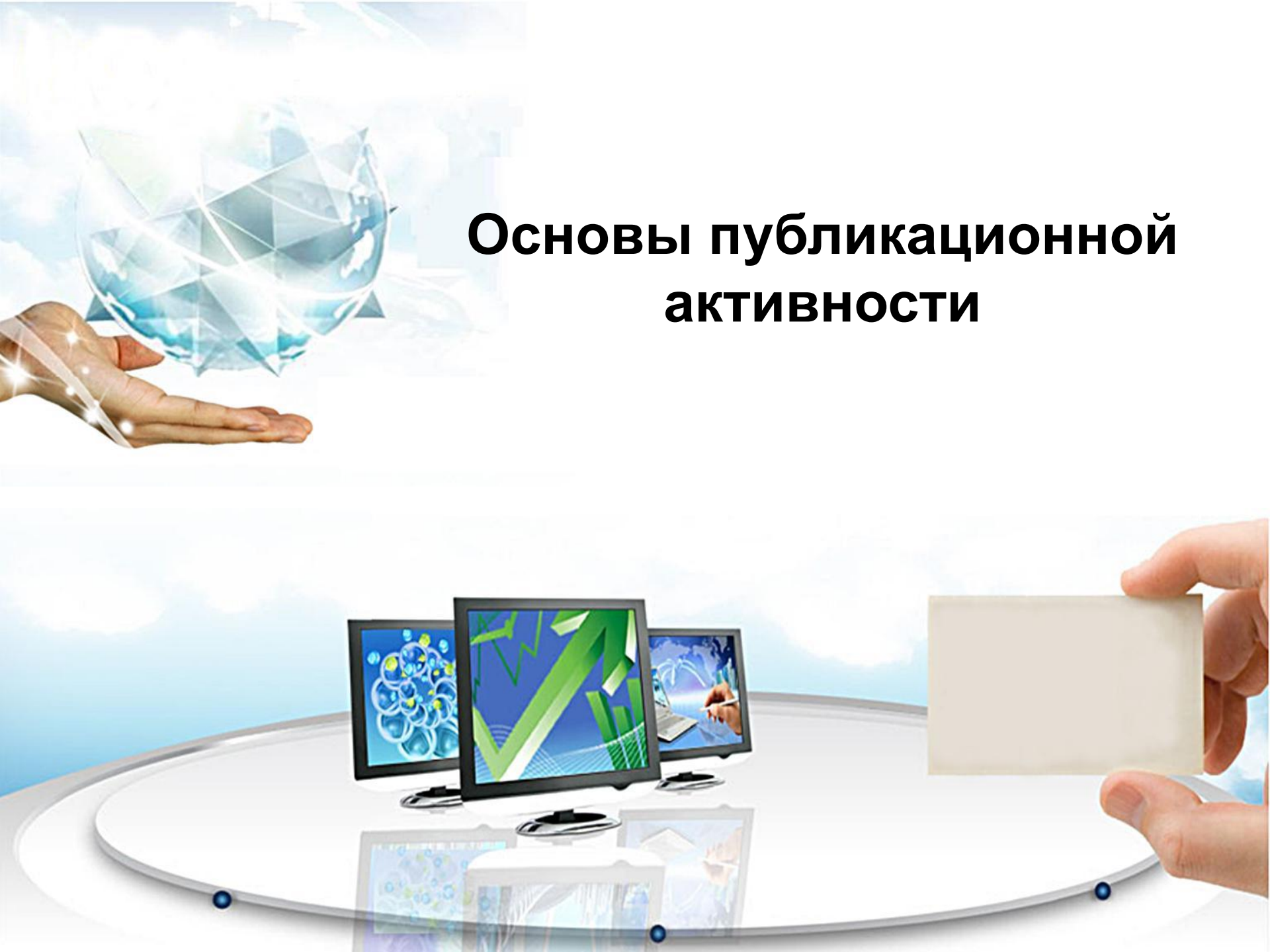


Основы публикационной активности



Общеизвестные факты

- Более 90% российских публикаций отсутствуют в зарубежных базах цитирования.
- Структура российских публикаций в зарубежных базах **кардинально** отличается от структуры в РИНЦ.
- Граф цитирований российских публикаций слабосвязный, цитируют плохо и мало.
- В российских базах много некачественных публикаций.
- Российские базы не состыкованы друг с другом.

Где хранятся научные тексты?

- РГБ (диссертации и авторефераты)
- ЦИТИС (отчеты о НИОКР)
- Дирекция ФЦП (заявки отчеты по ФЦП)
- РФФИ, РНФ и другие фонды (заявки и отчеты о НИОКР)
- ФИПС (база патентов)
- eLibrary, Киберленинка, MathNet (русские публикации)
- РУКОНТ, ИНФРА-М и др. (массивы научной литературы)
- ЭБС ВУЗов (ВКР, ВУЗовский контент)

В чем проблема и как ее решать?

Проблема:

- Слабая интеграция российской науки в мировую, разрозненность баз, отсутствие современных отечественных инструментов не позволяют решить ряд важных задач: картирование науки, оценка коллективов исследователей, оценка научных журналов и др.

Что делать:

- Интегрировать информацию из различных источников.
- Комбинировать анализ полных текстов и наукометрию.
- Не бояться придумывать что-то новое, чего нет в Scopus или WoS 😊

Что такое анализ полных текстов научных публикаций?

НАЗВАНИЕ ДОКЛАДА

Автор 1, Автор 2

Место выполнения работы

Адрес автора, электронный адрес автора

Аннотация

Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации. Текст аннотации.

Ключевые слова: слово1, слово2, слово3.

1. Введение

Текст введения. Текст введения. Текст введения. Текст введения. Текст введения. Текст введения. Текст введения. Текст введения.

2. Заголовок пункта

Текст статьи [Ссылка1]. Текст статьи [Ссылка 2]. Текст статьи [Ссылка 3]. Текст статьи.

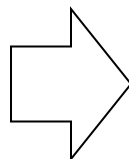
3. Заключение

Текст заключения. Текст заключения. Текст заключения. Текст заключения. Текст заключения. Текст заключения. Текст заключения. Текст заключения.

Литература

1. Источник 1.
2. Источник 2.

1. Выделение текста.
2. Лингвистический анализ (морфология, синтаксис, семантика).
3. Извлечение информации (название, авторы, аннотация, ключевые слова и тд.)
4. Индексация в базе данных поисково-аналитической машины.



Базовые алгоритмы для работы с полными текстами

- Семантический поиск.
- Поиск тематически похожих документов.
- Автоматическое выявление ключевых слов.
- Автоматическая классификация.
- Кластеризация.
- Тематический анализ.
- Семантический поиск текстовых заимствований.
- Извлечение информации из текстов.
- Анализ тональности высказываний.

Пример 1: выявление научных коллективов и анализ их активности

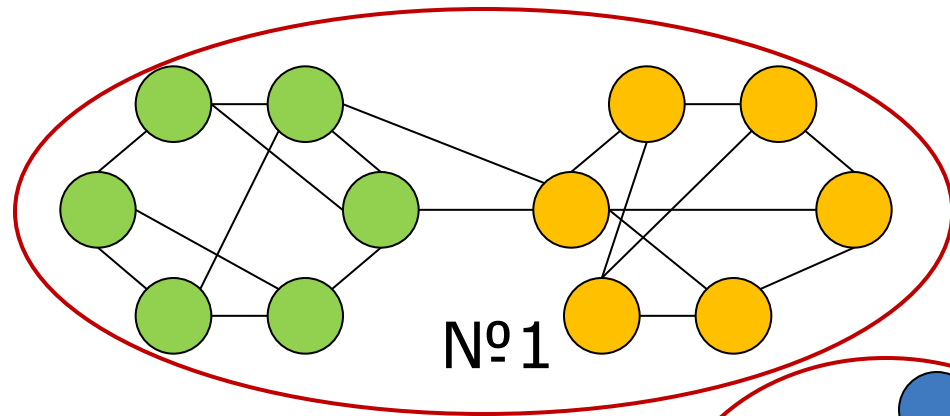
Что такое научный коллектив?

- Научный коллектив – группа исследователей, ведущих совместную научную деятельность в одной или нескольких близких предметных областях и имеющих совместные публикации в открытой научной печати за определенный период времени.

Как решается проблема на западе

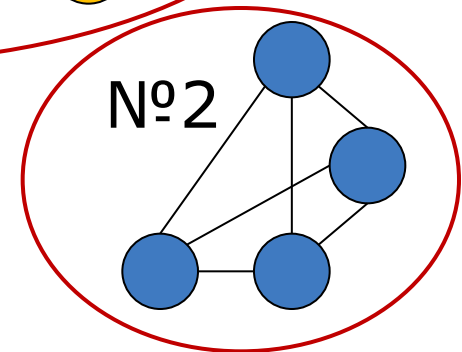
Анализ цитирований:

- Cytoscape
- igraph
- NetMiner
- ...



Западные инструменты:

- Не позволяют разделять тесно взаимодействующие коллективы.
- Не используют информацию о цитированиях.
Не используют полные тексты для определения предметной области коллектива.
- Не выделяют роли в коллективах (лидеры, основной состав, приглашенные авторы и др.)



С использованием западных инструментов задача выявления российских коллективов исследователей **не может быть решена!**

Формирование графа

1. Реляционно-ситуационный метод поиска и анализа текстов и его приложения

Авторы: Г.С. Осипов, И.В. Смирнов, И.А. Тихомиров.

Год публикации:2008.

2. СПОСОБ И СИСТЕМА СЕМАНТИЧЕСКОГО ПОИСКА ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Авторы: Осипов Геннадий Семенович (RU), Тихомиров Илья Александрович (RU), Соченков Илья Владимирович (RU), Смирнов Иван Валентинович (RU).

Год публикации:2013.

3. Увеличение разрешения полнопольного оптического когерентного томографа при использовании цветного датчика изображения

Авторы: А. Л. Кальянов а , В. В. Лычагов а , И. В. Смирнов а , В. П. Рябухо аб.

Год публикации:2013.

Приведение имени автора к единому виду:

1. Смирнов И.В.	} =>	Смирнов И.В.
2. Смирнов Иван Валентинович		
3. Смирнов Илья Владимирович		

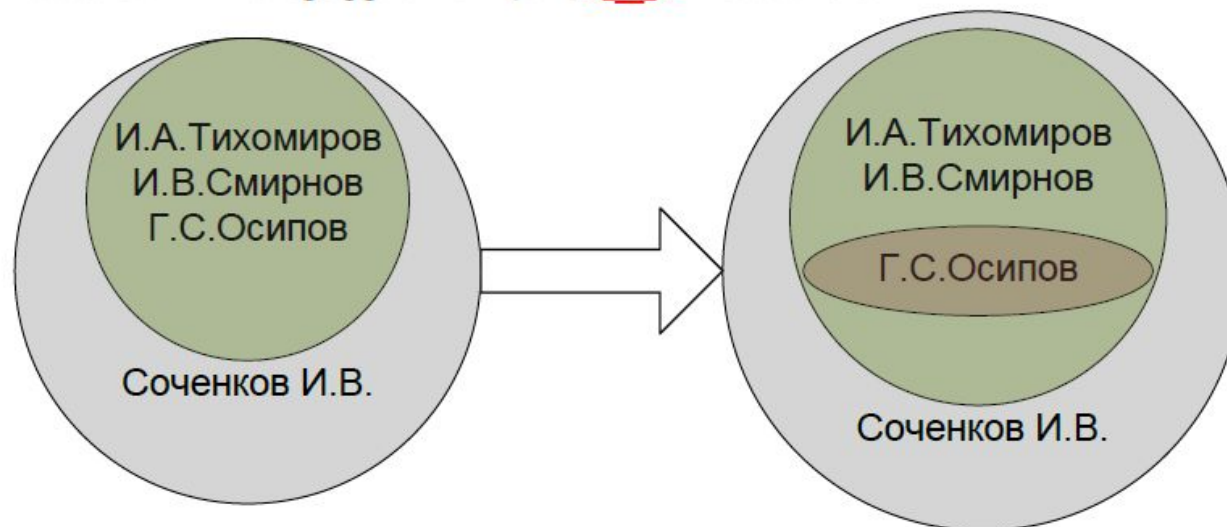
Как разделить авторов с одинаковыми именами?

- Построение коллективов для различных отраслей науки.
- Кластеризация построенных коллективов на основе анализа полных текстов.

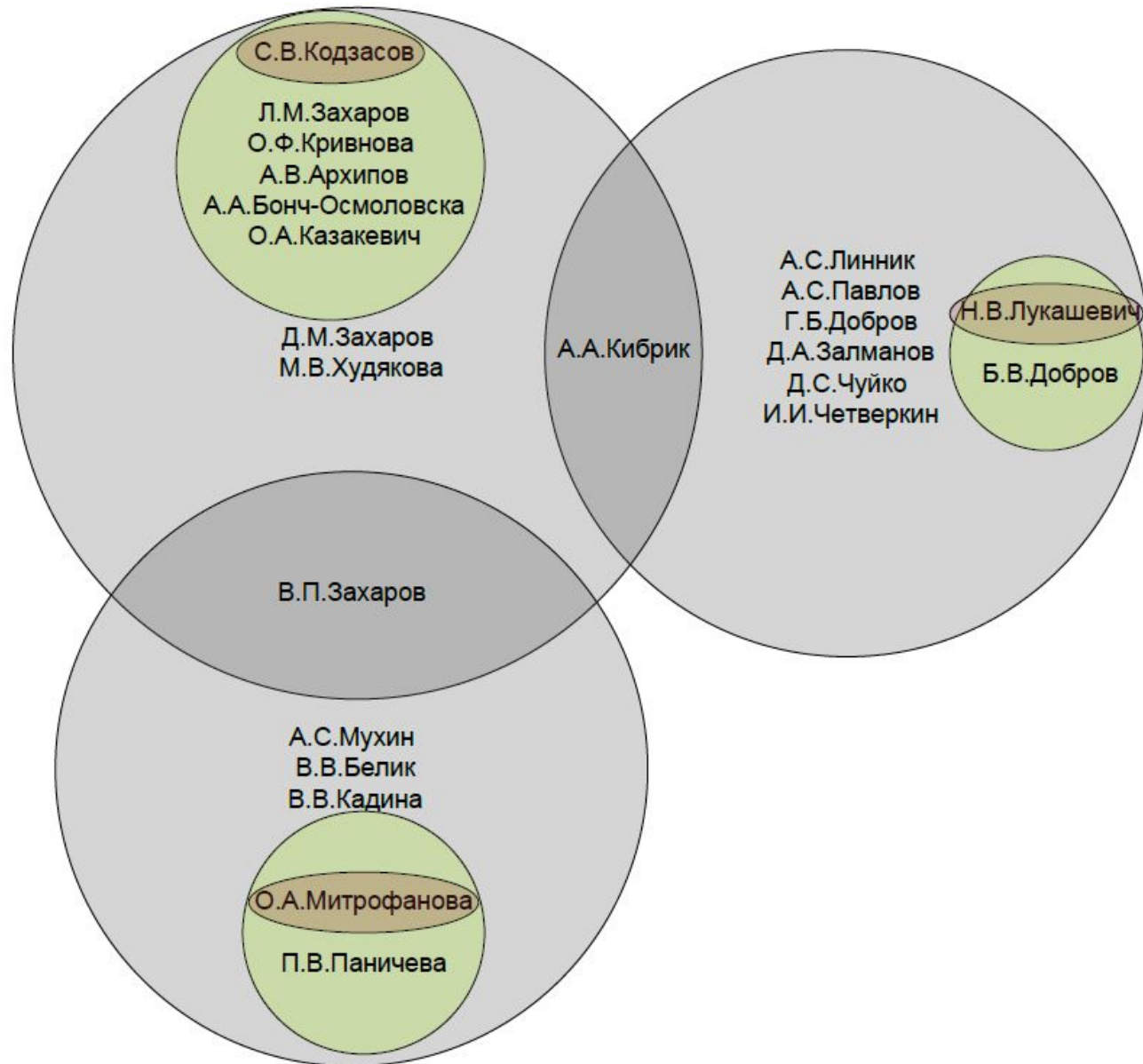
Учет цитирований

Литература

1. Gennady Osipov, Ivan Smirnov, Ilya Tikhomirov. Application of Linguistic Knowledge to Search Precision Improvement. // Proceedings of 4th International IEEE conference on Intelligent Systems 2008 Volume 2. – P. 17-2 – 17-5.
2. Тихомиров И.А., Смирнов И.В. Интеграция лингвистических и статистических методов поиска в поисковой машине Exactus // Труды международной конференции Диалог 2008 – С. 485–491.
3. Золотова Г. А., Онипенко Н. К., Сидорова М. Ю. Коммуникативная грамматика русского языка. – М. 2004 – 544 с.
4. Осипов Г. С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. – М.: Наука, Физматлит, 1997 – 112 с.
5. Смирнов И.В., Соченков И.В., Муравьев В.В., Тихомиров И. А. Результаты и перспективы поискового алгоритма Exactus. // Труды российского семинара по оценке методов информационного поиска РОМИП'2007-2008. Санкт-Петербург: НУ ЦСИ, 2008 – С. 66–76.



Пример коллективов



Лидеры коллектива

[Н.В. Лукашевич \(27\)](#)

Регулярные авторы

[М.С. Агеев \(12\)](#) [Б.В. Добров \(25\)](#) [А.В. Сидоров \(13\)](#)

Нерегулярные авторы

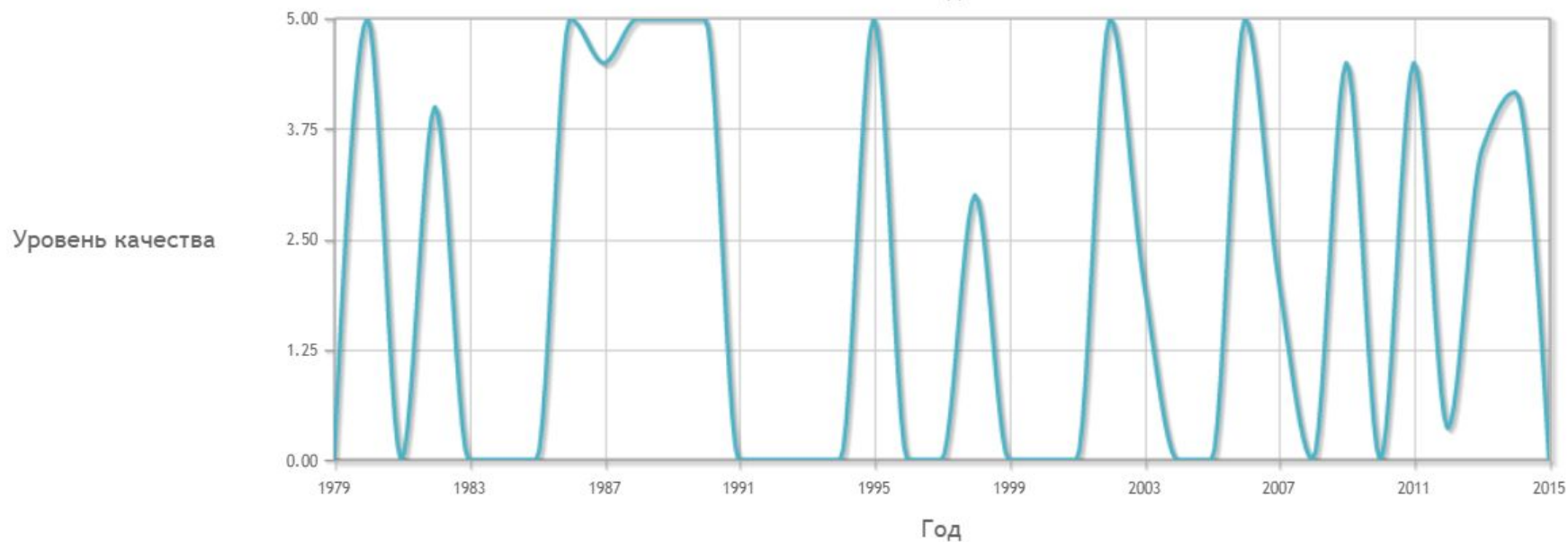
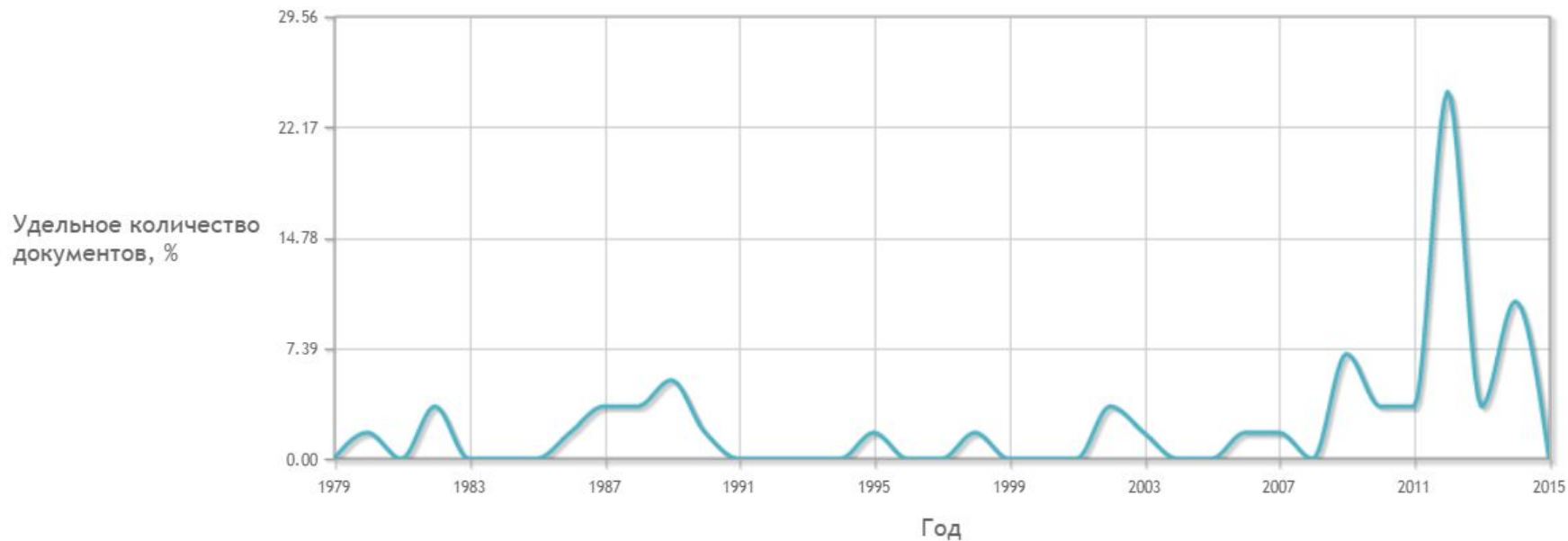
[А.А. Алексеев \(9\)](#) [П. Анд \(1\)](#) [Н.А. Богомолов \(5\)](#) [С.В. Журавлев \(2\)](#) [Д.А. Залманов \(1\)](#) [А.Д. Ковалев \(1\)](#)
[Д.С. Котельников \(2\)](#) [А.С. Линник \(2\)](#) [Ю.М. Логачев \(1\)](#) [Н.В. Макаров-Землянский \(1\)](#) [А.С. Павлов \(7\)](#)
[А.Д. Салий \(1\)](#) [Р.А. Сеннов \(1\)](#) [С.В. Сыромятников \(1\)](#) [И.И. Четвёркин \(2\)](#) [Д.С. Чуйко \(1\)](#)
[С.В. Штернов \(3\)](#) [О.А. Штернова \(1\)](#)

Приглашенные авторы

[А.В. Богомолова \(4\)](#) [Г.Б. Добров \(3\)](#) [О.И. Карасев \(3\)](#) [А.А. Кибрик \(6\)](#) [Т.Н. Юдина \(6\)](#)

"ahmad" "creditor" "debtor" "refrhet" "аккуратность предсказания" "англоязычный текст" "аннотация кластера"
"булевская формула" "внешняя характеристика" "генератор" **"генератор текста"**
"грамматическая роль" **"естественный текст"** "защита природы" "извлечение шаблона" "корпус
мнения" "кредит" "лист дерева" "логистическая регрессия" "массово" "метод обучения"
"модель выбора" "неоценочный" **"обнаружение текста"** **"общественно-
политический тезаурус"** "оент" "описание рубрики" **"основной
узел"** "очередное предложение" "получение кредита" "понятие текста" "понятие тысячи"
"порождение текста" "построение представления" "разный эксперт" "состав речи" "социально-
экономический процесс" "список словосочетания" "странность" "сумма вероятности" **"текстовый вход"**
"тематическая структура" **"тематический узел"** "терминологическое словосочетание" "тип группы"
"участник события" **"частотность"** **"частотность леммы"** "частотность
слова" "частотность словосочетания" "часть набора" "широкая область"

Анализ активности коллектива



Пример 2: оценка качества научного журнала

Как сейчас оценивается журнал?

- Экспертным путем.
- С использованием наукометрии.
- Комбинированный метод (RSCI).

Дополнительных характеристики журнала, которые можно получить на основе анализа полных текстов:

- качество статей;
- тематическая структура журнала;
- заимствования;
- выявление накруток цитирований;
- ...

Обнаружение накрутки цитирований

Спам. Массовая публикация коротких тезисов с большими списками цитирования.

Как определить:

- вычислить реальный размер публикации и соотносить его с количеством источников.

Ложное цитирование. Приписка цитирований статей, которые не используются в исследовании.

Как обнаружить:

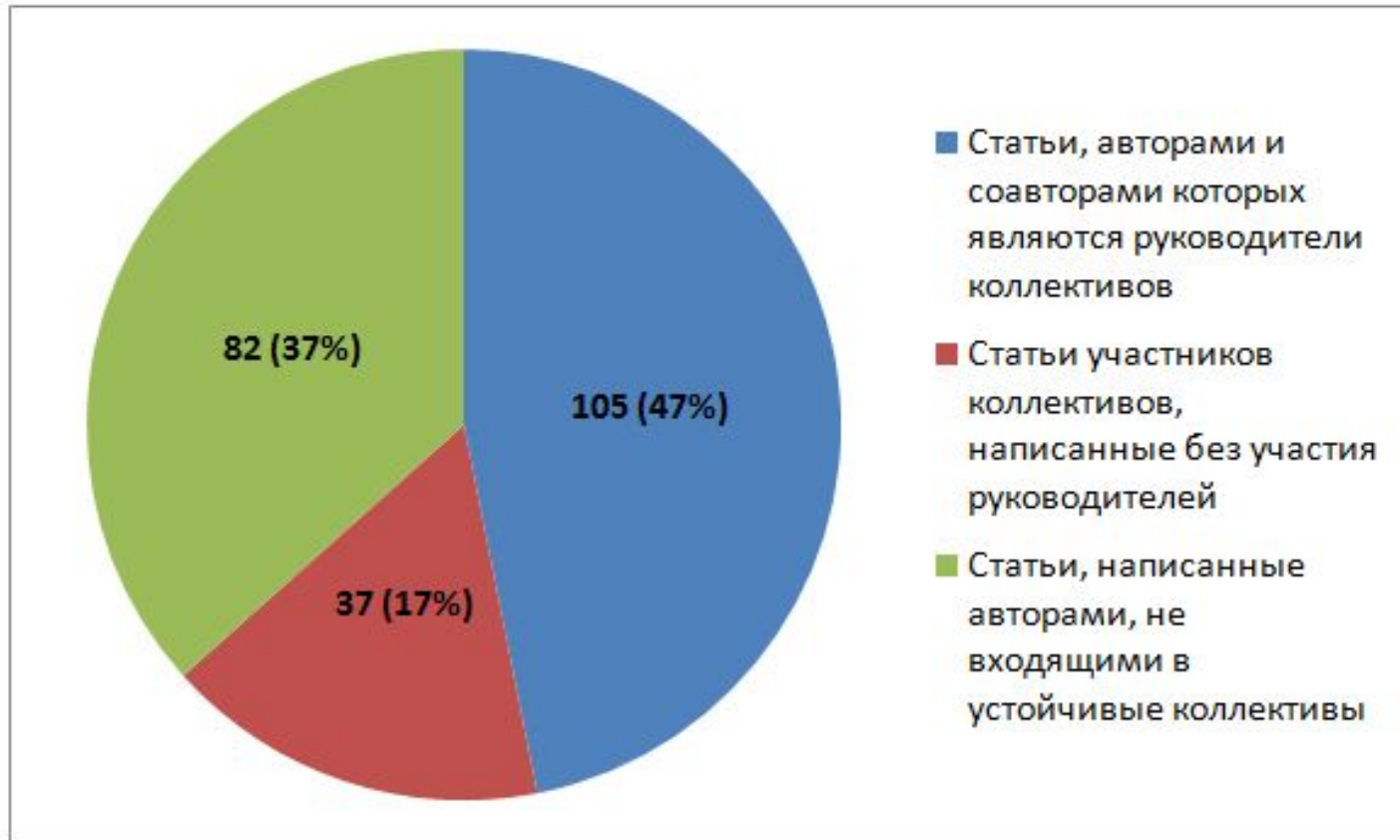
- определить тематическое соответствие контекстов, тематики и наличия реальных ссылок в текстах публикаций.

Клонирование и плагиат. Публикация нечетких дубликатов с измененным названием и текстом.

Что делать:

- Использовать средства поиска текстовых заимствований.
- Сопоставлять списки цитирования (см. Гуреев В.Н., Мазов Н.А. Анализ цитирования как основа для разработки дополнительного модуля в системах антиплагиата // НТИ. Сер.1.– 2013.–12.–С. 12-15

Соотношение статей в журнале по типу авторства



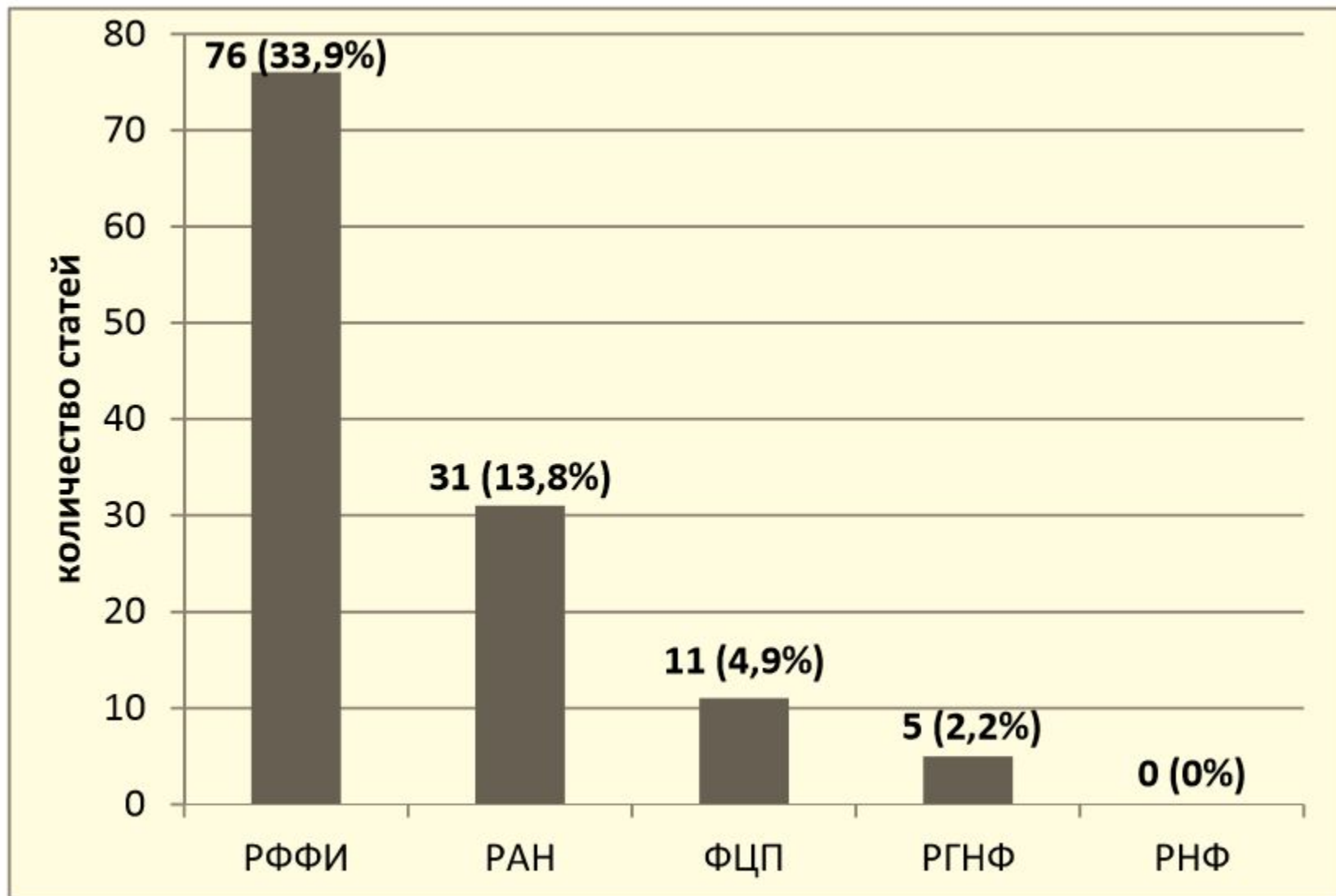
Тематическая структура журнала

№	Терминологическая лексика кластера	Авторы	N
1	<p>"аксиоматический подход" "аппроксимация оболочки" "бинарные отношения"</p> <p>"важность критериев" "векторные критерии" "выбираемые варианты"</p> <p>"выбираемый вектор" "готовность лпр" "граница парето" "диалоговые карты" "задача выбора" "иррефлексивный" "искусственный интеллект" "лпр" "многокритериальная оптимизация"</p> <p>"многокритериальная среда" "многокритериальные задачи"</p> <p>"многокритериальный выбор" "многомерная граница" "множество векторов" "множество парето" "недоминировать"</p> <p>"неизвестное множество" "отношение предпочтения"</p> <p>"парето-оптимальные векторы" "парето-оптимальный"</p> <p>"предпочтение лпр" "принцип эджворта-парето"</p> <p>"проблемы сужения" "сужение множества" "транзитивный"</p>	<p>В.Д. Ногин</p> <p>В.В. Подиновский</p> <p>А.В. Лотов</p> <p>А.И. Рябиков</p> <p>А.Л. Бубер</p> <p>А.В. Холмов</p> <p>О.Н. Климова</p> <p>Н.Б. Брусникина</p> <p>О.В. Басков</p> <p>А.О. Захаров</p> <p>А.П. Нелюбин</p>	14
2	<p>"именная группа" "интеллектуальный поиск"</p> <p>"информационный поиск" "коллекция документов" "компьютерная лингвистика" "морфологический анализ" "поисковые запросы"</p> <p>"семантический поиск" "синтаксический анализ" "слова запроса"</p>	<p>В.Ф. Хорошевский</p> <p>Г.С. Осипов</p> <p>И.В. Смирнов</p> <p>А.О. Шелманов</p> <p>Е.С. Кузнецова</p> <p>И.В. Храмоин</p> <p>И.А. Тихомиров</p> <p>Р.Е. Суворов</p> <p>И.В. Соченков</p>	8

Проверка оригинальности статей



Поддержка исследований грантами



Пример 3: анализ научных направлений

Как сейчас выявляют научные направления?

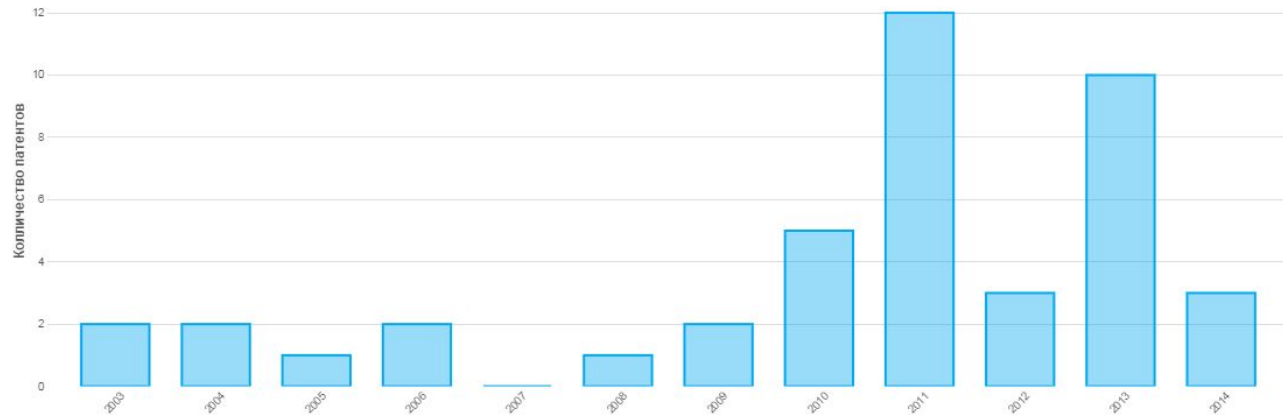
Фронты исследования по оптогенетике в ESI

<i>Fronts</i>	<i>Papers</i>	<i>Citations</i>	<i>Citations Per Paper</i>	<i>Mean Year</i>
BDNF SIGNALING MIMICS OPTOGENETIC CONTROL; ULTRAFAST OPTOGENETIC CONTROL; VTA GABA NEURONS DISRUPTS REWARD CONSUMPTION; AGRP NEURONS DRIVES FEEDING BEHAVIOR; APPLYING OPTOGENETIC TOOLS	36	2,864	79.56	2010.8
OPTOGENETIC TOOLS; NEURAL CIRCUITS; ANALYZING; BEHAVIOR; DEVELOPMENT	2	156	78.00	2011.5
TWO-PHOTON OPTOGENETIC TOOLBOX; SYNTHETIC MEMORY TRACE; FEAR MEMORY RECALL; TWO-PHOTON OPTOGENETICS; OPTOGENETIC STIMULATION	6	110	18.33	2012.5

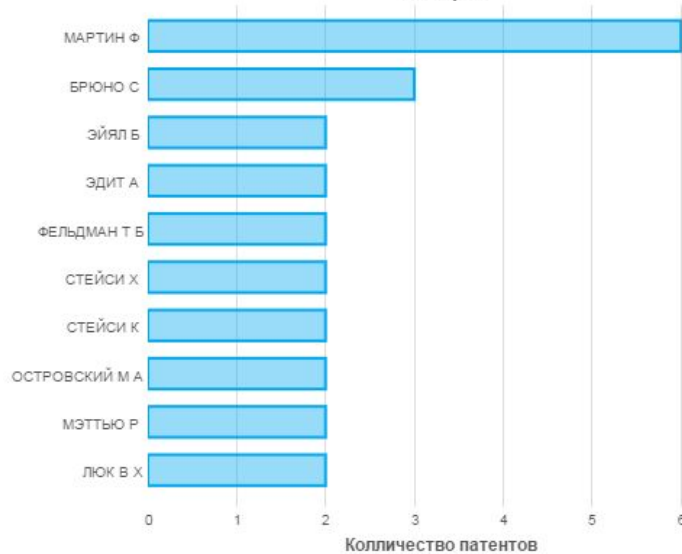
* Увы, российские публикации плохо цитируются и ни одного российского направления исследований мы не найдем.

Пример: российские патенты «оптогенетика» в Exactus Patent

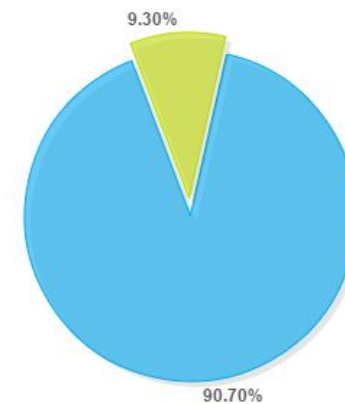
Динамика



Авторы



Коллекции



■ ФИПС. Изобретения
■ ФИПС. Полезные модели

Пример: коллектив по «оптогенетике» в Exactus Expert

Коллектив № 1

"bowden" "dowling" "durfee" "fedotov" "raman" "возбудить" **"волновод"** "восприимчивость"
"время задержки" "генерация импульса" "гладков" "гранат" **"групповая скорость"** "давление газа"
"изотропная среда" "квантовая электроника" **"короткий импульс"** "краткое
сообщение" **"накачка"** "основные данные" "синхронизм" "состояние поляризации"
"спектроскопия" "тензор восприимчивости" "частота импульса"

Лидеры коллектива

[Н.И. Коротеев \(19\)](#)

Регулярные авторы

[С.М. Гладков \(5\)](#) [А.М. Желтиков \(17\)](#) [А.Н. Наумов \(10\)](#) [А.Б. Федотов \(3\)](#)

Нерегулярные авторы

[Д.А. Акимов \(1\)](#) [Л.С. Асланян \(1\)](#) [А.М. Бродниковский \(1\)](#) [А.Г. Варданян \(1\)](#) [С.Н. Волков \(5\)](#)
[М.Г. Каримов \(1\)](#) [О.А. Колеватова \(1\)](#) [П.С. Ланда \(1\)](#) [О.В. Лобанов \(1\)](#) [М.В. Рычев \(2\)](#)
[Д.А. Сидоров-Бирюков \(2\)](#) [А.В. Тарасишин \(1\)](#) [М.Ф. Терновская \(1\)](#) [А.Б. Федоров \(2\)](#) [И.Л. Шумай \(1\)](#)

Приглашенные авторы

[Ф.В. Бункин \(71\)](#) [В.П. Жаров \(7\)](#) [В.Н. Задков \(4\)](#) [Б.Б. Кадомцев \(13\)](#) [В.Ф. Камалов \(3\)](#)
[Ю.Л. Климонтович \(18\)](#) [В.И. Кузнецов \(13\)](#) [С.А. Магницкий \(11\)](#) [В.А. Макаров \(24\)](#) [В.П. Маслов \(179\)](#)
[В.Б. Морозов \(7\)](#) [В.Я. Панченко \(18\)](#) [Ю.М. Романовский \(5\)](#) [Ю.С. Чилингарян \(10\)](#) [С.А. Лосев \(12\)](#)

[Подробнее !\[\]\(d3102649f02e825ddb76dc3de0190154_img.jpg\)](#) | [Публикации коллектива !\[\]\(55ca3a38dbb940110628e54e3ea7505d_img.jpg\)](#)

Пример: российские коллективы по «оптогенетике»

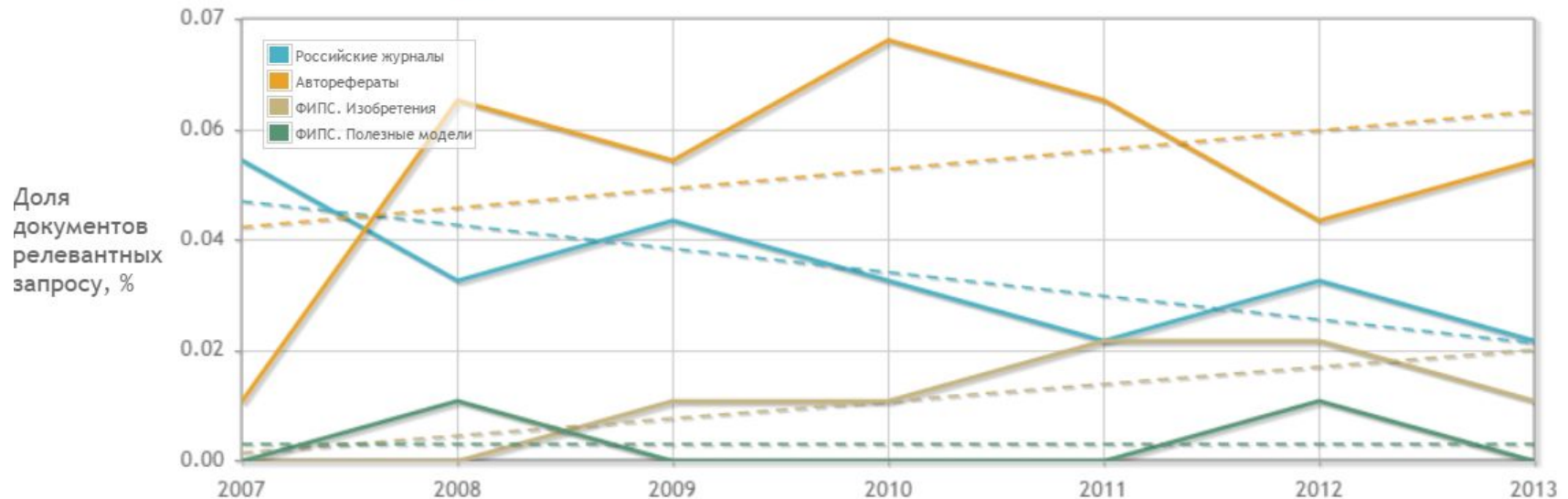
Российские научные коллективы, ведущие исследования по оптогенетике

№	Авторы	Ключевые слова
1	Н.И. Коротеев, С.М. Гладков, А.М. Желтиков, А. Н. Наумов, А.Б. Федотов	Волновод, изотропная среда, восприимчивость, состояние поляризации, тензор восприимчивости, синхронизм, время задержки
2	А.Г. Шмелев, Д.Д. Власова, В.А. Зуйков, К.В. Иванов, А.А. Калачев, К.Р. Кариммулин, А.В. Леонтьев, В.С. Лобков, В.Е. Мнускин, В.Г. Никифоров, С.В. Петрушкин, В.В. Самарцев, Г.М. Сафиуллин, А.Н. Токарева, Б.Ф. Тринчук, А.В. Шкаликов	Photon, спектроскопия, колебательная мода, внутримолекулярный, дифенил, накачка, четыреххлористый углерод, прогностическая валидность, регистрация эффекта, активный элемент
3	Т.Б. Фельдман, М.А. Островский, Ю.В. Кузнецов, Г.Р. Каламкаргов, О.В. Некрасова, М.П. Кирпичников, Х.П. Тахчиди, К.О. Муранов	Глазной, дедифференцированная клетка, дедифференцированный, клетка ретинального, нерв, офтальмология, пигментный, пигментный эпителий, ретинальный, ретинальный эпителия, сетчатка сосудистый стекловидное тело, стекловидный, энуклеация глаз, эпителий эпителий ретинального

Исходя из полученных данных можно сделать вывод, что в России существует минимум 3 коллектива, которые занимаются оптогенетикой.

«Квантовый компьютер»

Динамика по относительным показателям



Наибольший интерес российского научного сообщества к квантовым компьютерам приходится на 2007-2010 гг., после чего количество авторефератов и публикаций снижается. При этом за всплеском защит в 2008-2010 годах не последовал стремительный рост патентной активности. Вероятно, сейчас данное направление находится в стадии переосмысления, после чего может начаться как рост, так и дальнейшее угасание направления.

Прогноз: Вероятно, сейчас данное направление находится в стадии переосмысления подходов и поиска новых методов, далее может пойти как в новый всплеск, так и дальнейшее угасание направления.

Заключение

- Наукометрические методы не могут решить все проблемы в российских реалиях.
- Полнотекстовый анализ может дать действительно значимую информацию, которую невозможно получить с помощью наукометрических методов.
- Необходимо сочетать наукометрию и полнотекстовую аналитику.
- Необходимо работать над интеграцией различных информационных баз.