

Содержание канонического контента в агрегированных данных цифровой библиотеки Европеана

Научный руководитель – Кижнер Инна Александровна

Астрашабов Алексей Евгеньевич

Студент (бакалавр)

Сибирский федеральный университет, Гуманитарный институт, Красноярск, Россия

E-mail: astrashabow@gmail.com

Европеана - одна из крупнейших цифровых библиотек, содержащая десятки миллионов объектов, оцифрованных по всей Европе и за её пределами. Контент Европеаны, публикуемый агрегаторами, демонстрирует обширное культурное наследие каждой отдельно взятой страны [1]. Контентная стратегия ресурса также ставит целью представление национальной культуры всех стран Европейского союза [2]. Однако нас интересует вопрос о том, в какой мере Европеана демонстрирует канон европейской культуры? Краткое определение канона описывает его как список лучших и образцовых произведений, составленный лицом или группой лиц, часто с общей целью [3]. В ходе проведённого исследования был определен значительно низкий процент канонического присутствия в коллекциях ряда стран, чьи агрегаторы активно публикуют культурные данные в Европеане.

1. Методология

1.1. Выборка ключевых стран

Выбор страны определен количеством контента, который она публикует. Таким образом итоговый список составлен из стран с наибольшим количеством загруженных объектов: Нидерланды (9,864,661), Германия (6,747,870), Франция (4,901,356), Испания (4,830,085), Великобритания (4,793,417), Швеция (4,503,881). Число объектов, опубликованных перечисленными странами, составляет 35,641,270, более 50% контента Европеаны (60,478,742 объектов).

1.2. Формирование канонических списков

Анализ канонического контента проведен на основе списков канонических авторов литературных, живописных и музыкальных произведений. Источники списков канонических авторов включают в себя: списки для чтения, коллекции канонических произведений национальных библиотек, музейные и выставочные коллекции, а также зрительские и читательские рейтинги. Формирование списка, пригодного для проведения исследования, проведено по результатам нескольких пересечений конкретного автора в первоначальных списках, полученных из разных источников. В ходе сравнения списков удалось определить 229 авторов канонических произведений из разных стран.

1.3. Определение содержания канонического контента с помощью API

На данном этапе, используя полученные списки авторов канонического контента, удалось провести подсчеты количества объектов, публикуемых в виде произведений конкретного автора или контента относящегося к его личности и творчеству. Процент содержания канонического контента определен исходя из отношения количества объектов связанных с конкретным каноническим автором к количеству объектов, опубликованных страной, считающейся местом творчества и родиной данного автора. Для достижения наибольшей точности результата все вычисления и процесс поиска осуществлялись на базе Europeana API.

2. Результаты

2.1. В какой мере Европеана демонстрирует канон европейской культуры?

Таблица 1 демонстрирует малый процент содержания контента, относящегося к каноническим авторам. Однако агрегаторы некоторых стран публикуют канонический контент в значительно большем количестве (Испания), но несмотря на это, процент канонического контента этих стран остается довольно низким. Эти результаты демонстрируют ту модель формирования контентной базы Europeana, в условиях которой канонический контент не является объектом внимания со стороны агрегаторов. Страны же, публикующие наибольшее количество объектов, имеют одни из самых низких показателей канонического содержания.

3. Заключение

Результаты таблицы 2 демонстрируют нам, что Europeana отражает канон западной цивилизации в очень малом объеме. Несмотря на все культурное и количественное разнообразие данных, половина контента Europeana включает в себя около 12,5% канонического контента.

Вероятно такое положение канона в Europeana может быть основано на той контентной стратегии, в которой цифровая библиотека стремится к наибольшему объему культурных данных всех стран мира. Однако в таком случае возникают вопросы о том как агрегаторы формируют публикуемые массивы данных и какую Европу все-таки демонстрирует нам Europeana?

Источники и литература

- 1) Di Giorgio S. Culturaitalia, the Italian national content aggregator in Europeana // *Procedia Computer Science*. – 2014. – Т. 38. – С. 40-43.
- 2) Heijink K. *Europeana Content Strategy* // Europeana v1. 0. Europeana: Think Culture. – 2009.
- 3) Compton M.T. *Infinite Canons: A Few Axioms and Questions, and in Addition, a Proposed Definition*. 2016. toddmcompton.com

Иллюстрации

Процентное отношение канонического контента ко всему контенту, публикуемому страной. Страны представленные в таблице, опубликовали на момент исследования более 50% контента Европейцы.

Страна	Кол-во загружаемых объектов	% канонического контента
Нидерланды	9,864,661	0,2
Германия	6,747,870	1,45
Франция	4,901,356	1,7
Испания	4,830,085	7,6
Великобритания	4,793,417	1,5
Швеция	4,503,881	0,08

Рис. 1. Таблица 1

Количественные и процентные показатели содержания канонического контента в массиве данных, публикуемых шестью крупнейшими странами.

Кол-во объектов, загружаемых шестью первыми странами	Кол-во объектов, относящихся к каноническим	% канонического контента
35,641,270	648,743	12,5

Рис. 2. Таблица 2