

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

подпись

«___» _____ 2020 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 «Прикладная информатика в искусстве и гуманитарных науках»

Тенденциозность в агрегированном цифровом контенте: количественный анализ состава цифровых коллекций Europeana

Руководитель	_____	доцент, канд. техн. наук	А.В. Усачёв
	подпись, дата		
Консультант	_____	ст. преподаватель	И.А. Кижнер
	подпись, дата		
Выпускник	_____		К.И. Петрова
	подпись, дата		

Красноярск 2020

Продолжение титульного листа БР по теме Тенденциозность в агрегированном цифровом контенте: количественный анализ состава цифровых коллекций Europeana

Нормоконтролер

подпись, дата

Е.Р. Брюханова

СОДЕРЖАНИЕ

Введение	4
1 Корпус как инструмент исследования	6
1.1 Понятие корпуса и его задачи.....	6
1.2 Корпус визуальных данных	9
1.3 Цифровая библиотека Europeana.....	12
1.4 О неравномерности представления данных.....	14
2 Методы и материалы исследования	16
2.1 Методы исследования	16
2.2 Europeana как корпусный менеджер	17
2.3 Сбор данных коллекций стран мира	19
3. Результаты исследования.....	23
3.1 Анализ коллекций стран мира	23
3.1.1 Общие сведения.....	23
3.1.2 Культурное наследие и естественнонаучная коллекция	27
3.2 Точечный анализ коллекций Молдавии, Замбии, Уругвая	31
3.3 Обсуждение результатов.....	36
Заключение	38
Список используемых источников.....	40
Приложение А. Примеры таблиц Excel	44
Приложение Б. Примеры географических карт.....	46

ВВЕДЕНИЕ

На сегодняшний день в совершенно разных сферах научной деятельности используются большие массивы данных. В гуманитарных науках распространенным инструментом является корпус данных.

Europeana – это цифровая библиотека, содержащая более 50 миллионов оцифрованных экспонатов. Среди них 34 миллиона составляют изображения. Такое огромное количество данных, а также системы поиска и фильтрации потенциально позволяют использовать Europeana как ресурс для корпусных исследований.

Однако для того, чтобы результаты исследований на базе Europeana были корректными и достоверными, корпус данных должен соответствовать определенным критериям – таким, как репрезентативность и сбалансированность. Проблема заключается в том, что ресурсы, подобные Europeana, практически не исследованы на предмет репрезентативности и сбалансированности.

Выпускная квалификационная работа посвящена анализу тенденциозности цифровых коллекций портала Europeana.

Цель настоящей работы – выявить неравномерность представления данных о культурном наследии в Europeana на примере коллекций стран мира.

Задачи исследования:

- изучить понятие корпуса данных, ввести понятия репрезентативности, сбалансированности и неравномерности представления данных культурного наследия;
- собрать данные о цифровых коллекциях стран мира в портале Europeana;
- проанализировать полученные данные;
- выявить закономерности в получившихся данных;
- определить неравномерность цифровых коллекций стран мира в Europeana.

Научная новизна данного исследования заключается в недостаточной исследованности корпусов визуальных данных, предоставляющих инструментарий для работы в сфере гуманитарных наук.

Существование репрезентативного и сбалансированного корпуса визуальных данных позволило бы проводить различные исследования для широкого спектра гуманитарных задач. Проблема исследования заключается в том, что на сегодняшний день визуальных корпусов немного, их репрезентативность и сбалансированность не исследованы, и, соответственно, их нельзя использовать для получения достоверных результатов исследований.

Актуальность настоящей работы заключается в потребности получать достоверные результаты исследований в корпусах визуальных данных, в необходимости выявлять определенные параметры, влияющие на качество исследований и наши представления о мировой культуре.

Объект исследования – агрегатор цифровых коллекций, представляющих культурное наследие.

Предмет исследования – представленность данных в цифровых коллекциях портала Europeana.

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

Первая глава содержит информацию об особенностях корпусных исследований в гуманитарных науках, а также об искажениях в корпусах данных.

Вторая глава «Методы и материалы исследования» описывает методы, использованные при данном исследовании.

Третья глава «Результаты исследования» включает демонстрацию, описание и обсуждение полученных результатов анализа цифровых коллекций стран мира портала Europeana.

Заключение подводит итоги исследования, список литературы содержит библиографические описания цитируемых источников, в приложении представлены примеры собранных данных в виде таблиц, диаграмм, карт и прочих изображений.


[Извлечено 43 страницы]

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
подпись

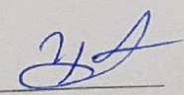
« 13 » 07 2020 г.

БАКАЛАВРСКАЯ РАБОТА

09.03.03.14 «Прикладная информатика в искусстве и гуманитарных науках»

Тенденциозность в агрегированном цифровом контенте: количественный анализ состава цифровых коллекций Europeana

Руководитель




подпись, дата

доцент, канд. техн. наук

А. В. Усачёв

Консультант




подпись, дата

ст. преподаватель

Н. А. Кижнер

Выпускник



подпись, дата

К. И. Петрова

Красноярск 2020

Продолжение титульного листа БР по теме Тенденциозность в агрегированном цифровом контенте: количественный анализ состава цифровых коллекций Europeana

Нормоконтролер



подпись, дата

Е. Р. Брюханова