Федеральное государственное автономное образовательное учреждение

высшего образования

«Национальный исследовательский университет «Высшая школа экономики»

Факультет гуманитарных наук

Цифровые методы в гуманитарных науках

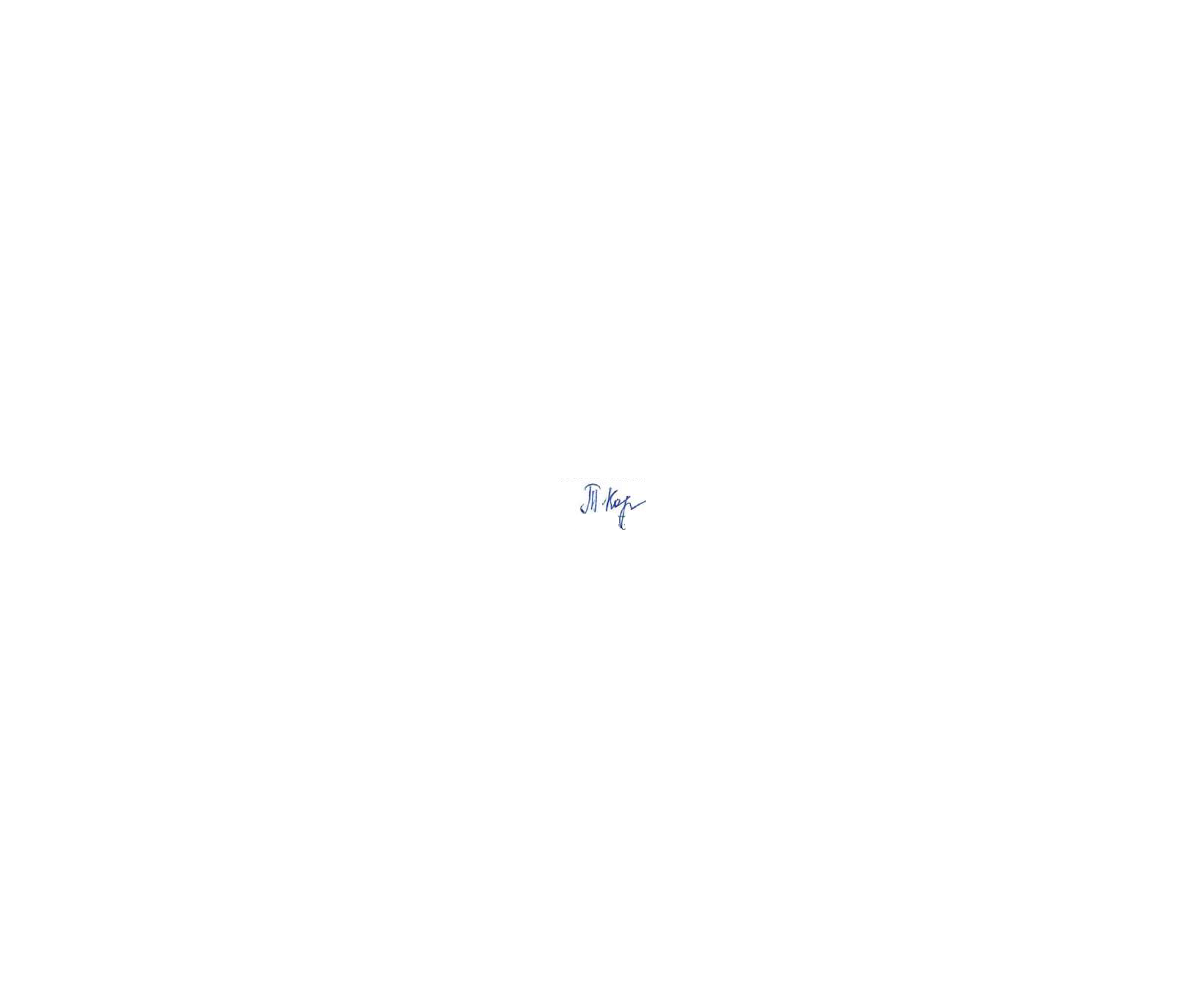
магистратура

О Т Ч Е Т

по производственной практике

Выполнила студентка гр. МЦМГН 191

Колпакова Т.А.



Проверили:

Скоринкин Д.А.,

доцент Факультета гуманитарных наук

\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(оценка) (подпись)

Скоринкин Д.А.,

руководитель магистерской программы

\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(оценка) (подпись)

02.04.2021

## Введение

Практика состояла в проведении научно-исследовательской работы на тему «Музейные сайты в исторической перспективе: поиск, организация данных, анализ» для Центра цифровой гуманитаристики ПГНИУ.

Было получено следующее техническое задание:

1. Нужно выбрать 5-6 крупных музеев (можно взять по списку посещаемости или размеру коллекций), взять музеи одного типа – художественные, например, исторические или титульные национальные.
2. Дальше было бы хорошо прокачать их версии с веб-архивов (интернет архив и национальные, например британский UKWA) берите версию примерно за конец года.
3. Дальше надо сверить тексты и выкинуть данные предыдущего года. Если будете писать скрипт - круто, если нет, можно попробовать сервисами онлайновыми на сверку текстов.
4. Нужно выделить текстовые блоки, например, новости, описание объектов, сторителлинг, блоги и т.д. Категории должны быть крупными, чтобы сравнивать музеи.
5. Выделяете топики 1) по отдельным текстовым блокам; 2) по всему корпусу за год; 3) по всему корпусу за 20+ лет.
6. Сравниваете музеи между собой, выявляете общее-особенное по блокам, по динамике, по корпусам в целом. Важно сравнить и динамику метаданных – обновляемость сайтов, структуру сайтов, количество адресов, которые они сменили и пр. Динамику по обновляемости страниц, количество текстовых страниц и загруженных изображений можно посмотреть в Интернет Архиве, например, для Метрополитен музея эти данные, [тут](https://web.archive.org/web/changes/metmuseum.org) и [тут](https://web.archive.org/details/metmuseum.org) и карта сайта [тут](https://web.archive.org/web/sitemap/metmuseum.org).

## Выполненные задачи

1. Получено техническое задание от организации;
2. Выбран фокус исследования – анализ блогов исторических англоязычных сайтов музеев;
3. Произведен поиск наиболее интересных для анализа сайтов;
4. Отобраны 5 сайтов музеев по следующим критериям:
   * крупный музей исторического типа;
   * есть англоязычная версия сайта;
   * есть раздел «Блоги», и он активно заполнялся на протяжении длительного периода.

Были выбраны для дальнейшего анализа блоги следующих музеев:

* Британский музей https://blog.britishmuseum.org/
* Музей Американской истории https://americanhistory.si.edu/blog
* Музей истории Канады https://www.historymuseum.ca/blog/
* Музей истории Шотландии https://blog.nms.ac.uk/archive/
* Национальный музей истории https://phm.org.uk/blog/

1. В процессе обсуждения с Надеждой Георгиевной Поврозик было откорректировано техническое задание – при написании скрипта для выкачивания информации напрямую с сайта использование веб-архивов не требуется. Кроме того, пункт 3 технического задания также потерял актуальность – данные не повторяются, следовательно, очистка не требуется.
2. Были написаны скрипты для выкачивания материалов блогов в формате csv. Для упрощения дальнейшего анализа таблицы всех сайтов создавались по стандартной схеме (title, date, tags, text) несмотря на то, что не везде присутствуют теги, например.
3. Был собран весь интересующий корпус записей блогов с подробными метаданными с 5 сайтов (более 1000 записей) за период сентябрь 2008 – март 2021.
4. Было проведено тематическое моделирование собранного корпуса и сравнительно-сопоставительный анализ блогов музеев (анализ метаданных, тематическое своеобразие блогов разных музеев).
5. На основе полученных результатов была написана научно-популярная статья.

## Заключение

По итогам практики все заявленные в техническом задании пункты были выполнены, а результаты оформлены в виде научно-популярной статьи. Все материалы размещены на сайте <https://github.com/TatiKolpakova/Museums>.

УЧЕТ ВЫПОЛНЕННОЙ РАБОТЫ

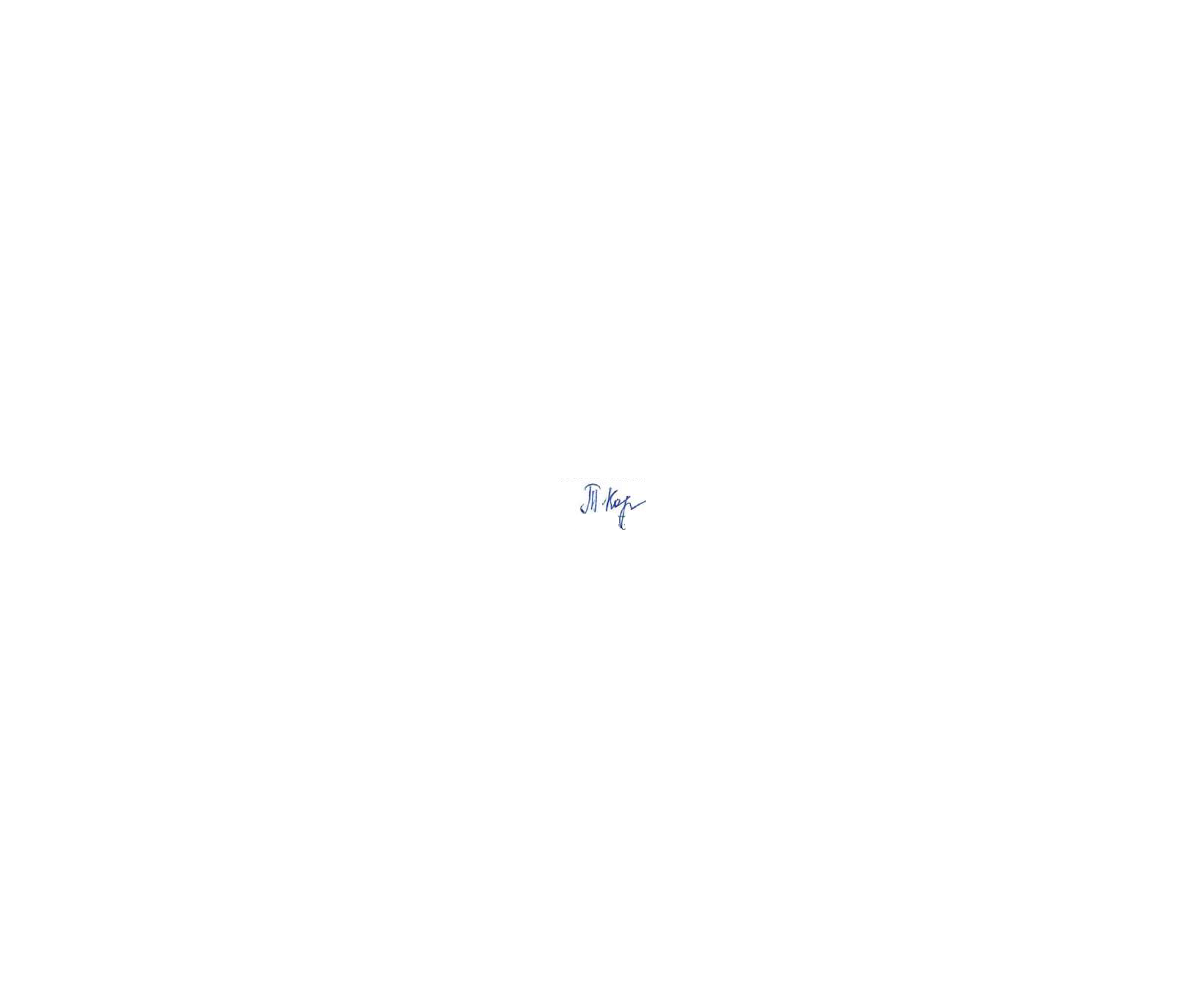
Название практики: Музейные сайты в исторической перспективе: поиск, организация данных, анализ

Организация: Центр цифровой гуманитаристики ПГНИУ

Руководитель от НИУ ВШЭ: Скоринкин Д. А.

Руководитель от организации: Поврозик Н. Г.

|  |  |  |  |
| --- | --- | --- | --- |
| Срок выполнения | Краткое содержание работы | Указания/комментарии руководителей практики | Отметка о выполнении  работы |
| Октябрь – ноябрь 2020 | * получено техническое задание от организации; * выбран фокус исследования – анализ блогов исторических англоязычных сайтов музеев; * произведен поиск наиболее интересных для анализа сайтов; * отобраны 5 сайтов музеев по следующим критериям:   + крупный музей исторического типа;   + есть англоязычная версия сайта;   + есть раздел «Блоги», и он активно заполнялся на протяжении длительного периода.   Были выбраны для дальнейшего анализа блоги следующих музеев:   * Британский музей https://blog.britishmuseum.org/ * Музей Американской истории https://americanhistory.si.edu/blog * Музей истории Канады https://www.historymuseum.ca/blog/ * Музей истории Шотландии <https://blog.nms.ac.uk/archive/> * Национальный музей истории https://phm.org.uk/blog/ * написаны скрипты для выкачивания материалов блогов в формате csv; |  | Выполнено |
| Декабрь 2020 – февраль 2021 | * эксперименты с тематическим моделированием собранного корпуса и сравнительно-сопоставительный анализ блогов музеев (динамика метаданных, тематическое своеобразие блогов разных музеев); * попытки расширения собранного корпуса за счет дополнительных сайтов музеев, анализ особенностей блогов исторических музеев в разных странах. |  | Выполнено |
| март 2021 | * собран весь интересующий корпус записей блогов с подробными метаданными с 5 сайтов (более 1000 записей) за период сентябрь 2008 – март 2021. * написаны научно-популярная статья и отчёт по итогам практики. |  | Выполнено |

Студент-практикант  / Колпакова Т. А.