

Министерство образования Республики Беларусь

Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.4

ЧУРИН  
Антон Павлович

## **РАСПРЕДЕЛЕННАЯ СИСТЕМА ДЛЯ АНАЛИЗА ЛОГОВ**

### **АВТОРЕФЕРАТ**

на соискание степени магистра информатики и вычислительной техники  
по специальности 1-40 81 04 Обработка больших объёмов информации

Научный руководитель  
Теслюк Владимир Николаевич  
канд. физ.-мат. наук, доцент

Минск 2018

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## **Актуальность темы исследования**

Составление прогностических моделей особенно актуально для сферы бизнеса, главной задачей, в которой является обладать знаниями, благодаря которым можно повысить эффективность, сократить расходы, и/или увеличить объемы продаж. Именно сфера «Больших данных» является поставщиком эффективных прогностических решений – по статистике, на данный момент анализируется только 0,5% накопленных цифровых данных, остальная же часть содержит в себе огромное количество «скрытых» знаний, которые потенциально могут быть источником огромной прибыли и превосходства над конкурентами. Примерами таких задач могут быть задачи построения клиентоориентированной практики, задачи прогнозирования, создания инсайтов и контекстов, а также формулирования рекомендаций к действию.

Цель любой организации – прибыль. Это особенно актуально при работе в условиях жесткой конкуренции и нестабильной экономической ситуации в мире.

Для продуктивной работы требуется все рабочее время уделять разработке программного обеспечения, но в современном мире мы не можем говорить о постоянной и стабильной работе приложения при его разработке.

В связи с этим предлагается использовать системы логирования для понимания и отслеживания ошибок и недочетов во время этапа разработки для более корректной и продуктивной работы над проектом.

В любом случае, использование системы логирования работает на пользу, но отнимает большое количество времени для поиска и исправления ошибок, когда проект начинает разрастаться.

Для продуктивной работы сотрудников подобная система должна соответствовать потребностям компании, быть удобной в использовании и предоставлять наиболее полный, широкий набор функциональных возможностей.

Актуальность темы подтверждается отсутствием на данный момент в Республике Беларусь внедрённых или находящихся в процессе внедрения схожих систем, выполняющих комплекс задач по анализу логов.

## **Цель и задачи исследования**

**Целью** диссертации является изучение современных средств взаимодействия с нейронной сетью и разработка способов для качественной обработки данных логирования.

Для выполнения поставленной цели в работе были сформулированы следующие **задачи**:

- изучить техническую литературу и публикации по теме магистерской диссертации;
- исследовать пути обработки данных посредством нейронных сетей для анализа и построения данных зависимости;
- разработать модель системы, выполняющей сбор и анализ данных;

– рассмотреть пути дальнейшего совершенствования разработанной системы.

**Объектом** исследования является автоматизированная система для анализа логов.

**Предметом** работы является комплексное использование алгоритмов хранения и обработки больших объёмов данных в автоматизированных системах.

**Область исследования.** Содержание диссертационной работы соответствует образовательному стандарту высшего образования второй ступени (магистратуры) специальности 1-40 81 04 «Обработка больших объёмов информации».

**Структура и объём работы.** Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, четырёх глав и заключения, библиографического списка и приложения. Общий объём диссертации – 59 страниц. Библиографический список включает 14 наименований.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние проблемы высокая интенсивность ошибок на этапах рабработки програмных продуктов используя системы логирования для решения проблемы, определены основные направления исследований, а также даётся обоснование актуальности темы диссертационной работы.

В **первом разделе** рассматриваются необходимость разработки новой системы, изучается проблема и формируются требования к разрабатываемой системе.

Во **втором разделе** рассматриваются архитектура системы, стек используемых технологий и способы взаимодействия с аппаратными средствами.

В **третьем разделе** приводятся алгоритмы работы с данными в отдельных модулях программного средства.

В **четвёртом разделе** приведены результаты разработки и тестирования программного средства.

## ЗАКЛЮЧЕНИЕ

В данной диссертационной работе была рассмотрена задача проектирования и реализации системы для анализа логов. Тема диссертационной работы обусловлена заказом предприятия.

В ходе работы над диссертационной работой проанализирована растущая популярность языка Python. Исследованы различные пути и подходы к решению задач, связанных с разработкой программного средства на данном языке программирования. Опробованы различные интегрированные среды разработки и способы создания приложений.

В результате работы над проектом были решены следующие подзадачи:

- Изучены существующие аналоги программного средства: их функции и причины невозможности использования.

- Сформулированы требования к разрабатываемой распределенной системы для анализа логов.

- После анализа требований, принято решение реализации распределенной системы для анализа логов с возможностью дальнейшей расширяемости.

- Разработана архитектура проекта в соответствии с правилами программирования на языке Python.

- С учетом сформулированных требований и предложенной архитектуры разработано программное средство, решающее поставленную задачу.

Перед выполнением экспериментальной части исследования была изучена техническая литература и научные публикации по теме диссертации, рассмотрен уже имеющийся опыт решения проблемы. В частности, рассмотрены современные технические устройства.

Поставленная задача выполнена в полном объеме. При этом при разработке системы была заложена возможность её дальнейшего совершенствования. Среди перспектив развития системы можно выделить следующие:

- построение графиков исходя из логов;

- применение предиктивной аналитики;

В ходе определения требований в качестве разрабатываемого программного средства послужила графовая база данных и нейронная сеть.

В ходе работы получилось раскрыть тему диссертационной работы «Распределенная система для анализа логов».

Основными достоинствами разработанной системы являются:

- высокая скорость доступа и обработки данных;
- возможность принятия решений в ручном и автоматическом режиме;
- модульность и расширяемость системы.

Реализация и внедрение системы в реальных условиях потребует незначительных финансовых и временных затрат (установка технических средств, создание дата-центра, настройка и калибровка алгоритмов обработки данных). Однако при правильном и грамотном внедрении всех компонентов системы её использование в перспективе неизменно принесёт значительный положительный экономический эффект для любой компании.

## **СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**

1-А. Графовые базы данных. Чури́н А.П., Владыко В.Д., Теслюк В.Н. БГУИР, 2018 г.

2-А. Кластеризация в Hadoop. Владыко В.Д., Чури́н А.П., Теслюк В.Н. БГУИР, 2018 г.