

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт инженерной физики и радиоэлектроники
Кафедра экспериментальной физики и инновационных технологий

УТВЕРЖДАЮ
Заведующий кафедрой
_____ Орлов В.А.
подпись
«__» _____ 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Внедрение технологии больших данных (Big Data) в сферу
здравоохранения Красноярского края

27.04.05 Инноватика
27.04.05.01 Управление инновациями

Научный руководитель	_____	канд. техн. н., доцент	А.В.Вершков
	подпись, дата	должность, ученая степень	инициалы, фамилия
Выпускник	_____		М.К.Кыргызбай
	подпись, дата		инициалы, фамилия
Рецензент	_____	канд. техн. н., доцент	А.В.Селиванов
	подпись, дата	должность, ученая степень	инициалы, фамилия

Красноярск 2020

СОДЕРЖАНИЕ

Введение	3
1 Технология больших данных	4
1.1 Появление технологии больших данных	5
1.2 Хранение и обработка больших данных	8
1.3 Применение технологии больших данных в частности в сфере здравоохранения.....	12
2 Логико-структурный подход при внедрении технологии больших данных в сферу здравоохранения.....	20
2.1 Анализ заинтересованных сторон	21
2.2 Дерево проблем и дерево целей.....	26
2.3 Анализ процессов при внедрении технологии больших данных в сферу здравоохранения.....	29
2.4 Система сбалансированных показателей	34
3 Внедрение технологии больших данных (Big Data) в сферу здравоохранения Красноярского края.....	40
3.1 Применение технологии больших данных для Регионального архива медицинских изображений.....	42
3.2 Моделирование обработки сервером данных с помощью Big Data	47
3.3 Экономическое обоснование внедрения технологии больших данных в сферу здравоохранения.....	54
Заключение	60
Список использованных источников	62

ВВЕДЕНИЕ

Внедрение современных информационных технологий в медицину становится одной из перспективных направлений последнего десятилетия. Одной из таких технологий является технология «больших данных». Большие данные в сфере медицины позволят вывести ее на новый уровень. С их помощью в медицине открываются все новые возможности для лечения пациентов с использованием современных и точных методов. Технология «больших данных» дает возможность обеспечить индивидуальный подход к лечению пациента. А именно, устройства, использующие технологии Big Data, выступают в качестве персонального помощника врача. На основе анализа данных о передовых научных исследованиях в соответствующей сфере, имеющимся опыте лечения различных заболеваний, клинических исследований, отдельных лекарственных препаратов, может быть разработан индивидуальный план лечения, учитывающий особенности организма пациента. Активное применение цифровых технологий в медицинских целях порождают большое количество данных о состоянии пациента, которые можно применить не только для отдельной диагностики узко направленным специалистом, но и для интеграционного использования данных о состоянии здоровья. [1]

Надобность «внедрения технологий масштабирования баз знаний и внедрения систем поддержки принятия врачебных решений в повседневную деятельность» была отмечена в государственной программе РФ «Развитие здравоохранения» 2014 года [2], а также национальной программы «Цифровая экономика Российской Федерации» от 24 декабря 2018 года.

На данный момент автоматизация в медицинских учреждениях локализована, например, данные медицинских карт, анализы, рецепты и выписки, даже переведенные в электронную форму не помогают осуществлять такие объемные работы с данными, а лишь облегчает работу административного персонала. Поэтому актуальность настоящей работы не вызывает сомнений.

Задачи поставленные в ходе выполнения работы:

- исследовать появление, хранение, обработку и применение технологии больших данных;
- рассмотреть применение логико-структурного подхода при внедрении технологии больших данных в сферу здравоохранения;
- рассмотреть применение больших данных в сфере здравоохранения;
- рассмотреть применение технологии больших данных для Регионального архива медицинских изображений;
- финансовое обоснование применения технологии больших данных.

Объектом исследования является большие данные в сфере медицины.

Предметом исследования является применение технологий больших данных для обработки медицинских данных.

1 Технология больших данных

Определение Большие данные (Big Data) обычно расшифровывают довольно просто – это огромный объем информации, часто бессистемной, которая хранится на каком-либо цифровом носителе. Однако массив данных с приставкой «Биг» настолько велик, что привычными средствами структурирования и аналитики «перелопатить» его невозможно. Поэтому под термином «Биг Дата» понимают ещё и технологии поиска, обработки и применения неструктурированной информации в больших объемах.

Большие данные и компьютерные технологии для их анализа названы одной из 10 лучших революций в ближайшее десятилетие. Предусматривается, что его влияние аналогично влиянию Интернета, облака и, в последнее время, цепочечных цепочек (известных в криптовалютах как биткойн).

Большие данные обычно определяют с точки зрения проблем управления данными, которые не удастся решить в рамках традиционных баз данных в силу объема, разнообразия данных и требований к скорости.

Большие данные открывают новые перспективы как для крупных, так и малых компаний по выявлению новых возможностей улучшения сервиса для клиентов, повышения эффективности различных процессов, выявления прежде незамеченных типовых шаблонов и т.д. [3]

Для того чтобы использовать все преимущества технологии Big Data, необходимо не только аналитическое программное обеспечение, но и соответствующая инфраструктура для хранения, обработки и передачи данных.

Из-за огромных объемов данных особое значение имеет инфраструктура, позволяющая хранить и обрабатывать эти данные. Большая часть организаций в мире не имеют необходимых технологических ресурсов необходимой для работы с такими объемами данных. Проблема заключается не только в отсутствии системы хранения данных и дисков для хранения, проблема в том, что не хватает мощного оборудования для работы с данными, а именно интеграция всей современной системы хранения данных.

Цифровая экономика начинается с данных, данные нужны во всех сферах — этот постулат никем не оспаривается. У каждого типа данных своя сфера и особенности регулирования, и на практике трудно разобраться, по каким правилам эти данные можно использовать.

Если говорить о таком модном сегодня типе, как большие данные, то они никак не урегулированы. У всех, кто утверждает, что работает с большими данными, при ближайшем рассмотрении обнаруживается целый букет персональных данных. Например, при распознавании лиц в метро. То, что в кадре много лиц, из которых выхватываются, например, лица преступников, может и делает эти данные большими, но совершенно точно не дает права обрабатывать их абсолютно без правил. В каждом конкретном случае надо разбираться, какие данные используются.

Ждать того, что оборот именно больших данных будет урегулирован, не приходится. Вся мировая практика исходит из этого: никто не регулирует большие данные и не дает им определение. По сути, большие данные — это

данные, персональные или не персональные, обезличенные или не обезличенные, собранные в большой массив.

Хотя многие люди и компании используют слово «большие данные», они могут не всегда означать одно и то же или интерпретировать их одинаково. У большинства из нас есть смутное представление о том, что это может быть («все, что не подходит к листу Excel»), но большие данные - это не просто синоним «большого количества данных».

1.1 Появление технологии больших данных

Термин «большие данные» появился в 2008 году благодаря Клиффорду Линчу. В спецвыпуске журнала Nature эксперт назвал взрывной рост потоков информации - Big Data. В него он отнес любые массивы неоднородных данных свыше 150 Гб в сутки. Из статистических выкладок аналитических агентств в 2005 году мир оперировал 4-5 эксабайтами информации (4-5 миллиардов гигабайтов), через 5 лет объемы Big Data выросли до 0,19 зеттабайт (1 ЗБ = 1024 ЭБ). В 2012 году показатели возросли до 1,8 ЗБ, а в 2015 – до 7 ЗБ. Эксперты прогнозируют, что к 2020 году системы больших данных будут оперировать 42-45 зеттабайтов информации. [4]

До 2011 года технологии больших данных рассматривались только в качестве научного анализа и практического выхода ни имели. Однако объемы данных росли по экспоненте, и проблема огромных массивов неструктурированной и неоднородной информации стала актуальной уже в начале 2012 году. Всплеск интереса к Big Data хорошо виден в Google Trends (рисунок 1).

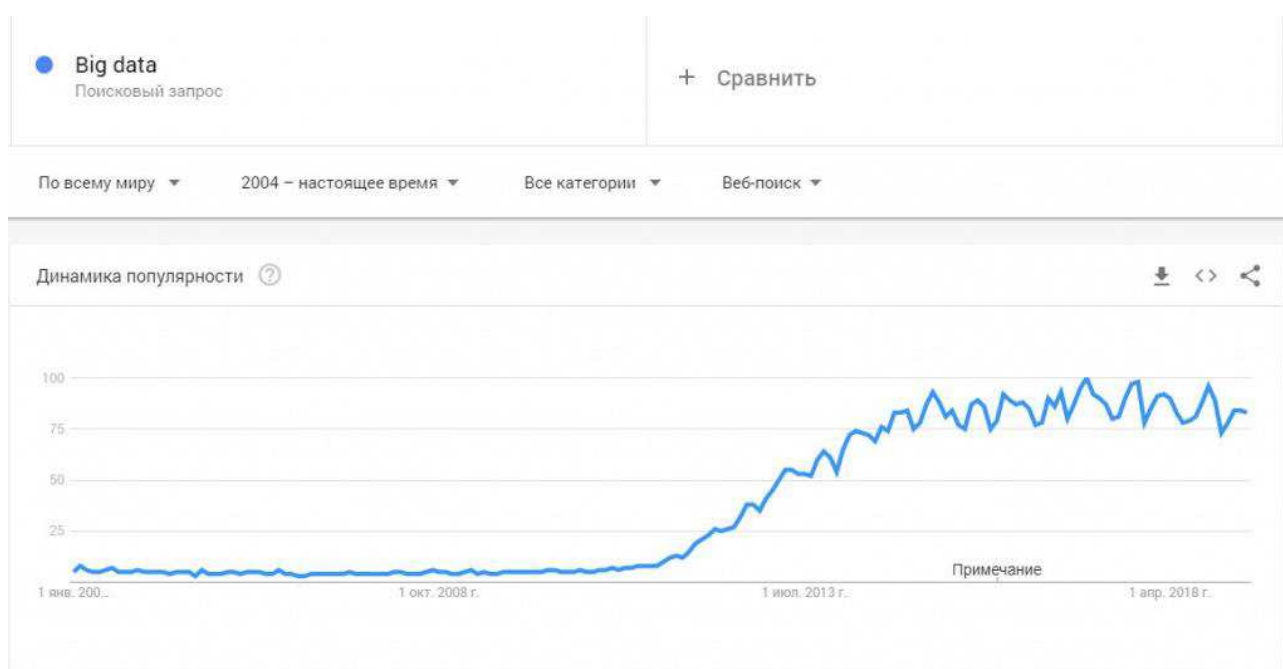


Рисунок 1 - Динамика популярности Big Data

К развитию нового направления подключились мастодонты цифрового бизнеса – Microsoft, IBM, Oracle, EMC и другие. С 2014 года большие данные изучают в университетах, внедряют в прикладные науки – инженерию, физику, социологию.

Чтобы массив информации обозначить приставкой «биг» он должен обладать следующими признаками, правило трех «V»:

- объем (Volume) – данные измеряются по физической величине и занимаемому пространству на цифровом носителе. К «Big» относят массивы свыше 150 Гб в сутки;

- скорость, обновление (Velocity) – информация обновляется минуту, секунду и для обработки в реальном времени необходимы специальные алгоритмы и технологии;

- разнообразие (Variety) – информация в массивах имеет неоднородные форматы, т.е. не структурированы, структурированы частично, полностью и скапливаться бессистемно. Например, социальные сети используют большие данные в виде текстов, видео, аудио, финансовых транзакций, картинок и прочего.

В современных системах рассматриваются два дополнительных фактора:

- изменчивость (Variability) – потоки данных могут иметь пики и спады, сезонности, периодичность. Всплески неструктурированной информации сложны в управлении, требует мощных технологий обработки;

- значение данных (Value) – информация может иметь разную сложность для восприятия и переработки, что затрудняет работу интеллектуальным системам. Например, массив сообщений из соцсетей – это один уровень данных, а транзакционные операции – другой. Задача машин определить степень важности поступающей информации, чтобы быстро структурировать. [5]

Большие данные полезны в различных областях. Отслеживание местоположения помогает логистическим компаниям снизить риски при транспортировке, скорости и надежности доставки. В финансовом мире Комиссия по ценным бумагам (SEC) использует аналитику больших данных для выявления возможных случаев мошенничества. Развлекательные компании, такие как Netflix и YouTube, используют прошлые просмотры и онлайн-поведение, чтобы увеличить вовлеченность и увеличить доходы. Рекламные компании, вероятно, являются крупнейшими игроками больших данных. Данные, проанализированные из Facebook, Twitter и Google, отслеживают поведение и транзакции, которые рекламодатели используют для запуска целевых кампаний. Селекционные компании используют беспилотники, летящие над полями, посылая данные изображений для информирования о процессе размножения. С большими данными, больницы могут улучшить мониторинг пациентов интенсивной терапии. Эффективность (дорогих) лекарств может быть измерена, и эпидемические вспышки могут быть предсказаны на ранней стадии.

Чем больше мы знаем о конкретном предмете или явлении, тем точнее постигаем суть и можем прогнозировать будущее. Снимая и обрабатывая потоки данных с датчиков, интернета, транзакционных операций, компании могут довольно точно предсказать спрос на продукцию, а службы чрезвычайных ситуаций предотвратить техногенные катастрофы. Приведем несколько примеров вне сферы бизнеса и маркетинга, как используются технологии больших данных:

- здравоохранение – больше знаний о болезнях, вариантов лечения, информации о лекарственных препаратах, все что позволит бороться с неизлечимыми болезнями прошлых и будущих столетий;

- предупреждение природных и техногенных катастроф – максимально приближенный к реальной угрозе прогноз, который спасает тысячи жизней людей. Задача стоит в следующем: собрать и обработать множество показаний датчиков и на их основе помочь определить максимально точное время и место возможного катаклизма;

- правоохранные органы – большие данные используются для прогнозирования всплеска криминала, терроризма в разных странах и принятия сдерживающих мер.

К основным способам анализа больших массивов информации относят следующие:

- глубинный анализ, классификация данных. Эти методики пришли из технологий работы с обычной структурированной информацией в небольших массивах. Однако в новых условиях используются усовершенствованные математические алгоритмы, основанные на достижениях в цифровой сфере;

- краудсорсинг. В основе этой технологии возможность получать и обрабатывать потоки в миллиарды байт из множества источников. Конечное число «поставщиков» не ограничивается ничем. Разве только мощностью системы;

- сплит-тестирование. Из массива выбираются несколько элементов, которые сравниваются между собой поочередно «до» и «после» изменения. A/B тесты помогают определить, какие факторы оказывают наибольшее влияние на элементы. Например, с помощью сплит-тестирования можно провести огромное количество итераций постепенно приближаясь к достоверному результату;

- прогнозирование. Аналитики стараются заранее задать системе те или иные параметры и в дальнейшей проверять поведение объекта на основе поступления больших массивов информации;

- машинное обучение. Искусственный интеллект в перспективе способен поглощать и обрабатывать большие объемы несистематизированных данных, впоследствии используя их для самостоятельного обучения;

- анализ сетевой активности. Методики Big data используются для исследования соцсетей, взаимоотношений между владельцами аккаунтов, групп, сообществами. На основе этого создаются целевые аудитории по интересам, геолокации, возрасту и прочим метрикам. [6]

В 2019 году важность понимания и главной работы с массивами информации возросла в 4-5 раз по сравнению с началом десятилетия. С массовостью пришла интеграция Big Data в сферы малого и среднего бизнеса, стартапы:

- облачные хранилища – технологии хранения и работы с данными в онлайн-пространстве для решения массы проблем малого и среднего бизнеса: дешевле купить услуги облачного провайдера, чем содержать дата-центр, персонал может работать с помощью удаленного рабочего стола, организуя себе хоум-офис;

- глубокое (глубинное) обучение, искусственный интеллект – аналитические машины имитируют человеческий мозг, то есть используются искусственные нейронные сети, где обучение происходит самостоятельно на основе больших массивов информации;

- Dark Data – сбор и хранение не оцифрованных данных о компании, не имеющие значимой роли для развития бизнеса, но необходимые в техническом и законодательном планах;

- блокчейн – упрощение интернет-транзакций, снижение затрат на проведение этих операций;

- системы самообслуживания – платформы для малого и среднего бизнеса для самостоятельного хранения и систематизирования данных (внедрены в 2016 году).

Несмотря на всю доступную информацию, многие организации не осознают, что сталкиваются с проблемой больших данных, или просто не готовы мыслить такими категориями. Организация может получить преимущества от применения технологий больших данных, если ее существующие приложения и базы данных больше не способны масштабироваться и справляться с внезапными увеличениями объема или разнообразия данных, или требований к скорости их обработки.

Если вовремя не найти правильного подхода к работе с большими данными, это может привести к повышению расходов, а также снижению эффективности работы и конкурентоспособности. И напротив, разумная стратегия по работе с большими данными может помочь организации сократить расходы и получить дополнительные преимущества при работе за счет осуществления текущих крупных рабочих нагрузок с помощью технологий больших данных, а также развертывания новых приложений для выгодного использования открывшихся возможностей.

1.2 Хранение и обработка больших данных

Технологии Big Data включают новые инструменты для всех этапов цикла обработки данных, использование которых доступно с точки зрения финансов и технических возможностей. С их помощью можно решать вопросы сбора и хранения больших массивов данных, выполнять их обработку для получения в дальнейшем ценной аналитической информации. В большинстве случаев работа

с большими данными подразумевает сбор необработанных данных и получения пригодной для использования информации.

Сбор необработанных данных (операций, записей звонков, история посещения сайтов, событий мобильных устройств и прочее) – это первая камень преткновения при работе с большими данными. Качественная платформа для работы с большими данными предоставляет разработчикам возможность сбора самых разнообразных данных, структурированных и нет, на любой скорости, от режима реального времени до пакетной обработки.

Каждая платформа для работы с Big Data обязана включать надежный, безопасный и масштабируемый репозиторий для хранения данных, до обработки и после. В зависимости от определенных требований возникает необходимость во временных хранилищах, а также резервных копий для перемещаемых данных.

При обработке и анализе выполняется неструктурированные данные структурируются и приводятся в пригодный для дальнейшего использования массив данных. Это осуществляется с помощью агрегации, сортировки, объединения или применения специальных расширенных функций и алгоритмов. Впоследствии сего итоговые пакеты данных сберегаются для последующей обработки, но бывают случаи, когда данные отправляются для бизнес-аналитики или визуализации.

Главная задача при работе с Big Data является получение на их основании ценных аналитических выводов для практического использования. На сегодняшний день, к сожалению, не для всех большие данные являются доступными для всех заинтересованных в применении данной технологии для собственных нужд в бизнес-аналитике. Это дало бы возможность пользователям технологии вне зависимости от типа аналитики получать готовые результаты в форме прогнозов или рекомендованных действий.

Технологии для работы с большими массивами данных продолжают активно развиваться. Уже сегодня у организаций есть выбор между различными типами аналитики для реализации различных функций.

Описательный анализ помогает пользователям ответить на вопрос: «Что произошло и почему?» В качестве примера можно привести традиционную среду для запросов и отчетов с панелями управления и системами оценок.

Прогнозирующий анализ позволяет пользователям оценить вероятность тех или иных событий в будущем. В качестве примеров можно привести системы прогнозирования, заблаговременных предупреждений и обнаружения мошенничества, а также приложения для профилактического обслуживания.

Предписывающий анализ формирует для пользователя определенные рекомендации (предписания). Он помогает ответить на вопрос: «Что делать, если произойдет событие X?».

Изначально инфраструктуры по работе с большими данными, например, Hadoop, поддерживали только пакетные рабочие нагрузки. Крупные пакеты данных загружались для обработки сразу, и процесс ожидания результатов растягивался на часы и даже дни. Но время ожидания результата постепенно стало критическим фактором, и требуемая скорость обработки больших данных послужила толчком к развитию таких новых инфраструктур, как Apache Spark,

Apache Kafka, Amazon Kinesis и т. д., способных поддерживать обработку потоковых данных в режиме реального времени.

Рассмотрим, как же именно происходит обработка и хранение данных. Большие данные хранятся как реестр данных в виде библиотеки со связанными элементами (рисунок 2).

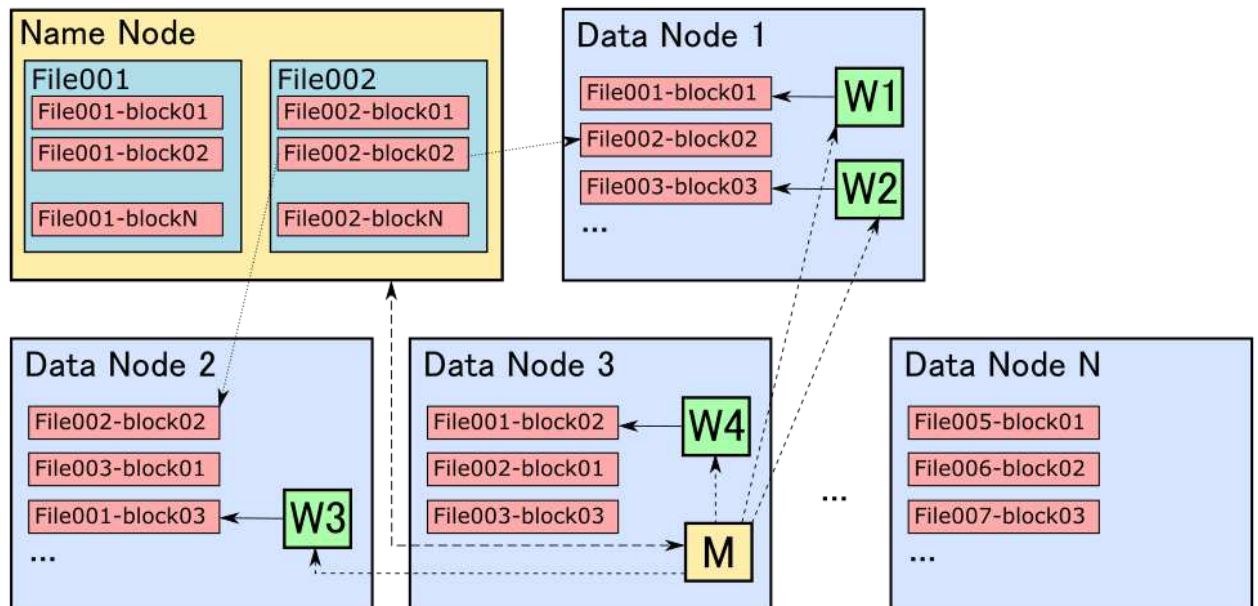


Рисунок 2 - Хранение нужных блоков данных

Каждая библиотека имеет свои полки в которых хранится обработанные данные (рисунок 3).

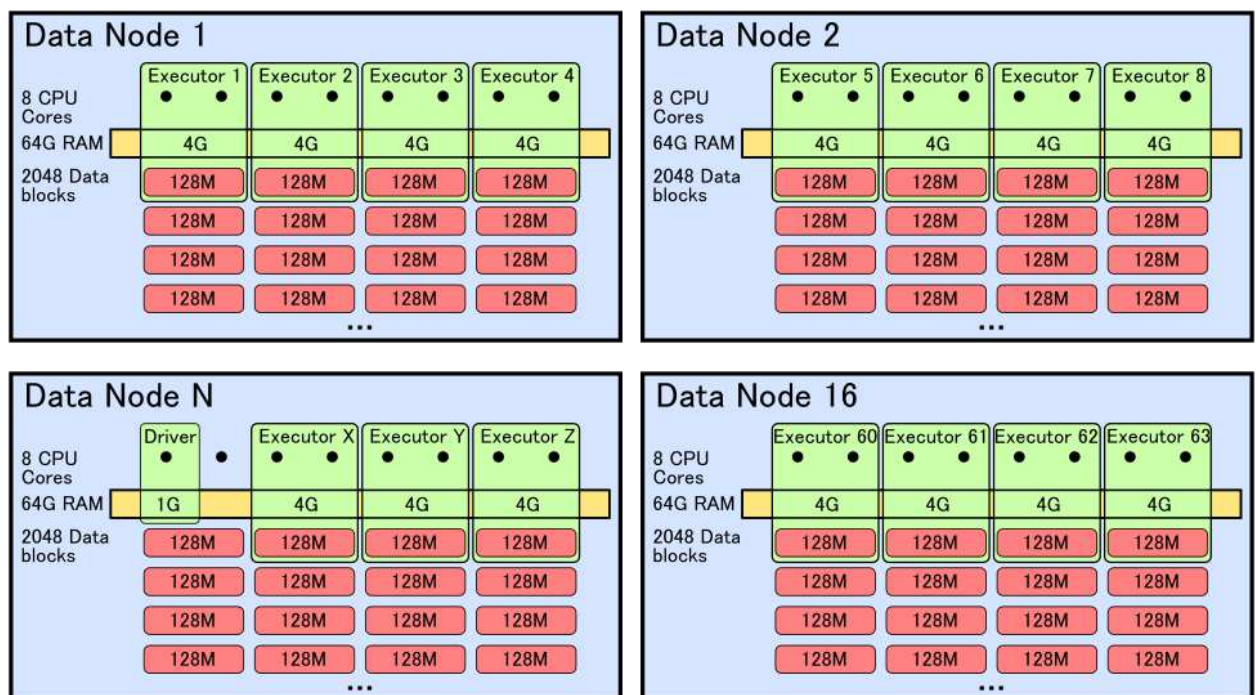


Рисунок 3 - Контейнеры или полки данных

И в каждой поле так же есть определённые ячейки заполненные данными. Большими данными считаются эти библиотеки с множеством единиц данных, которые не просто хранятся радом, но и связаны определёнными элементами и зависят друг от друга. [7]

Горизонтальная масштабируемость, обеспечивающая детальную обработку данных распределённых на сотни и тысячи вычислительных узлов без деградации производительности, принимают как базисный принцип при обработке Big Data, так как он более понятен и не сложен в применении. Из-за своей простоты в понимании специалистами данный принцип был включен в определение Big data. На этом методы и техники анализа не заканчиваются. Существует еще множество методов, которые применяются для больших данных, например:

- метод Data Mining: регрессионный анализ, ARL (обучение ассоциативным правилам), кластерный анализ, классификация;
- краудсорсинг;
- DFI (смешение и интеграция данных) – техника, позволяющая интегрировать разнородные данные из разнообразных источников для возможности глубинного анализа;
- машинное обучение, включая обучение с учителем и без учителя, а также Ensemble learning — использование моделей, построенных на базе статистического анализа или машинного обучения для получения комплексных прогнозов на основе базовых моделей;
- нейронные сети: сетевой анализ, оптимизация, генетические алгоритмы;
- распознавание образов – метод классификации и идентификации предметов, явлений, процессов, сигналов, ситуаций и прочих объектов;
- прогнозная аналитика; [8]
- имитационное моделирование: агентное, математическое, событийное, дискретно-событийное, системная динамика и многие другие виды моделирования применяются для больших данных, в зависимости от вида данных и связями между ними.;
- пространственный анализ;
- статистический анализ: A/B-тестирование, анализ временных рядов;
- визуализация аналитических данных — представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации как для получения результатов, так и для использования в качестве исходных данных для дальнейшего анализа.

В большей степени компании пользуются готовыми аппаратно-программными комплексами, которые устанавливаются в собственный или арендованный дата-центр в виде готовых серверных шкафов с необходимым программным обеспечением.

Иногда реляционные системы Exadata, Teradata, Netezza относят к основным решениям для больших данных. Хотя данные реляционные системы являются аппаратно-программными комплексами, то есть эффективными

методами обработки данных способные обрабатывать терабайты структурированной информации. [9]

1.3 Применение технологии больших данных в частности в сфере здравоохранения

В 2012–2013 годах технологии Big Data вышли за рамки предметной области информационных технологий и стали глубже проникать в структуры управления, бизнес, промышленность и науку. Пять лет назад использование Big Data стало тенденцией. Сегодня они активно изучаются специалистами IT и обсуждаются в профильной прессе. Оно и понятно: цифровые технологии стремительно развиваются, объем информации растет, и, одновременно, расширяются возможности для использования Big Data в реальных проектах. По результатам прогнозов IDC Digital Univers, до 2021 года общий объем данных на планете составит 40 зеттабайтов, что эквивалентно 5200 Гб на каждого жителя Земли, объемы хранящихся данных будут ежегодно увеличиваться на 40%, а рынок технологий и услуг Big Data в 2021 – 68,7 миллиард долларов США. [10]

Для российских реалий, к сожалению, большие данные на сегодняшний день не такие явные как в США. Затраты на проект по региональному архиву медицинских изображений стоила в 2020 году 19,5 миллионов рублей. И это только Красноярский край.

Суть метода больших данных в том, что массивы данных рожают новые данные имеющая свои определенные принципы:

- отказоустойчивость равная 99,99 %, практически идеальной;
- горизонтальная масштабируемость, если происходит увеличение объема данных в n раз, то количество серверов увеличивается в n раз;
- данные не должны иметь задержку при обработке, то есть должны быть локальны, обрабатываться и хранится в одной серверной стойки.

Впервые большие данные были использованы в крупных интернет-компаниях, таких как Facebook, Google. После этого большие данные стали использовать маркетинговые компании, для составления, персонально ориентированного или даже таргетного рекламного продвижения товаров, что на сегодняшний день успешно осуществляется. Далее развитие больших данных вышло за рамки «простых» использований и вошла в сферу финансов и наук о жизни, то есть в сферу медицины, генетики, геномики, биоинформатики и так далее. Последние несколько лет так же пытаются внедрить в государственный сектор в России (например, Госуслуги с интеграцией всеми сервисами, которыми мы пользуемся ФНС, банки, ПФР, МВД). Все те, кто используют большие данные делятся на несколько групп: поставщики, дата-майнеры, интеграторы, потребители, разработчики. Каждая группа очень важна для осуществления бизнеса в сфере больших технологий. Поставщики инфраструктуры решают задачи хранения и предобработки данных которые используют другие компании. Дата-майнеры разрабатывают алгоритмы для извлечения ценные сведения из больших данных. Интеграторы (системные) внедряют системы анализа больших данных на стороне клиента. И конечные потребители, которые покупают

программно-аппаратные комплексы и заказывают алгоритмы у консультантов. А разработчики готовых сервисов создают удобные для использования решения на основе доступа к большим данным.

Внедрение технологий в медицину становится одной из перспективных направлений последнего десятилетия. В сфере здравоохранения так же имеются огромное количество данных, источниками которых являются:

- клинические данные для поддержки принятия решений различной специализации (диагностическая, прогностическая, с элементами искусственного интеллекта, управления, уход за больными и т.д.), в виде стандартизированных данных из электронных историй болезни;
- данные с датчиков мониторинга здоровья;
- генерируемые экспертами конкретные показатели, письменные заметки и медицинские рецепты;
- звукозаписи и визуальные образы;
- данные специализированных исследований;
- данные о лекарственных препаратах;
- данные неотложной помощи;
- административно-паспортные данные;
- данные о страховании и медицинском страховании;
- данные об опыте и результатах использования методов нетрадиционной медицины и непрофессиональных инициатив в области здравоохранения и медицины;
- нормативные и законодательные документы из области социальной медицины общественного здравоохранения, рынка здравоохранения, политики и культуры;
- данные медицинской науки. [11]

Аналитика собранных сведений поможет сделать новые выводы о применяемых методах терапии и улучшить уход за пациентами. Ниже приведены лишь некоторые из преимуществ в использовании больших данных:

- планирование медицинского обслуживания физических лиц и групп населения, включая прогностическое управление течением заболеваний;
- определение и реализация максимально эффективных практических мер, способствующих сокращению числа повторных госпитализаций;
- снижение риска заражения крови и почечной недостаточности, вмешательство на ранних стадиях для уменьшения негативных последствий;
- оптимизация управления результатами лечения и затратами на лекарственные препараты;
- разработка инструментов, позволяющих повысить качество обслуживания пациентов.

Мировым лидером по разработке и внедрению технологий больших данных в здравоохранении на сегодняшний день являются США. Развитие больших данных в США произошло из-за экономической выгоды от их применения, т.е. большие данные сохранили огромное количество средств из государственного бюджета. По мнению аналитиков McKinsey Global Institute,

использование технологий Big Data в здравоохранении США будет формировать финансовый поток объемом 300 миллиардов долларов в стоимостном выражении, из которых две трети – за счет снижения расходов системы здравоохранения США. [12] Эксперты утверждают, что сравнительно небольшие инвестиции в массовое внедрение технологий Big Data в этой области могут в короткие сроки существенно повысить уровень качества жизни людей. [13]

Например, в США на сегодняшний день существует компания, которая ежегодно обрабатывает миллионы рецептов на препараты. Одни предоставляются лично пациенту, другие реализуются через фармацевтические аптеки. Том Хенри, вице-президент отдела по направлению решений в области знаний в Express Scripts. Он возглавляет подразделение, использующее лучшие инструменты аналитики для обработки больших данных. Они настолько эффективно анализируют информацию об отдельных пациентах, что вскоре смогут уведомлять медицинский персонал о серьезных побочных эффектах препарата задолго до того, как он вообще будет прописан больному. Это приведет к важным положительным изменениям в деле оздоровления нации:

- сотрудники медицинских учреждений будут узнавать, что пациент находится в группе риска возникновения зависимости, перед тем как выписать ему рецепт на болеутоляющие препараты. В такой ситуации можно будет выбрать другой план лечения или более тщательно проконтролировать потребление лекарств;

- анализ выписанных рецептов, данных о физиологии и другой медицинской информации позволит выявлять развитие хронического заболевания или болезни, которая еще не была должным образом диагностирована;

- анализ сведений о том, как пациент соблюдает предписания врача после выписки из больницы, поможет медицинскому учреждению прогнозировать вероятность повторной госпитализации в течение следующих 90 дней и принимать соответствующие меры, чтобы ее предотвратить.

Большие данные в медицине как технология обработки и связывания информации дает возможность не только записывать информацию об истории болезни пациента, но и анализировать все данные записанные при посещении пациента медицинского учреждения ранее. Таким образом система даст доступ для лечащего врача к истории болезни и поможет в разы быстрее ставить диагнозы, не запуская заболевания. Большие данные так же могут обучать будущих врачей на основе многопараметрических моделях, определять риски возникновения заболевания или осложнений, сократить время проведения научных исследований с года до нескольких недель.

Применение технологий Big Data для анализа все более сложных массивов медицинских данных открывает новые возможности в области здравоохранения. Основные задачи, стоящие перед разработчиками технологий Big Data в медицине, определяются, главным образом, особенностями циркулирующих в современном здравоохранении и биомедицине данных. Эти данные зачастую являются непреодолимыми для обработки с помощью традиционного

программного обеспечения не только из-за их объема, но и из-за разнообразия типов данных и скорости, с которой они должны анализироваться.

Наиболее остро необходимость новых программно-технических средств, опирающихся на методы анализа больших объемов данных, наблюдается в биоинформатике и биомедицине: методы полного геномного секвенирования, исследование микробиома.

В США проект по исследованию микробиома человека «Human Microbiome Project» был запущен одновременно с известным проектом по исследованию генома человека Human Genome Project. В ходе его реализации в рамках Национальных институтов здоровья США создан специальный центр Data Analysis and Coordination Center. Реализуется совместный китайско-европейский проект MetaHit, где ведутся активные исследования в этом направлении. В России в ряде проектов по исследованию микробиома участвует Центр исследований и разработок ЕМС. [14] Так же на сегодняшний день существует огромная база на национальном уровне в Нидерландах.

Основной проблемой является, то что большие данные активно применяются в западных странах, а на территории Росси только последние несколько лет. Задачи, которые позволяют решить технологии Big Data, контроль за процессом лечения, определение наиболее эффективных методов лечения, предотвращение эпидемий. Развитию технологий Big Data в здравоохранении способствует повсеместное создание межрегиональных медицинских баз данных.

Цифровые технологии – один из приоритетов развития сферы здравоохранения во всем мире, ежегодно этот рынок увеличивается на четверть. Процесс может обеспечить прорыв в доступности и качестве услуг без роста расходов на здравоохранение. А потому развитие цифровой медицины осуществляется при активном участии государства. [15]

В Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года технологии обработки «Больших данных» обозначены в числе «прорывных для мировой индустрии, в которых в перспективе 10–15 лет с высокой вероятностью может быть обеспечена глобальная технологическая конкурентоспособность России». [16]

По мнению аналитиков IDC, Россия является крупнейшим региональным рынком решений по аналитике Big Data. Рост объемов рынка таких решений в Центральной и Восточной Европе достаточно активный, каждый год этот показатель увеличивается на 11 %. К 2022 году в России достигнет в количественном отношении 5,4 миллиардов долларов. В 2018 году выручка от продажи соответствующих решений в РФ составила 40 % от совокупного объема инвестиций в технологии обработки Big Data всего региона.

Крупнейшим региональным рынком Big Data and Analytics (BDA) названа Россия. Вместе с тем, в России прогнозируются самые низкие темпы роста совокупного среднегодового темпа роста среди стран Центральной и Восточной Европы. Согласно оценке IDC, траты на BDA в России в период с 2017 по 2022 годы будут увеличиваться в среднем на 9,7 %. [17]

Современные ИТ-технологии создают принципиально новые возможности для медицины. Внедрение в практику здравоохранения информационных технологий стремительно изменяет способы диагностики и лечения, формы взаимодействия врачей с пациентами и коллегами, организацию лечения и восстановления здоровья. Медицинские информационные системы обеспечивают цифровизацию бизнес-процессов организаций здравоохранения и сбор необходимой информации для формирования электронного паспорта здоровья.

Медицина также, как и рекламные гиганты становится все более персонализированной. Однако рекламные гиганты используют таргетированную рекламу, а медицина заинтересована в персонализированном лечении, своевременной профилактике болезней, новых открытий в области неизлечимых болезней, а также предотвращение эпидемий. Стандартные медицинские услуги не успевают за развитием болезней и не удовлетворяют потребности пациентов, которые желают не только вылечить существующие болезни, но и контролировать свое здоровье на протяжении десятилетий.

Цифровизация медицины начиналась с устройств для измерения различных биофизических параметров, которые были применены для эксперимента, а в дальнейшем уже вошли в нашу жизнь как обыденность.

Значительный объем в медицинских информационных системах занимают изображения (рисунок 4), являющиеся важным источником данных при диагностике, оценке и планировании терапии. Компьютерная томография (КТ), магнитно-резонансная томография (МРТ), рентген, молекулярная визуализация, ультразвуковое исследование, фотоакустическая томография, рентгеноскопия, позитронно-эмиссионная томография с компьютерной томографией (ПЭТ-КТ) и маммография – вот лишь некоторые из методов визуализации, применяемых в клинических условиях. Объем данных медицинских изображений может варьировать от нескольких мегабайт на одно исследование (например, гистологические изображения) до сотен мегабайт на одно исследование (например, тонкие срезы КТ, включающие до 2500+ сканирований в одном исследовании). [18]. Для того что бы не потерять и хранить всю важную информацию приведенных выше необходимо иметь не только мощную систему хранения данных, но специалистов, которые будут осуществлять резервное копирование этих данных. Алгоритмы и техника являются основными при хранении и обработке, но также имеет значение опыт системных администраторов, безопасность центра обработки данных, его отказоустойчивость и возможность противостоять взлому.

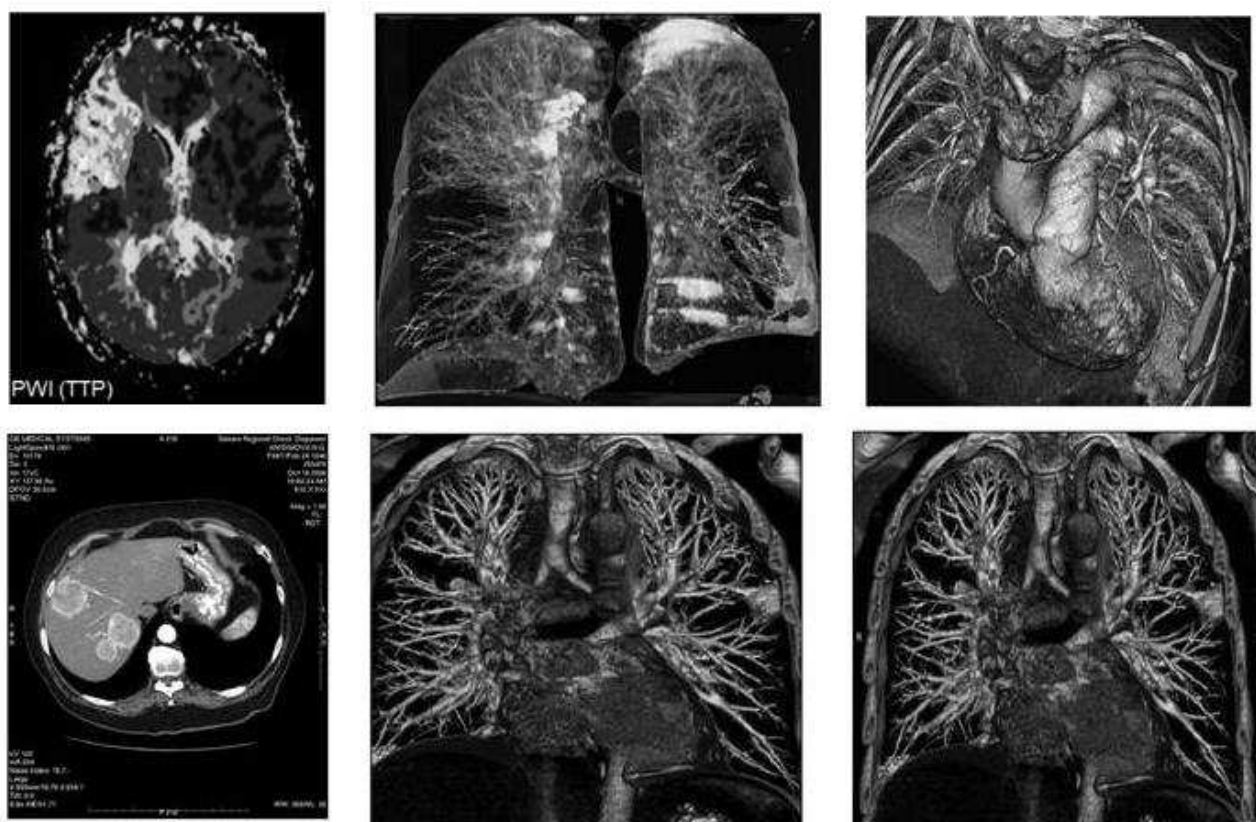


Рисунок 4 — Цифровые медицинские изображения

При применении больших данных имеют пользу не только медицинские учреждения и государство, но и страховые организации. В США такие данные помогают разоблачить преступников, которые обманывая систему страхования получают компенсационные выплаты за то, что пострадали в чрезвычайной ситуации, выявить людей, которые обманным путем получают лекарственные препараты содержащие наркотические вещества. А также это сохранит миллиарды долларов на лечение тяжело больных пациентов.

По направлению «Big Data в медицине» за последние 15 лет позволяет говорить о высокой динамике развития направления, выраженной в экспоненциальном росте числа предлагаемых технологических решений, начиная с 2007 года (рисунок 5), что коррелирует с тенденциями увеличения мирового спроса на технологии Big Data.

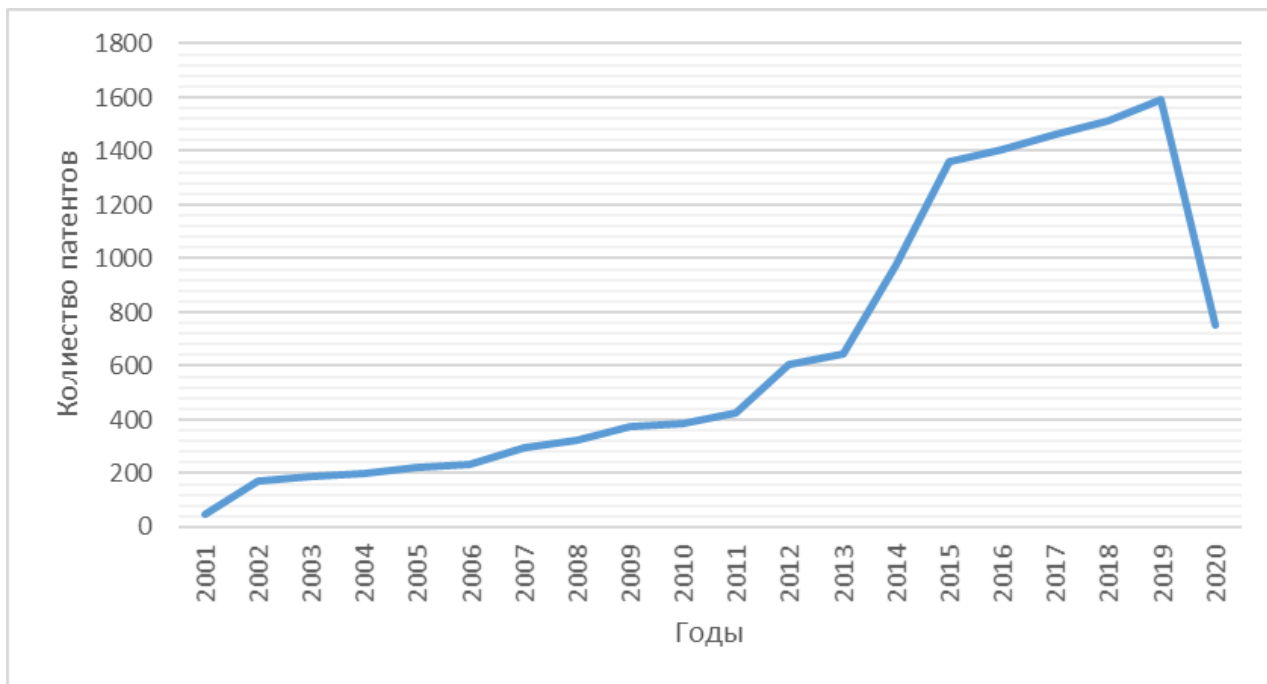


Рисунок 5 - Динамика патентования по направлению Big Data в медицине

В 2020 году такой низкий показатель так как данные приведены на первый квартал. [19]

В 1960-х медицина столкнулась с проблемой хранения данных о пациентах и исследований, все архивные комнаты были заняты, большая часть бумаг страдала от сырости, изнашивалась и приходило в негодный вид. Переписывать эти данные было затратно не только с точки зрения финансов, но и с точки зрения трудозатрат. С течением времени количество данных все больше и больше увеличивалось за счет увеличения населения Земли. В 1960-х годах количество земного населения равнялось 3 миллиарда, а в 2020 году это чисто возросло более чем в 2 раза. На каждого человека приходилось собирать данные на протяжении 70-80 лет. Несмотря на это с учетом использования современных технологий, данные собираются медленно, но эффективно. Сейчас данные в сфере здравоохранения уже считаются не терабайтами, а экзбайтами. По некоторым расчетам объем медицинских данных в 2014 г. достиг примерно 760 петабайт.

Последние исследования показывают, что более 30 % всех данных, хранящихся на планете, составляют медицинские данные, и в будущем ожидается быстрое увеличение этой доли за счет создания новой и оцифровки уже имеющейся информации. [20] Прогнозируется, что к 2021 г. количество медицинских данных будет достигать ~35000 петабайт. Важно то, что все эти данные не просто статистические, а представляют собой массив неструктурированных и разнородных данных, что определяет различные инструменты и методы для их сбора, хранения и анализа. [21]

В 2014 году развитие больших данных в России было под вопросом из-за отсутствия цифровых данных. С 2014 года началась оцифровка данных, а в 2018 году была принята программа «Цифровая экономика», которая дала толчок для развития технологии больших данных в сфере здравоохранения. Некоторых

медицинских учреждениях были пилотные проекты по цифровизации, которые показали хорошие результаты, но даже сегодня большая часть оборудования в больницах не имеет подключения к системе хранения данных.

Согласно дорожной карте «Healthnet», посвященной развитию технологий в медицине на период до 2035 года в сегменте «ИТ в медицине» планируется разработать системы поддержки принятия решений с использованием алгоритмов обработки больших объемов данных для локального использования в медицинских организациях. [22]

2 Логико-структурный подход при внедрении технологии больших данных в сферу здравоохранения

Логико-структурный подход (ЛСП) был разработан Агентством Международного развития в США в конце 60-х годов для оказания помощи в планировании, управлении и оценке процессов и мероприятий. Спустя определенное время ЛСП стали применять как инструмент не только для планирования проектов, но и для управления ими, например: Организация Международного развития ООН, Сельскохозяйственный банк развития, Германское общество технического сотрудничества, Датское общество международного развития, Международная организация труда, Служба развития зарубежных стран Норвежское общество международного развития, Шведская служба международного развития и так далее. В 1993 году было выпущено руководство по ЛСП которая называлась: Project Cycle Management. Integrated Approach and Logical Framework, European Commission, DG VIII.

Логико-структурный подход дает команде проекта определить заинтересованные стороны, цели, стратегии проекта и поставить себя на место каждого заинтересованного лица, определяя свои ожидания от проекта, таким образом команда глубже понимает свою роль в проекте и свои интересы реализации.

При четкой постановки задач, подзадач и мероприятий проекта можно получить «правильную иерархию» целей проекта. Мелкие цели подзадач помогают осуществить цели задача, а те в свою очередь цель мероприятий в проекте, которые определяют основную цель проекта, для чего он существует.

Этот подход помогает увидеть все риски и подводные камни которые ожидают на протяжении всего проекта. Таким образом можно просчитать где проект получит хороший толчок к реализации, а какие задачи будут его задерживать, а может даже не даст его осуществить. Но так жесть и второе преимущество применения ЛСП, на стадии планирования проекта можно сразу обозначить контрольные точки для мониторинга выполнения проекта, оценки работы при выполнении проекта и обозначить контрольные вехи. Что в будущем поможет сформировать систему сбалансированных показателей.

Для работы с логико-структурным подходом требуется не мало опыта ведь он не обеспечит 100 % реализации проекта. Стоит учитывать множество факторов: обозначение целей простыми словами, постройка несложных взаимоотношений, учет изменения проекта. Особенно изменение проекта, проект не стабильная структура, он меняется в зависимости и даже вместе с внешней средой, а внутренняя среда может привезти к разрешению проекта. Если в проекте нет точных и «простых» целей, значит команда или заинтересованные стороны не заинтересованы в реализации проекта.

Логико-структурный подход как единственный инструмент не поможет осуществить реализацию проекта, он работает только в совокупности с другими инструментами – социальными, экономическими и даже техническими для анализа среды.

Не смотря на все инструменты, которые помогают реализовывать проекты, ни один из них должным образом не разработает проект, который будет в точности такой же, как и в реалии. Проект имеет свойство меняться и подстраиваться под любые потребности окружающей ее среды.

ЛСП состоит из аналитической фазы и фазы планирования. В аналитическую входят три этапа:

- анализ заинтересованных сторон – для выявления основных заинтересованных людей или организаций для осуществления проекта;
- анализ проблем и построение «дерева проблем» – для того, чтобы выявить что именно мы решаем в ходе выполнения или осуществления проекта;
- анализ целей и построение «дерева целей» – для того, чтобы показать каких целей и показателей мы собираемся достичь при удачном запуске проекта.

После того как мы распланировали отправную точку для страта проекта переходим в фазу планирования. В фазу планирования входят пять этапов:

- построение внутренней логики проекта;
- определение допущений и факта риска;
- определение показателей и источников проверки;
- составление графика мероприятий;
- составление плана расходов.

От качества проведения аналитической фазы зависит не только фаза планирования, но и логика построения проекта, а также его успешность (таблица 1). [23]

Таблица 1 - Этапы аналитической фазы и фазы планирования

Аналитическая фаза	Фаза планирования
Этап 1. Анализ заинтересованных сторон	Этап 4. Построение внутренней логики проекта
Этап 2. Анализ проблем, построение «дерева проблем»	Этап 5. Определение допущений и факторов риска
Этап 3. Анализ целей, построение «дерева целей»	Этап 6. Определение показателей и источников проверки
	Этап 7. Составление графика мероприятий
	Этап 8. Составление плана расходов

2.1 Анализ заинтересованных сторон

Первым этапом при применении логико-структурного подхода является анализ заинтересованных сторон. Заинтересованные стороны, это все компании, люди, организационные структуры, инвесторы, которые причастны к реализации проекта. При формировании списка заинтересованных сторон следует начинать с команды проекта, ведь без людей у которых цель выполнить свою работу, проект останется на стадии идеи. Далее рассматриваются основные управляющие элементы, такие как инвестор, заказчики. После этого рассматривается третий круг лиц – партнеры. И в самый последний момент рассматриваются те, кто могут повлиять на изменение проекта уже в ходе его реализации: конкуренты.

Для того что бы начать анализ заинтересованных сторон необходимо ответить на ряд нижеследующих вопросов:

- на кого направлена деятельность по осуществлению проекта;
- кто будет участвовать в проекте;
- кто ещё может принять участие в проекте;
- от кого зависит реализация проекта, кто из сторонних лиц будет принимать ключевые решения, имеющие отношения к проекту (вышестоящие инстанции, органы власти и т.д.);
- от кого зависит реализация этих решений;
- кто не заинтересован в реализации проекта;
- для кого проект может представлять угрозу;
- кто может предоставлять сведения, необходимые для разработки и реализации проекта, поделится дополнительной информацией и опытом (коллеги, источники справочной информации, организации со схожей деятельностью и т.д.);
- какие организации, группы людей, отдельные лица могут оказать влияние на проект прямым или косвенным образом;
- помощь, консультации каких экспертов по определенным (узким) вопросам могут понадобиться при организации проекта;
- опишите целевые группы (пол, возраст, величина дохода, место работы, специальность и т.д.) и проанализировать эффект от планируемой деятельности на отдельных частях этих групп.

Самый краткий перечень держателей интересов в проектах выглядит следующим образом:

- менеджер проекта;
- команда проекта;
- инвесторы;
- общественные организации;
- органы власти;
- партнеры по бизнесу;
- потребители;
- конкуренты;
- заказчик. [24]

Полученные данные вносятся в таблицу 2.

Таблица 2 - Таблица анализа заинтересованных сторон

Группа заинтересованных сторон	Каков их интерес в проекте? (положительные и отрицательные аспекты)	При каких условиях стороны заинтересованы в долгосрочности проекта?	Возможное участие (возможная роль)

Для того что бы анализ заинтересованных сторон был полным необходимо ответить на ряд следующих вопросов:

- какие действия надо выполнять для того, чтобы цель проекта была достигнута;
- какими должны быть роли участников при выполнении этих действий;
- какие сведения необходимо иметь для разработки и реализации проекта (справочные сведения, организационные процедуры и т.д.) и кто их может предоставить;
- какие предположения относительно ролей заинтересованных сторон (целевые группы, потенциальные партнеры и оппоненты, сочувствующие наблюдатели и др.) или их реакций следует сделать для того что бы выполнение проекта было успешным. [25]

Отвечая на все вышеприведенные вопросы при формировании анализа заинтересованных сторон получили следующую таблицу 3.

Таблица 3 - Анализ заинтересованных сторон

Группа заинтересованных сторон	Каков их интерес в проекте?	При каких условиях стороны заинтересованы в долгосрочности проекта?	Возможное участие (возможная роль)
Команда проекта	1. Ускорение обработки больших данных в медицине. 2. Налаживание системы автоматизации бумажных работ. 3. Получение прибыли.	Получение постоянной прибыли за внедрение и поддержку системы.	– создание математической или имитационной модели обработки данных; – регулирование потока информации; – техническая поддержка; – заключение долгосрочных контрактов.
Медицинские учреждения	1. Уменьшение бумажной работы. 2. Экономия времени. 3. Быстрота обслуживания. 4. Удобство использования.	1. Быстрое обслуживание пациентов. 2. Увеличение «плана» лечения. 3. Экономия времени на бумажную работу. 4. Хранение истории болезни в одном централизованном месте.	– использование конечной технологии; – «работа над ошибками» для технических специальностей; предложения по совершенствованию системы.

Продолжение таблицы 3

Группа заинтересованных сторон	Каков их интерес в проекте?	При каких условиях стороны заинтересованы в долгосрочности проекта?	Возможное участие (возможная роль)
Конкуренты	<ol style="list-style-type: none"> 1. Предоставление своих услуг по обработке и хранению данных. 2. Внедрение усовершенствования в систему. 	<ol style="list-style-type: none"> 1. Возможность стать субподрядчиком. 2. Выигрыш в тендере как основной исполнитель. 3. Развитие ЦОД. 4. Получение аттестации. 	<ul style="list-style-type: none"> – участие в тендерах на предоставление услуги; – участие в качестве субподрядчика.
Министерство здравоохранения по Красноярскому краю	<ol style="list-style-type: none"> 1. Выполнение задачи заказчика. 2. Получение опыта как проекта в IT. 3. Получение прибыли. 4. Ускоренная работа по данным медицинских организаций. 5. Цифровизация медицины. 6. Доступное обслуживание в медицинских учреждениях (МУ). 7. Выполнение программы Цифровая экономика в медицине. 8. Уменьшение нагрузки на МУ. 9. Повышение качества жизни населения путем повышения качества обслуживания. 	<ol style="list-style-type: none"> 1. Уменьшение капиталовложений в МУ за счет экономии на ненужные траты в связи с внедрением системы. 2. Цифровизация медицины, ускоренное обслуживание и доступная медицина. 3. Защита данных МУ. 4. Повышение качества жизни населения 	<ul style="list-style-type: none"> – регламент работ для цифровизации медицины; – план по внедрению в МУ; – заключение долгосрочных контрактов.
Медицинский информационно-аналитический центр	<ol style="list-style-type: none"> 1. Обработка большого количества данных по всем МУ края. 2. Главные регуляторы в технической части для МУ в области их влияния. 	<ol style="list-style-type: none"> 3. Цифровизация медицины, ускоренное обслуживание и доступная медицина. 4. Защита данных МУ. 	<ul style="list-style-type: none"> – внедрение собственного ПО для МУ; – получение прибыли от внедрения ПО передающее данные в ЦОД; – заключение долгосрочных контрактов.

Окончание таблицы 3

Группа заинтересованных сторон	Каков их интерес в проекте?	При каких условиях стороны заинтересованы в долгосрочности проекта?	Возможное участие (возможная роль)
Пациенты	Ускоренное и возможное «безошибочное» лечение.	1. Экономия времени на бумажную работу. 2. Хранение истории болезни в одном централизованном месте.	Предложения по совершенствованию системы.
Разработчики и поставщики шифровально го/не шифровально го программно го оборудования /обеспечения	1. Продажа шифровального/ не шифровального оборудования и программного обеспечения. 2. Продажа оборудования для ЦОД. 3. Продажа ПО для ЦОД.	4. Прибыль от продажи шифровального/ не шифровального оборудования и ПО. 5. Продажа оборудования для ЦОД. 6. Продажа ПО для ЦОД. 7. Техническое обслуживание.	– разработка шифровального/ не шифровального ПО и усовершенствование по требованию закона; – заключение контрактов на использование только данного оборудования или ПО.

Главными заинтересованными сторонами являются министерство здравоохранения, медицинские учреждения и пациенты, получающие лечение.

Хотя пациенты, которые получают лечение ощущают изменение системы только с помощью следующих показателей как (быстрота обслуживания, точность диагностики, электронная база болезни с обширным доступом), которые будут рассматриваться в дальнейшем.

Для министерства здравоохранения и медицинских учреждения пользы от внедрения данной технологии больше, чем для пациентов.

На основе заинтересованных сторон выявлены следующие проблемы для построения дерева проблем:

- недостаточная заработная плата медицинского персонала;
- недостаток медицинского персонала;
- загруженность медицинских учреждений;
- отвлечение от основной деятельности медицинского персонала;
- отсутствие оборудования;
- низкая скорость обработки данных;
- большой объем медицинских данных;
- продолжительное лечение пациентов;
- недостаток льготных лекарств;
- недостаточное управление медицинскими данными;
- неверно поставленный диагноз;
- неудовлетворительное качество медицинского обслуживания;
- низкое качество жизни;

- отсутствие профилактического лечения;
- уменьшение продолжительности жизни населения;
- снижение экономических показателей трудоспособного население.

2.2 Дерево проблем и дерево целей

После того как определены основные заинтересованные стороны и их основные цели при реализации проекта строим дерево проблем. Дерево проблем представляет собой причинно-следственные связи между основными проблемами проекта. В дереве проблем существуют корневые проблемы, которые порождают основную проблему, которая в свою очередь определяет следствие этой проблемы. Следует запомнить, что проблема - это не отсутствие чего-либо нужного для дела, а вариант решения проблемы.

Дерево проблем строится по следующим трем правилам (рисунок 6).

1. Нижний уровень: проблема – это причина.
2. Верхний уровень: проблема – это следствие.
3. Средний уровень: проблема – не причина не следствие.



Рисунок 6 - Строение дерева проблем

Важно иметь в виду, что у ряда проблем может быть несколько причин, но для каждой причины характерна, как правило, одна проблема.

На основе проблем, выявленных в предыдущей главе, построим дерево проблем с причинно-следственными связями между ними (рисунок 7).

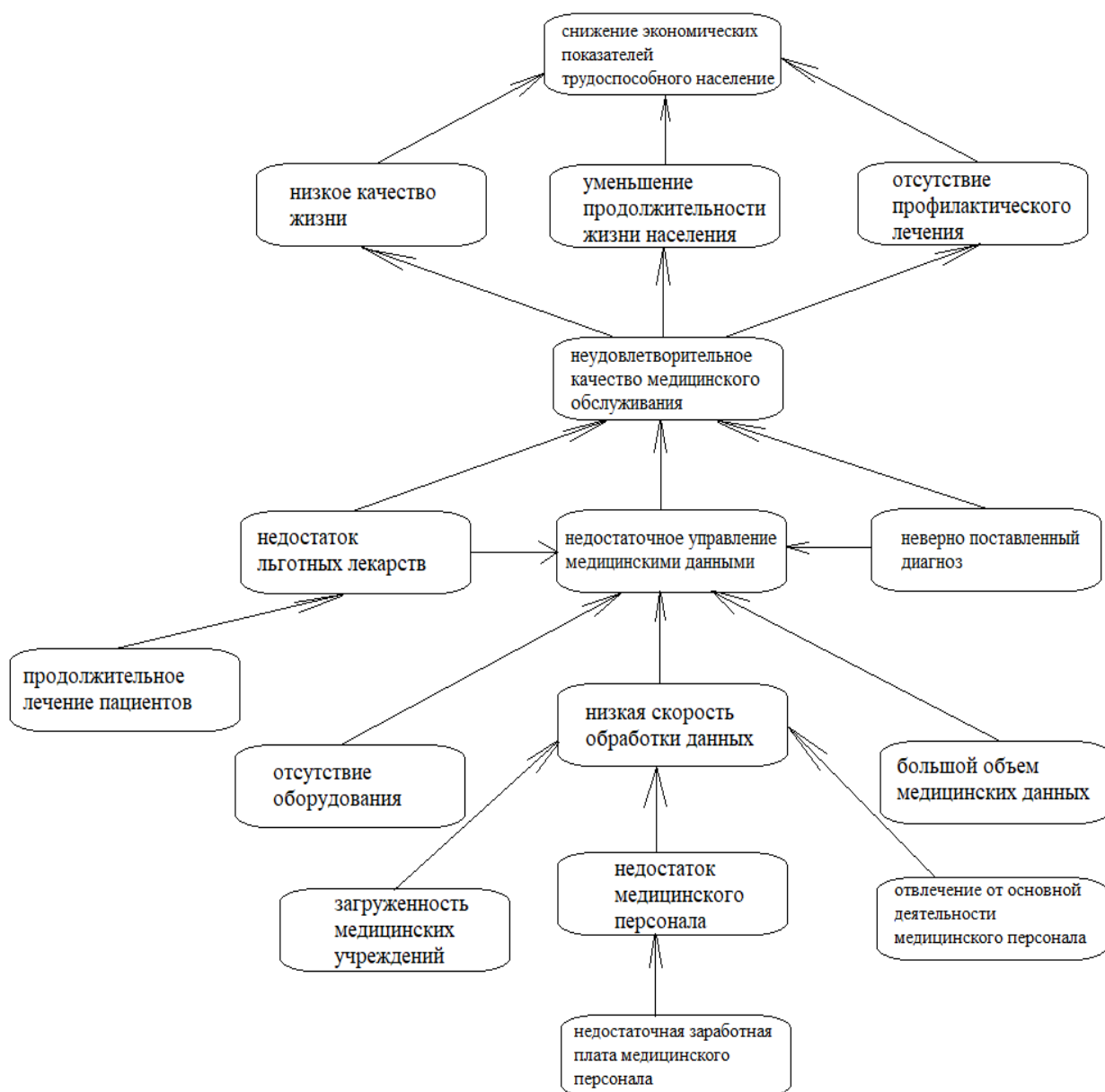


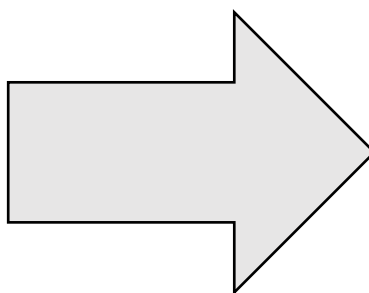
Рисунок 7 - Дерево проблем

Основные проблемы, которые существуют на данный момент приведены на рисунке. Корневыми проблемами являются недостаточность финансирования в сферу медицины в плане заработной платы и обеспечения необходимым оборудованием медицинские учреждения. Так же основная проблема заключается в недостаточном управлении медицинскими данными, которые рас сосредоточены по разным носителям, хранимые либо в бумажном архиве, либо в электронном. Для решения данной проблемы необходимо создать централизованную базу медицинских данных, которые будет удобно хранить, обрабатывать и использовать повторно.

Для построения дерева целей переформулируем проблемы в цели (таблица 4).

Таблица 4 - Переформулирование проблем в цели

Проблемы	Цели
недостаточная заработная плата медицинского персонала	достойная заработная плата медицинского персонала
недостаток медицинского персонала	достаточное кол-во медицинского персонала
загруженность медицинских учреждений	не загруженность медицинских учреждений
отвлечение от основной деятельности медицинского персонала	сосредоточенный на своей работе медицинский персонал
отсутствие оборудования	наличие оборудования
низкая скорость обработки данных	высокая скорость обработки данных
большой объем медицинских данных (на бумажных носителях)	Оцифрованные медицинские данные
продолжительное лечение пациентов	быстрое лечение
недостаток льготных лекарств	достаточное кол-во льготных лекарств
недостаточное управление медицинскими данными	хорошее управление медицинскими данными
неверно поставленный диагноз	корректный диагноз
неудовлетворительное качество медицинского обслуживания	удовлетворительное качество медицинского обслуживания
низкое качество жизни	высокое качество жизни
отсутствие профилактического лечения	наличие профилактического лечения
уменьшение продолжительности жизни населения	увеличение продолжительности жизни населения
снижение экономических показателей трудоспособного население	увеличение экономических показателей трудоспособного население



Рассмотрев поставленные цели в соответствии с таблицей 3 построим дерево целей (рисунок 8).

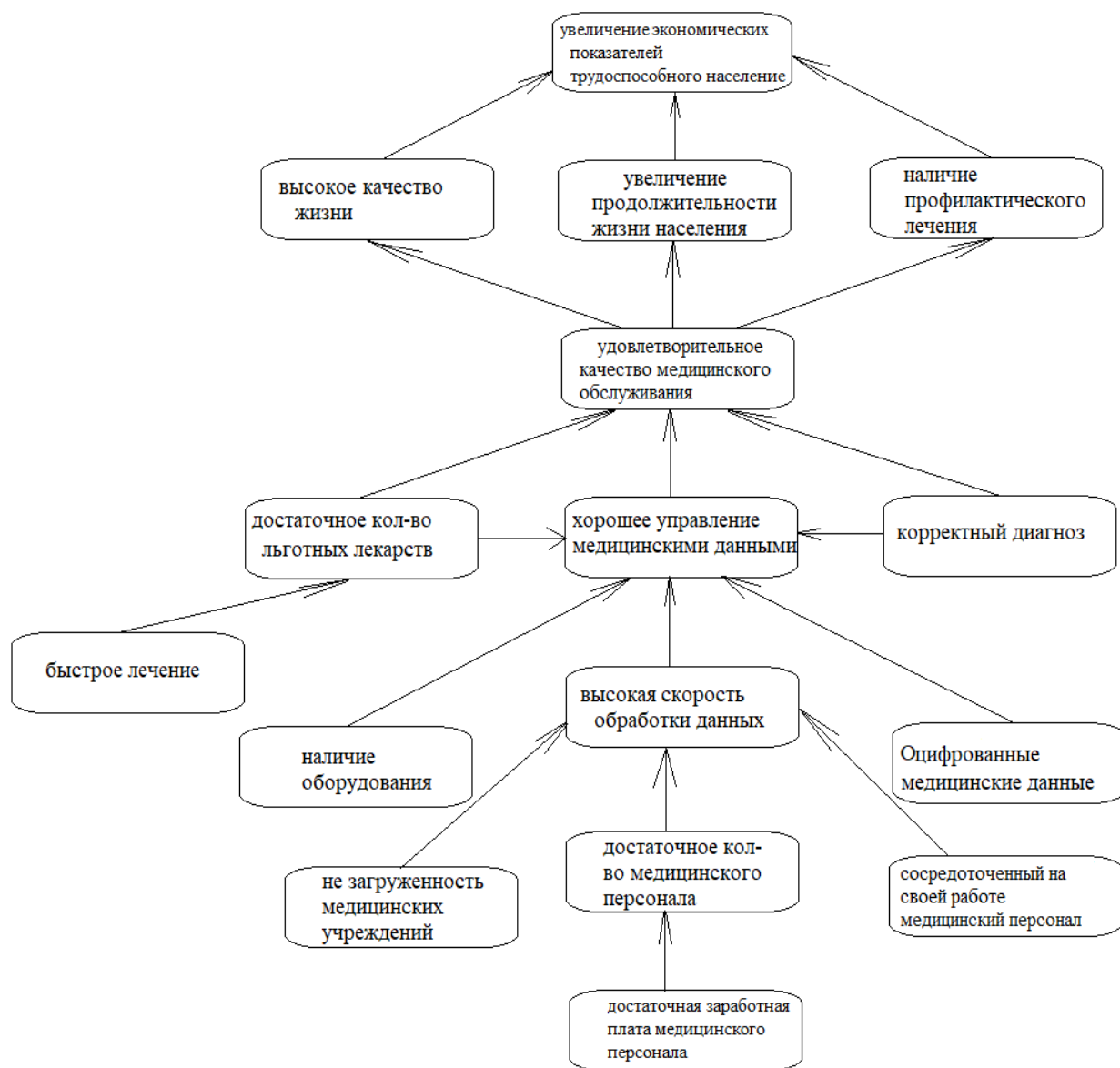


Рисунок 8 - Дерево целей

На данном рисунке показаны основные цели, достигаемые при внедрении технологии больших данных в сферу здравоохранения. Для достижения главной цели необходимо решить корневые, так как от нее зависят цели, поставленные выше.

2.3 Анализ процессов при внедрении технологии больших данных в сферу здравоохранения

Методология IDEF0 предписывает построение иерархической системы диаграмм – единичных описаний фрагментов системы. Сначала проводится описание системы в целом и ее взаимодействия с окружающим миром (контекстная диаграмма), после чего проводится функциональная декомпозиция - система разбивается на подсистемы, и каждая подсистема описывается

отдельно (диаграммы декомпозиции). Затем каждая подсистема разбивается на более мелкие и так далее до достижения нужной степени подробности. Каждая IDEF0-диаграмма содержит блоки и дуги. Блоки изображают функции моделируемой системы. Дуги связывают блоки вместе и отображают взаимодействия и взаимосвязи между ними. Взаимодействие работ с внешним миром и между собой описывается в виде стрелок, изображаемых одинарными линиями со стрелками на концах. В IDEF0 различают четыре типа стрелок. [26]

Первая стрелка называется «Вход» (рисунок 9). Входами в диаграмме являются результаты работ, объект, задание, любая материальная вещь, которая необходима для выполнения задачи. Так же следует знать, что любой выход может быть входом для другого функционального блока.

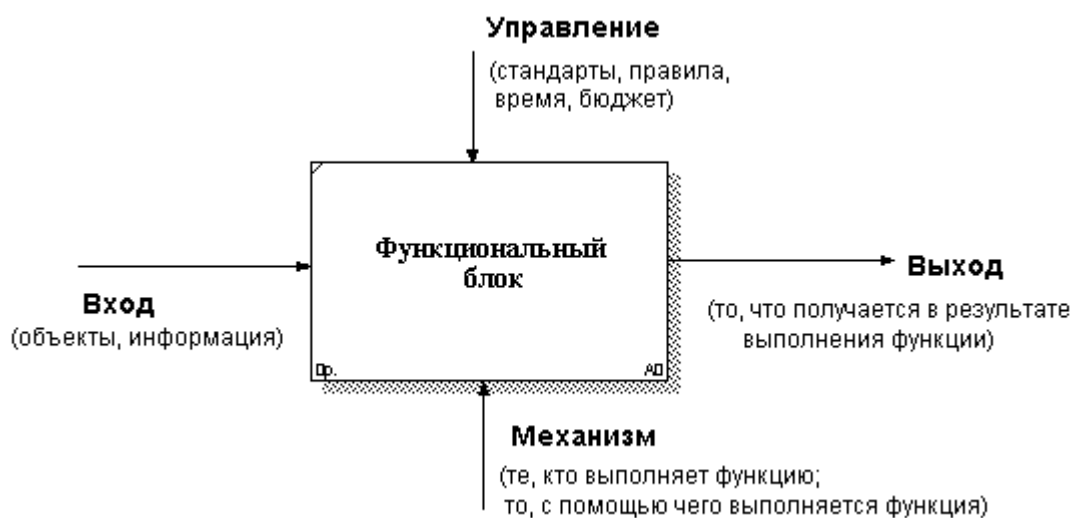


Рисунок 9 - Контекстная диаграмма (пример)

Вторая стрелка называется «Управление». Данная стрелка является основным исполнительным органом диаграммы, она регулирует выполнение и ход работы функционального блока или блоков. Без стрелки управления работа считается невыполненной или невыполняемой, так как нет никаких регламентов.

Третья стрелка «Выход», который является результатом работы, функциональный блок должен иметь хотя бы одну стрелку выходы, иначе работа считается невыполненной.

Четвертая стрелка «Механизм исполнения» или по-другому ресурсы с помощью которых выполняется работа в функциональном блоке.

Применение методологии IDEF0 на управление процессов внедрения технологии больших данных выглядит следующим образом.

Контекстная диаграмма, представленная на рисунке 10, показывает, что основной процесс является внедрение технологии в сферу здравоохранения, где входами является неавтоматизированная система медицинских данных, а выходом автоматизированная система медицинских данных.

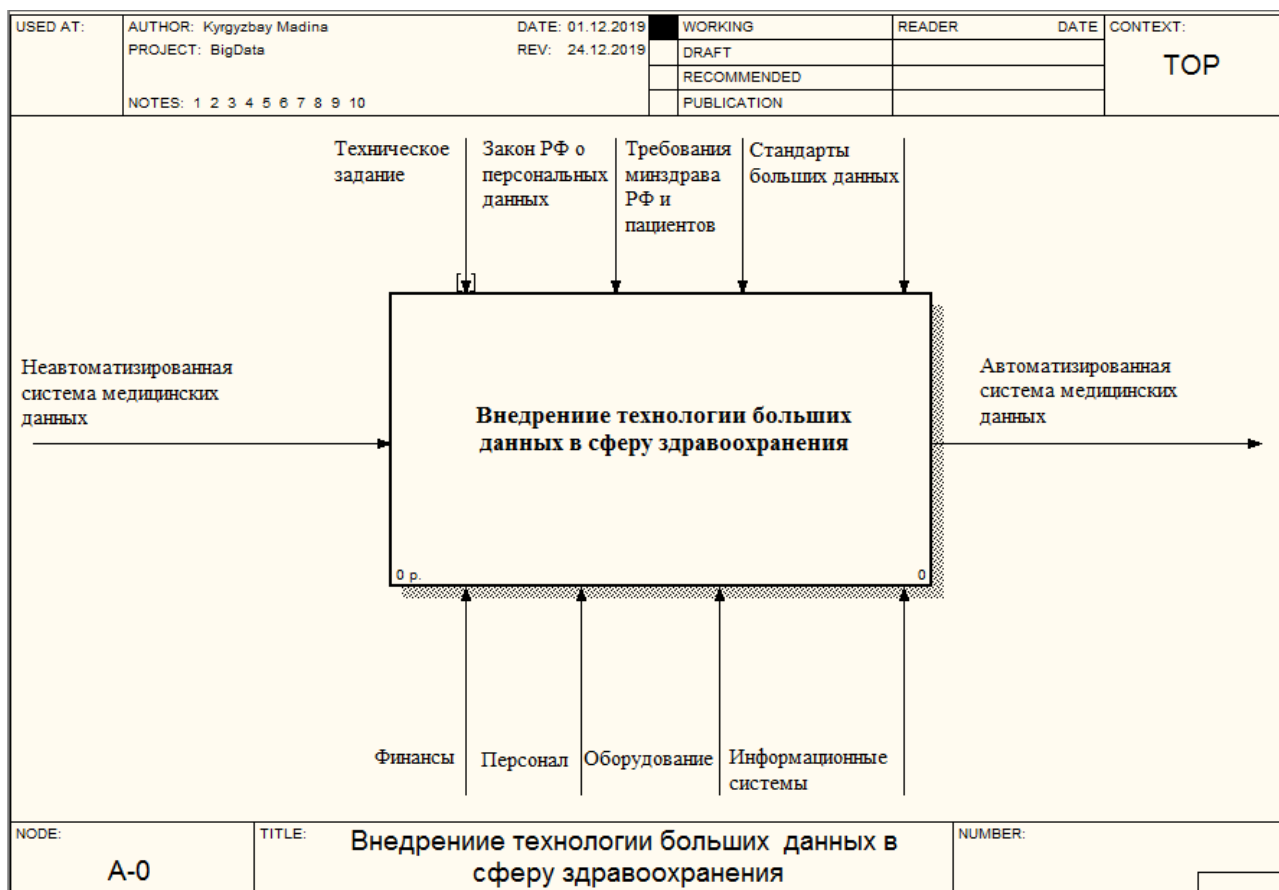


Рисунок 10 - Контекстная диаграмма

Управляющими элементами в данной диаграмме являются техническое задание (приложение А) на введение технологии больших данных в сферу здравоохранения, законы РФ о персональных данных, требования Министерства здравоохранения РФ и пациентов, а также стандарты больших данных.

Законы о персональных данных необходимо учитывать, так как большая часть медицинских данных состоит из данных людей, и обработка этих данных должно производиться в соответствии с законом и безопасностью, да бы исключить утечку информации и распространение данных.

Основными требованиями при внедрении технологии являются требованиями Министерства здравоохранения РФ, так как они являются заказчиками данной услуги в связи с национальной программы «Цифровая экономика Российской Федерации».

Для хранения и обработки так же существуют определённые стандарты, не все наборы данных могут являются большими данными, так же не всегда большой объем информации может входить в рамки технологии больших данных.

Так же основные ресурсы необходимые для реализации данного проекта: финансирование, персонал, оборудование и информационные системы.

Далее декомпозируем контекстную диаграмму на три составляющие, приведенные на рисунке 11.

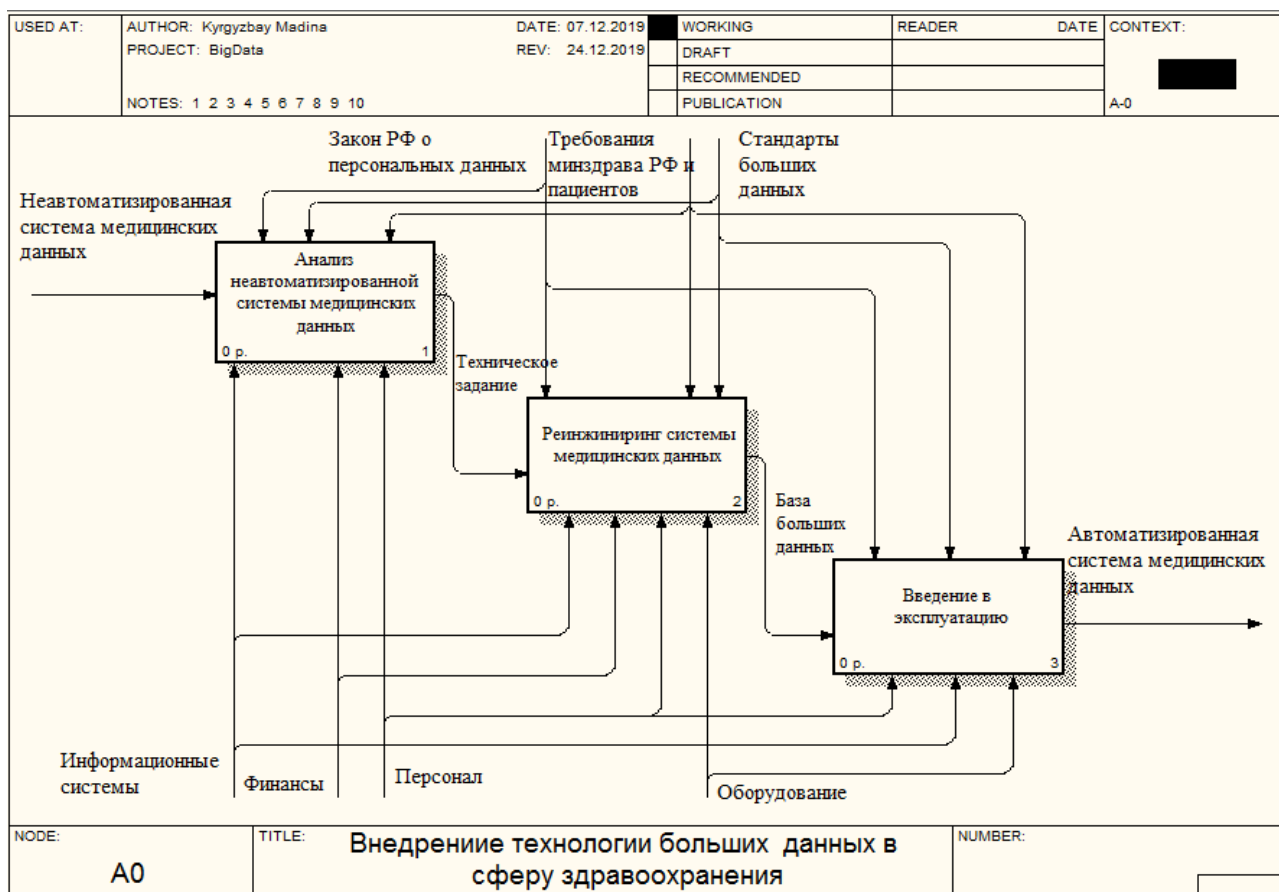


Рисунок 11 - Декомпозиция первого уровня

Контекстную диаграмму декомпозированы на три составляющие это:

- анализ неавтоматизированной системы медицинских данных – входом является неавтоматизированная система данных, выходом техническое задание, сформированное на основе анализа, управляющие элементы: законы о персональных данных, требования Минздрава РФ и пациентов, стандарты больших данных;

- реинжиниринг системы медицинских данных – вход техническое задание, выход база больших данных, управление за счет выше описанных элементов;

- введение в эксплуатацию – вход база данных, выход автоматизированная система медицинских данных, управляющие элементы остаются как у предыдущего.

На всех этапах работы ресурсы остаются стандартными: финансы, оборудование, персонал и информационные системы.

Декомпозируем каждый элемент на под процессы для точного понимания процесса внедрения технологии больших данных. Данные представлены на рисунках 12, 13, 14.

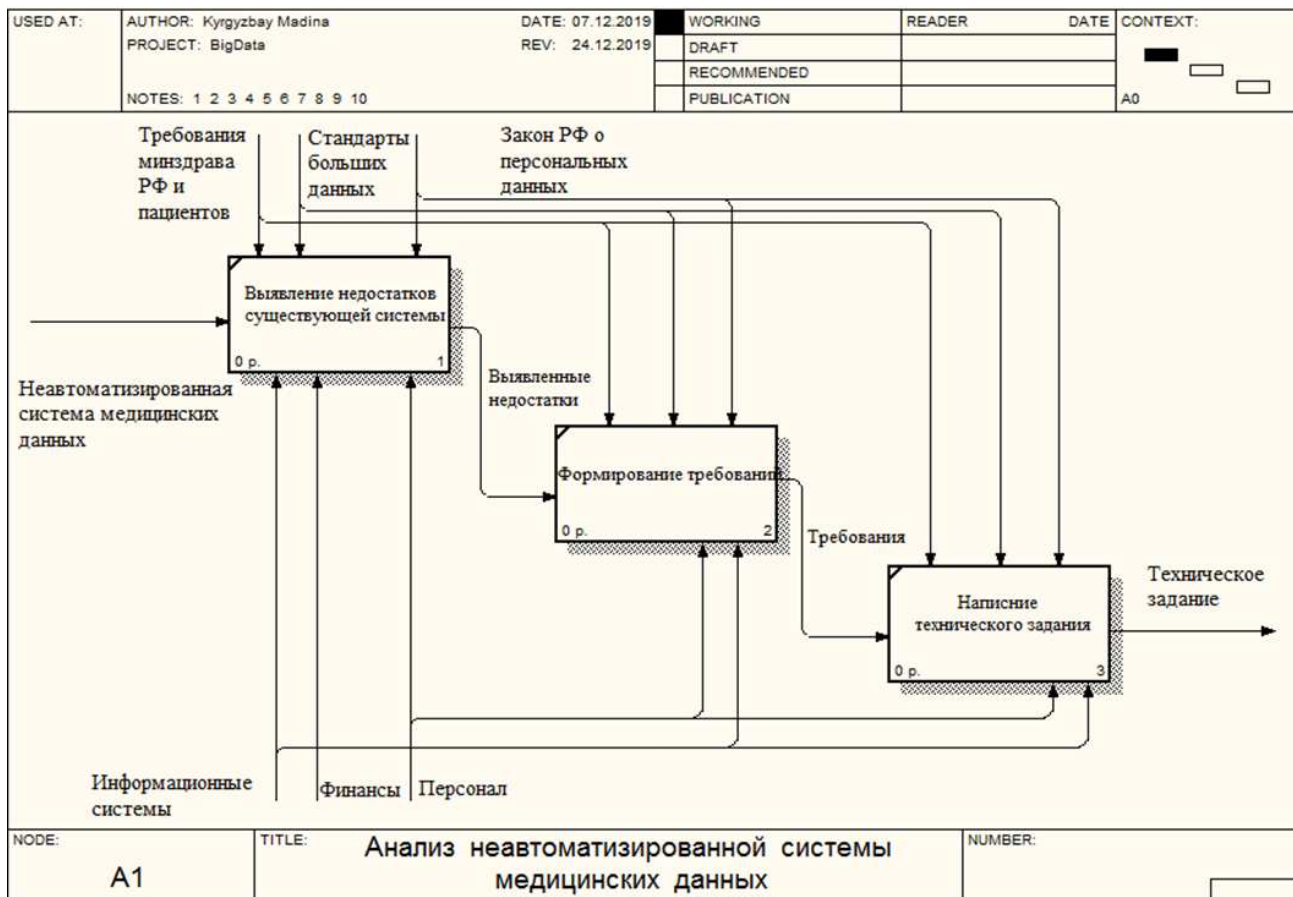


Рисунок 12 - Анализ неавтоматизированной системы медицинских данных

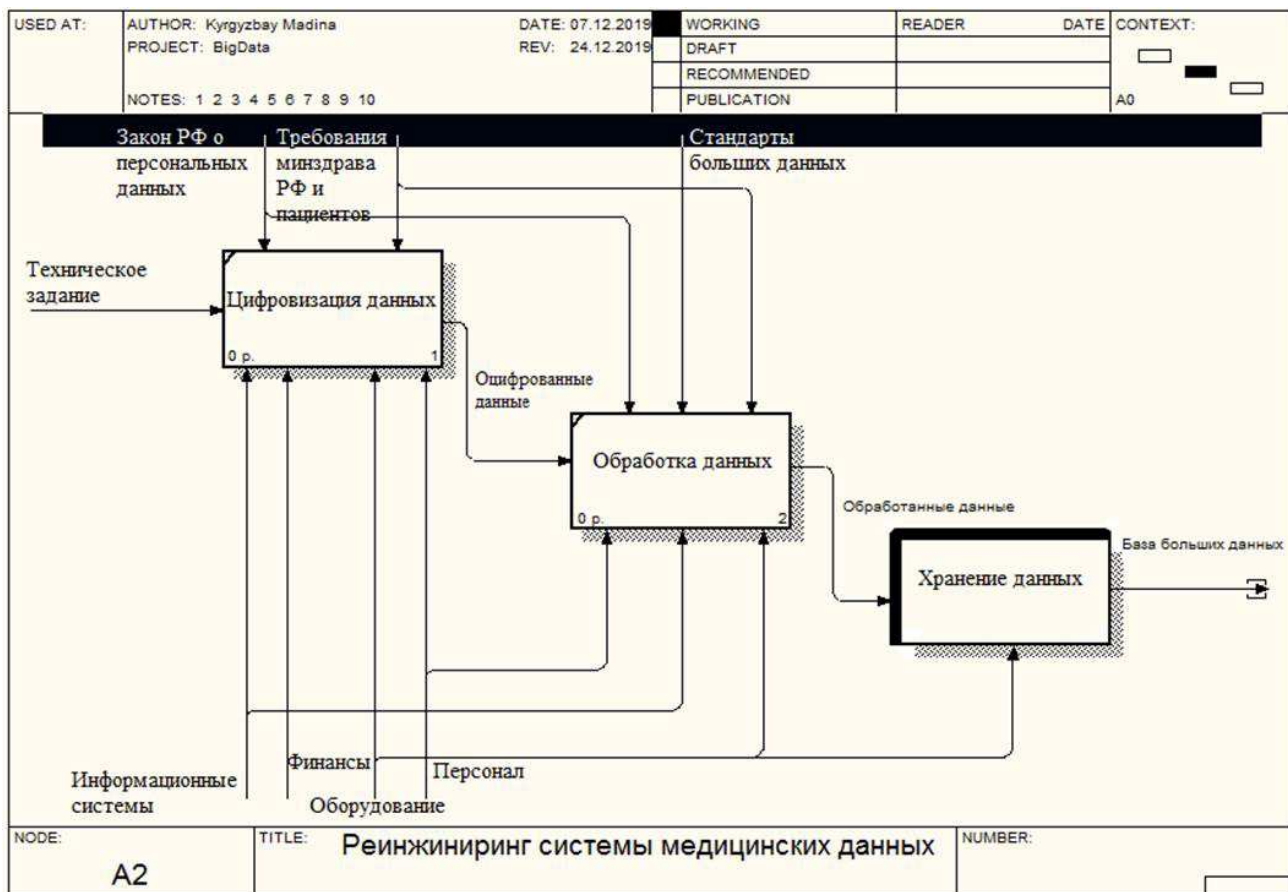


Рисунок 13 - Реинжиниринг системы медицинских данных

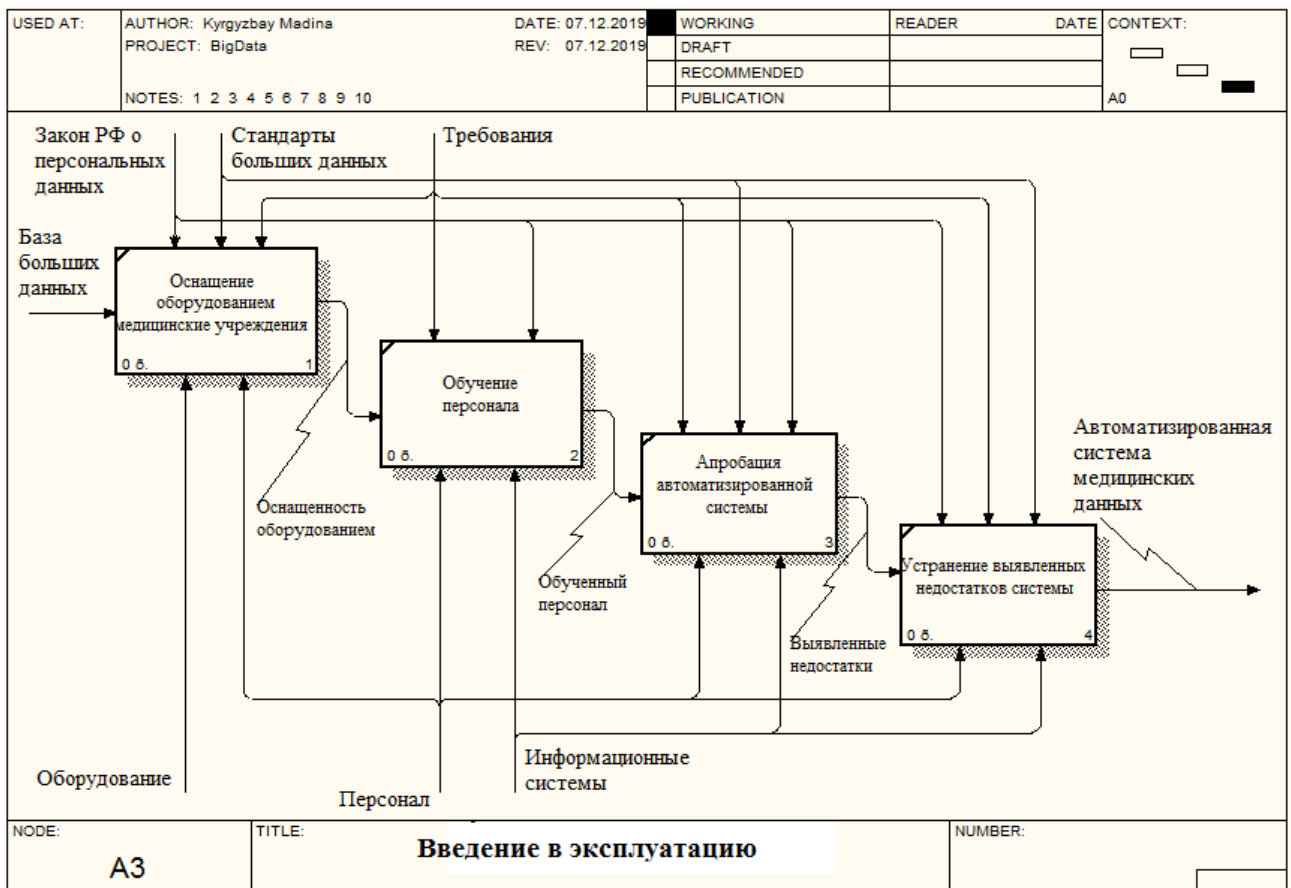


Рисунок 14 - Введение в эксплуатацию

Данные диаграммы необходимы для оформления системы сбалансированных показателей, а также являются минимальной иерархической структурой работ, которой руководствовались при написании технического задания и дальнейшего внедрения технологии больших данных в сферу здравоохранения.

2.4 Система сбалансированных показателей

Сбалансированная система показателей (ССП) появилась в конце 1980-х — начале 1990-х годов как инструмент для управления динамично развивающимися компаниями. [27]

Эти годы стали началом больших перемен во всем мире особенно в рыночной структуре, торговля стала либерализированной, в мире настал бум для технологических новшеств. Потребности потребителей менялись так же стремительно, компаниям было необходимо подстраиваться под существующий спрос.

В конце 1980-х годов Профессоры Роберт Каплан и Дейвид Нортон провели исследование 12 компаний и выяснили, что эти компании слишком сильно ориентируются на финансовые показатели. В ходе исследования стало известно, что компании уменьшали вклады в стратегически важные аспекты бизнеса. Персонал перестал обучаться, что привело к ухудшению качества товара, продукта и услуги. Перестал развиваться маркетинг, он больше не

привлекал покупателей, все делалось в основном за счет «сарафанного радио». Такие недочёты привели к тому что финансовые показатели, к которым стремились компании становились все хуже и хуже.

Это привело к тому что служащие компании не были заинтересованы в продвижении своей компании и продукции, их интерес был только в получении заработной платы. Без мотивации персонал сделал компанию убыточной.

В качестве способа решения выявленных проблем Нортон и Капланом была разработана концепция сбалансированной системы показателей.

В качестве эксперимента ССП ввели в несколько компаний в 1990 году и к концу календарного года результаты исследования показали, что система уникальна и помогает не только объединить финансовые и нефинансовые показатели, но и преумножать их.

Сбалансированная система показателей — динамично развивающаяся методика, которая все больше расширяет использование возможностей персонала и информационных технологий. [28]

ССП в классическом варианте состоит из четырех составляющих, которые отражают стратегически важные аспекты деятельности предприятия:

- финансы;
- клиенты;
- внутренние бизнес-процессы;
- обучение и развитие. [29]

Но в зависимости от направленности деятельности компании или ее стратегии, составляющие могут менять и дополняться. Между составляющими существует четкие причинно-следственные связи. (рисунок 15).



Рисунок 15 - Элементы ССП и их взаимосвязь

Первый этап, а именно финансовое направление рассматривает все что касается денежных средств: преумножение, уменьшение издержек, экономия на сырье и т.д. Но на этом не заканчивается, это направление так же рассматривает основные запросы инвесторов и акционеров, то есть управляющего органа проекта. Многие считают, что тут должны быть обозначены цели только инвестора и акционером, на самом деле эти финансовые показатели должны задевать весь проект в целом и его членов тоже. Так как команда проекта будет

выполнять основную работу при реализации проекта и это им достигать этих финансовых показателей.

Потребительское направление непосредственно связано с деревом целей, который получился благодаря дереву проблем. Там были указаны основные точки, которые могут быть преобразованы в ориентиры и цели потребительского направления.

Направление внутренних процессов больше необходимо для команды проекта, так как основные показатели и цели, ориентированные на улучшение продукта, услуги или товара. Чем быстрее и лучше выполняется работа команды проекта, тем лучше получается продукт, который будет удовлетворять потребителей и помогать достичь показателей потребительского направления.

Авторы ССП не рассчитывали, что их структура поможет описать большую часть бизнес-процессов, имеющих в компаниях. Совершенствовать внутренние процессы необходимо всегда, так как они могут вывести организацию на новый уровень. Но данное совершенствование зависит не только от изменения технологии и стремления к денежным показателям, но и от квалификации и опыта персонала, внешней и внутренней среды, а также других факторов.

Между целями тоже существуют причинно-следственные связи (рисунок 16).

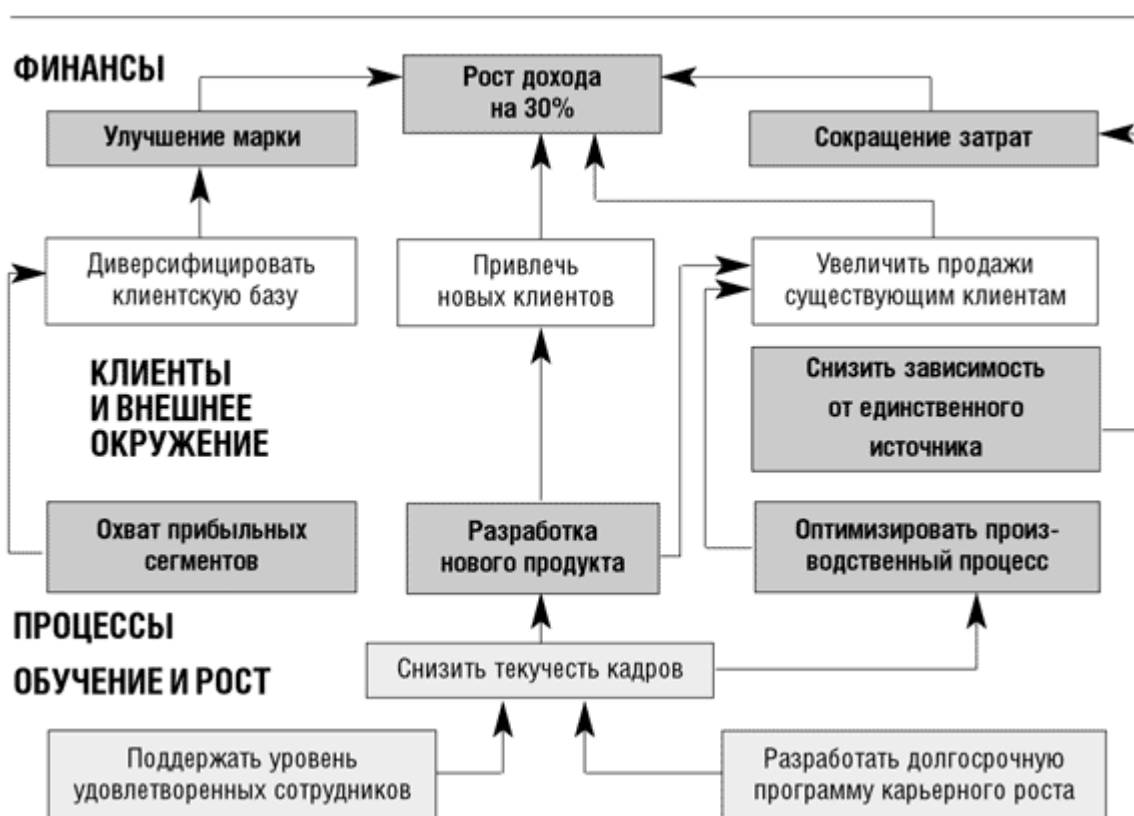


Рисунок 16 - Причинно-следственные связи между целями ССП

Основные показатели эффективности должны:

– отслеживать ход реализации стратегии и корректировать ее в соответствии с изменяющимися условиями;

– обеспечивать основу для планирования и оценки исполнения бюджета, а также деятельности каждого сотрудника.

Показатели могут быть самыми разными: ежедневными, ежеквартальными или даже ежегодными, все зависит от того какие показатели выберут управляющие элементы при осуществлении проекта. [30]

Для того что бы оценить степень реализации проекта необходимо ставить четкие промежуточные показатели, которые объединяются в основные целевые показатели, которые в дальнейшем можно посчитать.

Эти показатели не являются единственными верными, они имеют свойство меняться в зависимости от изменения проекта. Но стремиться к этим показателям одна из основных задач команды проекта. Они должны быть четкими и выполняемыми.

В зависимости от ситуации и от уровня амбиций в качестве ориентира можно выбирать конкурентов, признанных лучшими в том сегменте рынка, где работаем и мы, либо ближайших к нам по показателям. Во всех случаях тщательное изучение достижений конкурентов позволяет наметить целевые значения, которые будут служить эталонами для нас. Иногда полезным оказывается сравнение не с прямыми конкурентами, а с организациями, работающими в совершенно других областях бизнеса. [31]

Основными показателями внедрения технологии больших данных является ряд следующих показателей, представленных на рисунке 17. Которые были сформированы на основе декомпозиции процесса внедрения технологии больших данных в сфере здравоохранения.

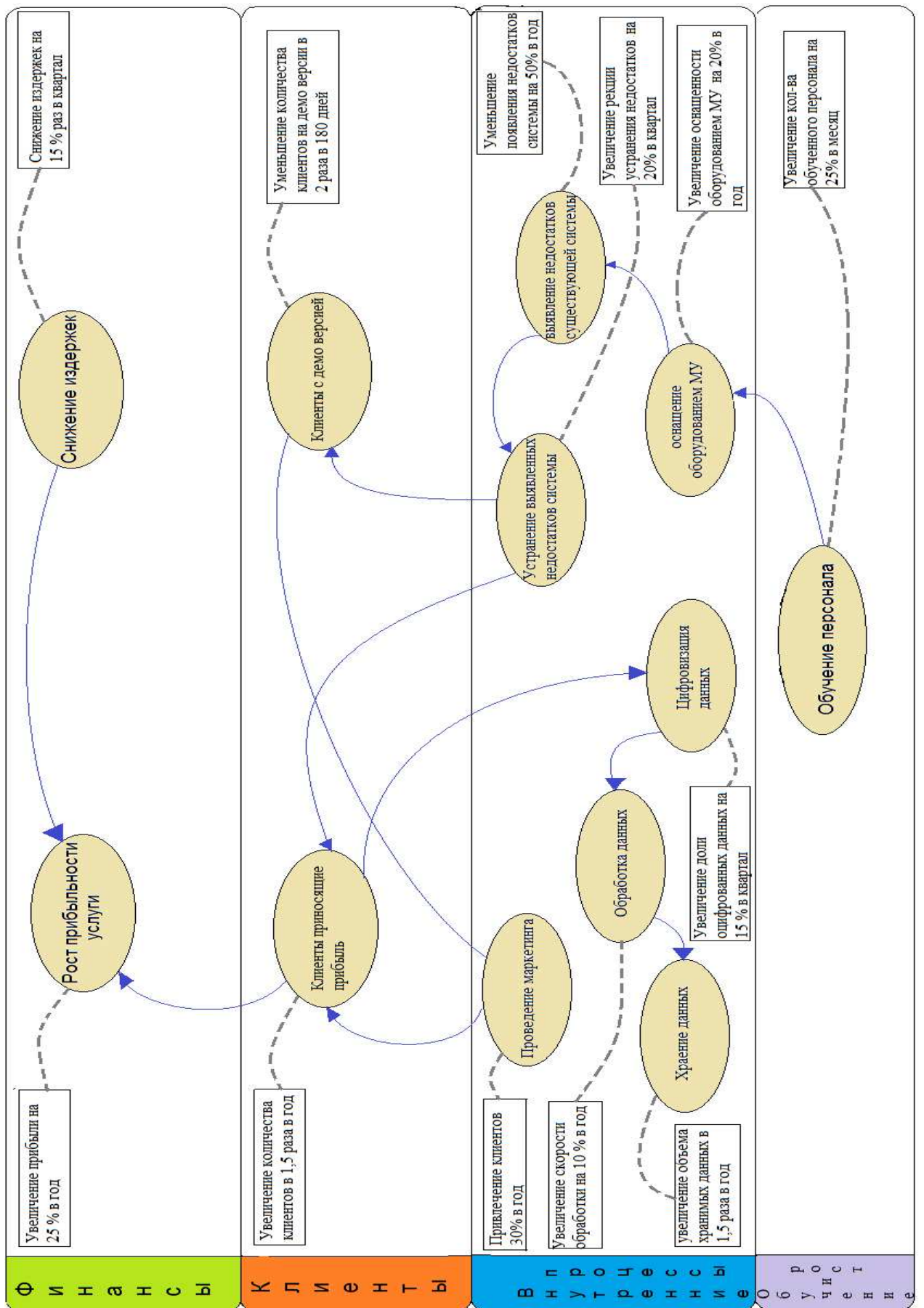


Рисунок 17 - ССП Внедрения технологии больших данных в сферу здравоохранения

Согласно рисунку 17 мы выделили следующие показатели в финансовой сфере: увеличение прибыли на 25 % в год, которая характеризуется тем что данная услуга расширяется по медицинским учреждениям и охватывает все больший объем территории, снижение издержек на 15 % в квартал, заключается в том, чтобы уменьшить траты медицинских учреждений на бумажные носители информации, нерациональное использование медицинского персонала и т.д.

В клиентах необходимо по большей части уменьшить количество мед. учреждений пользующиеся демоверсией данной услуги в два раза, так как тестовый режим предоставляется на период 180 дней и 30 дней на настройку и развертывание системы. В данный период с помощью услуги автоматизированной системы обрабатывается незначительная часть медицинских данных. А также необходимо привлекать все больше клиентов в 1,5 раза в год., что вполне осуществимо в связи с проведением государственных закупок данной услуги.

Далее на рисунке приведены внутренние процессы со следующими показателями:

- привлечение клиентов 30 % в год с помощью проведение качественного маркетинга;
- увеличение скорости обработки медицинских данных на 10 % в год;
- увеличить объем хранимых данных в централизованной системе в 1,5 раза в год, т. е. уменьшить данные, которые на бумажных или иных носителях;
- увеличит количество оцифрованных данных на 15 % в квартал;
- уменьшить проявление недостатков системы на 50 % в год, т.е. система должна быть отказоустойчивой;
- увеличение реакции устранения недостатков на 20 %, т.е. увеличить реакцию при появлении проблем;
- увеличение оснащенности оборудованием МУ на 20 % в год, с помощью предоставление серверного оборудования, каналов связей, тонких клиентов и т.д.
- увеличение количества обученного персонала на 25 % в год.

ЗАКЛЮЧЕНИЕ

Не подлежит сомнению, что технологии «Big Data» могут оказать существенное воздействие на медицину и здравоохранение, причем как в плане повышения точности диагностики, эффективности лечения и своевременного предупреждения болезней, так и в контексте улучшения работы самих медицинских учреждений. [39] Преимуществами при внедрении технологии больших данных станет не только увеличение продолжительности жизни населения страны, но и сохранит и даже поможет сэкономить бюджет, который ежегодно выделяется для сферы здравоохранения в России. Бюджет сохранится за счет экономии времени и ресурсов для постановки диагноза пациенту, данные денежные средства могут быть направлены целенаправленно использоваться в научных разработках и осуществления новых открытий в сфере медицины. Предполагается, что обширная база данных, которая накапливается и пополняется, даст возможность ставить быстрые и точные диагнозы и плана лечения. В практику врачей войдет широкое использование систем поддержки принятия решений и экспертных систем нового поколения, позволяющих предоставить врачам всеобъемлющий доступ к опыту коллег посредством анализа электронных медицинских паспортов пациентов как в локальном варианте, так и в масштабе страны. Это, в свою очередь, позволит минимизировать субъективный человеческий фактор при принятии врачебных решений о стратегии лечения пациента. [40]

В ходе работы были выполнены поставленные задачи. Технология больших данных появилась сравнительно недавно и становится одной из лучших технологий последнего десятилетия. Хранение и обработка больших данных требует не только хороших вложений денежных средств, но и специалистов, которых на данный момент не так много, как необходимо. Применение данной технологии возможно во всех сферах, начиная от коммерческого сектора, заканчивая медициной.

Логико-структурный подход в управлении проектами весьма эффективен на всех фазах жизненного цикла проекта, особенно при идентификации, разработке и мониторинге проекта и широко используется в разнообразных проектах, осуществляемых многими международными, правительственными, коммерческими организациями. Основной процесс является внедрение технологии в сферу здравоохранения, где входами является неавтоматизированная система медицинских данных, а выходом автоматизированная система медицинских данных. Управляющими элементами в данной диаграмме являются техническое задание на введение технологии больших данных в сферу здравоохранения, законы РФ о персональных данных, требования Министерства здравоохранения РФ и пациентов, а также стандарты больших данных. Так же основные ресурсы необходимые для реализации данного проекта: финансирование, персонал, оборудование и информационные системы.

Применение больших данных в сфере здравоохранения принесет не только упрощение работы медицинских организаций, но и так же даст возможность

открывать новые способы лечения и даст возможность сократить время исследований многих заболеваний.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Технология больших данных [Электронный ресурс]: Технологии и инновации // Журнал «Право. Журнал Высшей школы экономики». – Режим доступа: <https://law-journal.hse.ru/>
2. Государственная программа Российской Федерации «Развитие здравоохранения» [Электронный ресурс]: Программы // Министерство здравоохранения Российской Федерации. – Режим доступа: <https://www.rosminzdrav.ru/ministry/programms/health/info>
3. Где хранить Большие Данные [Электронный ресурс]: Технологии // Computerworld Россия. – Режим доступа: <https://www.osp.ru/cw/2017/07/13052036/>
4. Big data [Электронный ресурс]: AOV // Colltouch. – Режим доступа: <https://www.calltouch.ru/glossary/big-data/>
5. Обзор технологии «Большие данные» (Big Data) и программно-аппаратных средств, применяемых для их анализа и обработки [Электронный ресурс]: Статьи // КиберЛенинка. – Режим доступа: <https://cyberleninka.ru/article/n/obzor-tehnologii-bolshie-dannye-big-data-i-programmno-apparatnyh-sredstv-primenyaemyh-dlya-ih-analiza-i-obrabotki/viewer>
6. Big Data [Электронный ресурс]: Технологии и инновации // ITenterprise. – Режим доступа: <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>
7. Принцип обработки распределенных данных [Электронный ресурс] Классификация больших объемов данных на Apache Spark с использованием произвольных моделей машинного обучения // Хабр. – Режим доступа: <https://habr.com/ru/company/lanit/blog/413137/>
8. Сигель, Э. Просчитать будущее: Кто кликнет, купит, соврёт или умрёт. Предиктивная аналитика / Э. Сигель. — Москва: Альпина Паблишер, 2016. — 374 с.
9. Manyika, J. Big data: The next frontier for innovation, competition, and productivity / J. Manyika // McKinsey Global Institute. – 2016. – № 10. – С. 100-112.
10. Бабурин, В. А., Яненко, М. Е. Технологии Big Data в сервисе: новые рынки, возможности и проблемы / В. А. Бабурин, М. Е. Яненко // ТТПС. – 2014. – № 1 (27). – С. 100–105.
11. Уиллемс, С. М. Потенциальное применение технологии больших данных в онкологии / С.М. Уиллемс // Онкология ротовой полости. – 2019. – № 98. – С. 8-12.
12. Manyika J., Chui M., Brown B. et al. (2018) Big Data: The Next Frontier for Innovation, Competition, and Productivity / McKinsey Global Institute.
13. Савачук, И. Л. Big Data – технология, рождающая новый тип бизнеса / И.Л. Савачук // Бизнес & информационные технологии. – 2014 – № 3. – С. 36.
14. Big Data serving translational medicine [Электронный ресурс]: Технологии // Internet portal Computerworld Россия. – Режим доступа: <http://www.computerworld.ru/articles/Bolshie-Dannye-na-sluzhbe-translyatsionnoy-medsiny>

15. Цветкова, Л. А. Технология Больших Данных в медицине и здравоохранении России и мира (ФГБУ Дирекция научно-технических программ Минобрнауки России) / Л. А. Цветкова, О. В. Черченко // Мейнстрим – 2016. – №3. – С. 60-73.
16. Цветкова, Л. А. Внедрение технологий Big Data в здравоохранение: оценка технологических и коммерческих перспектив (Центр научно-технической экспертизы ИПЭИ РАНХиГС при Президенте РФ) / Л. А. Цветкова, О. В. Черченко // Экономика науки –2016. – №2 (том 2). – С. 138-150.
17. Россия названа крупнейшим рынком Big Data и бизнес-аналитики в регионе ЦВЕ [Электронный доступ]: В фокусе / DC DailyComm Коммуникации в ИТ-бизнесе. – Режим доступа: <http://www.dailycomm.ru/m/44819/>
18. Seibert, J. A. Modalities and data acquisition / J. A. Seibert // Practical Imaging Informatics – 2018. – № 6 (Chapter 1). – С. 49–66.
19. Всемирный указатель патентов Derwent [Электронный ресурс]: Derwent World Patents Index / DWPI. – Режим доступа: <http://thomsonreuters.ru/products/derwent-world-patents-index/>
20. Lederman, R., Managing Heals Information Delivery Processes for Better Medical Decision Making / S. Dreyfus, R. Lederman // Frontiers in artificial intelligence and applications. – 2017. – С. 329-340.
21. Попытки расчета количества информации на планете Земля [Электронный ресурс]: Редакционные статьи / NagNews. – Режим доступа: <https://nag.ru/articles/article/101906/popytiki-rascheta-kolichestva-informatsii-na-planete-zemlya.html>
22. Healthnet [Электронный ресурс]: Дорожные карты / НТИ. – Режим доступа: www.nti2035.ru
23. Логико-структурный подход в управлении проектами [Электронный ресурс]: Научные и учебно-методические разработки Института инноватики // Институт инноватики. – Режим доступа: <https://by.odt-office.eu/files/docs/Logiko-strukturnyj-podhod-v-upravlenii-proektami.pdf>
24. Стейкхолдеры процедур развития [Электронный ресурс]: Команда и мотивация // Projectimo. – Режим доступа: <http://projectimo.ru/komanda-i-motivaciya/zainteresovannye-storony-proekta.html>
25. Моноган, Ф. Взаимодействие с заинтересованными сторонами: практическое руководство по организации взаимодействия со стейкхолдерами / Ф. Моноган, Т. Крик, М. Форстрейтер. – Манхеттен: AccountAbility, 2015 – 148 с.
26. Методология IDEF0 [Электронный ресурс]: BPWIN // ITteach.ru. – Режим доступа: <https://itteach.ru/bpwin/metodologiya-idef0>
27. Сбалансированная система показателей [Электронный ресурс] История создания // JobGrade.ru. – Режим доступа: <https://www.jobgrade.ru>.
28. Разработка сбалансированной системы показателей. Практического руководства / под ред. А.М. Гершуна, Ю.С. Нефедьевой. – Москва: ЗАО «Олимп-Бизнес», 2004 – 88 с.
29. Сбалансированная система показателей (BSC) [Электронный ресурс]: Практический журнал по управлению финансами // Финансовый

директор. – Режим доступа: <https://www.fd.ru/articles/157266-kontseptualnyu-podhod-k-organizatsii-upravleniya-biznesom>

30. Сбалансированная система показателей [Электронный ресурс]: Экономика бизнеса // Экономика и жизнь. – Режим доступа: <https://www.e-online.ru/article/54460/>

31. О концепции сбалансированной системы показателей [Электронный ресурс]: Практический журнал по управлению финансами // Финансовый директор. – Режим доступа: <https://www.fd.ru/articles/157266-kontseptualnyu-podhod-k-organizatsii-upravleniya-biznesom>

32. Data mining – об интеллектуальном анализе данных [Электронный ресурс]: Технологии // TADVISER. – Режим доступа: <http://www.tadviser.ru/index.php/>

33. Боев, В.Д. Компьютерное моделирование: пособие для практических занятий, курсового и дипломного проектирования в AnyLogic7 / В.Д. Боев. — Санкт-Петербург: ВАС, 2014. — 432 с.

34. Модель обработки запросов сервером [Электронный ресурс]: Концептуальное проектирование систем в Anylogic 7 и GPSS World // ИНТУИТ национальный открытый университет. – Режим доступа: <https://www.intuit.ru/studies/courses/13846/1243/lecture/23951?page=3>

35. Должностная инструкция генерального директора ООО [Электронный ресурс]: Должностные инструкции // Общество профессионалов налог-налог.ру. – Режим доступа: https://nalog-nalog.ru/profstandarty_i_dolzhnostnye_instrukcii/dolzhnostnaya_instrukciya_general_nogo_direktora_ooo_-_obrazec/

36. Должностная инструкция менеджера по продажам, должностные обязанности менеджера по продажам, образец должностной инструкции менеджера по продажам [Электронный ресурс]: Работодателям // Работа.ру. – Режим доступа: rabota.ru/articles/hr/dolzhnostnaja_instruksija_menedzhera_po_prodazham_dolzhnostnye_objazannosti_menedzhera_po_prodazham_obrazets_dolzhnostnoj_instruksii_menedzhera_po_prodazham-3954


37. Менеджер по работе с корпоративными клиентами [Электронный ресурс]: Должностная инструкция // Pandia. – Режим доступа: <https://pandia.ru/text/77/165/6016.php>

38. Должностная инструкция главного бухгалтера [Электронный ресурс]: Должностные инструкции // Административно- управленческий портал AUP.RU. – Режим доступа: <https://clck.ru/P47Xi>

39. Цифровая революция в здравоохранении: достижения и вызовы [Электронный ресурс]: Технологии и инновации // ТАСС – Режим доступа: <https://tass.ru/pmef-2017/articles/4278264>

40. Программы для клиник и медицинских центров [Электронный ресурс]: Медицина // LiveBusiness. – Режим доступа: <https://www.livemedical.ru/tools/clinic>

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт инженерной физики и радиоэлектроники
Кафедра экспериментальной физики и инновационных технологий

УТВЕРЖДАЮ
Заведующий кафедрой

Орлов В.А.
подпись
« 9 » июня 2020 г.


МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Внедрение технологии больших данных (Big Data) в сферу
здравоохранения Красноярского края

27.04.05 Инноватика

27.04.05.01 Управление инновациями

Научный руководитель	 подпись, дата	канд. техн. н., доцент должность, ученая степень	А.В.Вершков инициалы, фамилия
Выпускник	 подпись, дата		М.К.Кыргызбай инициалы, фамилия
Рецензент	 подпись, дата	канд. техн. н., доцент должность, ученая степень	А.В.Селиванов инициалы, фамилия

Нормировано
10.07.2020


Красноярск 2020