

Разработка свободного программного средства обработки и структуризации библиографических данных

Кононова Александра Игоревна, к.т.н., доцент;

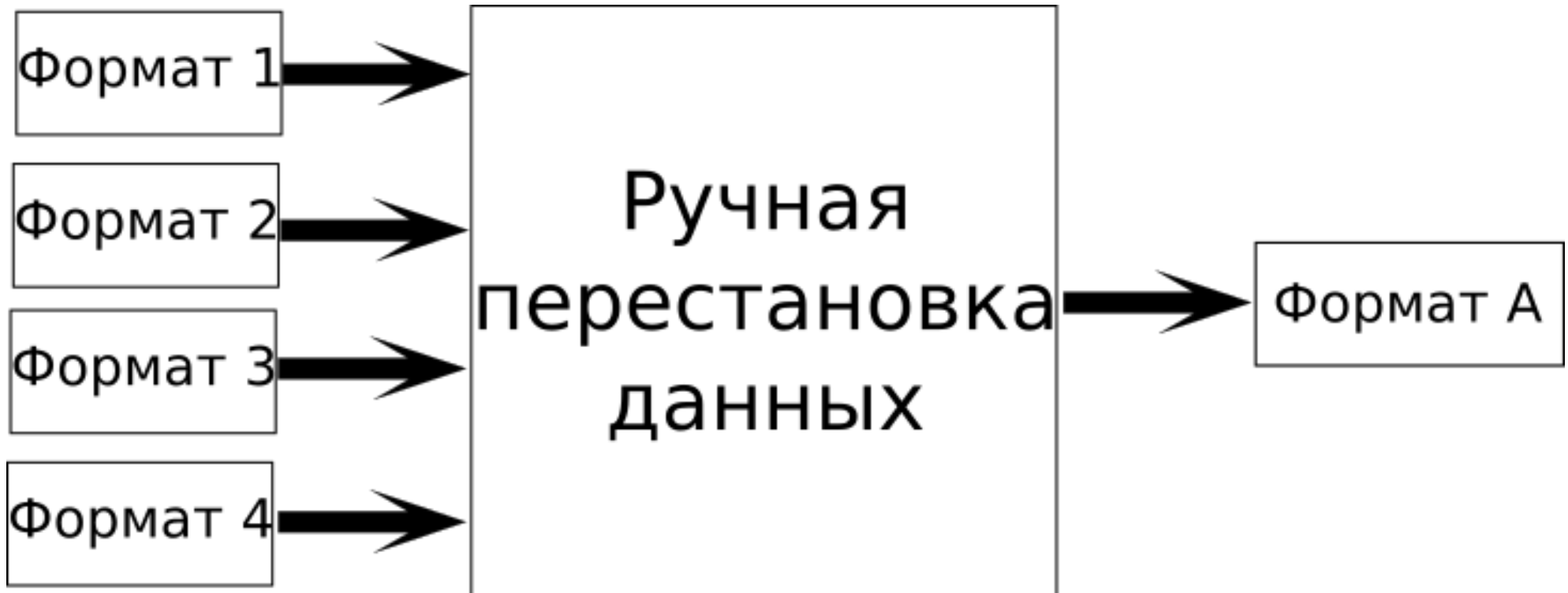
Петров Евгений Николаевич, студент

НИУ МИЭТ

Цель работы

Оптимизация процесса формирования сводной таблицы научных трудов кафедры посредством разработки программного средства обработки и структуризации библиографических данных, способного принимать информацию из различных источников и генерировать сводную отчетную таблицу, формат которой легко дорабатывать под возможные изменения требований оформления.

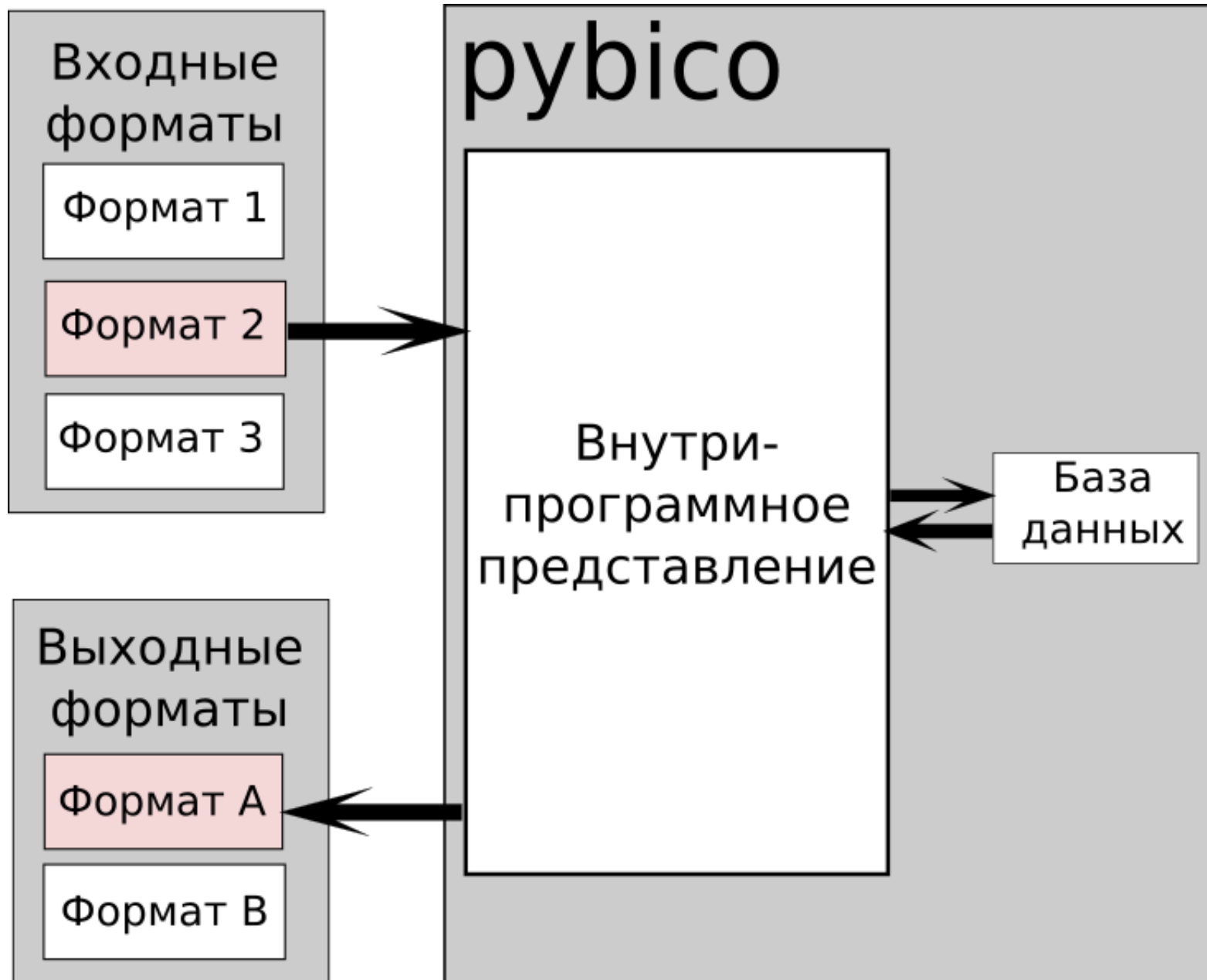
Проблемная ситуация



Примеры входных данных

- Алексеев **А. А.**, Иванов **Иван**, Петров П. П. Анализ интенсивности аварийных потоков в системах диспетчерского управления// Интернет журнал “Науковедение”, №6, 2004.
- Алексеев **А.А.**, Иванов **И.** Методика проектирования эффективных логических структур баз данных систем технической диагностики// **Современные проблемы науки, 2016.-№3.**

Функциональная схема



Аргументы командной строки

- h – вывод информации о поддерживаемых аргументах командной строки;
- v – подробная информация о ходе выполнения программы;
- l – путь к загружаемому файлу;
- s – путь к файлу сохранения;
- i – формат загружаемого файла (txt, id);
- e – формат файла сохранения (xlsx);
- u – логин пользователя базы данных;
- p – путь к файлу пароля от базы данных.

Входной формат txt

- Сидоров С. С., Иванов Иван, Петров П. П. Анализ интенсивности аварийных потоков в системах диспетчерского управления// Интернет журнал “Науковедение”, №6, 2004.
- Сидоров С.С., Иванов И. Методика проектирования эффективных логических структур баз данных систем технической диагностики// Современные проблемы науки, 2016.-№3.

Входной формат id

- 7676545
- 3243253
- 2342678

Выходной формат xlsx

Сводная таблица со следующими полями:

- номер раздела;
- автор (с возможностью подчеркивания сотрудников вуза, выделения жирным аспирантов и курсивом – студентов);
- название статьи;
- вхождение в РИНЦ, ВАК, Scopus, Web of Science; импакт-факторы;
- информация о городе, издательстве, годе, номере, томе и страницах;
- количество авторов, в том числе сотрудников вуза.

Регулярное выражение для обработки списка авторов

(?P<authors>

(

(\pL\.\ ?(\pL\.\)?\pL+,\?)| // И.И. Иванов, И. И. Иванов, И. Иванов

(\pL+ \pL\.\ ?(\pL\.\)?,\?)| // Иванов И.И., Иванов И. И., Иванов И.

(\p{Lu}\p{LI}+ \p{Lu}\p{LI}+,\?) // Иванов Иван, Иван Иванов

)+

)

База данных

author	publication	source
идентификатор	идентификатор	идентификатор
ИМЯ	название	название
должность	идентификатор источника	номер раздела
вуз	прочая информация	Scopus
кафедра	флаг учтенности	Web of Science
		РИНЦ
		ВАК

Программные средства

- Язык программирования: Python;
- регулярные выражения: модуль `regex`;
- http-запросы: модуль `requests`;
- работа с MySQL: модуль `PyMySQL`;
- работа с `xlsx`: модуль `xlsxwriter`.

Результат работы программы

	A	B	C	D	E	F	G	H	I
1	№ раздела	Автор (ФИО сотрудника МИЭТ, студента, аспиранта, center)	Название статьи, книги, монографии, уч. пособия и др.	Наименование журнала или конференции	Город, издательство, год, номер, том, страницы	Статус Web of Science Scopus РИНЦ ВАК	Импакт-фактор	Количество авторов	
2								Всего	В т.ч. сотрудников МИЭТ
3	1	2	3	4	5	6	7	8	9
4	0	<p>Авторы: <u>Авдеев А.Р.</u>, <u>Чук К.С.</u>, <u>Семеев А.В.</u></p>	<p><u>СОЗДАНИЕ МАСШТАБИРУЕМЫХ СРЕДСТВ ДЛЯ РЕШЕНИЯ ЗАДАЧ АНАЛИЗА БОЛЬШИХ ОБЪЕМОВ ДАННЫХ НА ОСНОВЕ СИСТЕМЫ УПРАВЛЕНИЯ БАЗЫ ДАННЫХ MONGODB</u></p>	<p><u>Вестник Поволжского государственного технологического университета. Серия: Радиотехнические и инфокоммуникационные системы</u></p>	<p><u>2016. №1(29). С. 55-63.</u></p>			3	3
5	0	<p>Авторы: <u>Чук К.С.</u>, <u>Исков Н.Г.</u></p>	<p><u>ОЦЕНКА И ВЫБОР УСЛУГ ПРЕДПРИЯТИЙ ЖКХ НА ОСНОВЕ ЭКОНОМИКО-МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ</u></p>	<p><u>Вестник Поволжского государственного университета сервиса. Серия: Экономика</u></p>	<p><u>2014. №2(34). С. 100-103.</u></p>			2	2
6	0	<p>Авторы: <u>Чук К.С.</u>, <u>В.Г.П.</u>, <u>Чук О.И.</u></p>	<p><u>УПРАВЛЕНИЕ СТОИМОСТЬЮ БИЗНЕСА ДЛЯ ПРИНЯТИЯ ОБОСНОВАННЫХ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ</u></p>	<p><u>Вестник Поволжского государственного университета сервиса. Серия: Экономика</u></p>	<p><u>2014. №1(33). С. 190-200.</u></p>			3	3
7	0	<p>Авторы: <u>Чук К.С.</u></p>	<p><u>СРЕДСТВА АУТЕНТИФИКАЦИИ И ИДЕНТИФИКАЦИИ БЕЗОПАСНОСТИ В КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ</u></p>	<p><u>Молодой ученый</u></p>	<p><u>2014. №4-2(63). С. 118-121.</u></p>			1	1

Результат работы

Было разработано программное средство обработки и структуризации библиографических данных, способное принимать информацию из различных источников и генерировать сводную отчетную таблицу, формат которой легко дорабатывать под возможные изменения требований оформления.

Результатом работы является оптимизация процесса формирования сводной таблицы научных трудов кафедры.

Дальнейшее развитие

- Увеличение числа обрабатываемых форматов;
- создание веб-сервиса в локальной сети вуза;
- добавление гибкой фильтрации.

Спасибо за внимание!