

# Исследовательский поиск научных публикаций. Проект SciNoon

---

Недумов Ярослав Ростиславович

20 марта 2019 г.

ИСП РАН

# План презентации

Что такое SciNoon?

Чем мы можем заниматься?

Что такое ИСП РАН?

Как к нам попасть?

Что такое SciNoon?

---

1. Система исследовательского поиска для R&D
2. База научных публикаций
3. Фреймворк для решения задач, связанных с анализом научных публикаций

# Задача исследовательского поиска<sup>1</sup>

В отличие от традиционного поиска:

## Контекст проблемы

- С открытым концом
- Постоянный
- Многогранный

## Процесс поиска

- Оппортунистический
- Итеративный
- Мультитактический

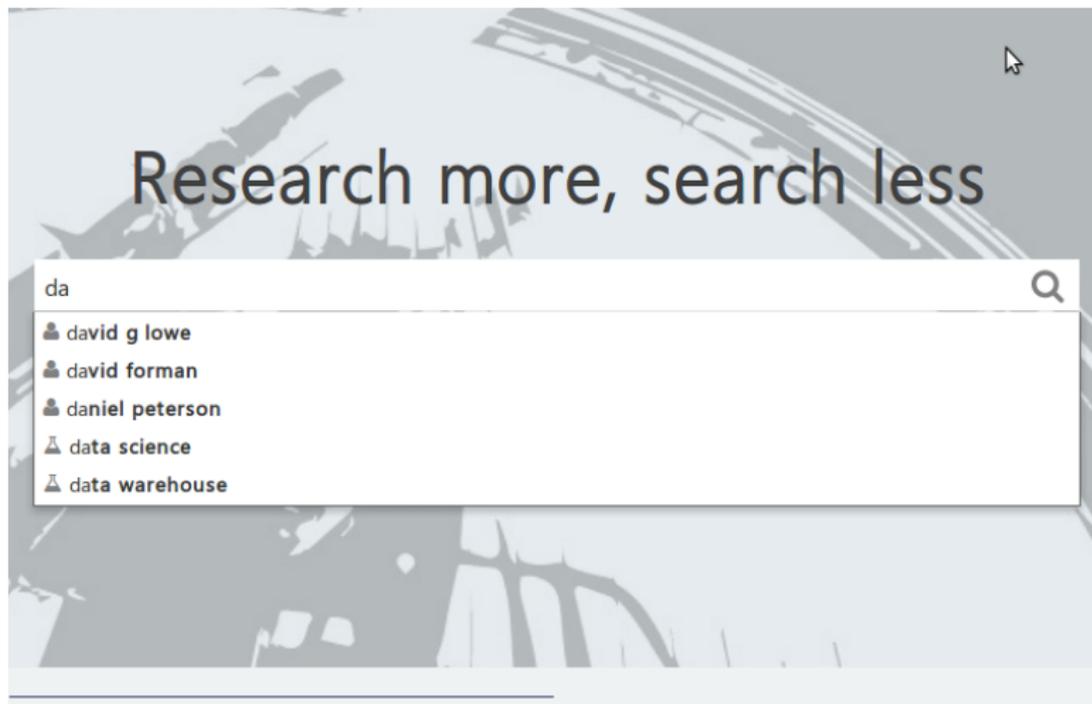
---

<sup>1</sup>White, R. W., Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. Synthesis lectures on information concepts, retrieval, and services, 1(1), 1-98.

# Виды инструментов исследовательского поиска

1. Инструменты помощи при формулировании запроса
2. Инструменты помощи при анализе результатов
3. Инструменты поддержки длительного поиска

# Инструменты помощи при формулировании запроса<sup>2</sup>

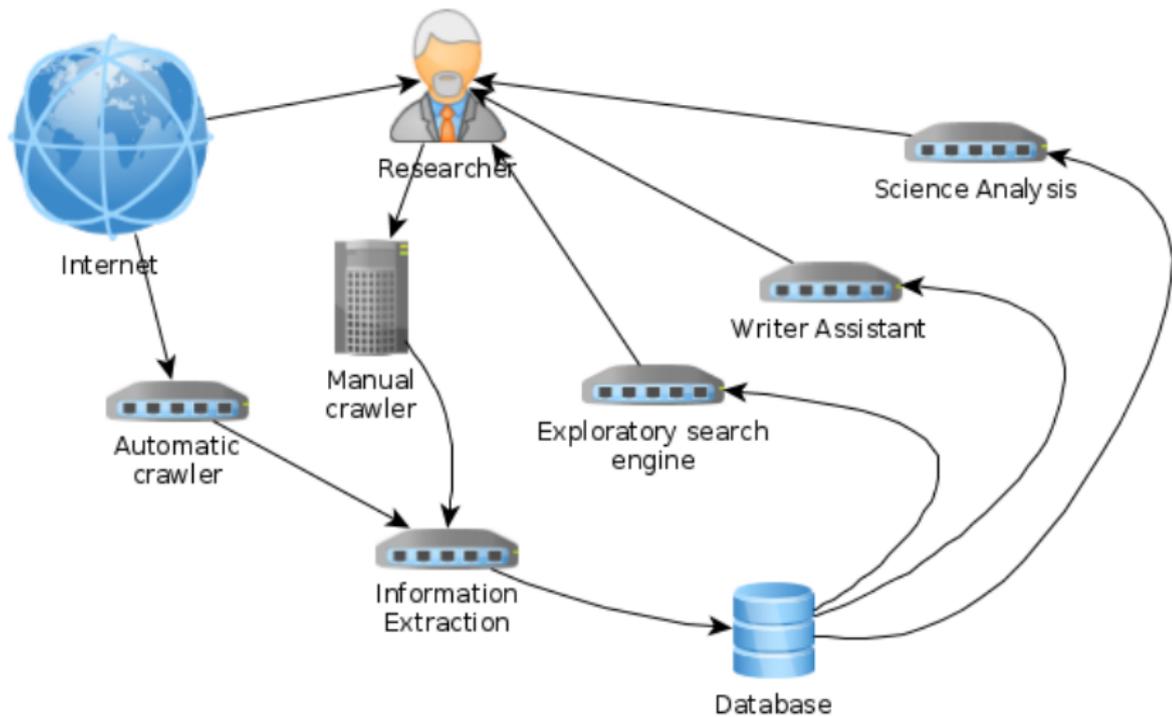


<sup>2</sup><https://academic.microsoft.com>





# SciNoon



Чем мы можем заниматься?

---

1. Системы поддержки проведения исследований
2. Графовые базы данных и интеграция данных
3. Обработка полуструктурированных данных

# Исследовательский поиск

## Помощь при запросе

1. Свой keyword search
2. Свой query suggestion
3. Учет контекста
4. Рекомендации

## Помощь при обзоре результатов

1. Карты науки для результатов поиска
2. Фасеты

## Поддержка длительного поиска

1. "Синхронная" карта для группового поиска
2. Gamification
3. Декомпозиция исследований

# Улучшения графовой базы данных

- Повышение удобства работы
  - Безопасность
  - Приватность
  - Object-Graph-Mapping
- Повышение скорости работы
  - Индексация
  - Большие данные

- Дедупликация статей
- Дедупликация авторов
- Интеграция исправлений от пользователей системы

# Извлечение информации

## Извлечение из PDF

- Повышение качества извлечения метаданных
- Извлечение отдельных секций и информации из них
- Извлечение фактов

## Извлечение из Web

- Адаптация к изменяющимся во времени форматам страниц
- Универсальный краулер

Что такое ИСП РАН?

---

1. Исследования
2. Инновации
  - Samsung, IBM, Huawei, NVidia, ...
3. Образование
  - МГУ, Физтех, ВШЭ

# Условия для студентов

1. Вовлечение в проекты
2. Рабочее место, доступ к оборудованию
3. Стипендия на третьем курсе, потом возможность устроиться в лабораторию

Как к нам попасть?

---

## Три этапа:

- Зачисление на кафедру
- Зачисление на семинар
- Зачисление к научному руководителю

## Критерии отбора

- Средний балл
- Баллы за программистские предметы
- Интерес к теме

# Что надо сделать?

Проявить инициативу и продемонстрировать самостоятельность и интерес:

- Выбрать или придумать тему
- Написать письмо на полстраницы-страницу с ответом на вопрос, чем вы хотите заниматься и почему?
- Приехать в ИСП РАН и поговорить

Время задавать вопросы