czyta cały plik linia po linii (również znaki końca linii, które są zapisywane jako „\n”).

1. Jeśli natrafia na linię, która zaczyna się od „@”, zapisuje sobie jej położenie, po to żeby później wstawić na dobre miejsce nowy klucz.
   1. Sprawdza jaki to jest typ rekordu, „article”, „inproceedings”, „incollection” i zapisuje tę informację
   2. Usuwa wszystko po znaku spacji
      1. „@article{„. „WOS:000670074400004,\n”
   3. Tworzy początek nowego klucza poprzez zapisanie pierwszej części przed spacją, np.:
      1. „@article{”
2. Jeśli natrafia na linię, która zaczyna się od „Author =”:
   1. Liczy wystąpienia „ and ” i zapisuje to
   2. Usuwa część przed nazwiskami i imionami a później dzieli imiona i nazwiska po „and”:
      1. „Ji, J. C.” ,„Luo, Quantian”, “Ye, Kan”
   3. Usuwa imiona dzieląc kolejne rekordy po przecinkach i wybierając pierwszą część:
      1. „Ji”, „Luo”, „Ye”
   4. Dodaje do klucza do 3 rekordów biorąc z każdego maksymalnie 3 znaki (w przypadku większej ilości and od 2 dodaje jeszcze „ETAL”):
      1. „@article{JiLuoYe”
3. Jeśli natrafia na linię, która zaczyna się od „Journal=” (lub „Booktitle” dla „inproceedings” i „incollection”):
   1. Jak w przypadku nazwy rekordku oraz autora usuwa pierwszą część, która nie jest istotna dla klucza.
   2. Następnie usuwa ostatnią część linii: „}},\n”
   3. Zamienia wszystkie litery na małe
      1. „mechanical systems and signal processing”
   4. Zamienia wszystkie pierwsze litery słów na wielkie
      1. „Mechanical Systems And Signal Processing”
   5. Zamienia wszystkie wystąpienia „And”, „With”, „\&”, „Of”, „On” na ich małe wersje, „-” na spację, a „:” na puste miejsce
      1. „Mechanical Systems and Signal Processing”
   6. Dzieli całość po spacjach aby otrzymać osobne pojedyncze słowa
      1. „Mechanical”, “Systems”, “and”, “Signal”, “Processing”
   7. Dodaje do tymczasowego klucza pierwsze litery każdego słowa (w przypadku jednosłownych tytułów bierze 8 liter, a w przypadku dwusłownych po 4 litery)(jeśli wcześniej jest zapisana informacja o tym że jest to „inproceedings” albo „incollection” to dodaje przed „Book”)
      1. „MSaSP”
4. Jeśli natrafia na linię, która zaczyna się od „Year=” (lub „Early Access Date =”):
   1. Usuwa przednią oraz ostatnią część (w przypadku „Early Access Date” dodatkowo usuwa człon zawierający miesiąc )
   2. Dopisuje rok do klucza
      1. „@article{JiLuoYe2021”
   3. Dopisuje wcześniej zapisaną w kluczu tymczasowym część z „Journal” lub „Booktitle”
      1. „@article{JiLuoYe2021MSaSP”
5. Jeśli natrafia na linię, która zaczyna się od „Article-Number=” (lub „Pages =”):
   1. Usuwa przednią oraz ostatnią część
   2. Dodaje do klucza literkę „a” („p” w przypadku „Pages”)
      1. „@article{JiLuoYe2021MSaSPa”
   3. Zamienia spacje na „space”
   4. Dopisuje do klucza całość
      1. „@article{JiLuoYe2021MSaSPa107945”
6. Jeśli natrafia na linię, która jest pusta, czyli jedyna jej zawartośc to „\n”
   1. Dopisuje do klucza „,\n”
      1. „@article{JiLuoYe2021MSaSPa107945,\n”
   2. Wpisuje nową linię klucza w miejsce starej
   3. Zapisuje rekord do pliku
7. Całość jest powtarzana aż do zakończenia pliku.