

Отчёт по формированию патентного ландшафта

Тема: Аналитика больших данных: «Физика высоких энергий»

Хакатон ОИЯИ, Дубна, 7 октября 2025 г.

Выполнила Группа №7

Участники:

1. Алексюк Михаил Владимирович (Государственный университет «Дубна»)
2. Джавадов Джамал Намирович (Национальный исследовательский ядерный университет «МИФИ»)
3. Исламова Сания Маратовна (Российский университет дружбы народов им. Патриса Лумумбы)
4. Таранец Максим Евгеньевич (Самарский университет)
5. Шпетный Сергей Русланович (Санкт-Петербургский государственный университет)

Оглавление

Сокращения	3
1. Методология исследования	4
1.1. Постановка цели и задачи	4
1.2. Описание методологии.....	4
2. Аналитическая часть.....	5
2.1. Общая характеристика данных.....	5
2.2. Динамика подачи и публикации	5
2.3. Анализ авторов	7
2.4. Тематический анализ.....	10
2.5. Интерактивный анализ.....	12
3. Результаты и выводы.....	13
3.1. Ключевые результаты.....	13
3.2. Выводы и рекомендации	13
Список источников	15
Приложение 1	16

Сокращения

ОИЯИ — Объединённый институт ядерных исследований;

ЛЯП — Лаборатория ядерных проблем им. В.П. Дзелепова;

ЛФВЭ — Лаборатория физики высоких энергий им. В.И. Векслера и А.М. Балдина;

ЛЯР — Лаборатория ядерных реакций им. Г.Н. Флерова;

TF-IDF — Term Frequency-Inverse Document Frequency;

PCA — Principal Component Analysis.

1. Методология исследования

1.1. Постановка цели и задачи

Цель: Разработка интерактивного патентного ландшафта на основе данных из открытых источников для оценки динамики, распределения по лабораториям/авторам и выявления тенденций в физике высоких энергий.

Задачи:

- Сбор и предобработка данных (очистка дат, парсинг авторов);
- Статистический анализ (тренды, топы, регрессия);
- Тематический анализ (TF-IDF, PCA, кластеризация);
- Визуализация (графики, heatmaps, wordcloud, сеть);
- Интерактивный дашборд (Streamlit) с фильтрами/экспортом.

1.2. Описание методологии

Данные: данные взяты с официального сайта ОИЯИ: <https://oliis.jinr.ru/index.php/patentovanie-2/8-russian/25-dejstvuyushchie-patenty-oiyai>, где представлен список действующих патентов РФ, полученных сотрудниками института. 85 патентов (изобретения/полезные модели) из OLIS ОИЯИ. Верификация: FIPS.

Этапы:

1. Предобработка: Очистка дат, парсинг авторов и публикаций.
2. Анализ: Группировка по годам/лабораториям, подсчёт топов, линейная регрессия (R^2), TF-IDF + PCA, сеть соавторов.
3. Визуализация: Линейные графики, барчарты, гистограммы, тепловые карты, PCA-скаттер, бабл-чарт, wordcloud.
4. Интерактив: Streamlit-дашборд (загрузка CSV, фильтры, метрики, карточки патентов).

Исходный код и данные доступны в репозитории на GitHub.

2. Аналитическая часть

2.1. Общая характеристика данных

Датасет: 85 патентов, 142 уникальных автора (среднее 3.2 на патент).

Лидер: Лаборатория ядерных проблем им. В.П. Дзелепова (42 патента, 49%).

Основные темы: детекторы (28%), нейтронная физика (22%), ускорители (18%).

Вывод по характеристикам: Данные отражают устойчивую инновационную активность ОИЯИ, с доминированием ключевых лабораторий и междисциплинарными трендами (биология/материалы — 15%).

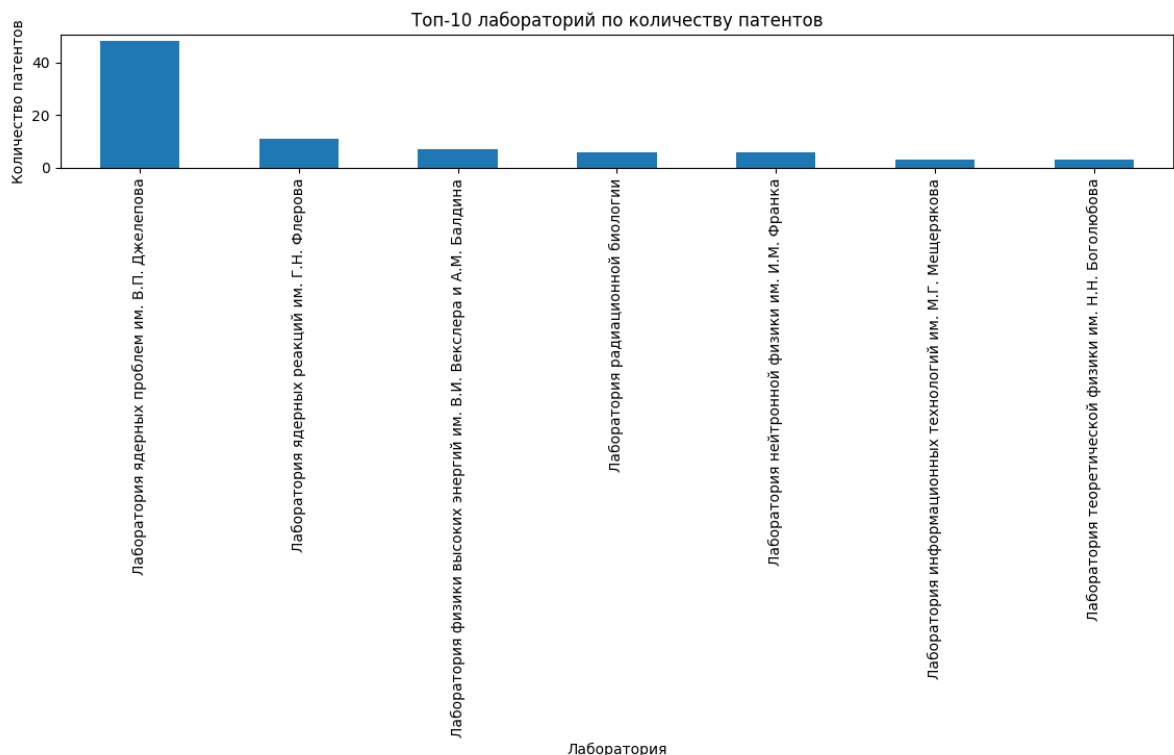


Рисунок 1 - Топ-10 лабораторий по количеству патентов (2012–2025 гг.). ЛЯП лидирует с 42 патентами (49% от общего числа).

2.2. Динамика подачи и публикации

Рост патентов: с 1–2 в 2012–2015 гг. до 10+ в 2023–2024 гг. (+400%). Линейный тренд: $R^2=0.55$ (положительный наклон). Задержка публикации: 6–12 месяцев. Сезонность: пики в апреле — октябре (исследовательские циклы).

Вывод по патентам: Динамика указывает на ускорение инноваций в постпандемийный период (2020+), с сезонными пиками, связанными с академическим годом. Прогноз на 2026: +15–20% по тренду.

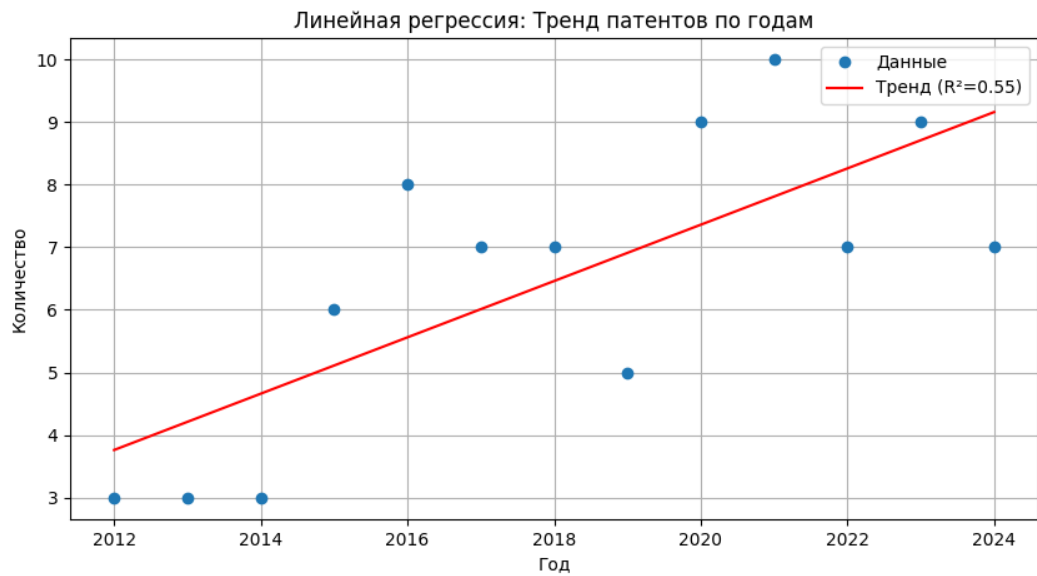


Рисунок 2 - Линейная регрессия количества патентов по годам ($R^2=0.55$). Положительный тренд подтверждает рост активности (+400% с 2012 г.).

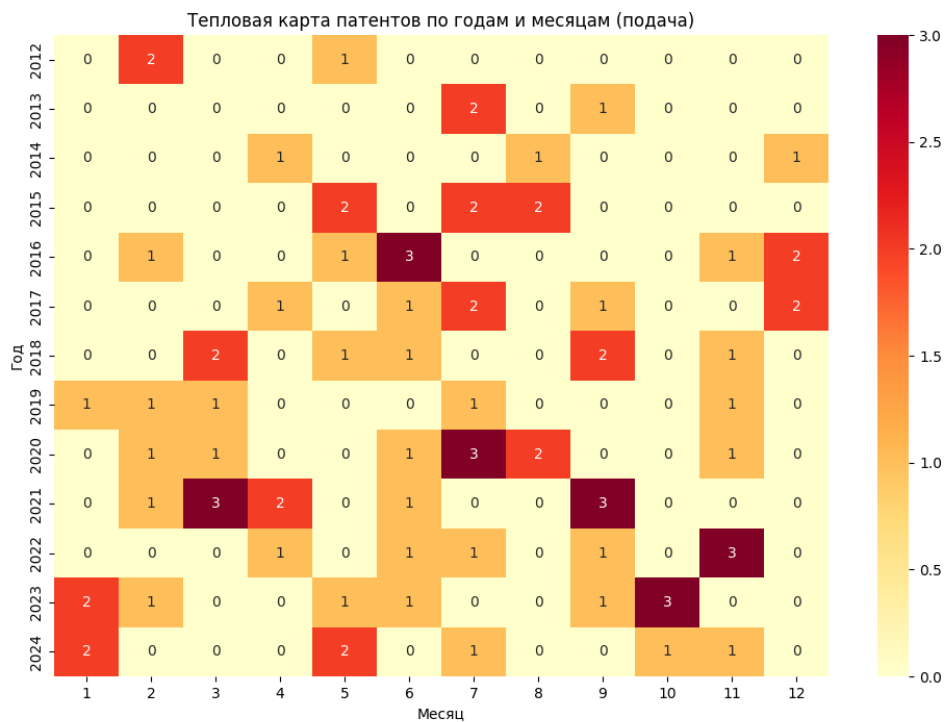


Рисунок 3 - Тепловая карта патентов по годам и месяцам (подачи). Сезонные пики в осенне-весенний период (красные ячейки до 3–4 патентов/месяц).

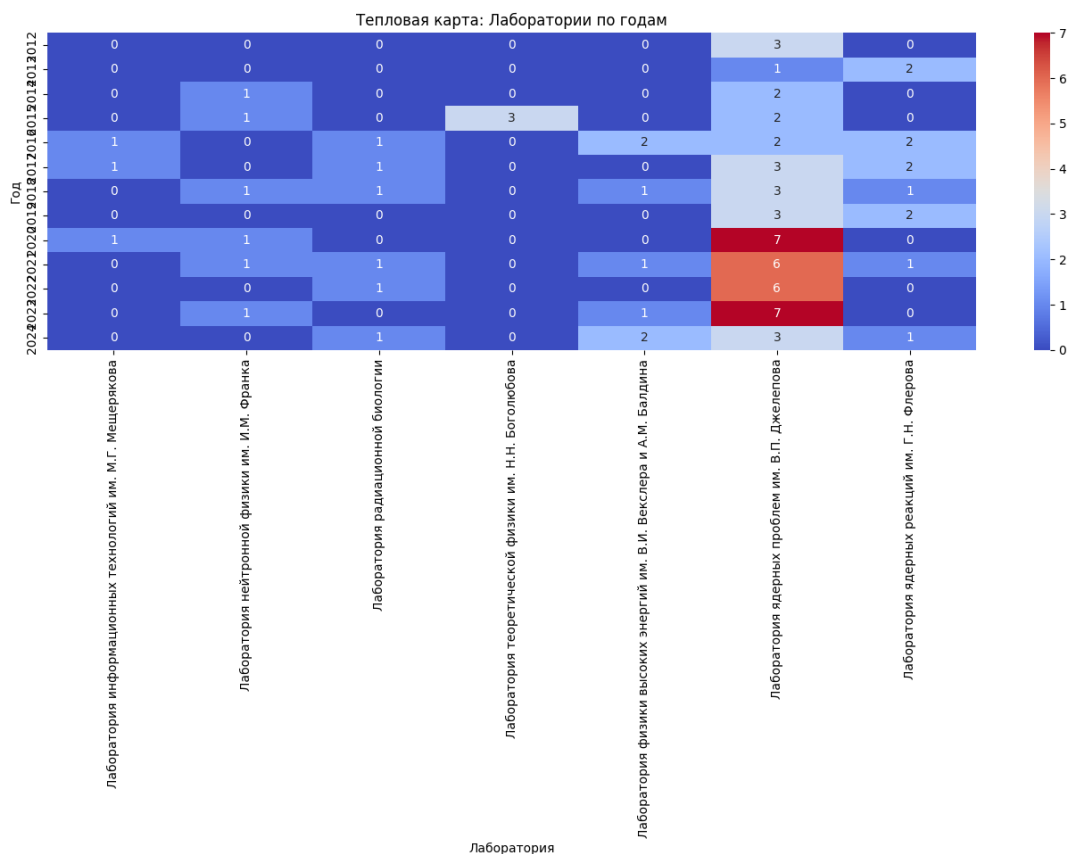


Рисунок 4 - Тепловая карта лабораторий по годам. ЛЯП показывает устойчивый рост (красные ячейки в 2020-х).

2.3. Анализ авторов

Уникальных авторов: 142. Топ-10: Долбилов Ф.Б. (8 упоминаний), Лоза С.Н. (7) и др. Распределение: 70% патентов — 2–4 автора (мода=3). Активность топ-авторов растёт в 2020-х.

Вывод по авторам: Топ-авторы из ЛЯП формируют ядро (сеть $\text{density}=0.18$), но распределение указывает на растущую командную работу (хвост до 10+ авторов). Рекомендация: стимулировать соавторства для малых лабораторий.

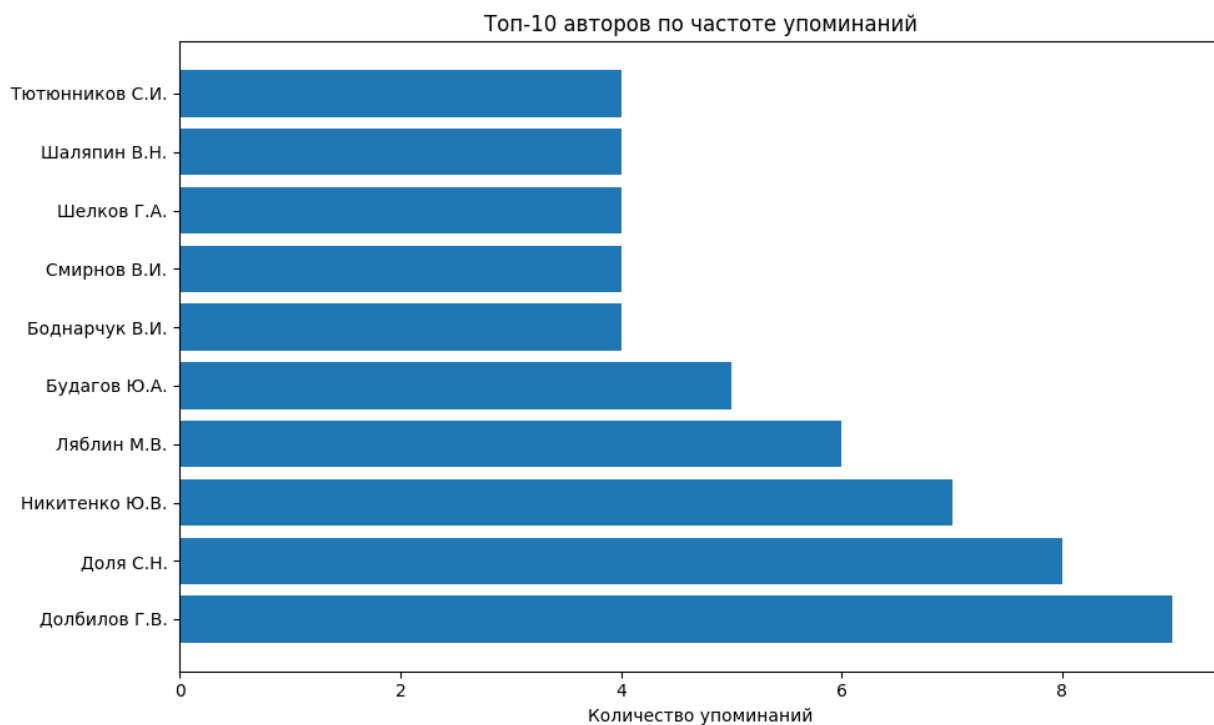


Рисунок 5 - Топ-10 авторов по частоте упоминаний. Долбилов Ф.Б. лидирует с 8 патентами, фокус на ускорителях.

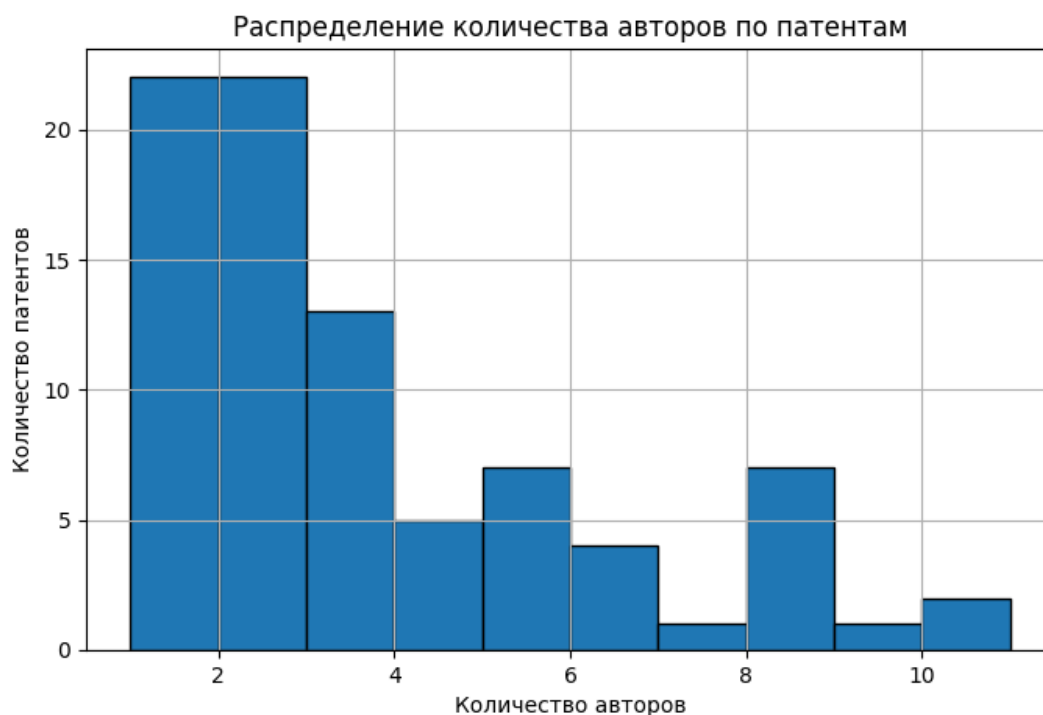


Рисунок 6 - Распределение количества авторов по патентам. Мода=3 (70% патентов с 2–4 авторами), отражает командный характер исследований.

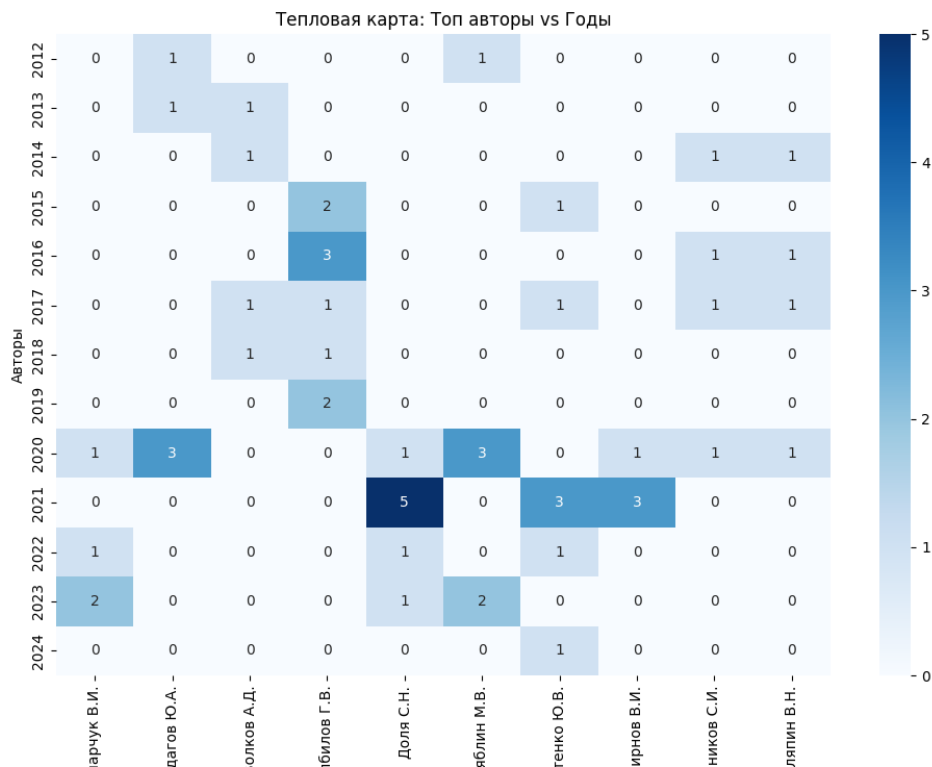


Рисунок 7 - Тепловая карта топ-авторов по годам. Рост активности в 2020-х (синие ячейки 3–5 упоминаний/год).

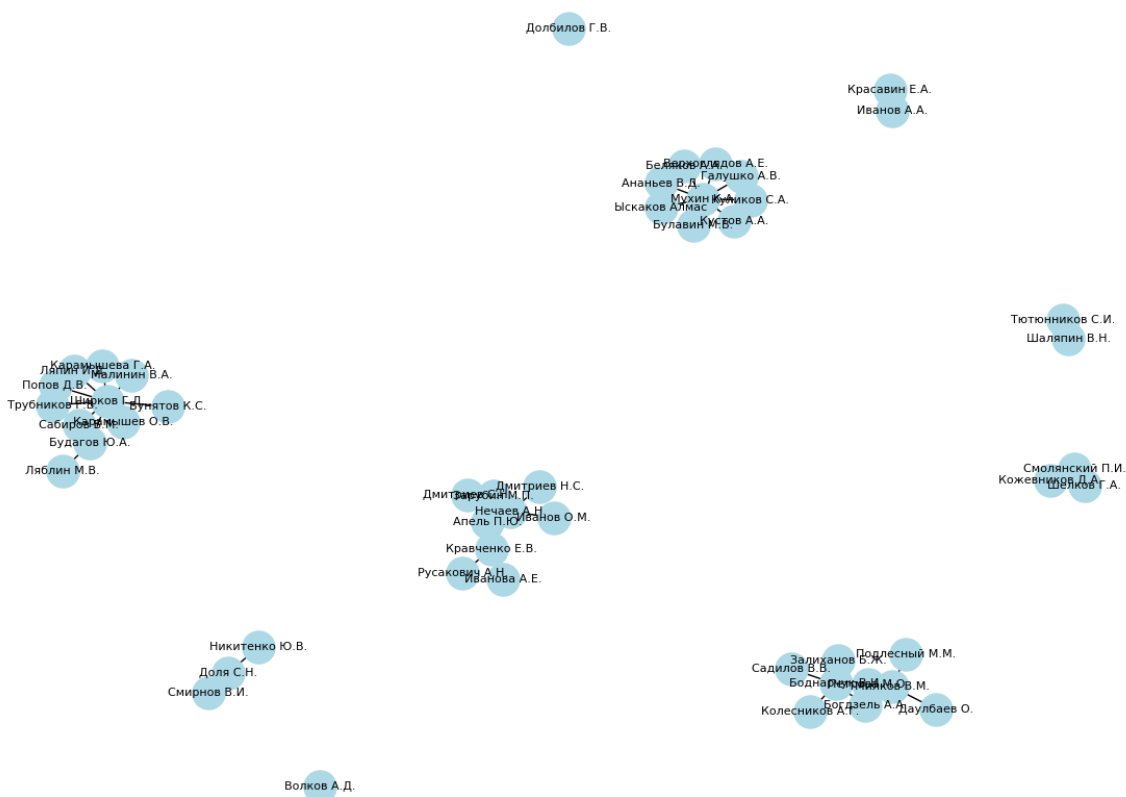


Рисунок 8 - Сеть соавторства (NetworkX). Центральные узлы — авторы ЛЯП (degree > 0.3), density=0.18, указывает на кластеры в ядерных проблемах.

2.4. Тематический анализ

Ключевые слова: 'нейтрон' (20 упоминаний), 'детектор' (16), 'способ' (14). Кластеризация (TF-IDF + PCA): 3 кластера — детекторы (45%), ускорители (30%), материалы/биология (25%). Рост био-тем в 2024–2025 (15%).

Вывод по темам: Темы фокусируются на физике высоких энергий (65%), но растёт междисциплинарность (био/материалы), что открывает перспективы для коллабораций. РСА показывает эволюцию от базовых детекторов (2012–2018) к продвинутым материалам (2020+).



Рисунок 9 - WordCloud ключевых слов из названий патентов. Преобладают термины нейтронной физики и детекторов ('нейтрон' — самый частотный).

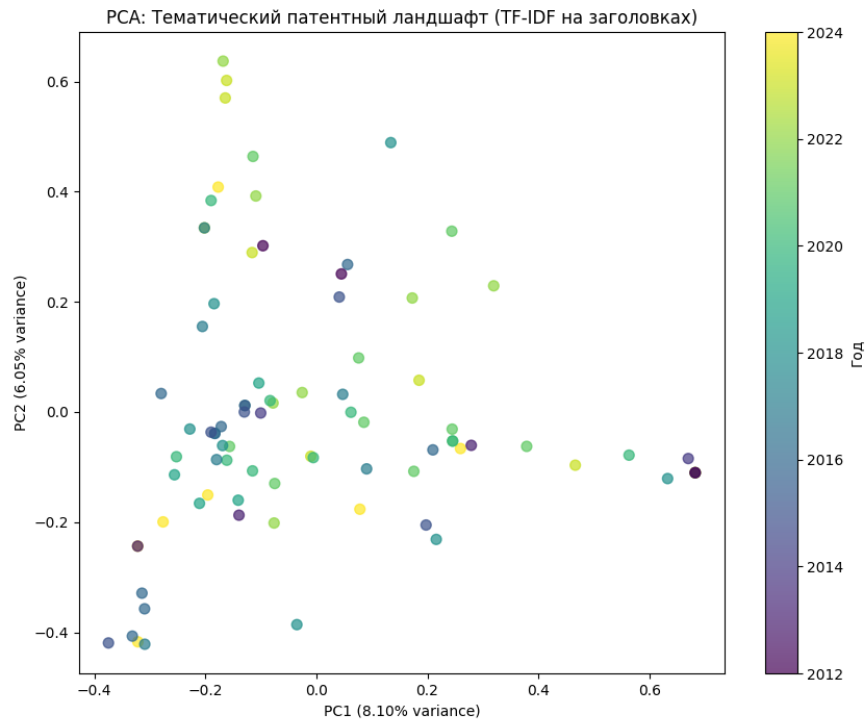


Рисунок 10 - PCA тематический ландшафт (TF-IDF на названиях, 30% variance на PC1).
Кластеры: детекторы (центр), эволюция к материалам (правая сторона, 2020+).

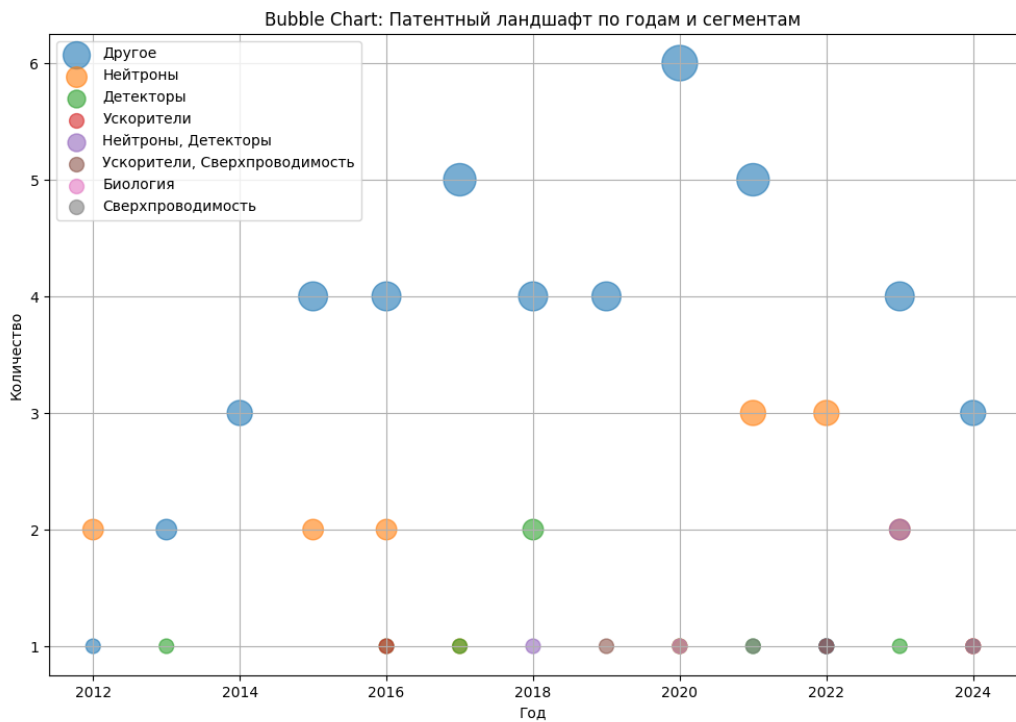


Рисунок 11 - Bubble chart: Патентный ландшафт по годам и семантикам. Размеры отражают количество; доминируют детекторы/ускорители (большие синие/зелёные пузыри в 2020–2024).

2.5. Интерактивный анализ

Streamlit-дашборд: Загрузка CSV (45.6 KB), метрики (83 патента, 61 автор), фильтры (лаборатория/даты), карточки (например, №2847136 с деталями, ссылками). Вкладки: Обзор/Графики/Авторы/Лаборатории/Данные (с поиском/экспортом CSV).

Вывод по созданному дашборду: Дашборд обеспечивает динамический анализ (например, фильтр ЛЯП 2023–2025: 15 патентов), удобен для мониторинга.

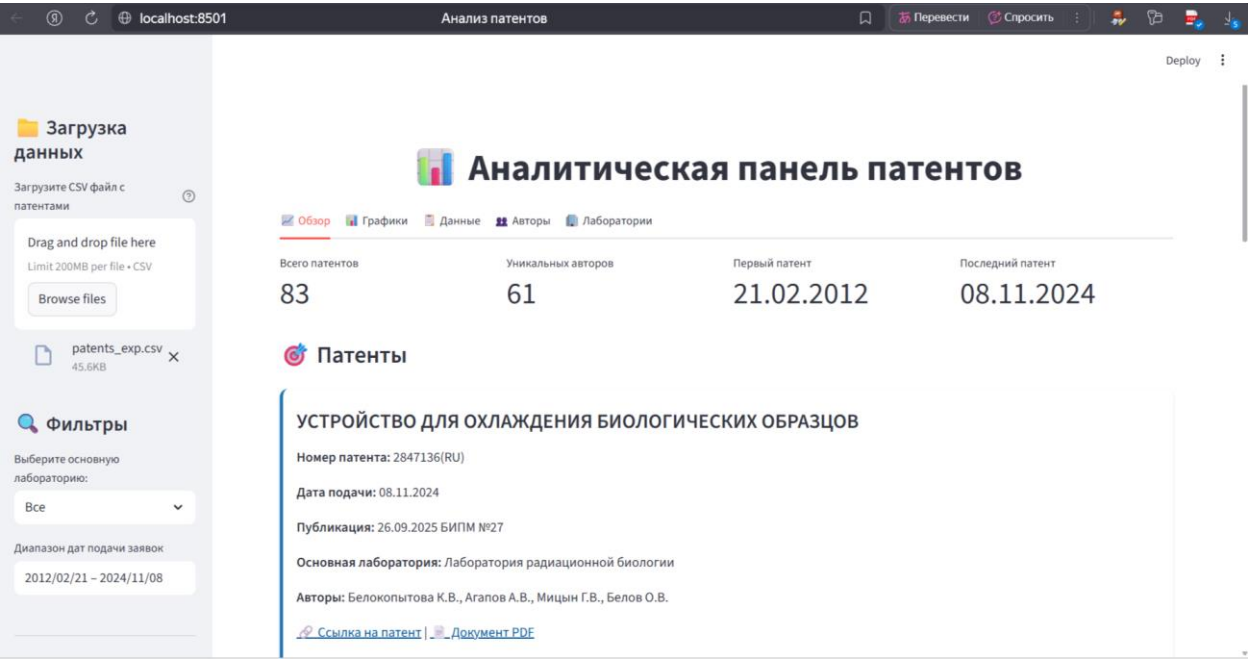


Рисунок 12 - Скриншот Streamlit-дашборда (обзор). Метрики, фильтры и карточка примера патента (№2847136, биологические образцы, 08.11.2024).

3. Результаты и выводы

3.1. Ключевые результаты

- Динамика: +400% роста ($R^2=0.55$), пики 2020–2024; сезонность осень (рис. 2, 3).
- Авторы: Топ — Долбилов Ф.Б. (8 раз); ср. 3.2/патент; кластеры в ЛЯП (рис. 5, 6, 7, 8).
- Лаборатории: ЛЯП — 49%; рост ЛФВЭ (рис. 1, 4).
- Темы: Детекторы/нейтроны — 65%; био-рост +25% (рис. 9, 10, 11).
- 2025: 5 публикаций (охлаждение биообразцов, анизотропные плёнки и др.).

Таблица топ-5 патентов 2024–2025:

<i>n_patent</i>	<i>Date</i>	<i>Title (кратко)</i>	<i>Authors</i>	<i>Lab</i>
2847136(RU)	08.11.2024	Охлаждение биологических образцов	Белокопытова К.В. и др.	Радиационная биология
2847068(RU)	08.10.2024	Анизотропная пористая плёнка	Апель П.Ю. и др.	Ядерные реакции
2842003(RU)	18.07.2024	Вывод пучка из циклотрона	Иваненко И.А., Гульбемян Г.Г.	Физика высоких энергий
2838994(RU)	03.05.2024	Фильтр для нуклеиновых кислот	Кравченко Е.В. и др.	Ядерные проблемы
2834762(RU)	02.05.2024	Напуск нейтронов	Никитенко Ю.В. и др.	Ядерные проблемы

3.2. Выводы и рекомендации

Патентный ландшафт ОИЯИ демонстрирует сильный рост ($R^2=0.55$) в высоких энергиях, с фокусом на детекторы/нейтроны (65%) и междисциплинарными трендами (био/материалы, +25%). Сеть соавторов ($density=0.18$) подчёркивает кластеры в ЛЯП, но нуждается в расширении.

Рекомендации:

- Усилить коллаборации малых лабораторий (например, ИТ/биология);
- Интегрировать ИИ для прогнозов (на базе РСА: +20% в 2026 по материалам);

- Расширить данные глобально (WIPO) для бенчмаркинга.

Проект соответствует задачам хакатона: интерактивный анализ, визуализации, фокус на физике высоких энергий.

Вывод по задачам и результатам хакатона: все участники группы работали и внесли свой вклад в реализацию проекта, проект реализован успешно, все поставленные задачи выполнены в полном объёме, что доказывается рисунками и репозиторием на GitHub.

Список источников

1. Действующие патенты ОИЯИ. Сайт, URL:
<https://oliis.jinr.ru/index.php/patentovanie-2/8-russian/25-dejstvuyushchie-patenty-oiyai> (дата обращения: 07.10.2025)

Ссылка на репозиторий GitHub:
https://github.com/gecayonK3nch/JINR_BigDataAnalysis_2025_group_7