Использование марковских цепей при анализе вариативности в русских диалектных корпусах

Георгий Алексеевич Мороз

23 марта 2023

Введение

Работа с журналами

Работа со статьями



Наукометрия

• Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.



Наукометрия

- Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.
- Оценивать принято через публикации, но тут обнаруживается, что какую бы меру не придумали, всегда находятся хакеры, которые ломают систему.



Наукометрия

- Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.
- Оценивать принято через публикации, но тут обнаруживается, что какую бы меру не придумали, всегда находятся хакеры, которые ломают систему.
- Появляется направление под названием наукометрия (scientometrics), которое аккамулирует в себе анализ данных в области цитирований, журналов и науки (ср. с другими частными направлениями: биоинформатика, дата журналистика и др.)



Чем занимаются наукометры

В основном они анализируют метаданные научных публикаций:

- авторов
- название, аннотации, ключевые слова
- список литературы (исходящие цитирование)
- где процитирована данная работу (входящие цитирования)
- где опубликовано (импакт фактор, квартили и другие характеристики журнала и т. п.)

В Вышке они еще и влияют на составление списка хороших и плохих журналов, за которые платят надбавки.

По лингвистике они выдали список из 2662 журналов, которые категорезированы по группам a > b > c > d.

Лингвистика входит в тег филология, лингвистика и медиакоммуникации, так что в списке много мусора.

Введение

Работа с журналами

Работа со статьями

Команда

- руководители
 - Г. Мороз
 - Б. Орехов
- студенты
 - А. Алексеева (б)
 - Т. Дедов (б)
 - A. Opexoв (м)
 - К. Сидоров (м)
 - А. Степанова (б)

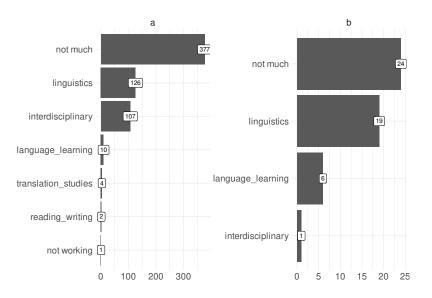
Что мы сделали

Мы начали рабоаться со списками а и ь:

- Таблица с журналами содержит поля
 - title: AAC: AUGMENTATIVE AND ALTERNATIVE COMMUNICATION
 - issn: 0743-4618; 1477-3848
 - hse_level: a
 - category: биология, медицина и здравоохранение;
 филология, лингвистика и медиакоммуникации
- Мы добавили поля:
 - only_PhLiMed: O
 - url: https://www.tandfonline.com/loi/iaac20
 - (ручная разметка) linguistics: not much
 - (ручная разметка) literary_studies: not much
 - (ручная разметка) аітя_scope: ...
 - url_domain: tandfonline.com
 - служебные мета-данные: finished, day, month, year



Разметка поля linguistics





Введение

Работа с журналами

Работа со статьями



Библиометрические базы

- Web of Science
- Scopus
- eLIBRARY.ru (РИНЦ)
- Google Scholar
- crossref, см. R пакет rcrossref
- openalex, см. R пакет openalexR
- Semantic Scholar
- The Lens



crossref

При помощи rcrossref мы извлекли для всех лингвистических журналов следующие поля:

- doi: 10.1515/ang-2012-0005
- author: Müller Patrick
- title: "The impediment that cannot say its name": Stammering and Trauma in Selected American and British Texts
- url: http://dx.doi.org/10.1515/ang-2012-0005
- volume: 130
- issue: 1
- page: -
- reference.count: O
- container.title: Anglia Zeitschrift f
 ür englische Philologie
- issn: 1865-8938,0340-5222



Наконец-то краулер!

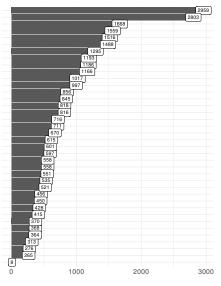
Мы разделили издательства между ребятами и они написали краулеры, чтобы достать следующие поля:

- abstract
- keyword
- year
- source



Что уже сделано (32 журнала):

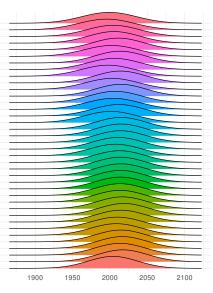
Journal of Child Language Journal of Linguistics Applied Psycholinguistics Journal of French Language Studies International Journal of the Sociology of Language Phonetica Linguistics Journal of the International Phonetic Association English Language and Linguistics STUF - Language Typology and Universals Natural Language and Linguistic Theory Russian Linguistics Multilingua Review of Cognitive Linguistics Discourse Processes indogermanische forschungen text and talk Language and Education linguistic review Journal of Multilingual and Multicultural Development Journal of Intercultural Studies Annual Review of Applied Linguistics Journal of Eastern African Studies Cognitive Linguistics Language and Intercultural Communication Journal of Quantitative Linguistics linguistic typology Applied Linguistics Review International Journal of Multilingualism Journal of Language, Identity & Duration Language and Cognition Language Assessment Quarterly Language Acquisition journal of african languages and linguistics Natural Language Semantics Corpus Linguistics and Linguistic Theory journal of english as a lingua frança Language Assessment Quarterly: An International Journal





Покрытие по годам

Journal of Child Language Journal of Linguistics Applied Psycholinguistics Journal of French Language Studies English Language and Linguistics Journal of the International Phonetic Association International Journal of the Sociology of Language Natural Language and Linguistic Theory Russian Linguistics Linguistics Discourse Processes text and talk Language and Education Journal of Multilingual and Multicultural Development Journal of Intercultural Studies Annual Review of Applied Linguistics Multilingua Journal of Eastern African Studies Language and Intercultural Communication Phonetica Journal of Quantitative Linguistics Review of Cognitive Linguistics linguistic typology International Journal of Multilingualism STUF - Language Typology and Universals Journal of Language, Identity & Ducation Cognitive Linguistics linguistic review Language and Cognition Applied Linguistics Review Language Assessment Quarterly Language Acquisition Natural Language Semantics Corpus Linguistics and Linguistic Theory journal of english as a lingua franca journal_of_african_languages_and_linguistics indogermanische_forschungen Language Assessment Quarterly: An International Journal





Введение

Работа с журналами

Работа со статьями



Репозиторий

https://github.com/linguistic-journals/journal_list_data

```
    ├─ crossref
    ├─ with_abstracts
    ├─ scripts
    ├─ abcd_list_annotated.xlsx
    ├─ helsenki.csv
    # папка со сделанными журналами
    ├─ cписок журналов
    ├─ paнкинг журналов из Хелсинской бд
```



Спасибо за внимание!

