

РЕФЕРАТ

<Ключевые слова>

СОДЕРЖАНИЕ

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем отчете о НИР применяют следующие термины с соответствующими определениями:

Когнитивный анализ данных — один из наиболее мощных инструментов исследования нестабильной и слабоструктурированной среды. Он способствует лучшему пониманию существующих в среде проблем, выявлению противоречий и качественному анализу протекающих процессов

Конвергенция — процесс сближения

Распределённые гетерогенные данные — это логически связанных между собой, разнородные данные, которые физически распределены по нескольким узлам в некоторой сети

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о НИР применяют следующие сокращения и обозначения:

SH — space hedgehog

SW — space walker

СДСН — синдром длительных статических нагрузок.

ВВЕДЕНИЕ

Актуальность исследования обусловлена необходимостью конвергенции передовых информационных технологий для решения проблемы «информационного взрыва», в науке и обществе. Проблема возникла, потому что в современном мире фундаментальная и прикладная наука, наряду с увеличением деятельности пользователей в интернете, приносит большие объемы данных, из-за чего стало накапливаться значительное количество информации, а огромное количество данных создает проблемы для их анализа и интерпретации. Последствия «информационного взрыва» приводят в конечном счете к развитию направления Big Data. Направление Big Data выступает альтернативой увеличения вычислительных мощностей, тем что дает новые информационные концепции и технологии, одной из которых является когнитивная визуализация данных.

Целью данной работы является, исследования особенности создания web-сервисов для работы с системами интеллектуального анализа больших данных в сетевых и облачных средах, исследовать возможности удалённого сетевого применения методики когнитивного анализа Big Data, предложить способы доступа к разнородным массивам данных, относящихся к одной предметной области, но которые получены из различных источников различными средствами, различными методами и хранятся в разных форматах под разными СУБД. В качестве программных систем когнитивной визуализации освоить и использовать программные системы SpaceWalker(SW) реализована на базе Java программная система когнитивной визуализации, создающая динамический 3D образ, которая позволяет оперативно проводить многомерный когнитивный анализ многомерных данных произвольной природы, и SpaceHedgehog(SH) методика и технология когнитивной 6D визуали-

зации многомерных данных, разработанная на основе исследования и развития технологии когнитивной визуализации, которая предоставляет возможность одновременно визуально воспринимать и анализировать в динамике шесть произвольных параметров многомерного объекта, проанализировать и предложить методику и программные средства разработки для них web-сервисов на различных платформах или мультиплатформенной основе, разработать web-сервис для использования программных комплексов когнитивного анализа Big Data, обеспечивающий не только быстрый и надёжный доступ к системам, но и релевантную визуальную интерактивность процесса когнитивного анализа данных в приемлемой для систем когнитивной визуализации форме, исследовать применимость и эффективность разработанного сервиса для систем когнитивного анализа данных.

В соответствии с поставленной целью были определены следующие задачи:

- изучить текущие состояние объекта,
- сформировать математическую модель объекта,
- проанализировать математическую модель и выявить причину проблемы и предложить улучшение объекта,
- спроектировать web-сервис по улучшенной модели объекта,
- реализовать web-сервис,
- оценить социальную значимость разработки,
- оценить технико-экономическое обоснование разработки,
- оценить безопасность и экологичность разработки,

Практическая ценность работы заключается в создании web-сервиса для когнитивного анализа распределённых гетерогенных данных

1 ОПИСАНИЕ И АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1 Организационная структура САО РАН

Цель деятельности САО РАН осуществлять научные исследования за объектами вселенной с земли, издание научных журналов и научных трудов, в которых публикуются результаты исследований ученых.

Для реализации своих целей у обсерватории существуют такие инструменты, как оптический телескоп БТА (Большой Телескоп Азимутальный) с диаметром главного зеркала 6 м и радиотелескоп РАТАН-600 (Радиотелескоп Академии Наук) с кольцевой многоэлементной антенной диаметром 600 м. Телескопы (инструменты) открытого коллективного пользования, и допускают широкую интеграцию с астрономическими сообществами по всему миру. Время наблюдений распределяется национальным комитетом по тематике российских телескопов сокращенно НКТРТ.

Обсерватория расположена в Зеленчукском районе Карачаево-Черкесской Республики Российской Федерации. БТА установлен на склонах г. Пастухова на высоте 2100 метров над уровнем моря. Здесь же находятся два малых телескопа диаметром 1 и 0.6 метров. РАТАН-600 сооружен в 20 км от БТА на окраине станции Зеленчукской на высоте 970 метров. Научный поселок Нижний Архыз (лабораторные и служебные корпуса САО и жилые дома сотрудников) построены на берегу реки Большой Зеленчук. Обсерватория имеет филиал в Санкт-Петербурге (Пулково). В САО работает 430 сотрудников, из которых чуть более 100 - научные работники. САО РАН по Распоряжению Правительства РФ от 27.06.2018 1293-р относится к организациям, подведомственным Министерству науки и высшего образования Российской Федерации (Минобрнауки России).[1]

САО РАН [2] [?] [?] имеет иерархический тип структуры управления с линейно-штабной организационной структурой. Условно можно выделить 5 объектов, представим их на рисунке 1.1



Рисунок 1.1 – Условная организационная структура САО РАН

1.2 Описание процессов



Рисунок 1.2 – Диаграмма компонентов

сы изменяются за счет обмена информацией между подсистемами. Для наглядности покажем граф связи между объектами, и объем информационного оборота между звеньями за ед. времени ТБ в день (рисунок 1.4).



Рисунок 1.4 – Граф показывающий связь между объектами и передаваемый объем информации

Нужно понимать, что информационный ресурс не переходит из одного объекта к другому, а накапливается и распределяется по системе, и общее количество семантической информации в системе остается неизменным. К примеру, если объект A передаст (поделится) информацией с другим каким-либо объектом, информация у объекта A не уменьшится, а у другого увеличится.

Рассмотрим пару объектов A и B , которые обмениваться информацией, пусть объект A передает информацию, а объект B её принимает.

Информация которая имеется только у объекта, то есть i ($i \in \{A, B\}$) обозначим как K_i . Количество переданной информации другому объекту обозначим J_A . Информация, которая была воспринята другой подсистемой B обозначим как I , а которая была не воспринята S_A . Для объекта A введем

функцию результативности решения это будет $P_A = P_A(K_A, J_A)$ данная функция показывает от чего зависит результативность решения у объекта A это количество информации, которая она имеет, и количество переданной информации. Для объекта B результативность будет зависеть от количества информации, которое в ней имеется, и принятой информации $P_B = P_B(K_B, I)$. Можем сократить аргументы, в уравнениях состояний, потому что система замкнута значит количество информации $J_A + K_A$ и K_B не изменяться во времени. Получаем итоговые формулы уравнений состояний:

$$P_A = P_A(J_A), \quad (1.1)$$

$$P_B = P_B(I). \quad (1.2)$$

За счет обмена информацией между объектами, происходит изменение величины $P_i = (i \in \{A, B\})$, обозначим формулой ценность информации :

$$v_A = \frac{dP_A}{dJ_A}, \quad (1.3)$$

$$v_B = \frac{dP_B}{dI}. \quad (1.4)$$

Значение v_i при $i \in \{A, B\}$ показывает как хорошо подсистемы взаимодействуют между собой. При $v_i > 0$ i -подсистема может положительно обмениваться информацией, а при $v_i < 0$ i -подсистема препятствует передаче или приему информации. Представим введем формулы интенсивность потока информации:

$$q_a(v_A, v_B) = \alpha(v_A + v_B). \quad (1.5)$$

Где α – это размерный коэффициент пропорциональности.

Составим формулу интенсивности в пределах времени:

$$q_a(v_A, v_B) = \frac{dJ_A}{dt} \quad (1.6)$$

Интенсивность информации принимаемого объекта или подсистемы, всегда меньше интенсивности отправителя, так как есть некоторая информация, которая была не воспринята получателем. Обозначим коэффициент (параметр) воспринятой информации принимаемым объектом (подсистемой) как $p(q_A)$. Обозначим формулу интенсивности для получателя:

$$q_B = p(q_A)q_A \quad (1.7)$$

Можем заметить что при высокой интенсивности $q_A \rightarrow \infty$ коэффициент $p(q_A)$ стремиться к нулю, и информация меньше воспринимается, а при $q_A \rightarrow 0$ почти вся преданная информация воспринимается, обозначим это формулами:

$$\lim_{q_A \rightarrow 0} p(q_A) = 1, \quad (1.8)$$

$$\lim_{q_A \rightarrow \infty} p(q_A) = 0. \quad (1.9)$$

Таким образом построив математическую модель мы нашли «узкое место», что в системе высокая интенсивность обмена информацией между подсистемами, и что некоторая информация может быть не воспринята подсистемой. Следовательно проблема заключается в большой интенсивности информационного потока между подсистемами.

1.4 Полная постановка задачи ВКР

Проведя анализ предметной области, рассмотрели основные выполняемые процессы в системе, схему организационной структуры, и в результате построения математической модели была выявлена проблема, что информационной оборот в системе очень интенсивный. Следовательно нашей целью является снизить интенсивность информации, и оптимизировать коммуникацию между подсистемами.

Для решения поставленной цели выделим ряд задач:

- Провести оптимизацию модели и процессов,

- Построить процессы по оптимизированной модели,
- Сформировать параметры и требования к разработке,
- Сравнить аналоги удовлетворяющие требованиям,
- Спроектировать разработку,
- Реализовать разработку.

2 СОЦИАЛЬНАЯ ЗНАЧИМОСТЬ РАЗРАБОТКИ

Добавление web-сервиса автоматизирует процессы на объекте, что в свою очередь увеличивает эффективность работы сотрудников, и сокращает число выплат сотрудникам участвующих на объекте. Основная социальная значимость разработки заключается в том, что становится более легко произвести точный анализ данных, из чего следуют такие преимущества социума как «предсказывание будущего», что может предотвратить, к примеру развитие эпидемий, на анализе ключевых слов-симптомов болезней, улучшить логистику движения транспорта, или предложить интересную музыку.

Большинство компьютеров исследователей не справляются с большим потоком информации. Размер, скорость накопления, и разнообразие данных превышает все современные возможные технические средства, разработанный web-сервис помогает исследователям решить проблему, без улучшения своих технических средств.

С помощью социальных сетей возможно стало проводить различные онлайн-эксперименты и эффекты появления новых технологий или тавров в социуме, или какое общественное мнение сформировано о происходящих событиях. Следовательно, разработка является актуальной для социума.

3 ТЕХНИКО-ЭКОНОМИЧЕСКОЕ ОБОСНОВАНИЕ РАЗРАБОТКИ

3.1 Описание целесообразности проектирования с точки зрения объекта

Внедрённый web-сервис позволит получать информации из внешних источников автоматизировано, так как основные функции web-сервиса сбор, хранение и предоставление гетерогенных данных, полученных из разных источников, следовательно, сократиться время получения данных, расширится возможность проведения различных аналитических действий с получаемой информацией.

3.2 Расчет затрат на разработку

Что бы рассчитать затраты на необходимые работы, для реализации разработки, определим продолжительность работ. Продолжительность работ можно определить, или по нормативам, или рассчитать среднее время продолжительности работ по формуле (3.1):

$$T_0 = \frac{3T_{min} + 2T_{max}}{5} \quad (3.1)$$

T_0 – ожидаемая длительность работ,

T_{min} – наименьшая длительность работ,

T_{max} – наибольшая длительность работ.

Совершим расчет по приведённым данным, расчеты ожидаемой длительности работ до разработки приведены в таблице 3.1, и расчеты ожидаемой длительности работ после разработки приведены в таблице 3.2.

Таблица 3.1 – Ожидаемая длительность работ до внедрения разработки

| Наименование работы | Длительность работы (час) | | |
|------------------------------|---------------------------|-----|-----------|
| | min | max | ожидаемая |
| поиск необходимой информации | 4 | 6 | 4,8 |
| получение информации | 3 | 5 | 3,8 |
| сохранение информации | 2 | 5 | 3,2 |
| Итого: | | | 11,8 |

Таблица 3.2 – Ожидаемая длительность работ после внедрения разработки

| Наименование работы | Длительность работы (час) | | |
|------------------------------|---------------------------|-----|-----------|
| | min | max | ожидаемая |
| поиск необходимой информации | 3 | 5 | 3,8 |
| получение информации | 0,5 | 1 | 0,7 |
| сохранение информации | 0,8 | 1 | 0,88 |
| Итого: | | | 5,38 |

Время доступа к нужным данным сократилось, из-за использования новой разработки, следовательно получаем, что ожидаемое время работы тоже сократилось, можно заметить, что почти в половину раз меньше, чем было до внедрения разработки.

Рассчитаем экономическую эффективность разработки, для этого рассчитаем сначала затраты на различные выплаты сотрудникам. Расходы на различные виды выплат сотрудникам рассчитываются по формуле (3.2):

$$Z = \sum_i \left(n_i \tilde{z}_i \left(1 + \frac{a_c}{100} \right) \right) \quad (3.2)$$

Z – суммарный расход на выплаты,

n_i – численность персонала,

\tilde{z}_i – среднегодовая заработная плата работника,

a_c – процент отчислений на социальное страхование, пенсионный фонд и фонд стабилизации.

В объекте задействованы следующие сотрудники:

- Ответственный за хранение данных: количество 1; среднегодовая заработная плата работника 360000; процент отчислений 72000,
- Ответственный за передачу данных на объект: количество 1; среднегодовая заработная плата работника 400000; процент отчислений 80000.

Совершим расчет по приведённым данным:

$$\begin{aligned} Z &= \sum_i \left(n_i \tilde{z}_i \left(1 + \frac{a_c}{100} \right) \right) = \\ &= \left(1 \cdot 360000 \left(1 + \frac{72000}{100} \right) \right) + \left(1 \cdot 400000 \left(1 + \frac{80000}{100} \right) \right) = \\ &= 320400000 (\text{рублей}) \end{aligned}$$

Теперь рассчитаем стоимость разработки, которая вычисляется по формуле(3.5):

$$S_p = Z + P \quad (3.3)$$

S_p – стоимость разработки,

P – прибыль.

Прибыль составит 42% от выплат работникам. Рассчитаем стоимость разработки:

$$P = 0,42 \cdot Z = 134570000 (\text{рублей})$$

$$S_p = Z + P = 320400000 + 134570000 = 454970000 (\text{рублей})$$

Далее рассчитаем рентабельность затрат на проект, воспользовавшись формулой 3.5:

$$R = \frac{P - P \cdot 0,2}{Z} \cdot 100\% \quad (3.4)$$

R – рентабельность затрат.

Рассчитываем рентабельность затрат:

$$R = \frac{P - P \cdot 0,2}{Z} \cdot 100\% = \frac{134570000 - 134570000 \cdot 0,2}{320400000} \cdot 100\% = 34\%$$

Наконец по рассчитанным данным можем показать результат экономического эффекта от усовершенствования объекта. Примем за E_n (норма

рентабельности) 25% . Должно соблюдаться условие что $R > E_n$, получаем $34\% > 25\%$ из этого следует, что усовершенствование экономически эффективно.

Рассчитаем срок окупаемости по формуле 3.5:

$$T_{per} = \frac{1}{R} \quad (3.5)$$

T_{per} – срок окупаемости.

Расчитываем значения срока окупаемости:

$$T_{per} = \frac{1}{R} = \frac{1}{0,34} = 2,9 \text{ (лет)}$$

Следовательно приблизительное время окупаемости:

$$2,9 \cdot 12 = 108 \text{ (месяцев)}$$

4 БЕЗОПАСНОСТЬ И ЭКОЛОГИЧНОСТЬ РАЗРАБОТКИ

Что бы эффективность труда была высокой, нужно учитывать условия работы пользователя web-сервиса. В основном его работа это интеллектуальная деятельность, поэтому для оценки условий труда нужно использовать показатели напряженности трудового процесса.

Напряженность труда пользователя оценивается методом анализа трудовой деятельности и её структуры, посредством наблюдения всего рабочего дня в течение определённого времени, к примеру на протяжении одной недели. За это время анализируются все производственные факторы, как благоприятные, так и неблагоприятные, которые происходят в процессе деятельности. Полученные показатели трудового процесса имеют качественную и количественную выраженность, а также делятся на виды нагрузок, таких как интеллектуальные, эмоциональные режимные, монотонные, сенсорные.

По степени опасности и вредоносности условия труда делятся на четыре класса:

- оптимальные (1 класс) — сохраняется здоровье работника, поддержание высокой степени работоспособности,
- допустимые (2 класс) – вредные факторы не превышают допустимые значения. В течение регламентированного отдыха организм полностью восстанавливается,
- вредные (3 класс) – вредные факторы превышают допустимые значения, и вызывают негативные воздействия на организм,
- опасные (4 класс) – воздействие вредных факторов создает угрозу для жизни, и высок риск развития профессиональных поражений организма.

Проведем оценку напряженности трудового процесса: интеллектуальные нагрузки в таблице 4.1, сенсорные нагрузки в таблице 4.2, эмоциональные нагрузки в таблице 4.3, монотонность нагрузок в таблице 4.4, режим работы в таблице 4.5

Таблица 4.1 – Оценка интеллектуальной нагрузки

| Фактор | Заключение | Оценка |
|---|---|--------|
| содержание работы | решении задач, с единоличным принятием решения | 2 |
| восприятие информации их оценка | восприятие информации с последующей коррекцией действий и выполняемых операций | 2 |
| распределение функций по степени сложности задания | выполнение конкретного задания осуществляется с обработкой, и выполнением с последующей проверкой выполнения задания | 2 |
| характер выполняемой работы | работа протекает по строго установленному графику с возможной его коррекцией по мере необходимости | 2 |

Таблица 4.2 – Оценка сенсорной нагрузки

| Фактор | Заклучение | Оценка |
|---|--|--------|
| длительность сосредоточенности наблюдения | 26-50% | 2 |
| плотность сигналов и сообщений за 1 час работы | меньше 100 | 1 |
| число производственных объектов одновременного наблюдения | до 5 объектов | 1 |
| размер объекта различения при длительности сосредоточенного внимания (% от времени смены) | 5-1,1 мм 50% | 2 |
| Работа с оптическими приборами при длительности сосредоточенного наблюдения (% от времени смены) | Отсутствует | 1 |
| Наблюдение за экраном видеотерминала (ч в смену) | 4-6 часов | 3,1 |
| Нагрузка на слуховой анализатор | помех нет, разборчивость слов равна 100% | 1 |
| Нагрузка на голосовой аппарат (суммарное количество часов наговариваемых в неделю) | меньше 16 часов | 1 |

Таблица 4.3 – Оценка эмоциональной нагрузки

| Фактор | Заключение | Оценка |
|--|---|--------|
| степень ответственности за результат собственной деятельности. Значимость ошибки | случае допущенной ошибки дополнительные усилия только со стороны самого работника | 1 |
| степень риска для собственной жизни | отсутствует | 1 |
| степень ответственности за безопасность других лиц | отсутствует | 1 |

Таблица 4.4 – Оценка монотонности нагрузки

| Фактор | Заключение | Оценка |
|--|------------|--------|
| число элементов, необходимых для реализации простого задания или многократно повторяющихся операций | 5-10 | 2 |
| продолжительность (с) выполнения простых производственных заданий или повторяющихся операций | 5-100с | 2 |
| время активных действий (в % к продолжительности смены) | больше 25% | 1 |
| монотонность производственной обстановки (время пассивного наблюдения за ходом техпроцесса в % от времени смены) | меньше 70% | 1 |

Таблица 4.5 – Оценка нагрузки режима работы

| Фактор | Заключение | Оценка |
|--|--|--------|
| фактическая продолжительность рабочего дня | 7-8 ч | 1 |
| сменность работы | односменная работа (дневная) | 1 |
| наличие регламентированных перерывов и их продолжительность (без обеденного перерыва) | перерывы регламентированы, больше 7% рабочего времени | 1 |

Посчитаем количество показателей в каждом классе результат приведен в таблице 4.6

Таблица 4.6 – Оценка нагрузки режима работы

| Показатели | Класс условий труда | | | | |
|--|---------------------|---|-----|-----|-----|
| | 1 | 2 | 3,1 | 3,2 | 3,3 |
| количество показателей в каждом классе | 16 | 8 | 1 | 0 | 0 |

От 1 до 5 показателей отнесены к 3.1 и/или 3.2 степеням вредности, а остальные показатели имеют оценку 1-го и/или 2-го классов, следовательно устанавливается допустимый (2 класс). Сотруднику необходимо соблюдать профилактические мероприятия, чтобы предотвратить негативные последствия профессиональной деятельности

Сотрудник работает мало подвижно, что влечет за собой риск возникновения синдрома длительных статических нагрузок (СДСН).

Форма СДСН может возникать из-за ухудшения кровообращения, чтобы их предотвратить, или хотя бы уменьшить появление, нужно правильно организовать рабочее место, также это поможет снизить развитие искривления позвоночника.

Опишем организованное рабочее место. Высота рабочей поверхности стола должна регулироваться в пределах 680 - 800 мм; при отсутствии такой возможности высота рабочей поверхности стола должна составлять 725мм. Рабочий стул должен быть регулируемым по высоте и углам наклона 79 сиденья и спинки, а также расстоянию спинки от переднего края сиденья. Регулировку высоты поверхности сиденья в пределах 400-550 мм и углам наклона вперед до 15 градусов, и назад до 5 градусов; высоту опорной поверхности спинки 300 мм, ширину не менее 380 мм; стационарные или съемные подлокотники длиной не менее 250 мм, шириной 50-70 мм. Клавиатуру следует располагать на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю или на специальной, регулируемой по высоте

Продолжительна работа за компьютером быстро утомляются глаза и развивается синдром сухого глаза. Это происходит, потому что испаряемость слезы из-за излучений от монитора повышается, и человек реже моргает, смотря в монитор. Для предотвращения ухудшения зрения, при работе за монитором следует придерживаться некоторых рекомендаций:

- дисплей не должен быть повернут экраном в сторону окна. В случае его расположения возле окна, необходимо расположить его перпендикулярно стеклу,
- свет от осветительных ламп не должен падать на дисплей с углом более 60 градусов от вертикали,
- освещенность рабочего места необходимо поддерживать в пределах порядка 170-250 Лк,
- при освещении рабочего места неприемлемо использование мигающих источников света,
- интерьер, на фоне которого установлен дисплей, должен быть неярким, не бросающимся в глаза. Соотношение яркости экрана и окружения не должно превышать 3: 1,
- необходимы перерывы через каждый час работы - 5-10 минут. Допускается непрерывная работы с компьютером 2 часа,
- использование специальных очков для работы за компьютером с прогрессивными линзами.

Web-сервис на прямую не может принести экологическую угрозу окружающей среде. При использовании разработки происходит сбережение человеческих и временных ресурсов. Используемая компьютерная техника должна придерживаться стандарта RoHS.

Что бы обеспечить электробезопасность должны соблюдаться следующие меры, электрическое разделение сетей, обеспечение недоступности токоведущих частей, зануление, защитное заземление, контроль и профилактика поврежденной изоляции.

Оператору требуется быть психологически подготовленным, к необходимости быть сконцентрированным на протяжении длительного наблюдения за работой перед монитором, во избежание ошибок пользователя.

ЗАКЛЮЧЕНИЕ

В ходе выполнения выпускной квалификационной работы был проведен анализ бизнес-процессов, построена и оптимизирована математическая модель, была обоснована необходимость создания web-сервиса.

Разработан web-сервис.

Внедрение web-сервиса увеличит эффективность работы в системе, ускорит обмен информацией, из-за снижения времени на обработку данных, аналитические возможности, для принятия решений.

В работе приведены технико-экономическое обоснование разработки и подтверждены безопасность и экологичность использования разработки.

Таким образом, поставленные цели и задачи в выпускной квалификационной работе, выполненной мной самостоятельно и с соблюдением правил профессиональной этики, были достигнуты и решены в полном объеме.

Затраты на разработку включают в себя оплату за энергопотребление, и временной ресурс разработчика.

В качестве дальнейшего направления исследования можно рассматривать развитие исследований в области big data и методов когнитивной визуализации, для улучшения алгоритмов анализа гетерогенных данных.