

Хранилище данных экспериментальных исследований в области естественных наук

Мануилов Георгий, Лесик Демьян, студенты гр. 13641/2

1 Техническое задание

1.1 Мотивация

В настоящий момент существуют некоторые проблемы организации структурированного хранения экспериментальных данных в небольших научных группах, ведущих исследования в области естественных науки (физика, химия, биология). Проблемы можно условно разбить на три категории:

- Локальность. Данные хранятся локально, без возможности какого-либо внешнего доступа к ним. Пример: результаты эксперимента хранятся на носителе одного из членов научной группы, в то время как другие члены группы работают рассредоточено, в результате чего передача данных может быть реализована только через личную встречу или через пересылку данных, непрозрачную для остальных членов группы.
- Неструктурированность. Зачастую даже в пределах одной научной группы не существует жестко заданного формата хранения метаданных об эксперименте (протокола эксперимента), в результате чего возможна неправильная трактовка результатов эксперимента, а также невозможно его воспроизведение. Кроме того, такой подход к хранению метаданных затрудняет подключение новых участников к научной работе.
- Потеря актуальности. В результате первых двух пунктов появляется проблема потери актуальности данных, так как при рассредоточенной работе нет возможности получать последние данные об эксперименте с малой задержкой.

1.2 Описание системы

Требуется разработать систему, предназначенную для хранения и первичной обработки данных экспериментальных исследований в области естественных наук. Система должна предоставлять следующие функции:

- Возможность накопления и сортировки экспериментальных данных. Под экспериментальными данными подразумеваются данные, полученные с экспериментальной установки (далее сырые данные), а также исчерпывающие метаданные, необходимые для их интерпретации (далее метаданные).
- Возможность управления форматами экспериментальных данных. Сырые данные могут представлять собой как файлы определенного расширения, так и бинарные файлы. Метаданные могут состоять из произвольной информации, то есть необходимо исключить искусственные (например, строго табличное представление) ограничения на формат метаданных.
- Возможность написания и "горячей" установки отдельных модулей, позволяющих проводить парсинг сырых данных, расчет необходимых метрик и их графическое представление (далее парсеров). Модули должны быть полностью обособлены от самой системы.
- Возможность генерации простейших отчетов.
- Возможность экспорта данных для их автономного перемещения и хранения.

1.3 Требования к системе

- Должна быть реализована клиент-серверная архитектура, причем клиент должен предоставлять GUI.
- Должна быть обеспечена иммутабельность сохраненных данных.
- Должна быть реализована система пользователей и прав доступа.
- Механизм разработки парсеров может быть реализован как с использованием разработанного DSL, так и с использованием уже существующего языка программирования.
- Должен быть реализован API, позволяющий производить накопление сырых данных непосредственно в хранилище в процессе эксперимента (путем вызова методов API из некоторого связующего ПО для экспериментальной установки).