

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
«Южно-Уральский государственный университет
(национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра «Электронные вычислительные машины»

Автоматизированный сбор данных для библиометрического анализа публикаций

Автор работы:
студент группы КЭ-406
Лекомцев А.Е.

Научный руководитель:
Доцент кафедры ЭВМ
Алтухова М.А.

Актуальность

С каждым годом к исследованиям в сфере наукометрии растет интерес, проводятся систематические обзоры литературы, выполняется структурирование метаданных. Но вместе с растущим интересом к этой науке стал очевиден недостаток данных о российском сегменте научных публикаций и сравнительных исследований российского и международного контекстов.

Чтобы устранить нехватку нужно воспользоваться реферативной базой eLIBRARY, чтобы получить данные из нее.

Цель и задачи

Цель работы.

Автоматизация сбора информации о публикациях в реферативной базе eLIBRARY для ее дальнейшего анализа.

Задачи:

- 1) изучить поисковую систему в сервисе eLIBRARY: возможности, ограничения, виды запросов;
- 2) провести сравнительный анализ существующих программных решений, реализующих похожий функционал;
- 3) провести выбор средств реализации на основе сравнительного анализа и разработать приложение;
- 4) протестировать разработанное программное обеспечение.

Поисковая система сервиса eLIBRARY

ПОИСК

Найти

Расширенный поиск

Данная система имеет ограничение на вывод результатов поиска по заданным пользователем критериям в 10000 публикаций(даже если результатов поиска будет больше 10 тысяч).

Что искать

Где искать

<input checked="" type="checkbox"/> - в названии публикации	<input type="checkbox"/> - в названии организации
<input checked="" type="checkbox"/> - в аннотации	<input type="checkbox"/> - в списках цитируемой литературы
<input checked="" type="checkbox"/> - в ключевых словах	<input type="checkbox"/> - в полном тексте публикации

Тип публикации

<input checked="" type="checkbox"/> - статьи в журналах	<input checked="" type="checkbox"/> - диссертации
<input checked="" type="checkbox"/> - книги	<input checked="" type="checkbox"/> - отчеты
<input checked="" type="checkbox"/> - материалы конференций	<input checked="" type="checkbox"/> - патенты
<input checked="" type="checkbox"/> - депонированные рукописи	

Тематика

Авторы

Журналы **Добавить**
Удалить

Искать в подборке публикаций

Параметры

<input checked="" type="checkbox"/> - искать с учетом морфологии
<input type="checkbox"/> - искать похожий текст
<input type="checkbox"/> - искать в публикациях, имеющих полный текст на eLibrary.Ru
<input type="checkbox"/> - искать в публикациях, доступных для Вас
<input type="checkbox"/> - искать в результатах предыдущего запроса

Годы публикации - **Поступившие**

Сортировка **Порядок** **Очистить** **Поиск**

Удалить

Добавить
Удалить

Анализ аналогичных проектов

Критерий \ Название	«СЕБЗЕР»	«Web of Science»	«Scopus»
Возможность вывода данных из сайта	+	+	+
Экспорт данных в наиболее часто используемых форматах	-	-	-
Возможность редактирования и сортировки	-	-	-
Удобство и простота использования	-	+	+
Возможность использовать на eLIBRARY.ru	+	-	-

Требования к приложению

- поиск научных публикаций по заданным параметрам;
- сбор информации о публикациях;
- наличие функции предварительного просмотра публикаций и возможность дополнительной обработки данных;
- наличие функции перевода результатов в различные форматы для программ наукометрического анализа;
- наличие функции сохранения результатов поиска.

Обзор средств реализации

Критерий \ Название	Python	Java	C/C++
Производительность	-	+	+
Наличие библиотек и инструментов	+	+	+
Порог вхождения	+	-	-

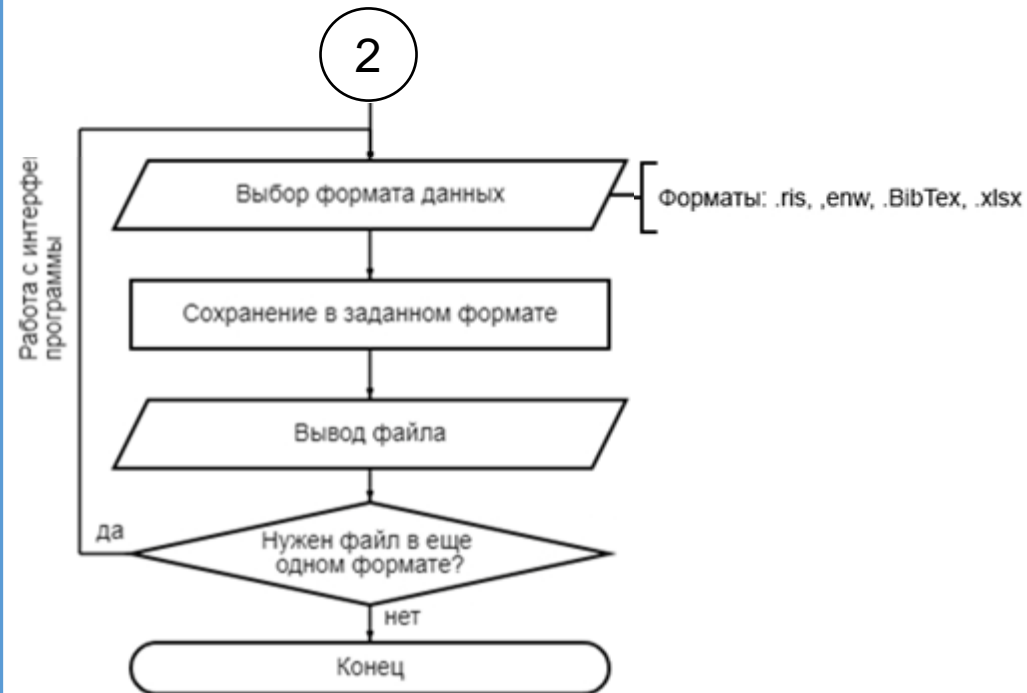
Выбор средств реализации

По результатам сравнительного анализа средств реализации и опыта разработки автора, был выбран язык программирования Python и его библиотеки для парсинга, такие как BeautifulSoup, Selenium, Selenium-stealth, а для разработки интерфейса была выбрана библиотека tkinter.

Интерфейс приложения

Что искать:	<input type="text"/>
Где искать:	<input checked="" type="checkbox"/> в названии публикации <input checked="" type="checkbox"/> в аннотации <input checked="" type="checkbox"/> в ключевых словах <input type="checkbox"/> в названии организаций авторов <input type="checkbox"/> в списках цитируемой литературы <input type="checkbox"/> в полном тексте публикации
Тип публикации:	<input checked="" type="checkbox"/> статьи в журналах <input checked="" type="checkbox"/> книги <input checked="" type="checkbox"/> материалы конференций <input checked="" type="checkbox"/> депонированные рукописи <input checked="" type="checkbox"/> диссертации <input checked="" type="checkbox"/> отчеты <input checked="" type="checkbox"/> патенты
Параметры:	<input checked="" type="checkbox"/> искать с учетом морфологии <input type="checkbox"/> искать похожий текст <input type="checkbox"/> искать в публикациях, имеющих полный текст на eLIBRARY.ru
Годы публикации:	<input type="text"/> - <input type="text"/> Поступившие: <input type="text" value="за все время"/>
Сортировка:	<input type="text" value="по релевантности"/> Порядок: <input type="text" value="по убыванию"/>
<input type="button" value="Очистить"/>	<input type="button" value="Поиск"/>
<input type="button" value="to Bibtex"/>	<input type="button" value="to RIS"/>
<input type="button" value="to ENW"/>	<input type="button" value="to Excel"/>

Алгоритм работы приложения



Тестирование

Для демонстрационного теста было введено ключевое слово «кум».

eLIBRARY ID: 35082075

EDN: URKYA 

НОМИНАЦИИ КУМ / КУМА В ВОСТОЧНОСЛАВЯНСКОЙ ПАРЕМИКЕ


МАРФИНА Ж.В.¹

¹ Луганский национальный аграрный университет, Луганск, Украина

Тип: статья в сборнике трудов конференции Язык: русский Год издания: 2018

Страницы: 111-115

eLIBRARY ID: 49818793

EDN: OVPKOK 

КУМ, КУМА И КУМЛЕНИЕ В КРОСС-ЖАНРОВОМ И КРОСС- КУЛЬТУРНОМ АСПЕКТЕ

ЩЕТИНИНА И.П.¹

¹ Московский городской педагогический университет

Научный руководитель: РАЙКОВА И.Н.¹

¹ Московский городской педагогический университет

Тип: статья в сборнике трудов конференции Язык: русский Год издания: 2022

Страницы: 158-165

КЛЮЧЕВЫЕ СЛОВА:

КУМ, КУМА, КУМЛЕНИЕ, КРОСС-ЖАНРОВЫЙ, КРОССКУЛЬТУРНЫЙ

АННОТАЦИЯ:

Статья посвящена изучению национальных и региональных вариантов обряда кумления у славян. Рассмотрено значение слов «кумиться», «кумушка» в песнях, исполнявшихся во время кумления. Особое внимание уделено многозначности слов «кум», «кума», показаны варианты употребления этих слов в различных жанрах классического фольклора.

Файлы полученные в результате работы программы

Результат работы программы в формате BibTeX

```
@article{МАРФИНА2018НОМИНАЦИИ,  
  title = {НОМИНАЦИИ КУМ / КУМА В ВОСТОЧНОСЛАВЯНСКОЙ ПАРЕМИКЕ},  
  author = {МАРФИНА Ж.В.},  
  year = {2018},  
  pages = {111-115},  
  institution = {Луганский национальный аграрный университет, Луганск, Украина},  
}  
  
@article{ЩЕТИНИНА2022КУМ,,  
  title = {КУМ, КУМА И КУМЛЕНИЕ В КРОСС-ЖАНРОВОМ И КРОСС- КУЛЬТУРНОМ АСПЕКТЕ},  
  author = {ЩЕТИНИНА И.П.},  
  year = {2022},  
  pages = {158-165},  
  institution = {Московский городской педагогический университет},  
  note = {Статья посвящена изучению национальных и региональных вариантов обряда кумления у славян. Р},  
  keywords = {КУМ, КУМА, КУМЛЕНИЕ, КРОСС-ЖАНРОВЫЙ, КРОССКУЛЬТУРНЫЙ},  
}
```

Файлы полученные в результате работы программы

Результат работы программы в формате . ris

```
TU - статья в сборнике трудов конференции
AU - МАРФИНА Ж.В.
TI - НОМИНАЦИИ КУМ / КУМА В ВОСТОЧНОСЛАВЯНСКОЙ ПАРЕМИКЕ
PY - 2018
SP - 111
EP - 115
PB - Луганский национальный аграрный университет, Луганск, Украина

TU - статья в сборнике трудов конференции
AU - ЩЕТИНИНА И.П.
TI - КУМ, КУМА И КУМЛЕНИЕ В КРОСС-ЖАНРОВОМ И КРОСС- КУЛЬТУРНОМ АСПЕКТЕ
PY - 2022
SP - 158
EP - 165
PB - Московский городской педагогический университет
N2 - Статья посвящена изучению национальных и региональных вариантов обряда кумления у славян. Рассмотрены
KW - КУМ
KW - КУМА
KW - КУМЛЕНИЕ
KW - КРОСС-ЖАНРОВЫЙ
KW - КРОССКУЛЬТУРНЫЙ
```

Файлы полученные в результате работы программы

Результат работы программы в формате .enw

```
%0 статья в сборнике трудов конференции
%A МАРФИНА Ж.В.
%Т НОМИНАЦИИ КУМ / КУМА В ВОСТОЧНОСЛАВЯНСКОЙ ПАРЕМИКЕ
%D 2018
%Р 111-115
%I Луганский национальный аграрный университет, Луганск, Украина

%0 статья в сборнике трудов конференции
%A ЩЕТИНИНА И.П.
%Т КУМ, КУМА И КУМЛЕНИЕ В КРОСС-ЖАНРОВОМ И КРОСС- КУЛЬТУРНОМ АСПЕКТЕ
%D 2022
%Р 158-165
%I Московский городской педагогический университет
%N Статья посвящена изучению национальных и региональных вариантов обряда кумления у славян. Рассмотрен
%К КУМ
%К КУМА
%К КУМЛЕНИЕ
%К КРОСС-ЖАНРОВЫЙ
%К КРОССКУЛЬТУРНЫЙ
```

Файлы полученные в результате работы программы

Результат работы программы в формате .xlsx

ID	EDN	Type	Title	Author	Data	Lang	Year	Pages	Keywords	Annotation
35082075	URKDYA	статья в с	НОМИНАЦИИ КУМ / КУМ	МАРФИНА Ж.В	Луганский нац	русский	2018	111-115		
49818793	OVRKOK	статья в с	КУМ, КУМА И КУМЛЕНИЕ	ЩЕТИНИНА И.Г	Московский го	русский	2022	158-165	КУМ, КУМ,	Статья посвящ

Основные результаты

- изучена поисковая система в сервисе eLIBRARY: возможности, ограничения, виды запросов;
- проведен сравнительный анализ существующих программных решений, реализующих похожий функционал;
- проведен выбор средств реализации на основе сравнительного анализа и разработано приложение;
- проведено тестирование разработанного приложения.

Спасибо за внимание!