

**ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата физико-математических наук Шайхисламова Дениса
Ильгизовича на тему: «Исследование и разработка методов для
сравнительного анализа суперкомпьютерных приложений на основе
технологий интеллектуального анализа данных»
по специальности 2.3.5. Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей**

Диссертационное исследование Д.И. Шайхисламова относится к проблеме анализа пользовательских заданий на суперкомпьютерах в целях оптимизации процессов их запуска и выполнения, а также для выработки рекомендаций пользователям по оптимизации своих заданий к конкретным условиям данного суперкомпьютера. Актуальность этой темы имеет перманентный характер — суперкомпьютерные технологии быстро развиваются, соответственно проблема приобретает новые аспекты, которые даже могут кардинально ее изменить. Более того, каждый суперкомпьютер технологически уникален. В этой связи отметим, что только в одном суперкомпьютере, *Ohio Supercomputer Center*, используется схожий подход к обнаружению пакетов ПО - *XALT+PBS*. При этом в диссертационном исследовании используется подход на основе *XALT+Slurm*, что препятствует прямому трансферу технологий. Представленный в диссертации подход к обнаружению пакетов ПО в пользовательских заданиях разработан для использования в суперкомпьютере *Ломоносов-2* (МГУ). Однако этот подход, в целом и в отношении отдельных его компонент, может использоваться в разработках аналогичных решений для других суперкомпьютеров.

В диссертационных исследованиях Д.И. Шайхисламова ключевым является понятие *схожести* прикладных заданий - они «обладают одними и теми же важными свойствами, связанными с поведением приложения во время его выполнения» (стр. 5 диссертации). Эти характеристики, «одни и те же» и «важные» свойства, получают разное математическое наполнение в зависимости от разрабатываемых методов выделения «схожих» заданий. В диссертации представлены два метода — известных в литературе как

статический и *динамический* методы. Далее мы даем описание содержания этих методов в конкретностях, оптимальных для суперкомпьютера *Ломоносов-2* (МГУ).

В первом методе используются статические характеристики исполняемых файлов. Задания считаются *статически схожими*, если их исполняемые файлы используют, в большой мере, пересекающиеся наборы имен функций и переменных. В качестве меры *схожести* предлагается *косинусное расстояние* между двум векторами фиксированной длины, сопоставляемых заданиям по NLP технологии Word2Vec в режиме PV-DBOW с распределенным набором слов. Для оценки точности *статического* метода использовалась иерархическая агломерационная кластеризация (sklearn, Python) для различных значений порога по *косинус-расстоянию* и ручная кластеризация. В качестве оценки точности *статического* метода вычислялся скорректированный индекс Рэнда (ARI). Лучшие значения ARI получились на уровне 0.75 при порогах 0.1-0.3, что говорит о хорошей точности *статического* метода.

Во втором методе используются динамические данные (временные ряды) о процессе выполнения задания, получаемые от системы мониторинга показателей датчиков вычислительных узлов — всего по 15 параметрам. Для оценки *динамической схожести* заданий используется расстояние между ними, вычисляемому по методу *динамической трансформации временной шкалы (DTW)*. Результаты динамического метода сравниваются с размеченной выборкой пар заданий — *схожи или нет*, выполненной вручную. Точность динамического метода определяется, как отношение правильно выделенных схожих и не схожих временных рядов пар заданий к общему числу пар. На тестовых выборках (~400 пар заданий) была получена точность на уровне 0.85. Уменьшение размерности временных рядов с 15 до 5 с использованием нейронной модели автокодировщика увеличило точность до 0.92. А введение масштабирования показателей мониторинговых датчиков, с коэффициентам 0.4 для загрузки CPU и 0.33 для загрузки GPU и

LoadAVG, дало увеличение точности с 0.85 до 0.95. Эти результаты говорят о весьма хорошей точности *динамического* метода выделения схожих заданий. В диссертации делается вывод, что *динамический* метод может уловить тонкие различия в отдельных классах прикладных заданий за счет обработки информации о поведении приложения во время исполнения с определенными параметрами запуска или входными данными.

Вышеприведенные конкретности содержания статического и динамического методов выделения схожих заданий получены в ходе диссертационных исследований Д.И. Шайхисламова в применении на суперкомпьютере *Ломоносов-2* (МГУ), что определяет практическую значимость полученных результатов. Однако мы отмечаем, что результаты диссертационного исследования в своей совокупности представляют системно проработанный подход к решению проблемы анализа пользовательских заданий на суперкомпьютерах в целях оптимизации процессов их запуска и выполнения, а также для выработки рекомендаций пользователям по оптимизации своих заданий к конкретным условиям данного суперкомпьютера. Добавим, что важной составляющей подхода являются результаты исследований эффективности множества математических инструментов интеллектуального анализа данных, представленных в Главах 2 и 3 диссертации. В таком расширенном формате, на наш взгляд, состоит научная новизна и актуальность разработанного подхода, а также научная значимость для разработчиков аналогичных методов выделения схожих заданий для других суперкомпьютеров.

Диссертация состоит из Введения, 4-х Глав, Заключения и Списка литературы, всего 99 страниц.

Во Введении сформулированы актуальность темы диссертации, цели и задачи диссертационного исследования, научная новизна и практическая значимость результатов выполненных работ.

Глава 1 посвящена обзору существующих методов для решения задачи выделения схожих приложений — в деталях обсуждаются статический и

динамический методы. Проводится анализ достоинств и недостатков существующих их реализаций.

В Главе 2 представлены результаты исследований по выбору математических инструментов для статического метода выделения схожих заданий, с оптимизацией для применения на суперкомпьютере *Ломоносов-2* (МГУ).

В Главе 3 представлены результаты исследований по выбору математических инструментов для динамического метода выделения схожих заданий, с оптимизацией для применения на суперкомпьютере *Ломоносов-2* (МГУ).

Глава 4 представлены результаты практического применения разработанных методов выделения схожих суперкомпьютерных приложений. Показано, что можно выявлять использование программных пакетов в новых, ранее не исследованных, заданиях, используя исторические данные. Апробация на суперкомпьютере «Ломоносов-2» показала, что разработанные методы позволяют выделить на 15% больше заданий, чем существующие решения, в некоторых случаях до 80% больше. Разработан новый метод кластеризации суперкомпьютерных приложений на основе совместного применения статического и динамического методов выделения схожих заданий. ARI оценка такого совместного метода, полученная на вручную размеченных данных, составила 0.8. Показано, что применение совместного метода кластеризации позволяет эффективно выявлять аномалии в сериях запусков пользовательских приложений, а также предсказывать оценки качества использования суперкомпьютерных ресурсов приложениями на основе исторических данных. По этим результатам можно говорить о научной новизне и практической значимости для повышения эффективности администрирования пользовательских заданий.

Достоверность полученных в диссертационных исследованиях результатов подтверждается системной проверкой работоспособности и корректности предложенных решений на размеченных экспертами данных,

апробацией на ранее неизученных данных, а также корректностью использования множества математических инструментов, получивших массовую апробацию в научных исследованиях, опубликованных в научной литературе.

К содержанию диссертационной работы имею следующие замечания:

- для статического метода выделения схожих заданий полезно было бы добавить сравнение используемой NLP модели со статистической мерой важности слов в документе, TF-IDF. А для динамического метода добавить сравнение DTW метрики, используемой для оценки точности, с метрикой EDR (Edit Distance on Real Sequences).

- для дополнительного акцента на практическое значение разработанных методов кластеризации пользовательских заданий, полезно было бы показать, какую помощь такие методы кластеризации могут оказать администраторам суперкомпьютерного комплекса.

Заметим, однако, что указанные замечания не снижают высокую оценку диссертации Д.И. Шайхисламова.

Текст диссертации написан ясным языком на хорошем научном уровне. По результатам диссертационной работы сделаны обоснованные выводы. Автореферат правильно отражает содержание диссертации и полученные результаты.

Диссертация «Исследование и разработка методов для сравнительного анализа суперкомпьютерных приложений на основе технологий интеллектуального анализа данных» Д.И. Шайхисламова отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к научно-квалификационным работам. Содержание диссертации соответствует паспорту специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени

М.В.Ломоносова. Диссертационное исследование оформлено согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В.Ломоносова.

Таким образом, соискатель Шайхисламов Денис Ильгизович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент:

доктор физико-математических наук, старший научный сотрудник,
главный научный сотрудник Курчатовского комплекса НБИКС-пт
ФГБУ НИЦ "Курчатовский институт"

Ильин Вячеслав Анатольевич

Дата: 24.02.2026

Контактные данные: тел. 7(915)0160048, e-mail: ilyin0048@mail.ru

Специальность, по которой официальным оппонентом защищена диссертация: 01.04.02

Теоретическая физика

Адрес места работы: 123182, Россия, Москва, пл. Академика Курчатова, д.1,

Рабочие контакты.: 7(499)1969539; e-mail: ilyin_va@nrcki.ru

Подпись сотрудника НИЦ «Курчатовский институт» Ильина В.А. заверяю

Заместитель директора - Главный ученый секретарь НИЦ «Курчатовский институт»

О.А. Алексеева

«_24_» _февраля_ 2026 г.