

# РАЗРАБОТКА ПОДХОДОВ И МЕТОДОВ ОБРАБОТКИ И ИЗВЛЕЧЕНИЯ БИБЛИОГРАФИЧЕСКОЙ ИНФОРМАЦИИ ИЗ НАУЧНОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКИ ELIBRARY

---

**Дырnochкин Александр**

Ульяновский государственный технический университет

*Целью работы является разработка подходов и методов обработки и извлечения библиографической информации, для последующего наукометрического анализа публикаций и формирования научных групп по заданной тематике, на основе данных об авторах.*

*Объектом исследования является информация по публикациям различных авторов из цифровых библиографических и реферативных базы данных. Такие базы представляют собой инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях. Также они являются одним из главных источников получения наукометрических данных для проведения оценочных исследований.*

*Предметом исследования является библиографическая информация по сотрудникам УлГТУ из научной электронной библиотеки elibrary.*

Достижение поставленной цели предполагает реализацию следующих задач:

- анализ и исследование существующих моделей, алгоритмов и методов интеллектуального анализа и кластеризации текстов
- проектирование алгоритма извлечения и структуризации библиографических данных с сайта [elibrary.ru](http://elibrary.ru)
- формализация и прототипирование алгоритма интеллектуального анализа текста, для формирования научных групп
- проведение экспериментов с полученным прототипом системы

*Научная новизна работы состоит в гибридизации алгоритмов интеллектуального анализа текстов и кластерного анализа для формирования научных (ученых) групп на основе библиографической информации.*

*Исходными данными системы являются научные идентификаторы сотрудников УлГТУ. На основании которых будет загружена информация по всем сатъям автора из elibrary.*

```
{
  "LastName": "Наместников",
  "FirstName": "Алексей",
  "PatronymicName": "Михайлович",
  "AuthorId": "http://elibrary.ru/author_profile.asp?id=392690"
},
{
  "LastName": "Ларионова",
  "FirstName": "Ольга",
  "PatronymicName": "Борисовна",
  "AuthorId": "http://elibrary.ru/author_profile.asp?id=813011"
},
{
  "LastName": "Киселев",
  "FirstName": "Евгений",
  "PatronymicName": "Степанович",
  "AuthorId": "http://elibrary.ru/author_profile.asp?id=150794"
},
{
  "LastName": "Крашенинников",
  "FirstName": "Виктор",
  "PatronymicName": "Ростиславович",
  "AuthorId": "http://elibrary.ru/author_profile.asp?id=15724"
},
}
```

Figure: Исходные данные

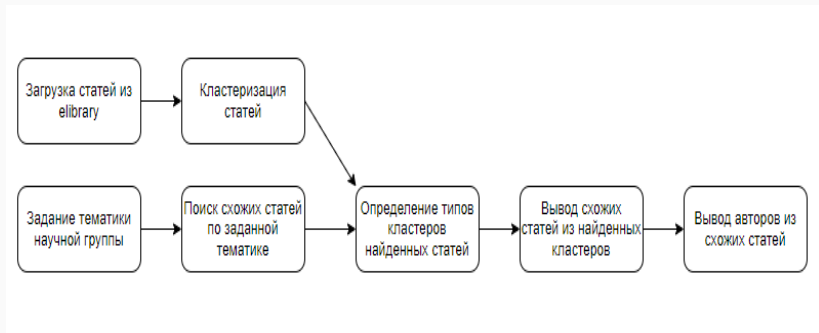


Figure: Схема формирования научных групп

- алгоритм извлечения информации по авторам с сайта elibrary
- алгоритм интеллектуального анализа текстов и кластерного анализа для формирования научных (ученых) групп на основе библиографической информации.
- механизм формирования графа соавторов



- Кластеризация текстовых документов из электронной базы публикаций алгоритмом FRiS-Tax. Н.Г. Загоруйко, В.Б. Барахнин, И.А. Борисова, Д.А. Ткачёв
- Тематическая кластеризации научной литературы. А.В. Николаев, В.В. Жуков
- Обзор и экспериментальное сравнение методов кластеризации текстов. П.А. Пархоменко, А.А. Григорьев, Н.А. Астраханцев