

ОТЗЫВ

официального оппонента, кандидата технических наук, старшего преподавателя, доцента, Мелеховой Анны Леонидовны на диссертационную работу Успенского Михаила Борисовича на тему «Разработка и исследование методов и моделей обработки диагностической информации для обнаружения и локализации неисправностей в системах хранения данных», представленную на соискание ученой степени кандидата наук по специальности 05.13.01 - «Системный анализ, управление и обработка информации»

Актуальность темы диссертационной работы

Современные системы хранения данных (СХД) становятся все более сложными год от года. Усложняется не только само дисковое хранилище как аппаратно-программный комплекс, увеличивается количество дополнительных компонентов и связей между ними. Устройства сетевого доступа, мониторинги, системы виртуализации – неисправность в любом узле может привести к отказу всей системы. Увеличивается диапазон возможных неисправностей. При этом гарантии по доступности данных в СХД лишь увеличиваются, а значит время на исследование инцидента, диагностирование проблемы и последующее её устранение уменьшается.

В условиях таких жестких временных ограничений становится все более важным качество автоматической диагностики неисправностей и их корректной локализации до компонента. Работа по установлению состояния отдельных компонентов СХД в зависимости от одних только журналов СХД, проведенная в рамках настоящего диссертационного исследования М.Б. Успенским, улучшает качество такой диагностики и имеет высокую практическую ценность.

Кроме практической значимости хотелось бы отметить широкий охват обзорной части и вдумчивую классификацию существующих подходов к диагностированию неисправностей СХД. Комбинирование нескольких подходов и неожиданное для данной области использование средств NLP делают данное исследование важным и с теоретической точки зрения.

Степень обоснованности научных положений, выводов и рекомендаций

Обоснованность научных положений, выводов, рекомендаций и достоверность полученных результатов подтверждается:

- обширным и глубоким исследованием предметной области, затронувшим не только научные работы российских и зарубежных авторов, но и практический инструментарий, существующий на рынке;
- использованием апробированного математического аппарата;

- получением значимого результата на рабочих, а не синтетических данных;
- публикацией полученных результатов и выводов в рецензируемых источниках;
- обсуждением полученных результатов и выводов на российских и международных конференциях.

Научная новизна диссертационной работы

Настоящее исследование интересно комбинированным подходом к процедуре диагностики неисправностей СХД, включающем и разработку онтологической модели, и использование методов NLP, и обоснованное применение методов машинного обучения (классификатора).

Промежуточные результаты, полученные автором, так же имеют научную ценность. Так, анализ и обработка журналов СХД (логов) не имеют аналогов и при этом могут быть использованы в других исследованиях, базирующихся на тех же исходных данных.

Говоря о научной новизне диссертационного исследования и полученных в нем результатов следует упомянуть следующие:

- новый метод построения диагностических моделей систем хранения данных, основанный на использовании онтологической модели и экспертной информации, где связь между объектами онтологии определяется методами машинного обучения;
- новый подход к обнаружению неисправностей в системах хранения данных, отличающийся от существующих использованием алгоритма классификации частично структурированных текстовых данных;
- новый многостадийный алгоритм обработки текстовой информации из журналов систем хранения данных, отличающийся возможностью обнаруживать неисправности на основании классификации частично структурированных текстов без детального анализа структуры, формата и порядка поступления сообщений мониторинга;
- новый метод обнаружения неисправностей в системах хранения данных, отличающийся от существующих комбинированием онтологической модели и алгоритмов машинного обучения для обработки текстовой информации, получаемой в процессе мониторинга систем хранения данных.

Практическая ценность результатов диссертационной работы

Практическая ценность данной диссертационной работы включает:

- разработку программного комплекса, способного диагностировать предаварийные состояния и идентифицировать причину неисправности (после обучения);
- получение решения, строящего диагностику на основании комбинации СХД специфичных журналов и общесистемных журналов;

- подготовку рекомендаций по настройке системы СХД для улучшения стабильности ее работы.

Оценка диссертационной работы и её завершенности в целом

Диссертационная работа Успенского М.Б. является целостной и завершенной научной работой, решающей задачу классификации неисправностей в системах хранения данных по средствам анализа журналов СХД. Оформление диссертации соответствует требованиям ГОСТ, работа написана грамотным литературным языком. Диссертация включает в себя введение, 5 глав, заключение, библиографический список и три приложения. Наполнение разделов соответствует общепринятым принципам написания диссертационных работ.

Во введении приводится актуальность темы исследования, формулируется задача и цели работы, приводятся положения, выносимые на защиту. Здесь же, в соответствии с общепринятыми принципами, обосновывается научная новизна, теоретическая и практическая значимость.

В первой главе описывается предмет исследования – системы хранения данных. Приводятся значимые для дальнейшего понимания ключевые характеристики данного класса систем. Далее проводится анализ существующих методов обнаружения и локализации неисправностей в СХД, выделяются критерии для сравнения. Сравнительный анализ на базе выбранных критериев обосновывается необходимость разработки новых методов, подходов и моделей.

Вторая глава исследования посвящена построению диагностической онтологической модели. Онтологическая модель сравнивается с семантической сетью, фреймовой моделью, продукционной и логической моделью. В модели выделены следующие классы объекта СХД: параметр СХД, компонент СХД, подсистема СХД, система, состояние, событие в СХД. Подробно описана модель отношений/свойства верхнего уровня в онтологической модели СХД. Рассмотрены два способа сбора диагностической информации – активный и пассивный. Подробно разобраны различные классы журналов (логов) как источников данных. Собраны как общесистемные журналы, так и журналы, специфичные для конкретного СХД.

Третья глава описывает алгоритм динамического обнаружения неисправностей в СХД. На вход этому алгоритму, строящему классификатор, передаются уже обработанные журналы СХД. При этом алгоритм обработки журналов СХД интересен сам по себе и включает несколько стадий: предварительную обработку, фильтрацию и выделение именованных сущностей (NER). Решая задачу классификации сравниваются два основных подхода: random forrest и градиентный бустинг (gradient boosting).

Четвертая глава представляет комбинированный алгоритм, построенный из комбинации предложенных ранее моделей и методов для локализации и обнаружения событий неисправностей в СХД. Здесь же даны рекомендации по модификации архитектуры использующего его программного обеспечения.

Пятая глава описывает экспериментальную проверку выдвинутых ранее теоретических предположений, сравнивает эффективность различных алгоритмов классификации текстов. Демонстрируются результаты работы всего алгоритма обнаружения и локализации неисправностей в СХД.

Заключительный раздел представляет полученные результаты, указывает направления дальнейших исследований, утверждает соответствие диссертационной работы паспорту специальности.

По содержанию диссертации можно сделать следующие замечания:

1. Хотелось бы увидеть дальнейшее развитие работы, когда диагностическая система будет не только идентифицировать источник проблемы по отработанным журналам, но и предоставлять экспертную оценку непосредственно во время работы СХД, связываясь с системами мониторинга (например, Zabbix).
2. Не совсем понятно, сколько времени займет перестроение онтологической модели и переобучение классификаторов при изменении параметров системы, которые происходят не реже нескольких раз в месяц.
3. Некоторые из рисунков в работе имеют невысокое разрешение, что усложняет чтение текста (например, рис. 2.2, рис. 3.3).
4. В работе не представлена архитектура СХД, на котором проводился экспериментальный прогон разработанного программного комплекса, и не приводится перечень компонентов. Возможно, это ограничено соглашениями о неразглашении, но исследование выиграло бы от этого.

Несмотря на данные недостатки, выполненная работа оставляет положительное впечатление.

Заключение

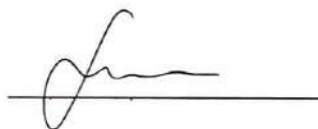
Диссертационная работа Успенского Михаила Борисовича на тему «Разработка и исследование методов и моделей обработки диагностической информации для обнаружения и локализации неисправностей в системах хранения данных» полностью соответствует требованиям п.п. 2.1-2.8 Положения о присуждении ученых степеней в ФГАОУ ВО «СПбПУ» (в действующей редакции) и требованиям государственных стандартов. Результаты и основные положения диссертации соответствуют паспорту специальности 05.13.01 «Системный анализ, управление и обработка

информации» в части пунктов пунктам 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации», 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации», 11 «Методы и алгоритмы прогнозирования и оценки эффективности, качества и надежности сложных систем.», 12 «Методы получения, анализа и обработки экспертной информации». Диссертация Успенского М.Б. является законченной научно-квалификационной работой, обладает научной новизной, имеет теоретическую и практическую значимость и решает актуальную научно-техническую проблему. Исходя из этого, Успенский Михаил Борисович заслуживает присуждения степени кандидата технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации».

Официальный оппонент, старший преподаватель кафедры теоретической и прикладной информатики ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)», архитектор программного интерфейса ООО «Акронис», глава команды разработки системы доступа к СХД «Акронис»:

Кандидат технических наук, доцент
кафедры Теоретической и
Прикладной Информатики МФТИ
(НИУ)

Мелехова Анна
Леонидовна



02.11.2020

Подпись руки
ЗАВЕРЯЮ:
ЗАВЕДУЮЩАЯ КАНЦЕЛЯРИЕЙ
АДМИНИСТРАТИВНОГО ОТДЕЛА
М.А.ГУСЕВА



ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)»
141701, Московская область, г. Долгопрудный, Институтский пер., 9
тел. +7 (495) 408-45-54
info@mipt.ru