

# Математическое и программное обеспечение сложных информационно-вычислительных систем

*Коллектив под руководством д.ф.-м.н., проф. Валерия Александровича Васенина*

Наш коллектив внедряет методы интеллектуального анализа больших данных в реальные, востребованные информационные системы.

Основные направления исследований — математические методы и средства:

- сбора больших объёмов данных и их верификации в разнородных источниках;
- анализа больших данных с использованием как классических алгоритмов, так и машинного обучения (социальные графы, текстовые коллекции, фото и видео);
- разработки и сопровождения крупных информационных систем (производительность и оптимизация высоконагруженных систем, верификация и анализ программ, сопровождение унаследованного кода и реинжиниринг, адаптация кода к новым системам управления базами данных, разграничение доступа в крупных системах).

В нашем коллективе можно получить практические навыки:

- полный цикл разработки информационных систем с web-интерфейсом;
- реализация алгоритмов анализа больших данных для «production»-систем;
- основные технологии: PostgreSQL, Python / Django, JavaScript / React JS.

Наши студенты публикуются в научных журналах, поступают в аспирантуру и защищают диссертации.

Наши выпускники работают в Яндексе, Банке России, Тинькофф, Huawei, ...

Докладчик: [Максим Александрович Кривчиков](#)

# Математическое и программное обеспечение сложных информационно-вычислительных систем

Результаты исследований коллектива апробируются и внедряются:

- ИАС «ИСТИНА» <https://istina.msu.ru> (полностью разработана нашим коллективом)  
— наукометрическая система для гибкой оценки эффективности научных организаций.



## ИСТИНА

Интеллектуальная Система Тематического Исследования НАукометрических данных

[Главная](#) [Поиск](#) [Статистика](#) [О проекте](#) [Помощь](#)

[Войти в систему](#)  
[Регистрация](#)

- Подсистема предэкспертного анализа информационно-аналитической системы РАН <https://nmr-ed.prn.ru>  
— система для проведения комплексной экспертизы всех гражданских научных исследований России.

**ИАСРАН** Информационно-аналитическая система  
Российской академии наук



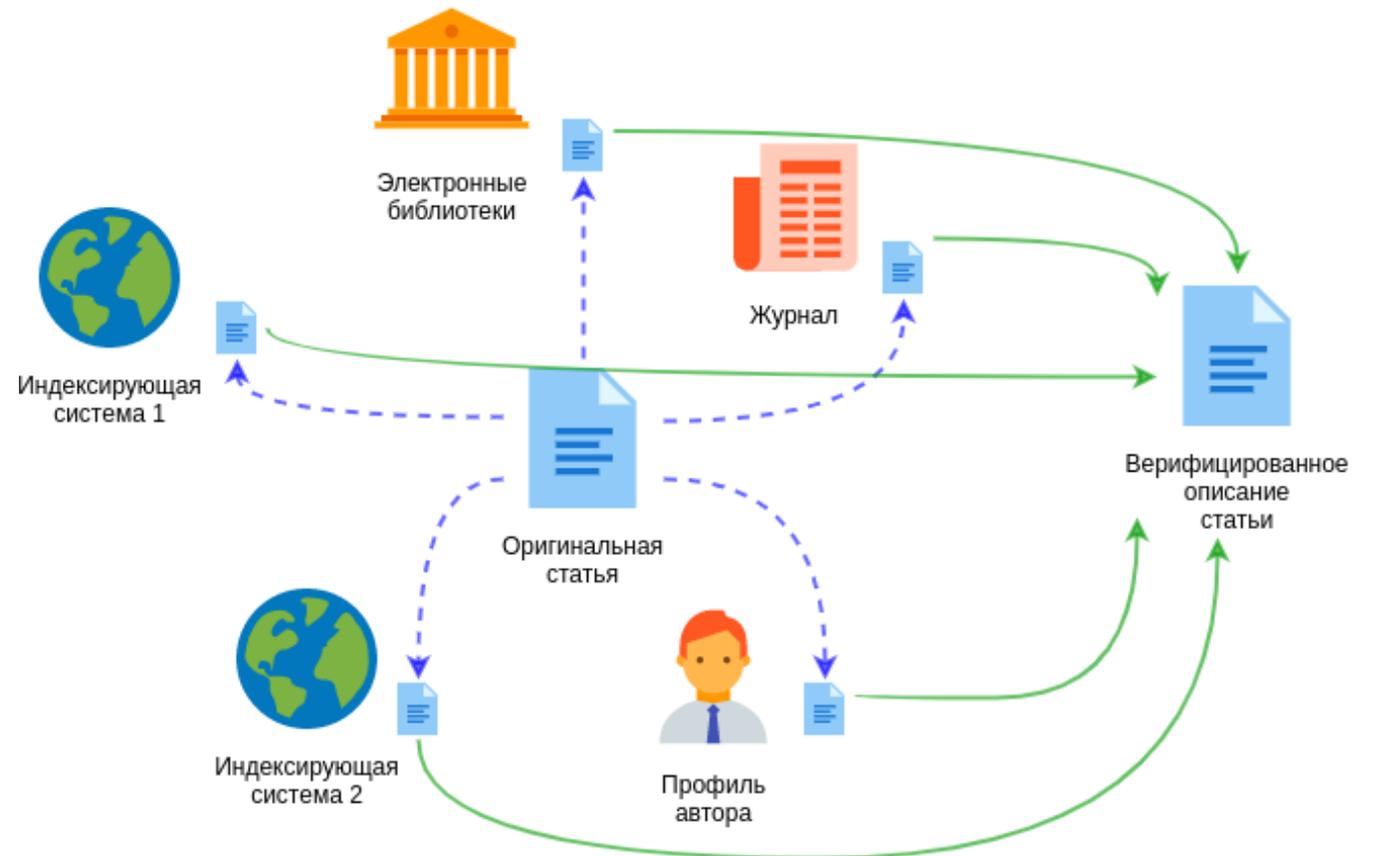
[Главная](#) [Корпус экспертов](#) [Конкурсы](#) [Контакты](#) [Документы](#)

Презентации с прошлогодних встреч — на сайте кафедры [2020](#), [2021](#), [2022](#) и [отдельная страница 2018 года](#).

# Ввод и верификация данных в больших сложно-организованных информационных системах

- Одно из направлений наших исследований - методы и средства автоматизации ввода и многоуровневой верификации данных в больших сложно-организованных информационных системах.

- **Пример:** данные о результатах научных исследований учёных
- Используются для поддержки управления организаций, стимулирования деятельности
- Большое количество видов результатов
- Большое количество источников со своими форматами данных, ошибками
- Данные появляются регулярно
- Большое количество связей между объектами



# Ввод и верификация данных в больших сложно-организованных информационных системах

## Верификация:

- Как сравнивать описания из разных источников?
- Как выявлять ошибки?
- Как объединять описания?
- Определение надёжности источников.
- Определение степени верифицированности объекта.
- Как автоматизировать процессы верификации?

## Ввод:

- Поиск информации в источниках
- Определение дубликатов
- Сопоставление описаний объектов
- Автоматизация процессов импорта

## Масштабирование:

- Как обеспечить высокую производительность механизмов при увеличении количества организаций и пользователей?
- Децентрализация данных

Патент #RU 2763458 С1: Способ ввода и многоуровневой верификации интенсивно поступающих данных в больших информационно-аналитических системах наукометрического содержания

# Семантические модели классификации данных в информационно-аналитических системах

- Во многих системах с объектами связаны **ключевые слова** и **рубрики** — элементы различных классификаторов (УДК, ГРНТИ и других).
- На основе этой информации можно классифицировать объекты и измерять **степень тематической близости** между двумя объектами.
- Множество ключевых слов и рубрик с весовыми коэффициентами будем называть *тематическим портретом*.
- Разработаны **метрики** для вычисления степени близости отдельных ключевых слов или рубрик (с учётом иерархии), а также степени близости двух портретов.
- Метрики показали способность справляться с тестовыми задачами, такими как выявление переводов. Также они могут быть использованы для **поиска экспертов**, рекомендации журналов и конференций и других прикладных задач.
- Разработан алгоритм для автоматического построения портрета по связанным объектам.

[Дмитрий Алексеевич Шачнев](#), [mitya57@mitya57.me](mailto:mitya57@mitya57.me), <https://t.me/mitya57>

## Представление сложной модели данных в виде набора предметно-ориентированных фильтров и показателей

- Базы данных больших систем имеют десятки типов объектов и сотни свойств и связей между объектами. Пример типов объектов в наукометрической системе: публикации, доклады, патенты, диссертации, учебные курсы, научно-исследовательские работы.
- Некоторые из свойств объектов могут быть использованы в отчётных, статистических материалах и рейтингах. Например: число цитирований публикации, объём финансирования НИОКТР, число академических часов в курсе.
- Идея: реализовать **конструктор правил**, который позволит автоматизировать построение таких отчётов и рейтингов.
- Пример правила: «*Статьи Web of Science, весовой коэффициент =  $10 \times \text{импакт-фактор} \div \text{число соавторов}$* ».
- Реализован **генератор SQL-запросов**, формирующий запросы на основе правил такого типа и онтологического описания структуры базы системы. Поддерживаются СУБД Oracle и PostgreSQL.
- Дальнейшие исследования: использование тематики исследования для фильтрации.

[Дмитрий Алексеевич Шачнев](#), [mitya57@mitya57.me](mailto:mitya57@mitya57.me), <https://t.me/mitya57>

# Системы типов в языках программирования

- Как сделать так, чтобы компьютер помогал писать программы без ошибок?

Основные направления исследований:

- **Разработка и реализация новых языков программирования или расширений существующих языков программирования**
- Средства автоматизированного анализа и трансформации программ
- Средства автоматизированного доказательства утверждений

## Децентрализованные открытые и частично-открытые web-системы

- Проблема крупных облачных информационных систем: централизованное владение данными. Все данные находятся под контролем оператора. Пользователь имеет только ограниченный доступ к своим данным. Такие системы неприменимы для пользователей, которым требуются расширенные средства защиты информации.
- (кроме того крупная облачная система сложна в реализации и сопровождении)
- Децентрализованный вариант исполнения: каждый пользователь имеет свой экземпляр сервиса-поставщика данных с нужной степенью защиты. Сервисы-потребители данных обеспечивают связность системы и поддерживают возможность глобального анализа.
- Целый ряд задач, как с большим объёмом программирования, так и более теоретических задач, связанных с доверием в децентрализованной системе.