
Интеллектуальный анализ открытых данных и здравоохранение

Арутюн Аветисян
академик РАН, директор ИСП РАН, arut@ispras.ru

8 октября 2020 г.

Основа ИСП РАН - постоянно развивающаяся научная школа

С.А. Лебедев

В.А. Мельников

В.П. Иванников

Л.Н. Королёв



25 лет
ИСП РАН

70 лет
ИТ в
России



Причины бурного роста ИТ в последние 10 лет (I)

Социум – основной заказчик инноваций.

Социальные медиа, поисковики, мобильные платформы.

Распространение СПО.

Linux, GCC, LLVM, QEMU, Valgrind, SFX, V8, Hadoop, Spark, Apache Ignite, Infinispan, Hazelcast.

Производство эффективной аппаратуры из компонентов общего назначения.

Высокопроизводительные системы, использование больших данных.



Причины бурного роста ИТ в последние 10 лет (II)

Игровые приставки → суперкомпьютеры

Графические акселераторы (**GPU**) используются в суперкомпьютерной области.

История процессора Cell (IBM, Sony, Toshiba):

- первое коммерческое применение — в игровой консоли Sony Play Station 3;
 - в 2009 году самым высокопроизводительным компьютером в мире был признан IBM Roadrunner (построен по гибридной схеме из процессоров AMD Opteron и IBM Cell 8i);
- разработка процессоров Cell была свёрнута.

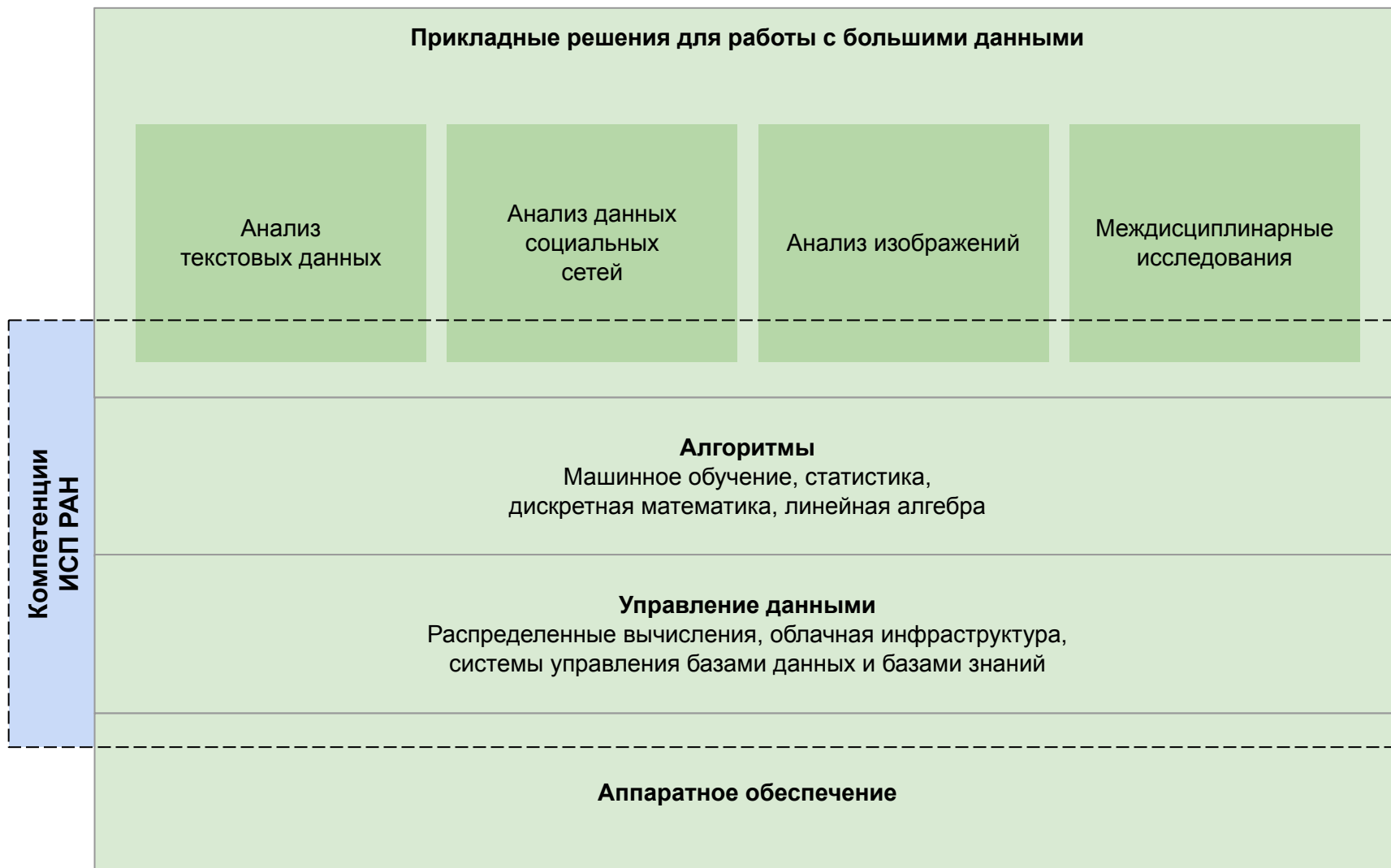
Поисковая машина → стек технологий для работы с большими данными (Hadoop)



Инфраструктура системы анализа данных

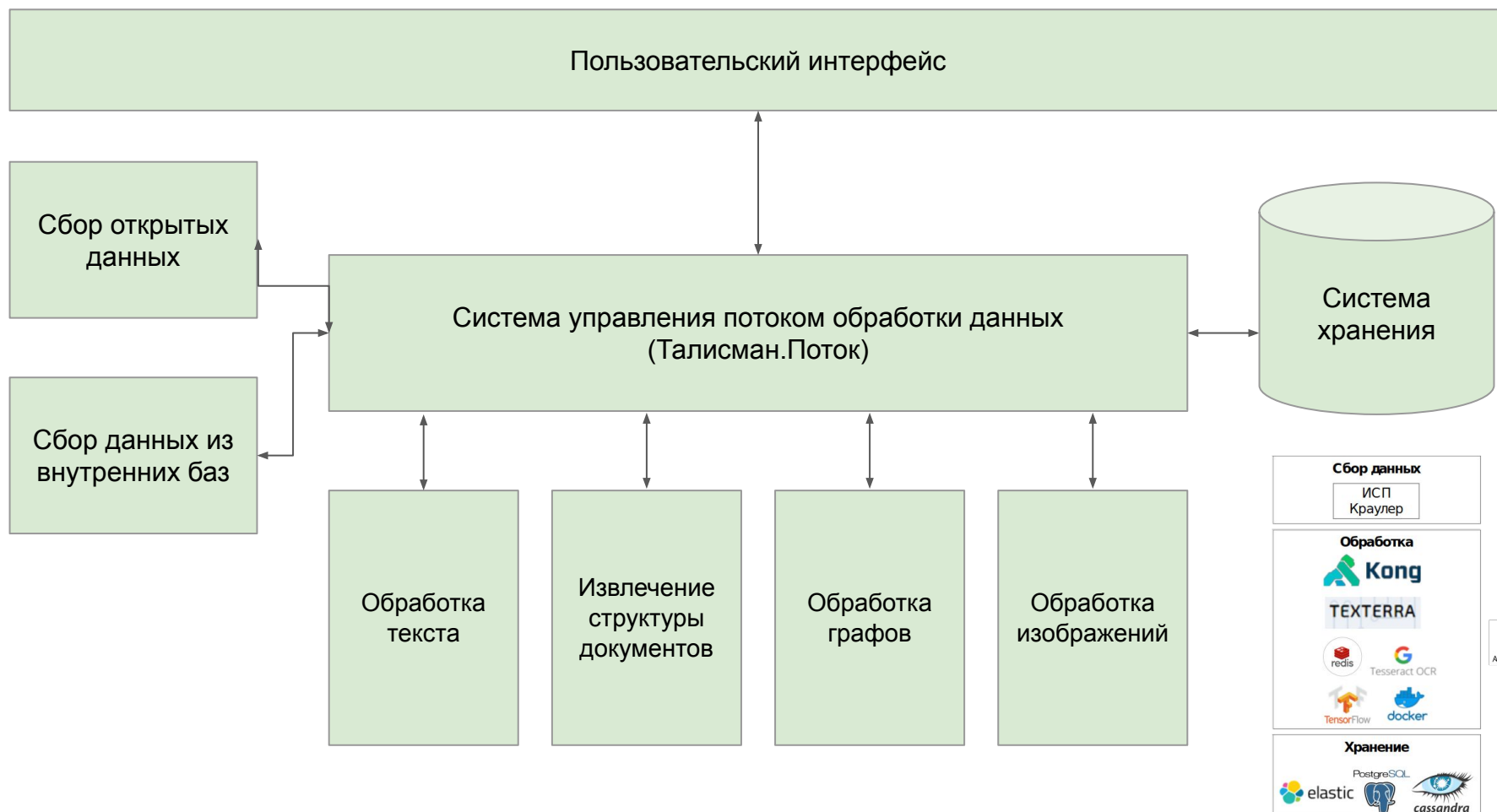


Стек технологий ИСП РАН



Талисман

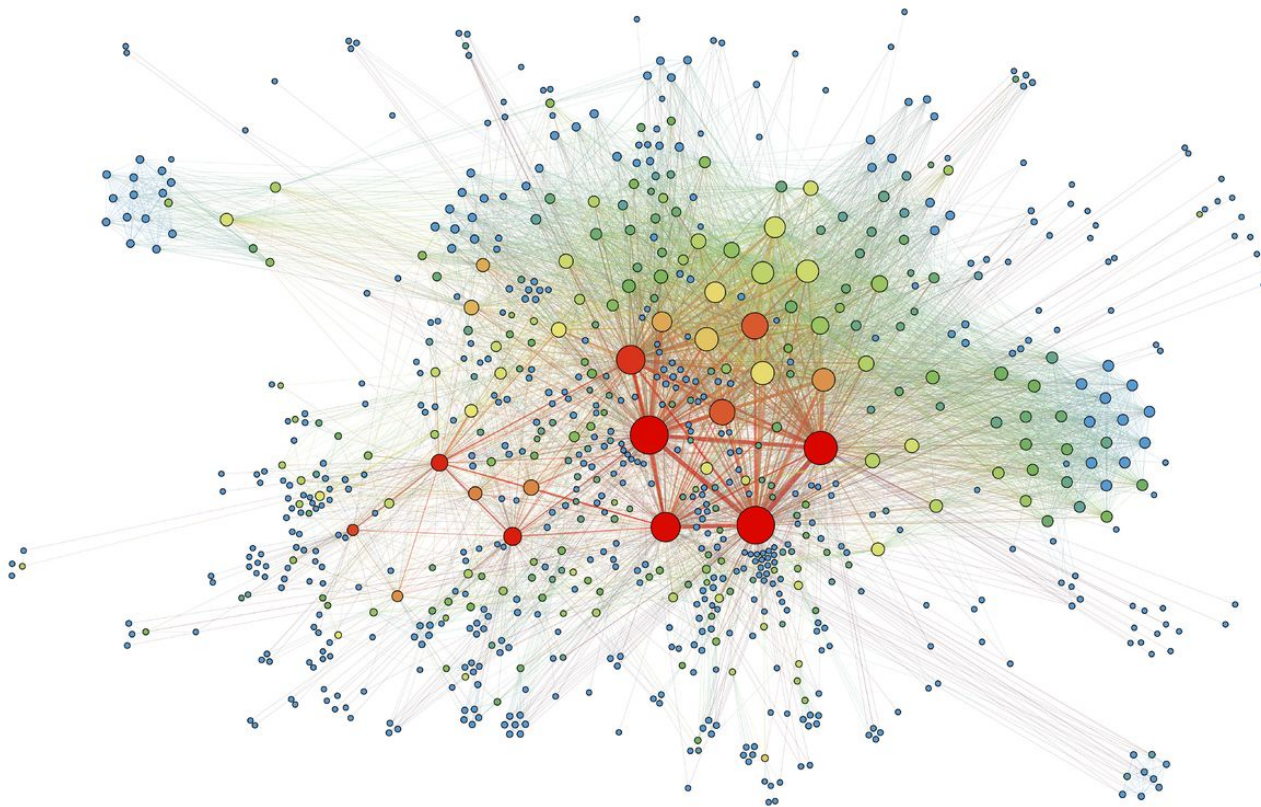
Среда анализа данных о людях, сообществах, продуктах, организациях и их взаимосвязях.
Внедрена в нескольких прикладных областях, в том числе в HR.



Медицинская статистика и большие данные

1. Медицинская статистика основывается на больших разнородных данных
2. Помимо задачи хранения и анализа медицинских данных, на любом субъектном уровне присутствует проблема определения их достоверности
3. Находясь внутри информационной системы невозможно сделать какие-либо заключения о достоверности данных
4. Один из подходов для оценки достоверности данных - обогащение внутренних баз данных (миграционных/налоговых/медицинских/...) открытыми данными

Использование открытых данных: раннее предсказание эпидемий



- Мониторинг людей с большим кругом общения, “party hosts”, позволяет, в том числе, пробовать предсказывать эпидемии на ранней стадии по графу социальных связей
- Граф социальных связей может быть получен с помощью сбора и анализа данных социальных сетей

НЦМУ в области медицины

Биомедицинские веб-лаборатории на базе масштабируемой облачной инфраструктуры

Практическая медицина

доказательная медицина

персонализированная медицина

Медицинское образование

доступ к современным знаниям


дистанционное обучение

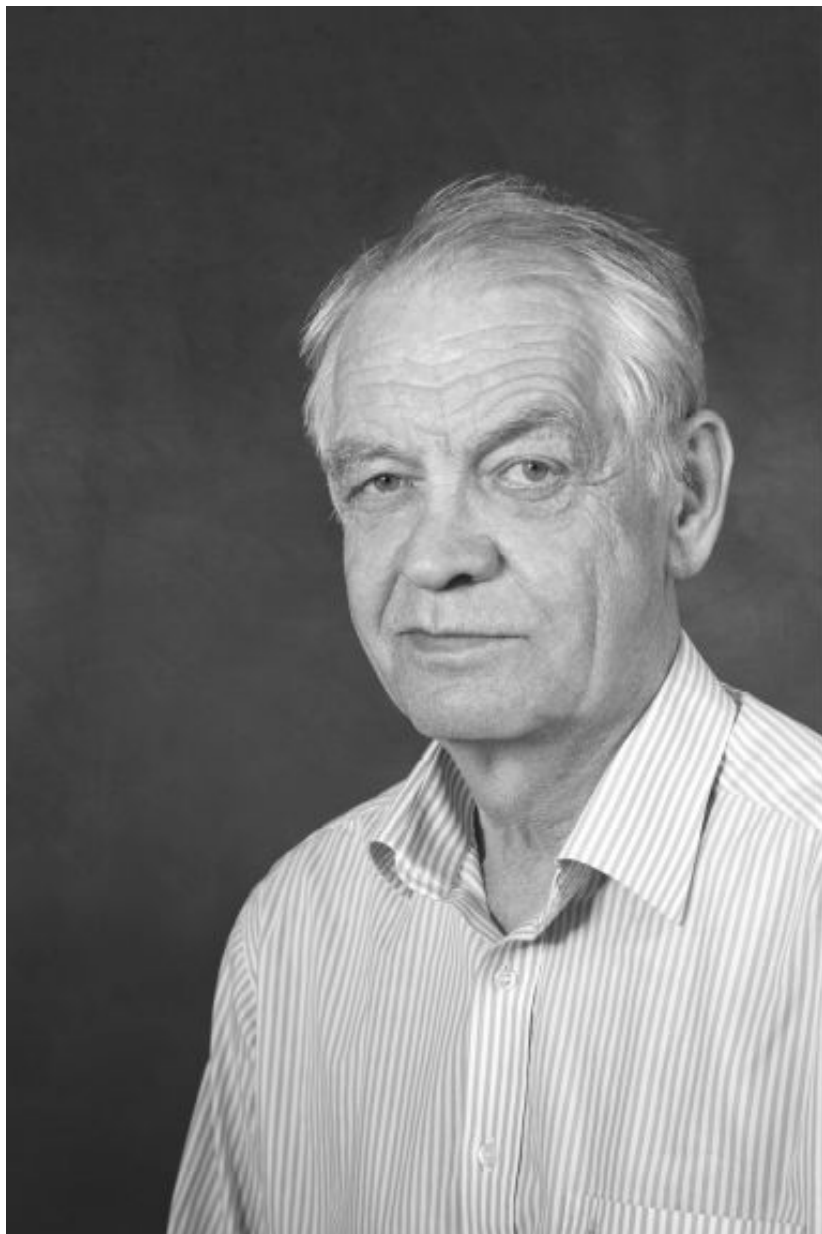
Научные исследования

анализ больших данных

облачные веб-лаборатории

Вместо заключения

1. Как в коммерческих так и в государственных структурах присутствует системная проблема достоверности данных.
 2. Обогащение открытыми данными - одна из мировых тенденций в data science.
 3. Создаваемые системы должны базироваться на современной инфраструктуре хранения и анализа данных. Подобные системы крайне крайне сложны и требуют наличия высоких компетенций разработчиков и экспертов.
- 
4. Необходимо создавать междисциплинарные команды (IT+Медицина).



ISPRAS Open 2020
Посвящается памяти
Виктора Петровича Иванникова

10-11 Декабря, 2020

Спасибо за внимание