

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНСТИТУТ МЕЖДУНАРОДНОГО БИЗНЕСА,  
ЭКОНОМИКИ И УПРАВЛЕНИЯ  
КАФЕДРА ЭКОНОМИКИ И УПРАВЛЕНИЯ

## ОТЧЕТ

по учебной практике по получению навыков  
исследовательской работы

Большие данные и их использование в современном  
бизнесе

Студент  
группы БМН-21-3

  
\_\_\_\_\_

А.А. Коновалов

Руководитель

канд. экон. наук, доцент

  
\_\_\_\_\_

Н.Р. Пашук

Нормоконтролер

канд. экон. наук, доцент

  
\_\_\_\_\_

Н.Р. Пашук



19.09.22



Владивосток 2022

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ЭКОНОМИКИ И СЕРВИСА  
ИНСТИТУТ МЕЖДУНАРОДНОГО БИЗНЕСА, ЭКОНОМИКИ И УПРАВЛЕНИЯ  
КАФЕДРА ЭКОНОМИКИ И УПРАВЛЕНИЯ  
ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ  
на учебную практику по получению навыков исследовательской работы

Студент: Коновалов Александр Александрович

Группы: БМН-21-3

Срок сдачи: 14.06.2022 – 25.06.2022

Содержание отчета по учебной практике по получению навыков исследовательской работы:

**Введение:** определить цель и задачи практики, основные методы, необходимые для их достижения (Объем – 1 страница)

**Раздел 1. Характеристика исследуемой проблемы по теме «Большие данные и их использование в современном бизнесе»**

Краткое содержание исследуемой проблемы и ее актуальность, степень разработанности Исследуемой проблемы (перечень авторов, внесших вклад в решение проблемы; отражение проблемы в государственных нормативных документах и т.п.); цель и задачи исследования (УК-1.1в, УК-1.3в).

**Раздел 2. Современное состояние исследуемой проблемы.**

Сущность исследуемой проблемы в авторском изложении с иллюстрацией, статистическим и аналитическим материалом, перспективы дальнейших исследований по данной теме (УК-1.1в). (Объем двух разделов – 10-12 страниц)

**Заключение.** В заключении обобщается изложенный в отчете материал, делаются выводы.

(Объем – 1-2 страницы)

**Список использованных источников** (включаются источники не старше 2017 года).

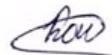
Составить список литературы с использованием Интернет-ресурсов, электронно-библиотечных систем, справочно-консультационных систем.

Руководитель  
канд.экон.наук, ст.преподаватель кафедры ЭУ



Н.Р. Пашук

Задание получил:



А.А. Коновалов

## РАБОЧИЙ ГРАФИК (ПЛАН) ПРОВЕДЕНИЯ ПРАКТИКИ

Студент – Коновалов Александр Александрович  
Кафедра - экономики и управления

гр. БМН-21-3

Руководитель практики – Пашук Н.Р.

Этапы практики	Виды работы	Срок выполнения	Отметка руководителя о выполнении
1. Подготовительный	Организационное собрание	07.02-21.02	<i>выполнено</i>
2. Исследовательский	Формулировка целей и задач исследования Анализ содержания исследуемой проблемы, степени ее актуальности и разработанности; формулировка цели исследования; формулировка задач исследования; разработка плана исследования	22.02-31.03	<i>выполнено</i>
3. Аналитический	Подбор и анализ информации по теме исследования Определение списка источников, необходимых для решения поставленных задач; сбор, систематизация и анализ информации; формулировка выводов; определение возможных направлений дальнейших исследований по выбранной теме	01.04-13.06	<i>выполнено</i>
4. Заключительный	Подготовка, оформление и сдача отчета по практике	14.06-25.06	<i>выполнено</i>

Руководитель учебной практики  
канд.экон.наук, ст. преподаватель кафедры ЭУ



Н.Р. Пашук

## Содержание

Введение	3
1 Характеристика исследуемой проблемы	4
2 Современное состояние исследуемой проблемы	6
Заключение	11
Список используемой литературы	12

## Введение

Ни для кого не секрет, что сегодня количество данных, которые необходимо хранить и обрабатывать, растет в геометрической прогрессии. Например, объем данных, хранящихся в Интернете, ежегодно увеличивается примерно на 35-40%. Интересно, что, с одной стороны, именно развитие современных информационных технологий позволяет и способствует тому, что объемы хранимых и обрабатываемых данных непрерывно растут. С другой стороны, для работы с быстро растущими объемами всевозможных данных требуется все больше ресурсов и все более сложных программных решений. Одной из самых современных и быстро набирающих популярность технологий является “большие данные”. Термин начал использоваться и быстро завоевал популярность лишь 7-10 лет назад. И сегодня крупнейшие мировые компании, занимающие лидирующие позиции в различных сферах бизнес деятельности, вкладывают миллиарды долларов в развитие этой сферы. В данной статье мы рассмотрим ключевые понятия, связанные с технологией больших данных, почему одни крупные компании готовы вкладывать огромные средства в развитие этой области, а другие нет. Каковы наиболее распространенные решения на основе больших данных на рынке сегодня, каковы проблемы и что ждет остальную часть ИТ-рынка с появлением новых технологий.

Цель – определить характеристику использования больших данных в современном бизнесе.

Задачи:

1. определить значение термина «большие данные»;
2. определить уровень разработанности проблематики исследования;
3. определить современное состояние исследуемой проблемы.

Методы исследования: анализ, общенаучный метод, описание.

## 1 Характеристика исследуемой проблемы

Основываясь на определении больших данных, можно сформулировать три ключевых принципа работы с такими данными: горизонтальная масштабируемость. Поскольку объем данных постоянно и быстро растет, а информация может быть столько, сколько вы хотите, система, требующая обработки этих данных, должна быть расширяемой. Например, если объем данных увеличивается в 2 раза, он должен иметь возможность увеличить пропускную способность оборудования в кластере в 2 раза, и система должна продолжать работать без потери производительности. Рассмотренный выше принцип горизонтального масштабирования подразумевает, что в кластере может быть много машин. Например, в компании Yahoo кластер содержит больше 45000 машин. При этом допускается, что доля этих машин будет постоянно выходить из строя. Методы работы с большими данными должны учитывать возможность таких сбоев и поддерживать работоспособность системы без каких-либо значимых последствий. В крупных распределённых системах, используемые данные хранятся на большом количестве машин. Но если данные находятся физически на одном сервере, а обрабатываются на другом, то ресурсы, требуемые на передачу данных, могут превзойти расходы на обработку данных. Соответственно при проектировании решений на big data одним из важнейших принципов является принцип локальности данных, сущность которого заключается в том, чтобы данные обрабатывались и хранились на одной и той же машине.

Исследования, связанные с большими данными, являются актуальной областью разработок в наше время. Они изучают большие данные и способы извлечения из них знаний. Они проводятся в различных дисциплинах и областях, таких как. Информатика, моделирование неопределенности, машинное обучение, статистическое обучение, распознавание образов, методы хранения данных, обработка сигналов и т.д. Исследования больших данных также имеют свои проблемы и задачи. Давайте рассмотрим ситуацию в

академических кругах, посвященных Big Data. Рассмотрим основные кластеры проблем исследования больших данных.

В рамках первой группы проблем рассматриваются проблемы, связанные с хранением и анализом больших данных. Их изучают З. Хуанг, Т. К. Дас, П. М. Кумар, Д. П. Ачарья и М. Р. Патра.

Проблема хранения связана с увеличением скорости создания новых данных в последние годы. Благодаря мобильным устройствам, Интернету вещей, растущей доступности Интернета и другим факторам объем производимой информации увеличивается в геометрической прогрессии. Из-за нехватки места для хранения они либо удаляются, либо не сохраняются. В связи с этим возрастает роль носителей информации и скорости ее записи и чтения для доступности больших данных с целью их анализа. Несмотря на достижения в этой области, такие как, например, распространение твердотельных накопителей, до сих пор не достигнута необходимая производительность накопителей для обработки больших данных.

Помимо количества производимой информации, в последние годы увеличилось и их разнообразие, что существенно усложняет задачи, связанные с анализом больших данных. Возникает необходимость сжатия или выборки обрабатываемых данных, так как существующие методы и алгоритмы не позволяют провести их анализ в разумные сроки. Автоматизация этого процесса, в том числе с помощью машинного обучения, — задача, стоящая перед учеными.

Новейшие технологии, такие как Hadoop и MapReduce, позволяют в разумные сроки собирать большие объемы частично отсортированных и несортированных данных. Чтобы их больше обрабатывать, их нужно выстроить в очередь. Разработка алгоритмов сортировки данных также является неотложной задачей.

Вторая группа исследовательских задач рассматривает проблемы извлечения знаний из больших данных и вычислительную сложность

обработки больших данных. Их изучают исследователи Аль-Джаррах, Мухаидат С. и Карагианнидим Г.К.

## 2 Современное состояние исследуемой проблемы

Мировой рынок технологий Big data растет на 25-30% в год, и к 2022 году, по данным IDS, он должен составить около 120 миллиардов долларов США. По данным другого аналитического агентства Wikibon, объем мирового рынка больших данных в 2018 году достигнет 127 миллиардов долларов США, а к 2024 году вырастет до 203 миллиардов долларов США.

Обработка «больших данных» позволяет выявлять модели, которые недоступны исследователю с использованием традиционных инструментов анализа. Широко известен эксперимент по сбору «больших данных», используемый производителем электромобилей Tesla, где источником информации являются покупатели электромобилей, а большая часть информации создается самим продуктом (он записывает ходовые качества, отказы оборудования, маршруты и др.). Также имеется опыт ритейлеров по определению характеристик покупателей на основе обработки данных групп покупателей в социальных сетях, изучения журналов активности с сайтов организации и анализа различных данных из сети Интернет.

Современные экономические исследования показали, что следующие отраслевые организации лучше всего подходят для использования больших данных.

1. Интернет-магазины со значительным количеством собственной информации в электронном виде намного больше, чем при реализации традиционных форм торговли. Таким образом, всемирно известных торговых площадок, такие как Amazon, eBay и другие, с момента своего создания распоряжаются информацией «больших данных», полученной непосредственно в информационных системах и ресурсах этих организаций, с целью увеличения продаж.

2. Ориентированные на инновации физические магазины, собирающие персонализированную массовую информацию о покупателях (на основе программ лояльности), а также информацию об ожидаемых изменениях внешней среды. Например, Tesco завоевала лидерство на рынке, внедрив

дополнительные программы лояльности, которые предотвращают отказ от целевых клиентов. Кроме того, эта компания активно использует информацию «больших данных» для оперативной корректировки текущего ассортимента (прогнозы погоды и т. д.).

Источниками «больших данных» являются различная информация, поступающая через: Интернет (социальные сети, сайты организаций и государственных органов, чаты, форумы, блоги и т. д.), операторы мобильной связи, интеллектуальные системы сбора данных (сенсоры, датчики, RFID-метки, звукозаписывающие устройства, носители видео- и фотоматериалов, спутники) и др.

Массив «больших данных» в сфере Интернета является наиболее значимой и формируется путем регистрации информации: посещения веб-сайтов, переходы на веб-сайты, по ключевым словам, источникам переходов, времени, проведенному на веб-сайтах, выбранным товарам, повторным покупкам, отказу от заказа, сервису, удовлетворенность, постоянные клиенты, данные о посетителях (регионы нахождения, пол, возраст, ежедневная активность, интересы) и т. д. При прямой покупке и послепродажном обслуживании программные продукты, обеспечивающие этот процесс, могут собирать более 500 видов различных данных.

Главной проблемой использования «больших данных» сегодня является несоответствие форматов различных источников информации, поскольку в состав этих данных входят текстовые документы, веб-журналы, видео, машинные коды, геопрограммная информация и т. д. Еще одна проблема — значительный объем секретной информации в части больших данных. В результате имеется огромное количество разного рода данных, но использовать их не представляется возможным даже для постановки задач, формулирования гипотез и оценки информации, позволяющей установить взаимосвязи и выявить закономерности.

Современные специалисты, работающие с «большими данными», делают некоторые попытки обработки имеющейся информации, что находит

отражение в публикациях по данной теме. Среди методов анализа «больших данных» относительно редко используются в экономических исследованиях A/B-тестирование, интеллектуальный анализ данных и т. д., а также широко известные экономические, математические и статистические методы, регрессия, анализ временных рядов, кластерный анализ и т. д.

Современные структуры компаний самостоятельно разрабатывают системы использования «больших данных», которые создают сами. В этих системах есть три компонента, в том числе:

- 1) инфраструктура (сервер, как правило, «облачный»);
- 2) функция автоматизированного сбора данных (поисковики для поиска данных, различные веб-ресурсы в Интернете, мобильные приложения, группы в социальных сетях и т. д.);
- 3) Инструменты анализа для обработки данных (программные продукты для анализа, поиска закономерностей, вывода и визуализации результатов).

Самым известным в мире инструментом для обработки «больших данных» является специальный программный комплекс Hadoop. На основе него создаются приложения для анализа и визуализации больших объемов информации. Еще один популярный инструмент для работы с большими данными — облачный сервис Microsoft Azure.

Стоит отметить, что существующие инструменты анализа данных, которые в настоящее время являются общедоступными, имеют определенные ограничения в их использовании. Например, известные инструменты анализа данных «Google Analytics» и «Яндекс Аналитика», принадлежащие одноименным поисковым системам, требуют от организации наличия веб-сайта для их использования и, следовательно, результатов аналитического исследования. ограничиваются информацией, сгенерированной на основе данных обработки информации на сайте. Более подробные исследования по изучению и систематизации поисковых систем на разных платформах содержатся в современных научных работах по этой теме. С момента появления «больших данных» особенно возрастает роль использования социальных сетей,

что нашло отражение в многочисленных работах на эту тему. В целях совершенствования формирования информации, развития предпринимательской деятельности решающую роль играют и другие многочисленные интернет-ресурсы, систематизированные в современных работах.

На сегодняшний день задача создания и предоставления «больших данных» заинтересованным пользователям в России является достаточно деликатной. Согласно отчету Департамента информационных технологий Правительства Москвы (далее - ДИТ Москвы) «Мониторинг развития интеллектуальных практик и технологий в мировых мегаполисах» за 2017 год, использование «больших данных» предприятиями может повысить эффективность предпринимательства и увеличить налоговые поступления в бюджет. В этом же документе отмечается, что "большие данные" Москвы имеют показатель проникновения в экономику 45%, что является лучшим показателем для России.

Следует отметить, что город Москва стал одним из первых субъектов федерации в России, где еще в 2013 году был создан государственный ресурс для анализа «больших данных», а именно портал «Открытые данные». Сейчас в России разрабатывается концепция создания и развития 50 умных городов России, которые должны стать основными центрами развития цифровой экономики, а значит, и развития и практического применения «больших данных» в нашей стране. Однако пока в России всего 18 городов, которые в той или иной мере могут быть отнесены к категории «умных городов», по данным Института развития Интернета (IRI). Эти города выбраны с учетом наличия необходимого уровня развития транспорта, жилищно-коммунального хозяйства, энергетики, промышленности и электронных государственных услуг.

При этом, согласно российским нормативным документам, около 85-90% муниципальных образований страны в настоящее время не соответствуют необходимым требованиям по уровню цифровизации. Поэтому возможности

субъектов Российской Федерации по развитию умных городов достаточно противоречивы.

Со стороны государства субъектам хозяйствования оказывается определенная информационная поддержка в виде предоставления различных видов данных, в том числе в виде «открытых данных». К такой информации относятся сведения о нормативно-правовой базе регулирования взаимоотношений хозяйственных контрагентов, системе электронного документооборота для взаимодействия с государственными органами и иными юридическими и физическими лицами, электронных государственных услугах, площадках для электронных платежей, данные о государственной поддержке бизнеса, данные наборы (датасеты), раскрывающие информацию о работе государственных органов (или направлений деятельности), и некоторые другие.

## Заключение

Технологии больших данных — это и вызов, и возможность. Полнота управления данными, возможности многоканальной интеграции данных и возможности анализа данных станут способностью предприятий достигать устойчивого развития. Понимание того, как большие данные увеличивают конкурентное преимущество компании и трансформируют бизнес-модели, является необходимым процессом для осознания ценности больших данных.

В современном бизнесе, практически независимо от специфики и отрасли, все более очевидными становятся ценность и высокая роль информации о потенциальных и текущих клиентах компании, ее конкурентах, будущих тенденциях рынка. Все чаще это становится необходимым условием для поддержания конкуренции в современном мире. В связи с этим и уже имеющимися примерами успеха внедрения больших данных крупными компаниями, заполняющими Интернет, можно предположить, что роль больших данных со временем будет только возрастать. В результате компании будут лучше знать и понимать потребности своих клиентов и предлагать им наиболее актуальные и подходящие решения, а потребители смогут пользоваться продуктами и услугами, которые лучше всего им подходят.

## Список использованной литературы

1 Карашук О.С., Майорова Е.А., Прохоров Ю.Н. "Большие данные" и перспективы их использования в предпринимательской деятельности // Вестник НГИЭИ. - 2018. - №10. - С. 77-87.

2 Коновалов, М. В. Big Data. Особенности и роль в современном бизнесе / М. В. Коновалов. — Текст: непосредственный // Технические науки: проблемы и перспективы: материалы VI Междунар. Науч. Конф. (г. Санкт-Петербург, июль 2018 г.). — Санкт-Петербург: Свое издательство, 2018. — С. 8-10.

3 Федорова Л.А., Ху Гуйюй, Хуан Сяоянь, Землякова С.А. Применение технологий big data в деятельности современных предприятий // Вестник Алтайской академии экономики и права. – 2020. – № 9-2. – С. 322-329.

4 Big Data – что такое системы больших данных? Развитие технологий Big Data [Электронный ресурс]. – URL: <https://promdevelop.ru/big-data/> (дата обращения: 15.07.2022).

5 Шпрингер, Е. Технологии big data: как анализируют большие данные, чтобы получить максимум прибыли [Электронный ресурс] / 120 Е. Шпрингер. – URL: <https://mcs.mail.ru/blog/tekhnologii-big-data-kakanaliziruyut-bolshie-dannye> (дата обращения: 18.07.2022).

6 Что такое Big data: собрали все самое важное о больших данных [Электронный ресурс]. – URL: <https://rb.ru/howto/что-такое-big-data/> (дата обращения: 19.06.2022).

7 Что скрывается в тени больших данных? [Электронный ресурс]. – URL: <https://blog.fts-eu.com/ru/2016/01/26/> (дата обращения: 19.07.2022).

8 Хранилище данных [Электронный ресурс]. – URL: <https://www.intuit.ru/studies/courses/599/455/lecture/10155> (дата обращения: 20.07.2022).

9 Революция Big Data : Как извлечь необходимую информацию из «Больших Данных»? [Электронный ресурс]. – URL: <http://statsoft.ru/products/Enterprise/big-data.php> (дата обращения: 20.07.2022).

10 Пять вещей, которые необходимо знать о Hadoop и Apache Spark [Электронный ресурс]. – URL: <https://www.osp.ru/news/articles/2015/49/13048137> (дата обращения: 21.02.2022).

11 BIG DATA 2017: Где хранить Большие Данные [Электронный ресурс]. – URL: <https://www.computerworld.ru/articles/BIG-DATA-2017-Gde-hranit-Bolshie-Dannye> (дата обращения: 25.07.2022).