

ОТЧЕТ О ПРОДЕЛАННОЙ РАБОТЕ С ИСПОЛЬЗОВАНИЕМ ОБОРУДОВАНИЯ ИВЦ НГУ

1. Аннотация

Для решения задач организации обработки нейрофизиологических данных на суперкомпьютерах от исследователя требуется квалификация в системном программировании. Возникает потребность в создании программного инструмента для поддержки пользователей в решении таких задач. Работа выполняется в рамках проекта “Ака” ИВМиМГ СО РАН по созданию системы высокоуровневого управления исследовательскими данными и организации вычислений на суперкомпьютерах.

Была достигнута цель – разработан проект и прототип программного инструментария для высокоуровневой организации обработки нейрофизиологических данных на суперкомпьютерах на основе концепции представления знаний о вычислениях в предметной области в виде вычислительных моделей. Вычислительные модели позволяют формально описать совокупности вычислительных сценариев в предметной области, представляя их в виде двудольных графов, содержащих в вершинах переменные (величины предметной области) и связывающие их вычислительные операции.

Для достижения цели были поставлены и выполнены следующие задачи:

- 1) Изучены проблемы, возникающие при обработке научных данных и сформулированы требования к программного инструментария;
- 2) Предложена архитектура программного инструментария на основе системы Ака и расширена функциональность этой системы;
- 3) Разработан прототип предметно-ориентированной среды для решения задач обработки нейрофизиологических данных и протестирован на реальной задаче.

Методы исследования: анализ требований к разработке программного обеспечения, обзор литературы и существующих программных средств; анализ пользовательских сценариев; применение средств языка моделирования UML; архитектурный стиль разработки сетевых сервисов REST.

Новизна работы заключается в предложенной модели организации удаленного выполнения операций вычислительных моделей, включая предложенную форму спецификации операций.

Полученные результаты могут использоваться разработчиками вычислительных моделей в Ака при расширении созданной среды для обработки нейрофизиологических данных и при разработке предметно-ориентированных сред для других областей науки.

2. Тема работы

Разработка программного инструментария для высокоуровневой организации обработки нейрофизиологических данных на суперкомпьютерах.

3. Состав коллектива

Налепова Елизавета Денисовна, студент НГУ ФИТ 4 курс группа 18203, кафедра параллельных вычислений.

Городничев Максим Александрович, ст. преп. каф. ПВ ФИТ НГУ.

4. Научное содержание работы

4.1. Постановка задачи

Для решения задач организации обработки нейрофизиологических данных на суперкомпьютерах от исследователя требуется квалификация в системном программировании. Возникает потребность в создании программного инструмента для поддержки пользователей в решении таких задач.

4.2. Современное состояние проблемы

Разработка моделей и программ для суперкомпьютеров является активной областью исследования и существующие прикладные пакеты не покрывают всех задач, с которыми сталкиваются пользователи высокопроизводительных вычислительных систем. Актуальна проблема создания предметно-ориентированных сред организации вычислительных экспериментов на высокопроизводительных вычислительных системах, что включает решение задач систематического управления данными и процессами моделирования и обработки в рамках научных исследований в различных предметных областях. Частные проблемы: спецификация сценариев расчетов (workflow), реализация сценариев при необходимости использования нескольких вычислительных систем и систем хранения данных одновременно, высокоуровневые интерфейсы для пользователей и др. Обзор проблемы и литературы по теме: Gorodnichev, M., Lebedev, D. Semantic tools for development of high-level interactive applications for supercomputers. *J Supercomput* 77, 11866–11880 (2021). <https://doi.org/10.1007/s11227-021-03731-6>

4.3. Подробное описание работы, включая используемые алгоритмы

Были проведены:

- обзор литературы и существующих программных средств, используемых как для решения прикладных задач, так и для обеспечения высокоуровневой организации обработки данных, в том числе на суперкомпьютерах;
- анализ пользовательских сценариев.

Сформулированы требования к разработке на основе изученных материалов.

Для обеспечения высокоуровневой организации обработки данных прикладному специалисту необходимо предоставить интерфейс, с помощью которого он сможет задавать параметры задач и ставить расчеты на ВВС. Такой интерфейс для одной конкретной задачи в совокупности с программными компонентами, обеспечивающими решение задачи на ВВС, называется мини-приложением.

Для каждой группы нейрофизиологов, в соответствии с задачами проводимого исследования, может требоваться разработка специальной коллекции мини-приложений.

Для повышения эффективности разработки мини-приложений целесообразно систематически переиспользовать программные компоненты, реализующие отдельные этапы или методы обработки данных, компоненты графического интерфейса и системные программные компоненты.

Таким образом, для реализации целей работы предложено создание инструментария, автоматизирующего разработку мини-приложений на основе базы знаний о расчетах в предметной области.

Знания о расчетах формализуются вычислительной моделью, представляющей собой ориентированный двудольный граф с операциями и переменными в вершинах.

Заданные множества входных и выходных переменных и нефункциональные требования к выводу алгоритма получения значений выходных переменных по значениям входных называются спецификацией задачи на вычислительной модели. Спецификация задачи формализует требования к генерации мини-приложения для конкретной задачи.

Элементы графического интерфейса, такие как поля ввода и вывода данных, а также программные модули, реализующие вычислительные операции, задаются на этапе разработки вычислительной модели и накапливаются для дальнейшего переиспользования. Для моделирования архитектуры программного инструментария применены средства языка UML.

Для удаленного выполнения операций вычислительных моделей предложена разработка сетевых сервисов в соответствии с архитектурным стилем REST.

Для реализации веб-интерфейса для прикладного специалиста использован язык JavaScript и библиотека Bootstrap v5.0.

4.4. Полученные результаты

В результате проделанной работы:

1. Разработан проект программного инструментария для высокоуровневой организации обработки нейрофизиологических данных на суперкомпьютерах: предложена разработка инструментария на базе системы Ака (ИВМиМГ СО РАН), которая предоставляет средства для систематического накопления и переиспользования сценариев в предметных областях. Предложено расширение Ака системой сервисов для реализации вычислительных операций на высокопроизводительных вычислительных системах. Для обеспечения высокоуровневой организации обработки нейрофизиологических данных в проекте предложена разработка вычислительных моделей, формализующих совокупность сценариев вычислений в этой предметной области и оставляющих возможность для расширения описаний вычислений, а также создание коллекции элементов графического интерфейса для визуализации данных.
2. Разработан прототип предметно-ориентированной среды для решения задач обработки нейрофизиологических данных на суперкомпьютерах: реализован сервис выполнения Python-операций; программные модули, реализующие операции обработки данных, включены в коллекцию; разработана вычислительная модель для решения одной из задач обработки нейрофизиологических данных; реализован графический интерфейс для взаимодействия пользователя с задачей. Прототип предметно-ориентированной среды сокращает время, затрачиваемое нейрофизиологами на решение задач обработки данных, поскольку предоставляет высокоуровневые инструменты для решения задач в терминах предметной области и понятный пользовательский интерфейс.
3. Проведено исследование разработанной среды на реальной задаче проведения спектрального анализа ЭЭГ-сигналов.

Внесенные изменения в архитектуру системы Ака, разработанные программный интерфейс сервиса выполнения Python-операций и спецификация операций позволяют разрабатывать предметно-ориентированные среды для других предметных областей и языков программирования.

5. Эффект от использования кластера в достижении целей работы

Предполагалось выполнение расчетных задач на кластере, но из-за невозможности установить Python библиотеки план изменился (недоступен репозиторий pypi.org:443 files.pythonhosted.org:443, скриншот прилагается).

```
(env) ednalepova@clu:~/eeg_processing/brain-data-processing> pip install mne
Collecting mne
  Retrying (Retry(total=4, connect=None, read=None, redirect=None)) after connection broken by 'NewConnectionError('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f21c1f65668>: Failed to establish a new connection: [Errno 101] Network is unreachable,')': /simple/mne/
  Retrying (Retry(total=3, connect=None, read=None, redirect=None)) after connection broken by 'NewConnectionError('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f21c1f657f0>: Failed to establish a new connection: [Errno 101] Network is unreachable,')': /simple/mne/
  Retrying (Retry(total=2, connect=None, read=None, redirect=None)) after connection broken by 'NewConnectionError('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f21c1f652e8>: Failed to establish a new connection: [Errno 101] Network is unreachable,')': /simple/mne/
  Retrying (Retry(total=1, connect=None, read=None, redirect=None)) after connection broken by 'NewConnectionError('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f21c1f65a20>: Failed to establish a new connection: [Errno 101] Network is unreachable,')': /simple/mne/
  Retrying (Retry(total=0, connect=None, read=None, redirect=None)) after connection broken by 'NewConnectionError('<pip._vendor.requests.packages.urllib3.connection.VerifiedHTTPSConnection object at 0x7f21c1f656d8>: Failed to establish a new connection: [Errno 101] Network is unreachable,')': /simple/mne/
^Z
[7]+  Stopped                  pip install mne
(env) ednalepova@clu:~/eeg_processing/brain-data-processing> █
```