

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Гуманитарный институт

институт

информационных технологий в креативных и культурных индустриях

кафедра

УТВЕРЖДАЮ

Заведующий кафедрой

_____ А. В. Усачев

« _____ » _____ 2022 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 «Прикладная информатика в области искусства и гуманитарных
наук»

код – наименование направления

Пропуски культурно значимых данных среди цифровых копий книг из

Тема

каталога Минусинской общественной библиотеки

Руководитель _____

подпись, дата

ст. преподаватель _____

должность, ученая степень

И.А. Кижнер _____

инициалы, фамилия

Выпускник _____

подпись, дата

И.А. Елистратова _____

инициалы, фамилия

Красноярск 2022

Продолжение титульного листа БР по теме Пропуски культурно значимых данных среди цифровых копий книг из каталога Минусинской общественной библиотеки

Нормоконтролер

Е. Р. Брюханова

СОДЕРЖАНИЕ

Введение.....	3
1 Проблема пропусков и визуализация культурно значимых данных в цифровых копиях.....	5
1.1 Литературный корпус как основа для исследования пропусков культурных данных.....	6
1.2 Визуализация неопределенностей.....	6
1.3 Метод визуализации неопределенностей Оливии Вейн.....	8
1.4 Проблема неопределенности в гуманитарных науках	9
2 Культурно значимые данные в цифровых копиях книг Минусинской общественной библиотеки	10
2.1 Минусинская общественная библиотека.....	11
2.2 Анализ стратифицированной выборки при помощи агрегаторов.....	14
2.3 Метод визуализации пропусков культурно значимых данных раздела «Книги для детского чтения»	27
3 Результаты исследования.....	29
3.1 Яндекс таблицы.....	29
3.2 Процентное соотношение «доступности» книг	30
3.3 Визуализация данных раздела «Книги для детского чтения».....	35
Заключение.....	40
Список использованных источников.....	41
Приложение А «Яндекс таблицы».....	43

ВВЕДЕНИЕ

Минусинская общественная библиотека - культурно-образовательное учреждение, основанное в 1877 году. На основе анализа оцифрованных книг, представленных в каталоге библиотеки, можно выяснить, что интересовало сибирских читателей, какие идеи, люди и места были популярны в этой части мира далекой от культурных центров. Однако, действительно ли полученные цифровые копии книг станут основой цифрового анализа в гуманитарных науках? В этой работе я утверждаю, что доступных нам данных недостаточно для того, чтобы составить целостную историческую картину. Также данных не будет достаточно для того, чтобы определить, какие страны интересовали читателей, что они знали о мире.

Выпускная квалификационная работа посвящена пропускам культурно значимых данных среди цифровых копий книг Минусинской общественной библиотеки.

Достоверность результатов, полученных в ходе научного исследования, обусловлена использованием официальных ресурсов и источников.

Проблема исследовательской работы заключалась в том, что в записях каталога Минусинской общественной библиотеки были неполные данные об авторах книг. Это затрудняло поиск, а в некоторых случаях книги так и не удалось найти.

Объект исследования – цифровые копии книг из каталога Минусинской общественной библиотеки.

Предметом исследования являются пропуски культурно значимых данных среди копий книг из каталога Минусинской общественной библиотеки.

Цель настоящей работы – выявить пропуски культурно значимых данных среди цифровых копий книг из каталога Минусинской общественной библиотеки.

Задачи исследования:

- проанализировать стратифицированную выборку разделов библиотеки с помощью агрегаторов;
- выявить процентное соотношение категорий доступности в каждом из исследуемых разделов;
- визуализировать процентные соотношения и результаты;
- сравнить полученные результаты с результатами других исследований.

Выпускная квалификационная работа состоит из введения, трёх глав, заключения, списка использованной литературы и приложения.

Первая глава «Проблема пропусков и визуализация культурно значимых данных в цифровых копиях» кратко предоставляет информацию научных работ и диссертаций, которые были взяты за основу при работе над этой квалификационной работой.

Вторая глава «Культурно значимые данные в цифровых копиях книг Минусинской общественной библиотеки» описывает структуру записей в каталоге Минусинской общественной библиотеки, используемые агрегаторы для поиска книг, программное обеспечение, в котором велись все таблицы с данными, а также где и по какому методу создавалась визуализация.

Третья глава «Результаты исследования» включает описание и обсуждение сформированных таблиц с данными по книгам, их классификации доступности, а также процентное соотношение доступных книг в каждом из исследуемых разделов и визуализацию.

Заключение подводит итоги исследования. Список литературы даёт библиографическое описание цитируемых в работе источников. В приложении представлены иллюстрации: сформированные табличные данные о книгах.

Актуальность исследования заключается в росте темпов оцифровки, которая открывает возможности для новых исследований документов, к которым ранее не было доступа. Это также способствует необходимости выявлять новые и актуальные способы визуализации этих данных.

1 Проблема пропусков и визуализация культурно значимых данных в цифровых копиях

1.1. Литературный корпус как основа для исследования пропусков культурных данных

Для того, чтобы проводить гуманитарные исследования, в наше время доступны компьютерные технологии. Они позволяют обрабатывать огромные объемы текстов, сокращая время для их изучения. «Сочетание компьютерных технологий и нового взгляда на языковые феномены привело к возникновению новой науки – корпусной лингвистики». [1] Этот раздел языкознания занимается разработкой, созданием и использованием текстовых корпусов.

Корпус – это собрание текстов, ориентированное на изучение и анализ языка. Корпуса позволяют решать поисковые задачи по подбору материала, технологические задачи для сравнения версий текстов, а также корпуса помогают в проведении количественных исследований. Зачастую корпуса хранятся в репозитории. Репозиторий — это ресурс для хранения и публикации научных данных, корпусов оцифрованных литературных текстов и библиографических баз данных, которые авторы предоставляют в свободный доступ другим исследователям. Главными задачами репозитория являются повышение видимости и доступности данных, а также поддержание культуры цитирования данных. Все публикуемые дата сетсы проходят рецензирование и техническую подготовку для того, чтобы обеспечить достоверность, и полноту данных.

Публикация открытых данных в машиночитаемых форматах расширяет возможности цифровых и количественных исследований литературы, делает более доступной кропотливо собранную справочную информацию. Что немаловажно, те же данные могут послужить удобным материалом для

студенческих проектов в области Digital Humanities. [2] Одним из примеров репозитория является сайт Национального корпуса русского языка¹.

Национальный корпус русского языка содержит тексты периода от первых восточнославянских памятников (XI век) до первых десятилетий XXI века. Этот корпус включает в себя язык прошлых эпох и современный в разных вариантах — литературном, разговорном, просторечном, диалектном. В корпус включены и представляют большую часть произведения художественной литературы (проза, драматургия, поэзия, звучащая речь кино), имеющие культурную значимость, а также представляющие интерес с точки зрения языка. Однако, Национальный корпус включает в себя и другие жанры литературы: мемуары, публицистика, научно-популярная и научная литература, публичные выступления и частная устная речь, переписка, дневники, частные документы. Все тексты имеют свою особую разметку, которая зависит от специфики задач для определенного жанра литературы.

Так, например, поэтический корпус служит основой для стиховедческих исследований, поэтому в нем есть особая разметка, связанная с ключевыми для стиховедения понятиями — метром и ритмом. Акцентологический корпус посвящен истории русского ударения, играющего ключевую роль в его разметке. Аннотация устного корпуса также отражает ударения и иную специфику звучащей речи. [2]

Корпуса предоставляют большие массивы данных для исследований, однако мы не можем быть уверены в том, что эти массивы являются абсолютно полными и репрезентативными. Дело в том, что в массивах данных могут содержаться пропуски и неопределенности. Мы не можем утверждать, что обобщение, которое делается на их основе является абсолютно точным и валидным. Есть несколько способов, которые позволяют выявить и рассмотреть неопределенности, один из них — это визуализация.

¹ <https://ruscorpora.ru/>

1.2 Визуализация неопределенностей

Темпы оцифровки коллекций культурными учреждениями растут, следовательно, объем доступных культурных данных тоже увеличивается. [3] Хотя основная причина, по которой учреждения оцифровывают свои фонды, заключается в расширении и увеличении доступа к материалам, также существует гипотеза, что, будучи цифровыми, эти материалы можно читать, просматривать, изучать и т.д. новыми и творческими способами. [4] Следовательно, вместе с этим растет понимание того, что нам нужно извлечь из них как можно больше пользы. Простая оцифровка не делает коллекции доступными для анализа, а стандартные интерфейсы не обязательно поддерживают тот тип взаимодействия, который пользователям удобен.

С этими проблемами помогает справиться визуализация данных. При помощи интерактивного визуального представления, созданного с помощью программного обеспечения, визуализация позволяет сделать краткий обзор, увидеть закономерности и продемонстрировать богатство оцифрованных коллекций. [5] Визуализация может упростить анализ, изучение и презентацию коллекций для различных аудиторий: научных исследователей, работников культурных учреждений и широкой общественности. [6] Однако, существуют различные подходы к визуализации, выбор которых зависит от типа визуализируемых данных. Визуализация неопределенности - тема, которая становится все более актуальной и интересной для сообщества визуализации.

Неопределенность является стандартным условием, при котором работает большая часть коллекционеров, коммуникаторов и хранителей культурных коллекций. [7] Причинами неопределенности могут быть: первоначальная неопределенность в документации объекта или неточность при идентификации и интерпретации объекта. Практически каждый тип информации об объекте может быть неточным, его можно оспорить, или он может отсутствовать полностью. Очевидно, что такие недостатки затем влияют на каждый дальнейший этап обработки данных.

Тем не менее, известно, что неопределенность также связана с культурными объектами на уровне интерпретации и исследования: что на самом деле представляет собой объект? Какова его ценность или почему он имеет отношение к делу? Визуализация неопределенности стремится к тому, чтобы пользователи могли получить более полное представление о данных, чтобы можно было принимать во внимание ограничения, связанные с исследованиями. [8]

1.3 Метод визуализации неопределенности Оливии Вейн

В своей докторской диссертации [9] Оливия Вейн уделяет особое внимание визуализации культурных данных, представленных на шкале времени. Принцип визуализации данных, предложенный Оливией Вейн, основан на расположении изображений экспонатов на линии времени. Он подходит для исторических данных, так как учитывает возможные временные неопределенности в их описании.

Пример визуализации коллекции Купера Хьюитта из Смитсоновского музея на рисунке 1 демонстрирует авторский принцип О. Вейн. Большинство предметов из этой коллекции, имеют неточную дату, поэтому был выбран оптимальный подход, который позволил избежать визуализации, которая искажает даты происхождения предметов, для которых они расплывчаты. Выбранный подход заключается в расположении изображений в случайном месте в пределах диапазона их дат.

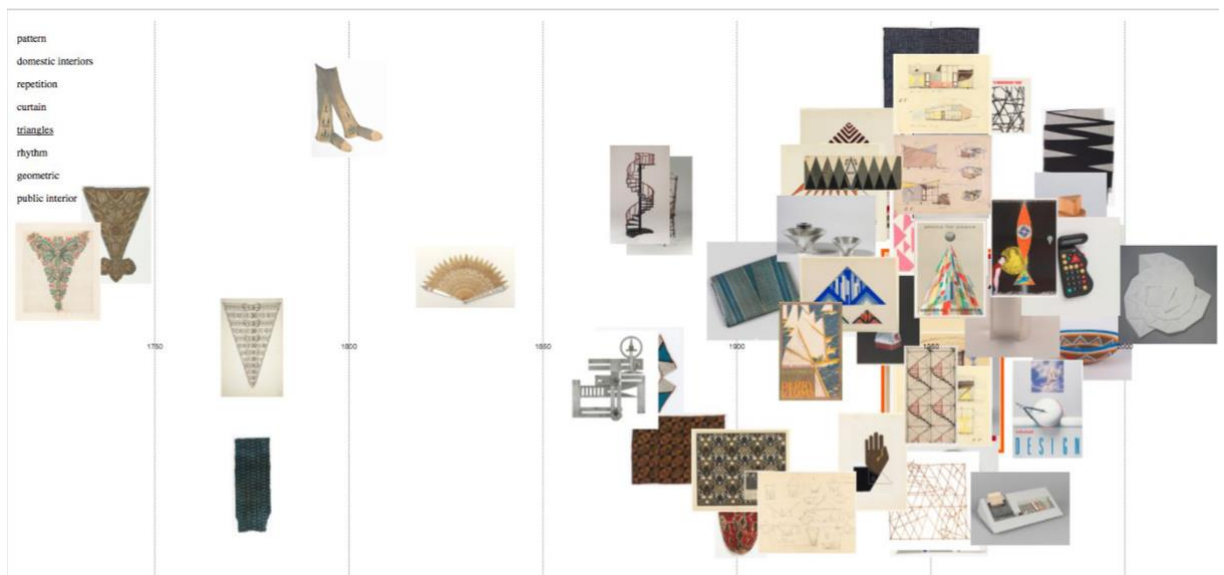


Рисунок 1 – Прототип визуализации коллекции Смитсоновского музея. Экспонаты расположены горизонтально по дате, в случайном месте в пределах диапазона их дат. Источник: Olivia Vane. Timeline design for visualising cultural heritage data.

Однако, результат больше напоминает коллаж, чем график, побуждая пользователя только смотреть, а не изучать. Поэтому в визуализации О. Вейн, пример которой представлен на рисунке 2, добавляет возможность просмотра изображения в большем размере. При этом действии также появляется информационное окно с названием объекта, датой, тегами и гиперссылкой на его полную учетную запись на веб-сайте коллекции.



Рисунок 2 – Прототип визуализации коллекции Смитсоновского музея. Подробная информация о выбранном экспонате (Название, страна, дата, теги и гиперссылка на его полную запись на веб-сайте коллекции). Источник: Olivia Vane. Timeline design for visualising cultural heritage data.

1.4 Проблема неопределенности в гуманитарных науках (Пропуски культурно значимых данных)

Зачастую, гуманитарные исследования подвергаются сомнениям и вызывают вопросы доверия к данным и вычислениям. Это связано с наличием субъективности, отсутствием точности и конфликтов мнений, которые свойственны изучаемым материалам и гуманитарным наукам. [10]

Например, проблемы упущений и предвзятости в данных проявляются в истории культуры в связи с политикой и социальными особенностями времени, к которому они относятся. М. Мольтруп [11] в своей работе указывает на недостаточную представленность женщин в архивах графического дизайна, а Л. Кляйн [12] обращает внимание на рабыню Томаса Джефферсона Хемингс, которая никогда не фигурирует в переписке Т. Джефферсона напрямую, но о которой много пишут через контекст. И хотя не все упущения и предвзятости в культурных данных являются настолько политическими, они по-прежнему

имеют большое значение для того, чтобы свести к минимуму процент неточностей.

Когда субъективность и неточность являются определяющими характеристиками, это особым образом влияет на доверие. Когда ученые предоставляют данные, их обязанность состоит в том, чтобы выявить степень, в которой они могут быть ненадежными. [13]

Когда мы сталкиваемся с цифровым отображением данных, как мы можем убедиться в том, что оно включает в себе абсолютно все данные?

Например, в Лондонском музее Виктории и Альберта Оливия Вейн работала над коллекцией Королевского фотографического общества, которая находилась в процессе оцифровки. Коллекция включала более 270 000 фотографий. Вейн начала работать с коллекцией достаточно рано, в то время было оцифровано всего 2% фотографий. В результате преобладали альбуминовые печати² и дагерротипы³ просто потому, что они были оцифрованы первыми.

Результаты визуализации не отражали общую картину коллекции. Ее могли знать только люди, непосредственно работающие над ней. Визуализация Оливии Вейн не была неправильной, но то, что она демонстрировала, требовало дополнительных сведений и уточнений.

Другой пример – пятилетний проект «Living with Machines», использующий цифровые коллекции для изучения жизненного опыта индустриализации в Великобритании XIX века. [14] Исследование должно было происходить на основе оцифрованных газет в коллекции Британской библиотеки. Онлайн-доступ к более чем 33 миллионам газетных страниц может означать репрезентативную выборку, но это всего лишь около 6% от физической коллекции газет (к тому же неизвестен процент газет, опубликованных

² Альбуминовая печать — фотографический процесс, основанный на использовании альбумина, получаемого из белков яиц, для удержания светочувствительных веществ на бумаге.

³ Дагерротипия — ранний фотографический процесс, основанный на светочувствительности иодистого серебра.

изначально). Таким образом обнаружилась существенная недостаточность представления прессы.

Зачастую, при разработке визуализации, возникает стремление опустить данные для того, чтобы создать более четкую картину. [15] Такой метод рискует представить ложный взгляд на исторические события. Для того, чтобы повысить уровень доверия к данным, необходимо указывать, где к данным или их представлению следует относиться с осторожностью; предоставлять пользователям информацию о процентных представлениях.

2 Культурно значимые данные в цифровых копиях книг Минусинской общественной библиотеки

2.1 Минусинская общественная библиотека

Минусинская общественная библиотека была основана в 1877 году почетным гражданином и ботаником-исследователем Мартьяновым Николаем Михайловичем. Основной причиной для основания библиотеки являлось отсутствие каких-либо аналогов и разбросанность книжного материала. В основном книги в единичных экземплярах были у населения в домашнем пользовании, а люди, достаточно обеспеченные для того, чтобы иметь собственную библиотеку, пользовались ей исключительно в частном порядке. Соответственно, большинству читающей публики были недоступны книжные материалы в том объеме, в котором они нуждались.

Таким образом, главной целью Минусинской общественной библиотеки стало соединение всех книг, которые люди могли бы пожертвовать, в одно целое, систематизировать их и дать возможность пользоваться ими всем желающим. Было решено основать библиотеку при музее, который был в управлении Николая Мартьянова, и впоследствии перевести ее в самостоятельное существование. Осуществление этой цели взял на себя Комитет местного музея во главе с самим Мартьяновым.

Николай Михайлович Мартьянов - провизор, ботаник и почётный гражданин Минусинска. Николай Михайлович родился 27 июля 1844 года в Виленской губернии в семье лесного объездчика. Получить полное образование он не мог из-за небольших доходов отца. Поэтому, Мартьянов сразу после второго класса гимназии был вынужден пойти в ученики к аптекарю, чтобы помогать семье материально. Он успешно сдал экзамены при Петербургской медико-хирургической академии на помощника провизора в 1866 и стал работником аптеки Царского села. Получив специальность, Мартьянов увлекся ботаникой и проводил все свое свободное время за научными исследованиями и собиранием ботанических и минералогических коллекций. Затем, Мартьянов получил предложение минусинского доктора приехать и поработать в городской аптеке, на что он сразу же согласился. Это было прекрасной возможностью исследовать уникальную флору. [16]

Минусинск в 1874 году имел прочные культурные традиции и обладал обширным разнообразием неизученных растений. Поэтому Николай Михайлович нашел в этом городе благоприятные возможности для продолжения своей исследовательской деятельности в ботанике.

«Вокруг него сплотился кружок энтузиастов из местной интеллигенции и ссыльных народников. Опираясь на их поддержку, Николай Михайлович в 1877 году подал прошение о создании музея.» [17]

Мартьянов получил поддержку от Городской думы и открытие музея состоялось. С помощью упорного и самоотверженного труда основателя музей из далекой сибирской провинции вскоре становится известным в научных кругах всей страны и за рубежом. Мартьянова по праву можно считать основоположником музейного краеведения в Сибири. В музее Мартьянов выставлял свои коллекции растений и делился своими открытиями со всеми желающими. Спустя время ему пришла идея о создании библиотеки, так как тяга жителей Минусинска к знаниям была очень велика.

Библиотеку нужно было наполнять книгами. Выполнение этой задачи не было сложным, благодаря помощи минусинского общества, которое жертвовало

значительное количество литературы и денег. Первое время библиотека существовала только благодаря общественным пожертвованиям.

«За первые два месяца на учреждение библиотеки пожертвовали до 1550 книг и около 500 рублей денег. За пять лет существования библиотеки было принесено в дар жителями 3324 сочинений в 4478 томах на сумму 5201 руб. и прислано из других городов 321 сочинение в 414 томах. Кроме этого, разными учеными обществами и учреждениями было пожертвовано 407 книг по истории, астрономии, географии и археологии.» [18]

Спустя время библиотека получила поддержку от Городской думы. От нее выделялась постоянная субсидия, деньги с которой шли на выписку журналов и книг, на переплет и брошюровку, жалование библиотекаря, отопление, освещение и мебель. Плата за чтение с посетителей не взималась, так как она считалась преградой к самообразованию. Плата взималась в редких случаях, когда посетитель был гостем в городе, либо же посетителя никто из жителей не знал.

Благодаря безвозмездным пожертвованиям и субсидии, Минусинская общественная библиотека могла удовлетворить потребности читателей любых сфер интересов, учитывая их индивидуальные запросы. В ее каталоге насчитывалось 17 разделов, а в общей численности на ее полках хранилось почти 5 тысяч книг.

Таблица 1 демонстрирует количественные показатели книг в каждом разделе каталога Минусинской общественной библиотеки в алфавитном порядке.

Таблица 1 - Количество книг в каждом разделе каталога Минусинской общественной библиотеки.

Название раздела	Количество книг	Название раздела	Количество книг
География, этнография, путешествия	229	Правоведение	224
Естествознание и антропология	479	Словесность (романы и повести)	1354

История и теория словесности, критика и библиография	281	Сочинения о Сибири	205
История с вспомогательными науками	548	Справочные книги и смесь	35
История церкви и книги священные и духовно-нравственные	75	Театр	205
Книги для детского чтения	222	Технология, архитектура, механика, математика и сельское хозяйство	261
Медицина и гигиена	95	Философия, психология и педагогика	147
Периодические издания	128	Языкознание	47
Политические и социальные науки	451		

Качественная статистика в заметках каталога тоже присутствует. Например, в самом начале работы библиотеки чаще всего читались книги по беллетристике и истории, а затем приходили запросы на выписку сочинений Д.Ф. Купера. Таблица 2 демонстрирует на каких авторов в разные года ориентировалась библиотека при выписке книг. Все это делалось по запросам читателей.

Таблица 2 – Авторы и книги, запрашиваемые читателями у библиотеки.

Год	Авторы
1878 (1-ая половина)	Пушкин А.С., Гоголь Н.В., Белинский В.Г., Добролюбов Н.А., Костомаров Н.И., Сеньковский Ф.А., Булгарин Ф.В.
1878 (2-ая половина)	Толстой Л.Н., Некрасов Н.А., Тургенев И.С., Щедрин М.Е., Достоевский Ф.М., Островский А.Н., Писемской А.Ф., Ч. Диккенс, В. Скотт, В.М. Гюго
1880	Т. Майн Рид, Д.Ф. Купер, Ж. Верн, Чистяков М.Б.
1881	
1882	Кавелин К.Д., Беляев А.Р., Ф.К. Шлоссер, Ф. Минье

Для того, чтобы выяснить, что интересовало сибирских читателей, какие идеи, люди и места были популярны в этой части мира далекой от культурных

центров, мы можем обратиться к цифровым копиям книг из каталога Минусинской общественной библиотеки. Их необходимо разбить по разделам, затем отдельно классифицировать метаданные и занести результаты в таблицу, представить процентное соотношение доступных и недоступных книг и визуализировать результаты исследования. Однако действительно ли полученные копии книг могут стать основой цифрового анализа в гуманитарных науках? Гипотеза исследования заключается в том, что для составления целостной исторической картины полученных данных не будет достаточно.

В своей работе я продемонстрирую, насколько велики пропуски культурных данных, что препятствует исследованиям и искажает представление о том, какие знания проецировали в своих работах ученые того времени.

2.2. Анализ стратифицированной выборки при помощи агрегаторов

Анализ основывался на четырех разделах каталога Минусинской общественной библиотеки. Разделы были выбраны исходя из их принадлежности к разным направлениям. В выборку, использованную для анализа, вошли один из художественных разделов («Книги для детского чтения»), представленный на рисунке 3, в количестве 222 книг, естественно-научных разделов («Медицина и гигиена») в количестве 95 книг, справочный раздел («Справочные книги и смесь») 35 книг и один их гуманитарных разделов («Языкознание»), который продемонстрирован на рисунке 4, 47 книг. Разделы в общем включали в себя около 400 книг, это 8% от общего числа книг в каталоге.

Будет ли достаточно такого объема полученной информации, чтобы понять, какие художественные книги интересовали людей и какие под жанры были популярны больше всего? В какой степени была развита медицина и какие методы освещались в научно-популярных работах?

1656. ВУЛКАНЫ и землетрясенія. Спб. 1880 г.
 4198. ГАДКІЙ утенокъ. Андерсона.
 1437. ГОДЪ на сѣверѣ. Максимова. Спб. 1871 г.
 4454. Г. А. Потемкинъ, Фурмана.
 1489. ГУДЗОНОВЪ заливъ. Майнъ-Рида. Спб. 1870 г.
 2475. ДАВИДЪ Копперфильдъ. Диккенса. Спб. 1875 г.
 1496. ДАРЫ океана. Майнъ-Рида. Спб. 1869 г.
 4496. ДВА адмирала. Купера.
 3356. ДОМЫ и сынъ. Диккенса. Спб. 1849 г.
 4637. ДОНЪ-Кихоть для дѣтей. Спб. 1867 г.
 1510. ДОНЪ-Кихоть Лаханчскій. Сервантеса. Спб. 1870 г.
 4490. ДОЧЕРИ скватера. Майнъ-Рида.
 564. 2836. ДИКАЯ Африка. Дю-Шалью. Спб. 1873 г.
 1488. ДѢВСТВЕННЫЕ лѣса. Майнъ-Рида. Спб. 1870 г.
 2452. ИСТОРИЯ Россіи въ разсказахъ для дѣтей. Ишпювой. Спб. 1841 г.
 2578. ИСТОРИЯ шотландскаго натуралиста Томаза Эдвардса. Смайльса. Спб. 1877 г.
 КАКЪ я отыскалъ Ливингстона? Стэнли. Спб. 1874 г.
 2581. КАРТИНЫ древняго міра. Фигье. Спб. 1866 г.
 409. КАРТИНЫ изъ жизни насекомыхъ. Ганнке Спб. 1873 г.
 1546. КАТАКОМБЫ. Е. Туръ. Спб. 1878 г.
 1516. КВЕНТИНЪ Дорвардъ. Вальтеръ Скотта. Спб. 1867 г.
 1468. КВИЧИ. Вессерель 2 т. Спб. 1868 г.
 1451. КНИГА для первоначальнаго чтенія. Водовозова. Спб. 1878 г.
 1577. КНЯЗЬ Италійскій графъ Суворовъ-Рымникскій. Спб. 1873 г.
 1478. КОЛОНИЯ на кратерѣ. Купера. Спб. 1867 г.
 4197. КОНЕКЪ-горбунокъ. Ершова. Спб.

Рисунок 3 – Цифровая копия страницы раздела «Книги для детского чтения» из каталога Минусинской общественной библиотеки.

Х. Языкознание.

994. АЛЕЙСКІЙ. Опытъ грамматики русскаго языка. Спб. 1859 г.
 3307. АНТОНОВЪ. Грамматика. Спб. 1869 г.
 4766. БЕРТЕ. Нѣмецкая хрестоматія. Спб. 1874 г.
 1874. БУДАГОВЪ. Практическое руководство турецко-татарскаго языка. М. 1837 г.
 2332. БУСЛАЕВЪ. Опытъ исторической грамматики русскаго языка. М. 1858 г.
 1887. ВОСТОКОВЪ. Русская грамматика. Спб. 1842 г.
 2450. ВОСТОКОВЪ. Сокращенная русская грамматика. Казань 1848 г.
 2699. ГРАММАТИКА латинская въ пользу юношества. М. 1817 г.
 474. ДЮВАРЪ. Нѣмецкая азбука. С.-Петербургъ. 1863 г.
 3011. ЗОЛОТНИЦКІЙ. Отрывки изъ чувашско-русскаго словаря. Казань 1872 г.
 3695. КИРПИЧНИКОВЪ. Слитакосъ русскаго языка. М. 1876 г.

Рисунок 4 - Цифровая копия страницы раздела «Языкознание» из каталога Минусинской общественной библиотеки.

Для определения объема недостающих данных и поиска цифровых копий книг для восстановления библиотеки использовались такие агрегаторы, как Google Книги⁴, Национальная электронная библиотека (НЭБ)⁵ и Проект «Гутенберг»⁶.

⁴ <https://books.google.ru/>

⁵ <https://rusneb.ru/>

⁶ <https://www.gutenberg.org/>

Агрегатор Google Книги (англ. Google Books) представляет сервис полнотекстового поиска по оцифрованным компанией Google книгам. Этот сервис позволяет искать по названию и цитатам из письменных работ. Последнее исследование количества оцифрованного контента проводилось в 2010 году, результаты показали около 130 миллионов книг. [19]

Агрегатор под названием Национальная электронная библиотека (НЭБ) — это «проект Министерства культуры Российской Федерации, предоставляющий пользователям доступ к оцифрованным документам, размещённым в российских библиотеках, музеях и архивах. Фонд библиотеки содержит в себе редкие издания, различные рукописи, диссертации, ноты, патенты, периодическую литературу. На сегодняшний день в каталоге национальной электронной библиотеки около 46 миллионов записей.

Проект «Гутенберг» (англ. Project Gutenberg, или PG) — это некоммерческое сообщество, в работе которого принимают участие волонтеры. Целью этого проекта является перепечатывание и цифровизация классической литературы для ее доступности. На 2021 год в коллекции Проекта «Гутенберг» есть более 60 тысяч книг.

Для анализа полученных данных из цифровых копий книг, были определены следующие параметры собираемых данных:

- номер в каталоге;
- название;
- автор;
- год издания в каталоге;
- ближайший год издания к оригинальному среди доступных книг;
- в какой библиотеке хранится оригинал;
- агрегатор, с помощью которого найдена цифровая копия.

Параметры «номер в каталоге», «год издания в каталоге», «название» и «автор» брались из каталога Минусинской общественной библиотеки.

Определение места хранения оригинала книги при сборе данных основывалось на метаданных «Библиотека» из агрегаторов, упомянутых выше.

Определение даты издания книги основывалось на метаданных из агрегаторов, упомянутых выше «Год», «Год издания». При отсутствии информации о дате создания, проводился дополнительный просмотр цифровой копии книги при ее наличии, для поиска упоминания года в ней. Если просмотр оказывался безрезультатным, то год указывался как «-».

Все данные вносились в Google таблицы, у каждого раздела таблица была отдельная. Название таблицы совпадало с названием раздела, а параметры являлись заголовками к столбцам.

Таким образом, в раздел «Медицина и гигиена» вошли сборники о народных лекарственных растениях, употребляемых в различных губерниях, медицинские статистики, книги о гигиене, пособия для практикующих врачей и акушеров, одноименные справочники о болезнях, а также домашние лечебники и советы по проветриванию жилых помещений, фрагмент из которых представлен на рисунке 5. Самые упоминаемые авторы: Личинкин Е., Мечников Л.И., Писарев М., Португалов В. О. и Слюнин Н.

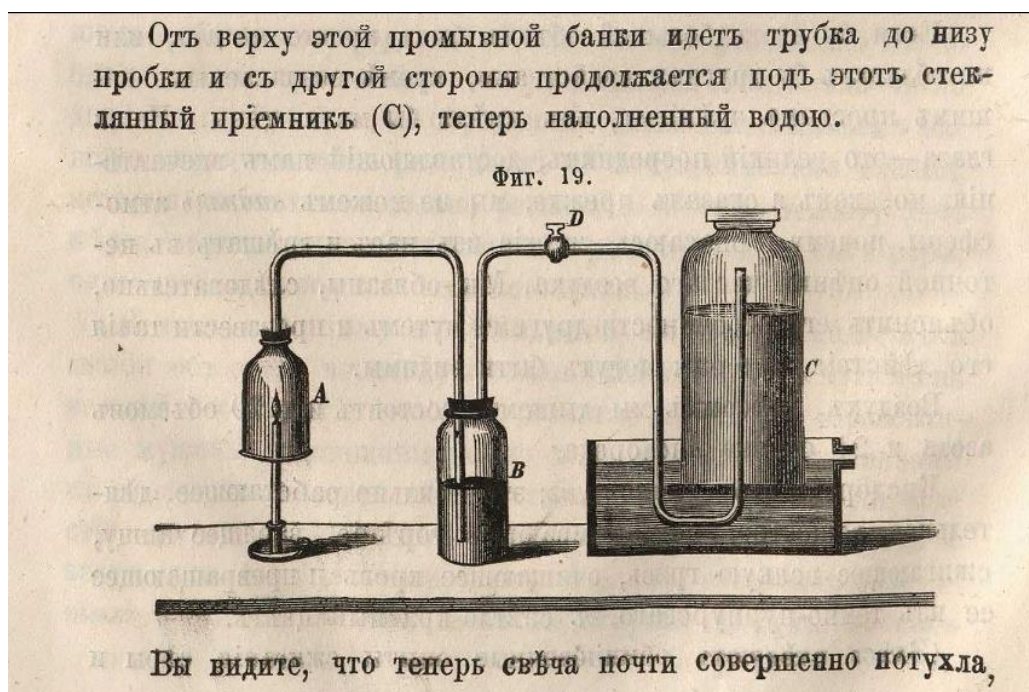


Рисунок 5 – Фрагмент из книги 1874 г. «Проветривание общественных и частных жилых помещений как неперемное условие сохранения здоровья и жизни» Л. Лидса раздела «Медицина и гигиена».

Раздел «Справочные книги и смесь» представляет художественные альбомы и иллюстрированные описания различных событий, кавказские, духовенческие и всеобщие календари, фрагмент из которых представлен на рисунке 6, домашние справочные книги, каталоги по волшебным фонарям, правила для скачек, пособие для преподавателей танцев и многое другое, что не вошло в остальные разделы.



Рисунок 6 – Фрагменты из книги «Всеобщий календарь на 1879 г.» раздела «Справочные книги и смесь».

Раздел «Языкознание» наполнен грамматическими учебниками по русскому, фрагмент из которых представлен на рисунке 7, и монгольскому языку, различными словарями (чувашский, французский, тунгусский),

руководствами к изучению таких древних языков, как греческий и латинский, а также самоучителями.



Рисунок 7 – Фрагмент из книги 1870 г. «Родное слово для детей младшего возраста». Ушинского К.Д., раздела «Языкознание».

Самым многочисленным разделом является «Книги для детского чтения». Он включает в себя большое количество художественной литературы, которая в свою очередь делится на жанры. Было решено разбить раздел на жанры и исследовать количественное соотношение. Это было сделано для того, чтобы лучше разобраться, какой тип художественной литературы интересовал читателей больше всего. Валидность результатов подтверждается цитатой из каталога Минусинской общественной библиотеки: «С появлением у библиотеки денежных средств, она стала значительно пополняться, сообразно потребностям читателей».

Всего раздел «Книги для детского чтения» включил в себя 12 категорий:

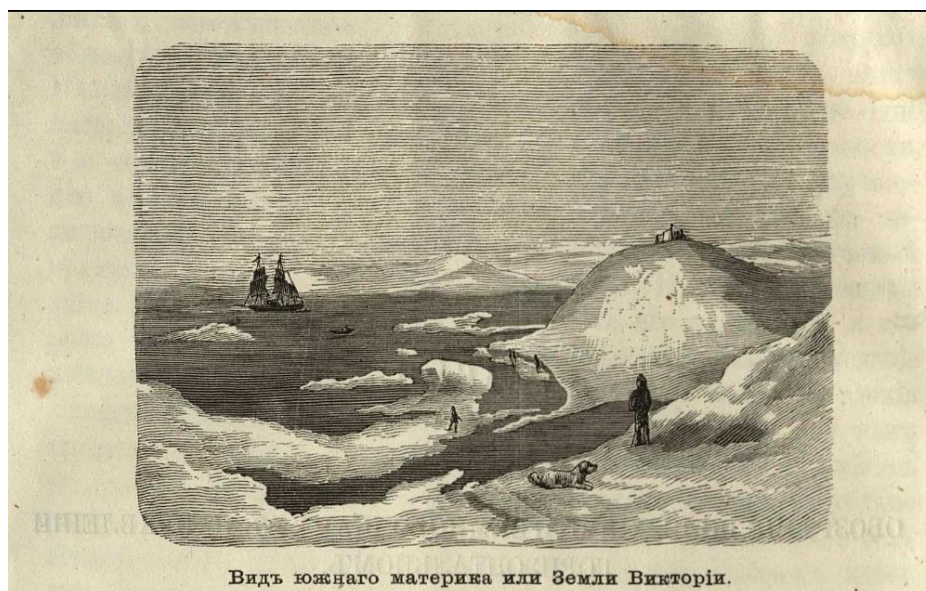
- историческая литература – книги, события которых связаны с историческими событиями, а главными героями или действующими лицами могут выступать известные в истории личности. В эту категорию вошли,

например, «Русская история в жизнеописаниях ее главнейших деятелей» Костомарова Н.И., «Абрагам Линкольн, освободитель невольников» Гофмана В.В., «Князь Григорий Александрович Потемкин-Таврический» Фурмана П.Р., фрагмент книги представлен на рисунке 8, и др.;



Рисунок 8 – Фрагмент из книги 1910 г. «Князь Григорий Александрович Потемкин-Таврический» Фурмана П.Р.

- география – учебники и рассказы путешественников о своем опыте странствий и других народах, подкрепленные фактами. В эту категорию вошли «Открытие Америки, Камчатки и Алеутских островов рассказы для детей», «Год на Севере», «Иллюстрированная популярная физическая география» фрагмент книги представлен на рисунке 9, и др.;



Видъ южнаго материка или Земли Викторіи.

Рисунок 9 – Иллюстрация из книги 1873 г. «Иллюстрированная популярная физическая география» Зуева Н.И.

- очерки – сюда вошли небольшие литературные произведения с краткими описаниями жизненных фактов без вымыслов. В эту категорию вошли «Фрегат „Паллада“» — книга очерков Гончарова И.А. и «Тропический мир в очерках животной и растительной жизни» Г. Гартвига;

- обучающая литература – учебники, книги для первоначального чтения и хрестоматии. В эту категорию вошли такие книги как «Астрономия», «Беседы о земле и тварях, на ней живущих» Бекетова А. фрагмент книги представлен на рисунке 10, «Воздух и его свойства», «Краткий учебник естествознания», «Первое знакомство с тем, что мы видим каждый день», «Растение и его жизнь», и т.д.;

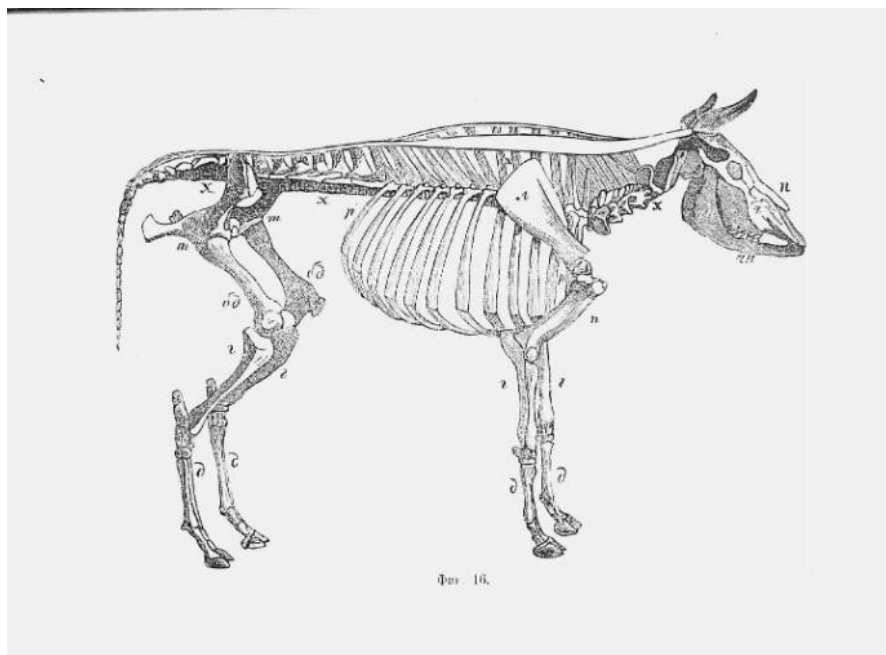


Рисунок 10 – Фрагмент из книги 1898 г. «Беседы о земле и тварях, на ней живущих» А. Бекетова.

- периодические издания – журналы, издаваемые ежегодно. В эту категорию вошли такие издания как «Вокруг света», «Детское чтение», «Задумчивое слово» фрагмент книги представлен на рисунке 11, «Иллюстрированный журнал для детей», «Природа и люди», «Семейные вечера», «Час досуга»;



Рисунок 11 – Фрагмент из журнала «Задумчивое слово» выпуск 1909 года.

- приключенческая литература – художественное произведение, основой которого является рассказ о вымышленных событиях, приключениях. В эту категорию включены произведения Д.Ф. Купера, фрагмент книги представлен на рисунке 12, М. Рида, Ж. Верна Сюда же включены книги Д. Дефо и С. Смайла;



Рисунок 12 – Фрагмент из книги 1876 г. «Американский пакетбот "Монтаук"» Д.Ф. Купера.

- сказки – один из жанров фольклора, либо литературы. В эту категорию включены произведения Григоровича Д.В., Х.К. Андерсена, Авенариуса В.П., Пушкина А.С., Одоевского В.Ф., Ершова П.П., Гофмана В.В., Вагнера Н.П., Афанасьева А.Н. и Ж. Санд;

- поэмы – литературные произведения в стихах. В эту категорию вошло произведение Лермонтова М.Ю. «Песня про Царя Ивана Васильевича, молодого опричника и удалого купца Калашникова»;

- повести – литературный жанр, описывающий естественное течение жизни. В эту категорию вошли «Тарас Бульба» Гоголя Н.В., фрагмент книги представлен на рисунке 13, «Детство и отрочество» Толстого Л.Н., «Антон Горемыка» Григоровича Д.В.;

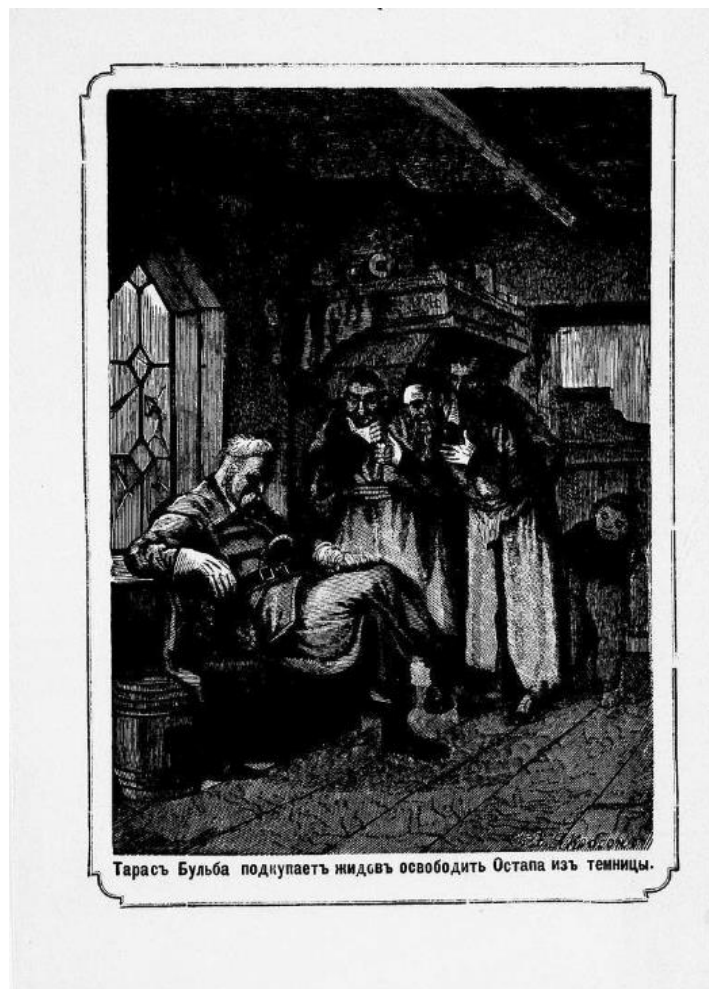


Рисунок 13 – Иллюстрация из книги 1878 г. «Тарас Бульба» Гоголя Н.В.

- социальная литература – книги, сюжет которых затрагивает социальные проблемы. В эту категорию вошли произведения «Зверобой» и «Следопыт» авторства Д.Ф. Купера, «Хижина дяди Тома» С.Г. Бичер;

- научная фантастика – книги, сюжет которых основывается на ненастоящих и выдуманных научных открытиях. В эту категорию вошли произведения «Паровой дом», «Путешествие к центру земли», фрагмент книги представлен на рисунке 14, «Вокруг луны», «Путешествие в страну пушных зверей, или Приключения на плавучем острове», «Путешествие по небу и звездам и вокруг земного шара в 1920 часов», «Открытие Земли» и «Приключения китайца» Ж. Верна;

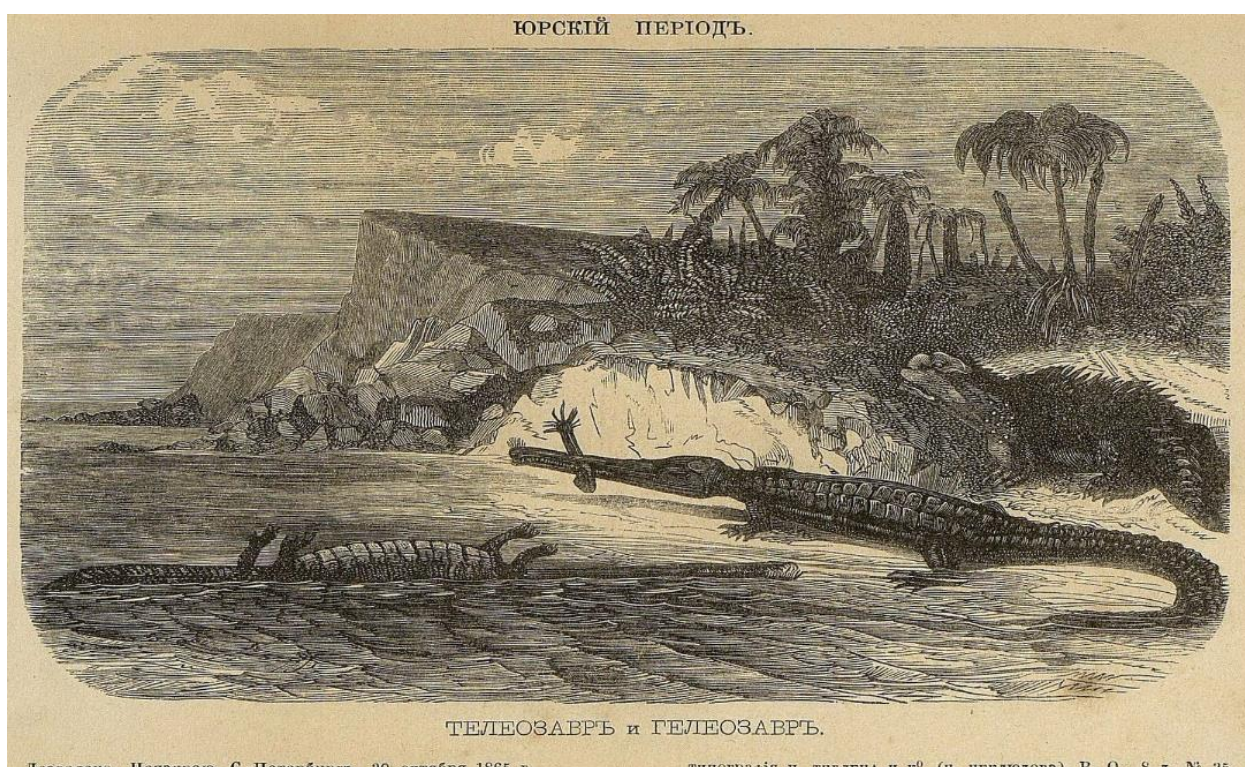


Рисунок 14 – Иллюстрация из книги 1865 г. «Путешествие к центру земли» Ж. Верна.

- без жанра – книги, которые не удалось отнести к какому-либо жанру.

Жанр книги определялся на основе метаданных агрегаторов, упомянутых ранее, и дополнительном описании, а затем был соотнесен с категориями. При отсутствии информации о жанре книги, она вносилась в категорию «Без жанра».

Когда поиск книг был завершен, то полученные из агрегаторов данные уже можно было бы использовать для изучения, однако, сложно распространить

выводы на весь дата сет, потому что книги в других разделах могут быть доступны в большей или меньшей степени.

Во время поиска выяснилось, что условия доступа к цифровым копиям были разными, поэтому, в зависимости от них, книги определялись в одну из четырех классификационных групп:

- «Доступны для сохранения на локальный диск» - в одном из используемых для поиска цифровых копий книг агрегаторов есть доступ к скачиванию файла в формате PDF или ePub;

- «Доступны по ссылке» - в агрегаторе нет доступа к скачиванию, но есть доступ к просмотру полного текста книги;

- «Доступны в альтернативном издании» - в агрегаторе искомая книга найдена в издании другого года, в другом переводе;

- «Недоступны совсем» - книги, которые не удалось найти ни в одном агрегаторе.

Также была проведена оценка количества книг в каждой категории, названной выше. Таблица 3 демонстрирует количественное соотношение книг из каждого исследуемого раздела и классификационных групп.

Таблица 3 - Количественное соотношение книг из каждого исследуемого раздела и классификационных групп.

Название раздела	Доступно для сохранения на локальный диск	Доступные по ссылке	Доступные в альтернативном издании	Недоступные
Книги для детского чтения	114	17	15	76
Медицина и гигиена	10	2	22	61
Справочные книги и смесь	10	0	4	21
Языкознание	17	0	6	24

После количественного подсчета книг в классификационных группах доступности каждого раздела, было определено процентное соотношение «доступных» и «недоступных» книг. Оно было выражено через круговые

диаграммы. Чем выше процент доступности книг – тем лучше. Процентные диаграммы помогают определить, результаты по какому из разделов можно считать самыми валидными для заключения на их основе каких-либо выводов.

2.3 Визуализация данных раздела «Книги для детского чтения».

Визуализация данных позволяет нам сжать большое количество информации до одной картинки, что увеличивает шанс увидеть закономерности и упрощает заключение на ее основе выводов. Основное внимание уделяется визуализации распределения культурных данных по времени.

Ситуацию с пропусками и «недоступностью» данных необходимо рассматривать с точки зрения ее неизбежности. Неопределенность является неотъемлемым условием при работе с культурными данными, так как практически каждый тип информации об объекте может быть неточным [20].

Тем не менее, визуализация неопределенности является необходимым способом демонстрации того, какие части информации или данных являются точными, полными, последовательными, определенными и наоборот. Это помогает создать более полную картину, делая изложение точнее. «Визуализация неопределенности стремится сделать все аспекты явными, чтобы пользователи могли получить более правдивое представление о данных, что обеспечивает единственную надежную основу для создания выводов». [21].

В качестве основы метода визуализации был взят принцип Стефана Янике. [22] Он визуализировал данные о музыкальных исполнителях, выровняв и распределив их портреты по временной шкале. Это дизайнерское решение было принято для того, чтобы было проще сравнивать распределение временных данных. Элементы в виде портретов масштабируются в соответствии с выбранной характеристикой. Эта характеристика может быть любой, если она поддается количественной оценке: на рисунке 15, например, показаны портреты музыкантов, масштабированные по “популярности”, измеряемой "общим количеством упоминаний о музыканте в онлайн и печатных СМИ". Точность

позиционирования даты отменяется в пользу эстетически приятных мозаичных изображений.

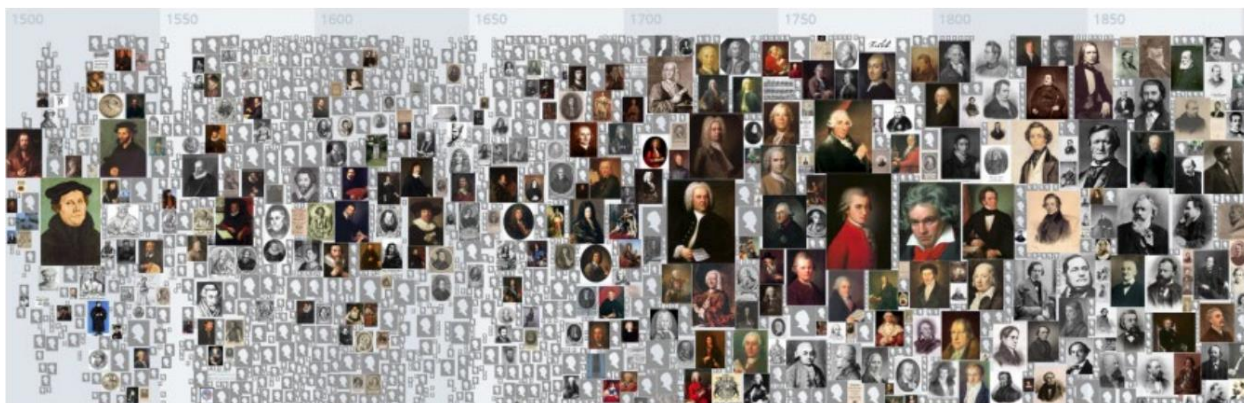


Рисунок 15 - Визуализация портретов музыкантов из музыкальной базы данных. Стефан Янике, 2018. Источник: Jänicke S. Timages: Enhancing time graphs with iconographic information, 2018.

Полученный объем информации слишком большой, поэтому визуализация раздела «Книги для детского чтения» необходима. С ее помощью указать на пропуски данных можно намного проще и эффективнее.

3 Результаты исследования

3.1 Яндекс таблицы

Для каждого раздела были созданы отдельные таблицы, внутри которых было по 4 листа – категории доступности. Таким образом, например, в таблице «Книги для детского чтения» был лист «Доступные книги (сохраненные локально)» (Рис. 16), «Доступные по ссылке», «Доступные в печатном виде» (Рис. 17) и «Недоступные». Таблицы необходимы для группировки метаданных из каталога и дополнительных из агрегаторов. Это помогло отследить количество книг в каждой из категории доступности.

	A	B	C	D	E	F	G
1	Доступные книги (сохраненные локально)						
2							
3	Номер	Название	Автор	Год издания в каталоге	Год издания	Где хранится	Агрегатор
4	не указан	Как я отыскал Ливингстона	Стэнли Генри Мортон	1874	1874	Российская государственная библиотека (РГБ)	https://rusneb.ru/
5	676	Народные русские сказки	Афанасьев Александр Николаевич	1873	1873	-	https://rusneb.ru/
6	801	Тарас Бульба	Гоголь Николай Васильевич	1865	1878	Российская государственная библиотека (РГБ)	https://rusneb.ru/
		Русская история в жизнеописаниях ее главнейших	Костомаров Николай			Российская государственная	

Рисунок 16 – Лист «Книги, доступные для скачивания на локальный диск» в таблице раздела «Книги для детского чтения».

	A	B	C	D	E	F	G
1	Доступные в печатном виде						
2							
3	Номер	Название	Автор	Год издания в каталоге	Где хранится	Агрегатор	Ссылка
4		Немецкая 4766 хрестоматия	Берте		1874	Российская национальная библиотека (РНБ)	https://rusneb.ru/catalog/000200_000018_v19_rc_1193800/
5		Немецкая 474 азбука	Дювуар Анастасия Ивановна		1863	Российская государственная библиотека (РГБ)	https://rusneb.ru/catalog/000199_000009_003566266/
		Начальные правила обучения французского	Ольдеко			Российская национальная библиотека	https://rusneb.ru/catalog/000200_000018_v19

Рисунок 17 – Лист «Доступны в печатном виде» в таблице раздела «Языкознание».

3.2 Процентное соотношение «доступности» книг

На рисунке 18, показаны проценты «доступных» и «недоступных» книг в среднем по всем исследуемым четырем разделам: «Медицина и гигиена»,

«Справочные книги и смесь» и «Языкознание», «Книги для детского чтения». Процент «доступности» 57% означает, что больше половины цифровых копий книг доступны (для сохранения на локальный диск, по ссылке и в альтернативных изданиях), а недоступные цифровые копии занимают 43%, это немного меньше половины. Проблема заключается в том, что данные показатели не отражают доступность цифровых копий по всем разделам в равной степени, поэтому необходимо рассмотреть каждый отдельно.



Рисунок 18 – Процентное соотношение «доступных» и «недоступных» книг в среднем по всем четырем исследуемым разделам («Книги для детского чтения», «Медицина и гигиена», «Справочные книги и смесь» и «Языкознание»).

Доступно 54%, недоступно 46%.

Рассмотрим показатели естественно-научного раздела «Медицина и гигиена». На рисунке 19 круговая диаграмма демонстрирует большой процент недоступных цифровых копий книг – 64%. Доступных книг 36%, из которых для ознакомления и сохранения на локальный диск всего 11%, по ссылке 2%, а в печатном виде 23%. Такой высокий процент «недоступности» связан с утратой актуальности научных работ. В них используется устаревшая терминология, а

современные системы диагностики и лечения отличаются от систем XIX века. Также произошли изменения международной классификации болезней. С практической точки зрения актуальны только справочники по анатомии, а также история болезней, их симптоматика и различные виды.



Рисунок 19 - Процентное распределение всех категорий «доступности» в разделе «Медицина и гигиена». Недоступных книг 64%, доступных 36% (для сохранения на локальный диск 11%, по ссылке 2%, в альтернативных изданиях 23%).

Цифровые копии из раздела «Медицина и гигиена» важны, так как с их помощью можно исследовать, какие методы применялись в медицине того времени, насколько была развита гигиена, какие заболевания были самыми распространенными. Среди всего прочего можно отследить статистику по вспышкам заболеваний в различных губерниях.

Переходим к гуманитарному разделу «Языкознание». На круговой диаграмме (Рис. 20) мы видим соотношение почти 50/50 – доступных цифровых копий 49%, а недоступных 51%. Ситуация с гуманитарным разделом лучше, чем с естественно-научным, потому что лингвистика по сравнению с медициной подверглась меньшим изменениям. Научные работы из раздела «Языкознание» продолжают цитироваться и в современных учебниках.



Рисунок 20 - Процентное распределение всех категорий «доступности» в разделе «Языкознание». Недоступных книг 51%, доступных 49% (для сохранения на локальный диск 36%, а в печатном виде 13%).

Гуманитарный раздел важен для исследования, так как дает возможность рассмотреть методы и подходы того времени к изучению языков.

Процентное соотношение «доступности» в справочном разделе «Справочные книги и смесь» представлено на рисунке 22. Ситуация как в предыдущих разделах - недоступных книг (60%) больше, чем доступных (40%). Это связано с тем, что большинство книг имеет тематику, которая была актуальна только в то время (Справочный словарь при чтении газет, Настольный словарь для справок). Или же это литература, не предусматривающая

переиздания (Календарь для духовенства за 1884 г., Отчет о состоянии Иркутской общественной гимназии за 1868 г. и 1869 г.).



Рисунок 22 - Процентное распределение всех категорий «доступности» в разделе «Справочные книги и смесь». Недоступных книг 60%, доступных 40% (для сохранения на локальный диск 29%, а в печатном виде 11%).

Справочный раздел важен тем, что он помогает лучше изучить локальные явления и бытовую жизнь того времени. Раздел наполнен книгами по уборке, советам по быту, справочниками по волшебным фонарям, которых в современном мире уже нет.

Переходя к самому объемному из всех четырех разделов – художественному (Рис. 23), мы видим, что доступных книг (66%) намного больше, чем недоступных (34%). Для сохранения на локальный диск доступно 51% книг, просмотр полного текста по ссылке 8%, а в печатном виде 7%. Большой процент доступности обусловлен тем, что художественные книги имеют более обширную целевую аудиторию, нежели книги из гуманитарных и естественно-научных разделов. Художественная литература издается большим тиражом и в различных переводах, что позволяет ей лучше сохраниться.



Рисунок 23 - Процентное распределение всех категорий «доступности» в разделе «Книги для детского чтения». Недоступных книг 34%, доступных 66% (для сохранения на локальный диск 51%, по ссылке 8%, а в печатном виде 8%).

Художественный раздел важен тем, что позволяет предположить, какие жанры и авторы были популярны в то время, какое представление о мире, формирующееся через литературу, складывалось у читателей.

3.3. Визуализация данных раздела «Книги для детского чтения».

Визуализация, изображенная на рисунке 24, демонстрирует количество доступных и недоступных авторов для раздела «Книги для детского чтения». Портреты авторов расположены по шкале в зависимости от количества книг, представленных в каталоге. Прозрачные портреты указывают на то, что большая часть книг автора недоступна. Размер портрета зависит от количества произведений, содержащихся в каталоге Минусинской общественной библиотеки. Например, рассмотрим Д.Ф. Купера. Его книги упоминаются в каталоге 23 раза, доступны из них 20, следовательно, портрет не прозрачный, находится чуть левее границы шкалы «20» и самый большой по размеру.

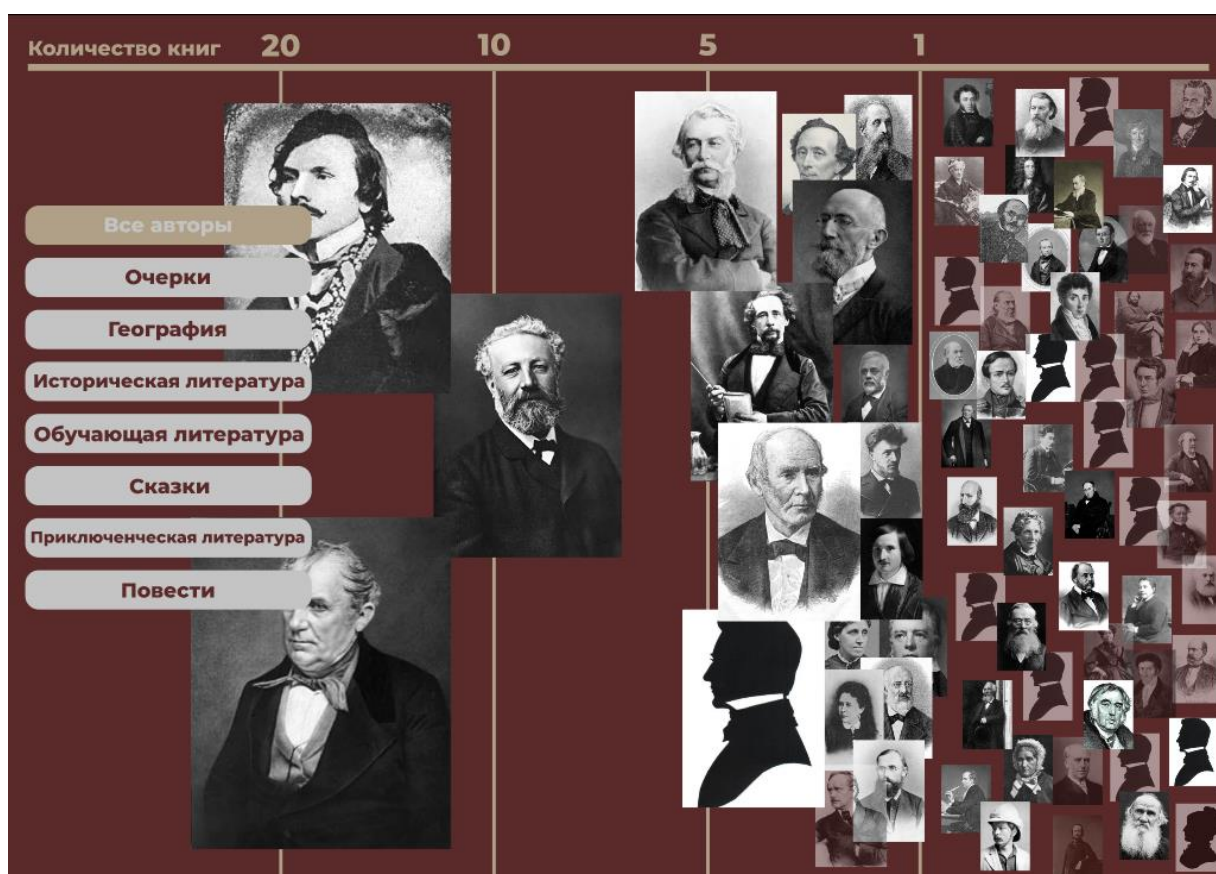


Рисунок 24 – Распределение авторов раздела «Книги для детского чтения» в зависимости от количества книг, упоминаемых в каталоге библиотеки.

Писатели, чьи работы в большинстве своем «недоступны» имеют полупрозрачные портреты.

Также в визуализацию была добавлена дополнительная информация о каждом авторе (Рис. 25). В ней содержатся имя, фамилия, отчество (при его наличии), годы жизни, страна рождения, краткая характеристика писательского направления и деятельности, список «доступных» (черным цветом) и «недоступных» (красным цветом) книг из каталога библиотеки. Узнать эту информацию можно кликнув на портрет писателя. Это было сделано для того, чтобы можно было быстро узнать о вкладе писателя в библиотечную коллекцию.



Рисунок 25 – Профиль автора Ж. Верна. На баннере предоставлена главная информация о его годах жизни, стране, жанровом направлении. Ниже главной информации располагается список «доступных» (черным цветом) и «недоступных» (красным цветом) книг его авторства из каталога библиотеки.

При отображении портретов возникла проблема отсутствия информации об авторе. Например, в каталоге Минусинской общественной библиотеки указывалась только фамилия автора. Если дополнительный поиск не давал результатов, то вместо портрета помещалась картинка (Рис. 26) с безымянным мужчиной или женщиной. Недоступными оказались 26 из 71 авторов.



Рисунок 26 – Профиль автора А. Ганике, информация о котором недоступна. Его портрет был заменен на нейтральное изображение, а описание состоит из вопросительных знаков.

В визуализацию также были добавлены кнопки, позволяющие переключать фильтры по жанрам. Например, при нажатии на кнопку «Обучающая литература» (Рис. 27) пропадут портреты всех авторов, кроме тех, чьи книги относятся к учебникам. При отборе произведений на жанры возникла проблема неочевидности их в некоторых книгах. Распределение в таком случае происходило, опираясь на краткое содержание книги и биографию автора. Фильтры по жанрам важная часть визуализации, так как демонстрирует количественное соотношение авторов в различных жанрах.



Рисунок 27 – Результат фильтрации по жанру «Обучающая литература». На шкале изображены портреты только тех авторов, чьи книги относятся к учебникам, справочникам и т.п.

ЗАКЛЮЧЕНИЕ

Главным результатом этой работы является уникальный метод для обнаружения неопределенностей и неточностей в текстовом корпусе. Уникальным метод считается из-за применения визуализации для представления пропусков данных культурного характера среди художественной литературы. Методы визуализации неопределенностей Оливии Вейн и Стефана Янике применялись на археологические экспонаты и музыкальных исполнителей, а литература оставалась без внимания. Таким образом, взяв за основу их подход визуализации неопределенностей и приспособив под особенности изучения метаданных художественной литературы, удалось получить новый метод. Обнаружить неточности получилось при сравнении цифровых копий, имеющихся в агрегаторах, со списком физических книг из каталога

Минусинской общественной библиотеки. Это дало нам возможность систематически и широко посмотреть на большой корпус текстов и создать новый подход. Он дополняет более узкий и глубокий анализ, указывая на ограничения в нем, а также этот подход указывает, где конкретно находятся пропуски данных.

Если говорить о конкретных результатах использования метода, то стоит упомянуть о том, что художественные произведения фиксируются в корпусе, а произведения научные, справочные и лингвистические пропадают. Таким образом, наши результаты наводят на мысль о том, что дальнейшее распространение нашего метода, корпуса и результатов дадут более подробное понимание работы с пропусками данных в научной, справочной или других жанрах художественных произведений.

Созданный в этой работе метод можно использовать при исследовании других сибирских и даже европейских библиотек. Еще одним направлением исследования может быть определение популярных тем и географических названий. Это позволит еще лучше понять, что читали и узнавали люди в XIX веке.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Н. Булгакова. Что такое корпус? [Электронный ресурс] – Режим доступа: <https://lingvomaster.org/ru/cto-takoe-korpus->
2. О корпусе. Национальный корпус русского языка. 2022. [Электронный ресурс] - Режим доступа: <https://ruscorpora.ru/page/corpora-about/>
3. Nauta G. J., van den Heuvel W. Cultural institutions are increasingly digitising their holdings, 4, 2015.
4. Hughes, L. M. Digitizing collections: strategic issues for the information manager. Facet Publishing, 2004
5. Friendly, M. The Golden Age of Statistical Graphics, Statistical Science, 2008, 23, 502–535. <https://arxiv.org/pdf/0906.3979.pdf>

6. Brehmer, M., Lee, B., Bach, B., Riche, N.H. & Munzner, T. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE transactions on visualization and computer graphics*, 23(9), pp.2151-2164, 2017
7. Windhager F., Salisu S. & Mayr E., and others. *Uncertainty of What and for Whom -And Does Anyone Care? Propositions for Cultural Collection Visualization*, 2019.
8. Gershon, N. Visualization of an Imperfect World. *IEEE Computer Graphics and Applications*, 18(4):43–45, 1998.
9. Vane, O. Timeline design for visualising cultural heritage data. PhD dissertation. Royal College of Art, UK, 2019. <https://www.oliviavane.co.uk/phd>
10. Nyhan, J. and Flinn, A. *Computation and the Humanities. Towards an Oral History of Digital Humanities*, 2016. https://www.researchgate.net/publication/310795050_Computation_and_the_Humanities_Towards_an_Oral_History_of_Digital_Humanities
11. Moltrup, M. Herstory of Graphic Design: Elaine Lustig Cohen. *Collections: A Journal for Museum and Archives Professionals* 15 (2-3): 167–177. 2019.
12. Klein, Lauren F. The Image of Absence: Archival Silence, Data Visualization, and James Hemings. *American Literature* 85 (4): 661–688. 2013.
13. Boyd Davis, S, and Kräutli F. The Idea and Image of Historical Time: Interactions Between Design and Digital Humanities. *Visible Language* 49 (3): Special issue ‘Critical Making: Design and the Digital Humanities’ 100–119, 2015.
14. LwM. 2019a. *Living with Machines: Rethinking Research to Illuminate Our Past*. Living with Machines, 2020. <http://livingwithmachines.ac.uk/>.
15. Boyd Davis, S., Vane, O. & Kräutli, F. *Can I believe what I see? Data visualization and trust in the humanities*, 2021.
16. Мартьянов Николай Михайлович — краевед, основоположник музейного краеведения в Сибири [Электронный ресурс]. – Режим доступа: <http://region.krasu.ru/node/525> – Дата доступа: 2022.

17. Мартьянов Николай Михайлович [Электронный ресурс]. – Режим доступа: <https://www.krasplace.ru/martyanov-nikolaj-mixajlovich>. – Дата доступа: 2022.
18. Каталог Минусинской общественной библиотеки, 1883. – 6 – 8 с.
19. Taycher L. Books of the world, stand up and be counted! All 129,864,880 of you, 2010. <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>
20. S. H. Lee and W. Chen. A comparative study of uncertainty propagation methods for black-box-type problems. Structural and Multidisciplinary Optimization, 37(3):239, 2009.
21. D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. IEEE Transactions on Visualization and Computer Graphics, 22(1):240–249, 2016
22. Jänicke S. Timages: Enhancing time graphs with iconographic information, 2018.

ПРИЛОЖЕНИЕ А

Скриншоты таблиц в программе Excel. Подробные данные о книгах, разбитые по категориям доступности цифровых копий.

	A	B	C	D	E	F	G
1	Доступные книги (сохраненные локально)						
2							
3	Номер	Название	Автор	Год издания в каталоге	Год издания	Где хранится	Агрегатор
4	не указан	Как я отыскал Ливингстона	Стэнли Генри Мортон	1874	1874	Российская государственная библиотека (РГБ)	https://rusneb.ru/
5	676	Народные русские сказки	Афанасьев Александр Николаевич	1873	1873	-	https://rusneb.ru/
6	801	Тарас Бульба	Гоголь Николай Васильевич	1865	1878	Российская государственная библиотека (РГБ)	https://rusneb.ru/
7	1139	Русская история в жизнеописаниях ее главнейших деятелей	Костомаров Николай Иванович	1874-1880	1880	Российская государственная библиотека (РГБ)	https://rusneb.ru/

Рисунок А.1 – Скриншот таблицы раздела «Книги для детского чтения» листа категории доступности «Доступны для сохранения на локальный диск».

	A	B	C	D	E	F	G	H
1	Доступные книги (по ссылке)							
2	Номер	Название	Автор	Год в каталоге	Год издания	Где хранится	Агрегатор	Ссылка на чтение
3	1510	Дон Кихот в пер. Ламанчского	Мигель де Сервантес Сааведра	1870	1935	Российская государственная библиотека	НЭБ	https://rusneb.ru/catalog/000199_000009_005145683/
4	1490	Ползуны по скалам	Рид Томас Майн	1876	1927	Российская государственная детская библиотека (РГДБ)	НЭБ	https://rusneb.ru/catalog/000207_000017_RU_RGDB_BIBL_0000425675/

Рисунок А.2 – Скриншот таблицы раздела «Книги для детского чтения» листа категории доступности «Доступно по ссылке».

	A	B	C	D	E	F	G	H
1	Доступные в печатном виде							15
2								
3	Номер	Название	Автор	Год издания в каталоге	Год издания	Где хранится	Агрегатор	Ссылка
4	3469	Задушевное слово	-	за 1880	за 1880	Российская государственная библиотека (РГБ)	https://rusneb.ru/	https://rusneb.ru/catalog/000199_000009_60000291492/
5	3470	Задушевное слово	-	за 1880	за 1880	Российская государственная библиотека (РГБ)	https://rusneb.ru/	https://rusneb.ru/catalog/000199_000009_60000291492/
6	2416, 3626	БАБУШКИНЫ сказки	Жорж Санд	1874	1897	-	https://rusneb.ru/	https://rusneb.ru/catalog/000199_000009_003945202/
		Рассказы из						https://rusneb.ru/catalog/000

Рисунок А.3 – Скриншот таблицы раздела «Книги для детского чтения» листа категории доступности «Доступно в печатном виде».

	A	B	C	D
1	Недоступные произведения			76
2				
3	Номер	Название	Автор	Год издания
4	4451	весна	-	1860
5	2357	БАШКИРЫ	-	не указан
6	2366	НОГАЙЦЫ	-	не указан
7	1383	СИБИРСКИЕ и среднеазиатские инородцы	-	не указан
8	1705	ТАТАРСКОЕ племя	-	не указан
9	4507	осень	-	1872
10	1693	ФИНСКОЕ племя	-	не указан
11	2456	ЧЕЧЕНЦЫ и лопари	-	не указан
12	2367	ИМЕРЕТИНЦЫ	-	не указан
13	1658	КАВКАЗСКИЕ Жители	-	не указан
14	2575	КАЛМЫКИ и буряты	-	не указан

Рисунок А.4 – Скриншот таблицы раздела «Книги для детского чтения» листа категории доступности «Недоступные».

	A	B	C	D	E	F	G
1	Доступные книги (сохраненные локально)						
2							
3	Номер	Название	Автор	Год издания в каталоге	Год издания	Где хранится	Агрегатор
4	4010	Профессиональная гигиена	Эрисман Фёдор Фёдорович	1877	1877	Российская государственная библиотека (РГБ)	https://rusneb.ru/
5	509. 3252	Популярная гигиена	Реклам Карл	1869	1869	Российская государственная библиотека (РГБ)	https://rusneb.ru/
6	4111	Материалы для медицинской статистики Смоленской губернии	Почтарев П.Ф.	1882	1882	Российская государственная библиотека (РГБ)	https://rusneb.ru/
		Наставление против появления и распространения повальных и				Государственная универсальная научная	

Рисунок А.5 – Скриншот таблицы раздела «Медицина и гигиена» листа категории доступности «Доступны для сохранения на локальный диск».

	A	B	C	D	E	F
1	Доступные книги (по ссылке)					
2	Номер	Название	Автор	Год в каталоге	Где хранится	Ссылка на чтение
3	4118	Письма к матерям об уходе за ребенком	Галанин М. И.	1882		http://az.lib.ru/n/nekrasowa_e_s/text_1882_pisma_k_materyam_oldorfo.shtml
4	1985	Безпредельно стиль гигиены	Португалов Вениамин Осипович	Д. 69 г. VIII		http://az.lib.ru/p/portugalow_w_o/text_1869_gigena_oldorfo.shtml

Рисунок А.6 – Скриншот таблицы раздела «Медицина и гигиена» листа категории доступности «Доступно по ссылке».

	A	B	C	D	E	F	G
1	Доступные в печатном виде						
2							
3	Номер	Название	Автор	Год издания в каталоге	Где хранится	Агрегатор	Ссылка
4	4099	О всасывании желчи в кишечном канале	Алексеев	1882	Российская государственная библиотека (РГБ)	https://rusneb.ru/	https://rusneb.ru/catalog/000199_000009_003594896/
5	4090	Материалы к вопросу о различии проявления (гесп. выражения) сил мышцами верхней и нижней конечностей	Варавин В.С.	1882	Российская государственная библиотека (РГБ)	https://rusneb.ru/	https://rusneb.ru/catalog/000199_000009_003595032/
		Записки			Российская государственная библиотека (РГБ)		https://rusneb.ru/

Рисунок А.7 – Скриншот таблицы раздела «Медицина и гигиена» листа категории доступности «Доступно в печатном виде».

	A	B	C	D
1	Недоступные произведения			76
2				
3	Номер	Название	Автор	Год издания
4		А. Х. Болотная лихорадка	-	(Пр. и Ох. 78 г. I)
5		Диететика детей	Беднарь	(Учит. (62 г.)
6		Организмъ человека Физиологические очерки	Бернштейн	(Учит. (62 г.)
7		Новая теория транс	Берд	(Пр. и Ох. 78 г. II)
8		Гигиенический конгресс в Париже	Бертенсон	В. Е. 79 г. V
9		4564 фармакологии	Бинц Карл	1873
10		498 Будьте здоровы!	Бок	1863
11		499 Человеческое тело: его жизнь, строение и холя	-	1873

Рисунок А.8 – Скриншот таблицы раздела «Медицина и гигиена» листа категории доступности «Недоступные».

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Гуманитарный институт

Кафедра информационных технологий в креативных и культурных индустриях

УТВЕРЖДАЮ

И. о. заведующего кафедрой

 А. В. Усачёв

« 4 » 07 2022 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 «Прикладная информатика в области искусства и гуманитарных наук»

Пропуски культурно значимых данных среди цифровых копий книг из каталога Минусинской общественной библиотеки

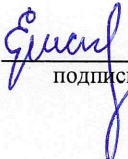
Руководитель


подпись, дата

канд. культурологии,
доцент

И.А. Кижнер

Выпускник

 — 04.07.22
подпись, дата

И.А. Елистратова

Красноярск 2022

Продолжение титульного листа БР по теме Пропуски культурно значимых данных среди цифровых копий книг из каталога Минусинской общественной библиотеки

Нормоконтролер



подпись, дата

Е. Р. Брюханова