

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

Научный руководитель – Кижнер Инна Александровна

Елистратова Илона Александровна

Студент (бакалавр)

Сибирский федеральный университет, Гуманитарный институт, Красноярск, Россия

E-mail: prlvnr.10080@gmail.com

Минусинская общественная библиотека - культурно-образовательное учреждение, основанное в 1877 году, насчитывающее 4543 книг в 6373 томах. 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. Данных не будет достаточно для того, чтобы определить, какие страны интересовали читателей, что они знали о мире.

1. Методология

1.1. Выборка разделов

Проведенный анализ основывается на четырех разделах каталога Минусинской общественной библиотеки. Это один из естественно-научных разделов (“Медицина и гигиена”), один из гуманитарных разделов (“Языкознание”), справочный раздел (“Справочные книги и смесь”) и один из художественных разделов (“Книги для детского чтения”). Разделы включают около 400 книг это 8% от общего числа книг в каталоге. Для определения объема недостающих данных для восстановления библиотеки был проведен поиск книг в агрегаторах, таких как Google Книги¹, Национальная электронная библиотека (НЭБ)² и Проект «Гутенберг»³. Полученные данные вносились в таблицу с полями, такими как название, автор, год издания, в какой библиотеке хранится и с помощью какого агрегатора получены данные. Затем, в зависимости от того, при каких условиях доступна цифровая копия книги, данные о ней вносились в соответствующий лист таблицы. Листы включали следующие классификационные группы: доступные для скачивания, доступны по ссылке, доступны в альтернативном издании, недоступны. После этого было вычислено процентное соотношение для каждой категории и определен процент недоступных книг. Представление оцифрованных книг как данных открывает возможность применения методов обработки данных к ним, включая визуализацию.[1] Была проведена визуализация для каждого из проанализированных разделов, с помощью онлайн-сервиса для разработки интерфейсов и прототипирования Figma⁴. Благодаря созданию визуальных отображений данных «мы можем видеть закономерности, тенденции или аномалии, постоянство или вариации способами, которые другие формы — текст и таблицы — не позволяют»[2].

2. Результаты

Рисунок 1 показывает 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. Недоступно около 40% изданий.

Таблица 1 демонстрирует количественное соотношение книг из каждого исследуемого раздела и классификационных групп.

Рисунки 2, 3, 4, 5 показывают процентное соотношение классификационных групп в разделах “Справочные книги и смесь”, “Медицина и гигиена”, “Языкознание” и “Книги для детского чтения”.

3. Заключение

Несмотря на то, что было найдено 38% цифровых копий для всех четырех разделов, большее количество копий оказалось недоступным, за исключением книг для детского чтения (найдено около 60% изданий). Таким образом, научная и справочная литература в большинстве случаев недоступна. Это значит, что мы имеем неполное представление о том, какие знания проецировали в своих работах ученые того времени. С другой стороны, агрегаторы предоставляют возможность частично восстановить Минусинскую общественную библиотеку. Однако, чтобы повысить вероятность валидного результата исследования, необходимо учитывать пропуски в данных и эксплицировать недостающие знания о библиотеке при публикации данных.

Источники и литература

- 1) 1. Vane O. Timeline design for visualising cultural heritage data. PhD dissertation. Royal College of Art, UK, 2019. <https://www.oliviavane.co.uk/phd>
- 2) 2. Friendly, M. The Golden Age of Statistical Graphics, *Statistical Science*, 2008, 23, 502–535. <https://arxiv.org/pdf/0906.3979.pdf>

Иллюстрации

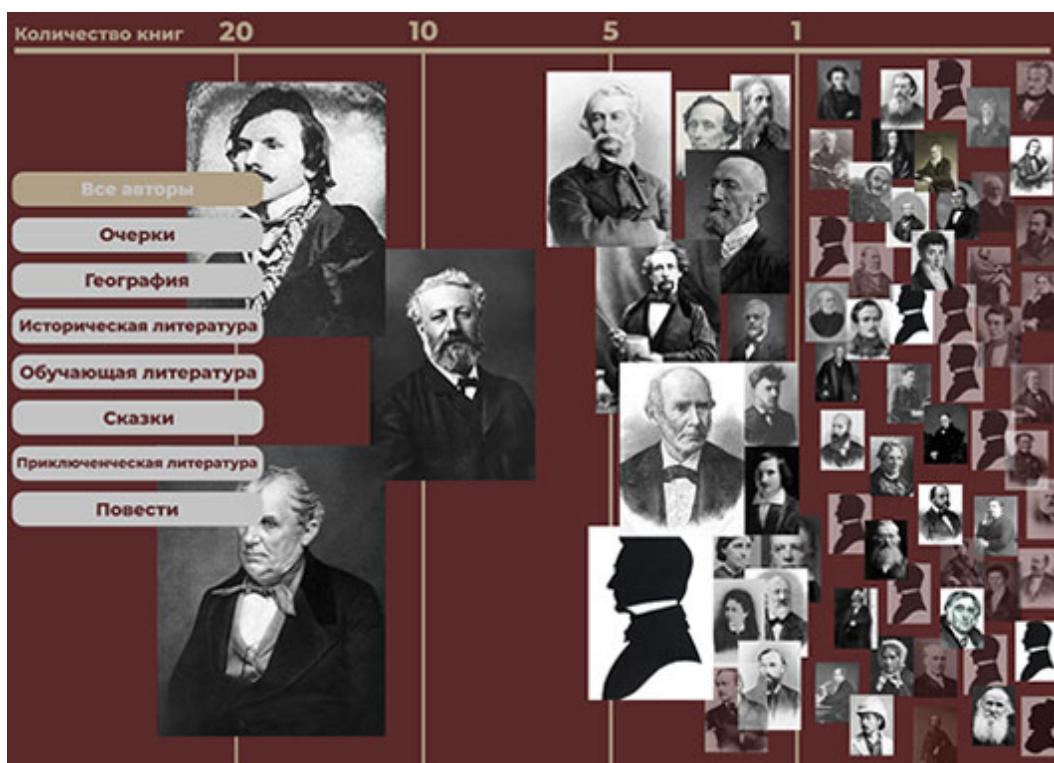


Рис. 1. Рисунок 1 - Авторы расположены на шкале в зависимости от количества книг данного автора, имеющихся в библиотеке.

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

Рис. 2. Таблица 1 - количественное соотношение книг из каждого исследуемого раздела и классификационных групп.



Рис. 3. Рисунок 2 - Процентное соотношение классификационных групп в разделе "Справочные книги и смесь".



Рис. 4. Рисунок 3 - Процентное соотношение классификационных групп в разделе "Медицина и гигиена".



Рис. 5. Рисунок 3 - Процентное соотношение классификационных групп в разделе "Языкознание".

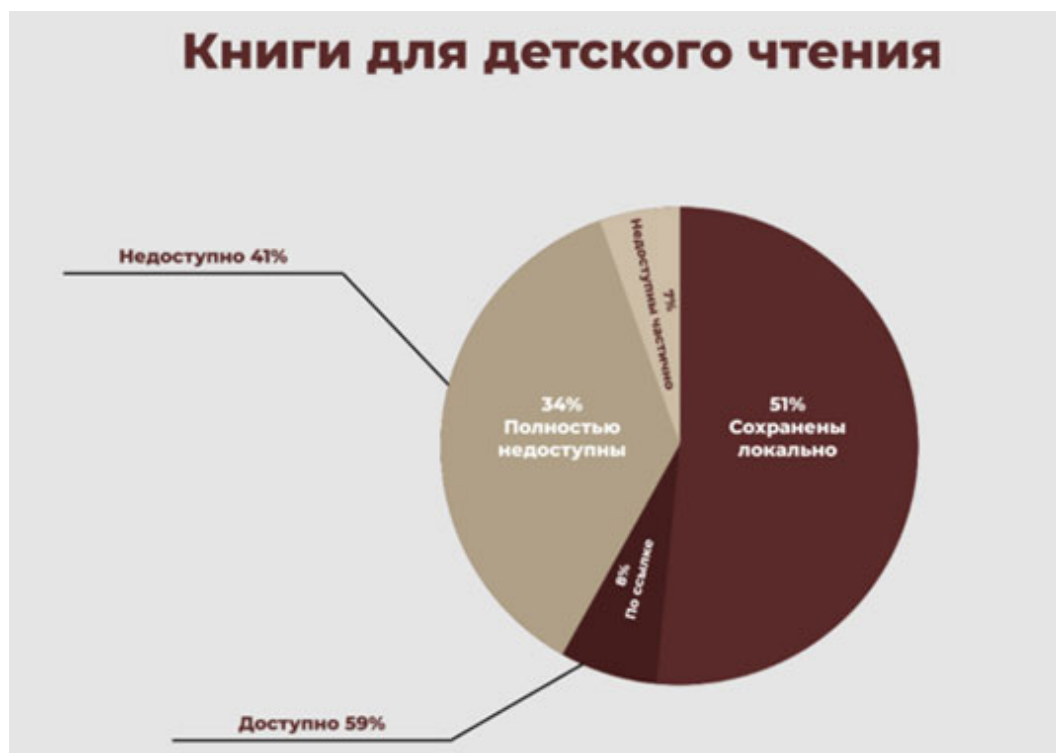


Рис. 6. Рисунок 3 - Процентное соотношение классификационных групп в разделе "Книги для детского чтения".