

Инстаграм Толстого

Как сделать популярными 26 000 фотографий

Анна Моргунова

Дарья Максимова (daria.maximova.m@gmail.com)

магистерская программа «Цифровые методы в гуманитарных науках»
НИУ ВШЭ (Москва)

Digital History Meetup — IBV Minsk 

14 марта 2020 г.

Что будет

- Откуда взялся этот проект
- Данные, с которыми мы работаем
- Что мы хотим сделать
- Что мы уже сделали
- Вызовы и как мы с ними справлялись
- Уверенный взгляд в будущее

История появления

- Государственный музей Л.Н. Толстого и магистратура «Digital Humanities» ВШЭ
- **Цель 1:** рассказать историю при помощи фотографий
 - история фотографии на примере фотографий ЛНТ
 - визуализация биографии
- **Цель 2:** открытие фотографий для исследователей
 - историки
 - филологи
 - data science (компьютерное зрение)
 - и т.д.

Как это работает

Государственный музей Л.Н. Толстого	Цифровые гуманитарии ВШЭ
данные	обработка, очистка, преобразование в машиночитаемый вид
экспертиза по содержанию архива	новые форматы подачи

Что планируем в итоге:

- ресурс с машиночитаемой информацией о фотографиях
- тематические подборки фотографий
- выявление трендов при помощи компьютерного зрения

Что было: фотографии

№ п/п	Изображения	Учетные обозначения	Название, датировка, материал, техника,размеры
1.		Ф-35 ГМТ КП-4001/23	Л.Н.Толстой Автор: Левицкий С.Л. (1819-1898 гг.). 1856. С.-Петербург 13,2 x 9,8.
2.		Ф-50 ГМТ КП-2210/2	Л.Н.Толстой Автор: Жерюзе И. 1861. Брюссель 3,0x5,33 (бланк).
3.		Ф-64 ГМТ КП-4230	Л.Н.Толстой Автор: Автофотопортрет. 1862. Ясная Поляна. На бланке автограф С. А. Толстой: «1862 г. Сам себя снял» (в левом верхнем углу) и «Гр. Л. Н. Толстой Фотография Ясной Поляны» (в левом нижнем углу и внизу посередине). 11,0 x 9,1 (фотография; овал) 10,7 x 13,3 (бланк).

...и так примерно для 8000 фотографий

Что было: рубрики

<i><u>Фонд фотоматериалов</u></i>	
1000	Фотографии основного фонда
1000-00	Прочее
1100	Фотографии с Л.Н. Толстым
1110	Л.Н.Толстой один (хронология по годам)
1849	1849 г.
185х	1850-е годы (1850, 1851, 1852 и т.д.)
186х	1860-е годы
187х	1870-е годы
188х	1880-е годы
189х	1890-е годы
190х	1900-е годы (юбилей в 1908г – отдельная рубрика)
1910	1910 г.
1120	Л.Н.Толстой в группах (с родственниками)
1121	Л.Н.Толстой и С.А.Толстая
1122	Л.Н.Толстой со своими детьми
1123	Л.Н.Толстой с внуками
1124	Л.Н.Толстой с другими родственниками (с одним лицом)
1125	Л.Н.Толстой в кругу семьи
1126	Л.Н.Толстой в кругу родных
1127	Л.Н.Толстой в кругу родных и знакомых
1130	Л.Н.Толстой в группах (прочее)
1131	Л.Н.Толстой с разными лицами (не родственниками)
1132	Л.Н.Толстой с одной персоной
1133	Л.Н.Толстой с деятелями культуры
1134	Л.Н.Толстой среди рабочих
1135	Л.Н.Толстой и крестьяне
1136	Л.Н.Толстой с детьми (среди крестьянских детей)
1137	Л.Н.Толстой с детьми (с тульскими школьниками)
<u>1138</u>	Последний приезд Л.Н. Толстого в Москву (проводы на Курском вокзале).

Как действовать

Описание, структура

- определение схемы описания
- выделение полей, значения которых надо доставать отдельно
- конвертация описания в TEI
- соотнесение описаний и фотографий
- выстраивание структуры БД

Обогащение

- новые признаки и поля:
 - поза
 - face detection
 - техника съёмки
- сбор в базу данных
- соединить письма, фотографии и дневники

Вебдев

публикация итогов — или заведём страничку в Инстаграме, или сделаем свой ресурс

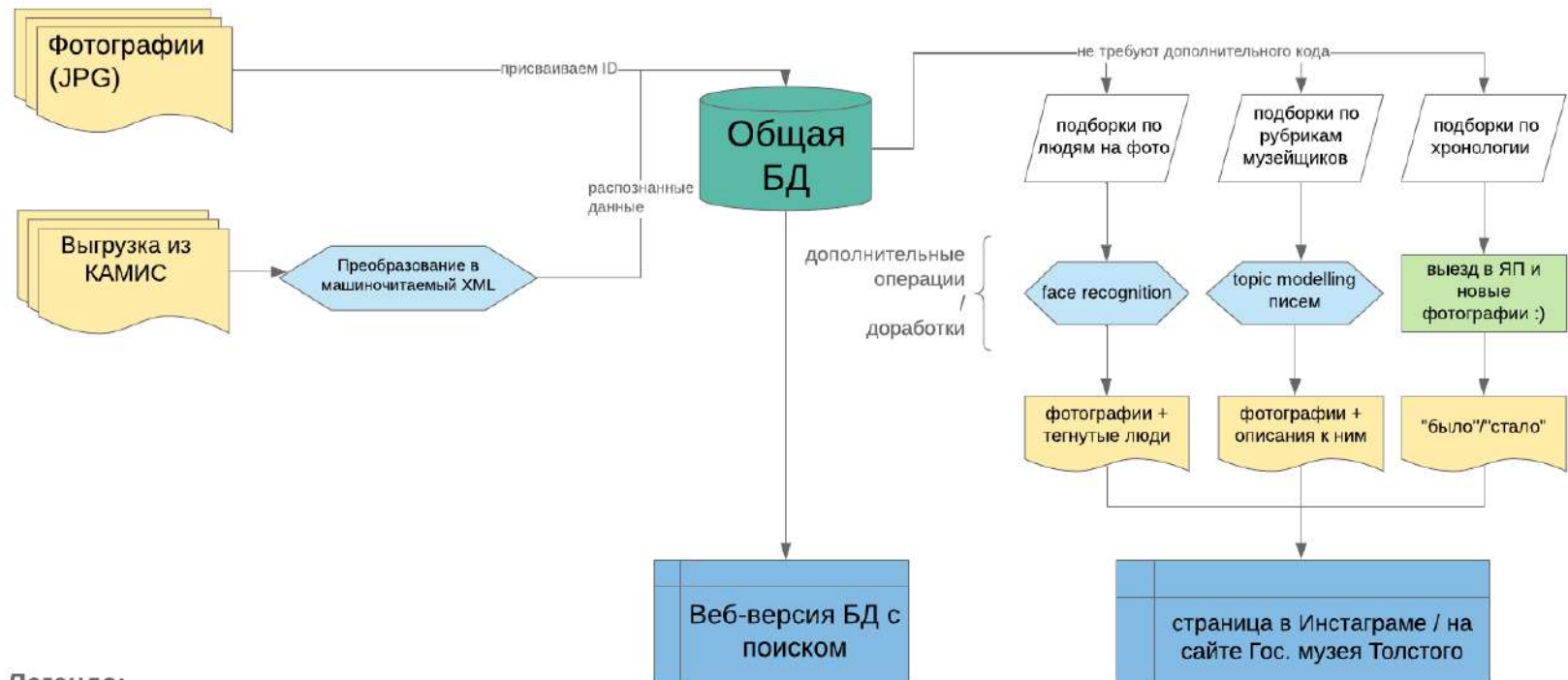
КАМИС и ТЕІ

КАМИС — система, которой пользуются сотрудники музеев России.

ТЕІ — более универсальный стандарт, к которому мы хотим прийти.

Теги КАМИС	Значение	Образец	Тег ТЕІ	Кол-во файлов, в которых есть тег (из 6)
AUTHFUL	полное имя автора		<author>	6
AUTHOR	автор			6
COMPNAME	подробное (?) описание			1
CREAT	полная дата	1.01.1895, 1 января 1895, ...		5
CREAT1				6
CREAT2				6
DOPAVT				6
EVECOD				6
ID				6
IDENTIF				6
MUSCODE		27		6
NFOND	название фонда, в котором лежит документ	<p>Фотофонд</p>		6
PAICODE		00000072		6
PNAM	название		<title>	6
SIZES	размеры фотографии, материалы			6
SPIEX				2
TYPOLOG			<sourceDesc>	6
V_PAICODE				6
VLAD1	владелец	<p>ФГБУК...</p>	<publisher>	6
WAYINF				5
	Обозначения:			
	id всех сортов			
	автор			

Процессы



Легенда:

- документы, данные (I/O)
- промежуточные результаты
- код / программные преобразования
- действия / непрограммные преобразования
- база данных
- веб-отображение

Что уже успели решить

- Непонятные теги КАМИСа заменяем на понятные теги TEI:
`<CREAT> => <date type="created">`
- Одноуровневое XML-дерево КАМИСа преобразуем в многоуровневое TEI-описание

```
<?xml version="1.0" encoding="UTF-8"?>
<document museum="27" paicode="00000107" version="1">
  <NFOND><p>Фотофонд</p></NFOND>
  <PAICODE><p>00000107</p></PAICODE>
  <AUTHF><p>Ходасевич Фелицан Иванович</p></AUTHF>
  <MUSCODE><p>27</p></MUSCODE>
  <SIZES><p>8,6 x 5,7 (фотография) 10,0 x 6,3 (бланк)</p></SIZES>
  <SPIEX><p>Выставка &quot;Старший сын&quot; (С.Л. Толстой)</p></SPIEX>
  <AUTHOR><p>Ходасевич Ф.И.</p></AUTHOR>
  <MUSEUM><p>ФГБУК «Государственный музей Л.Н. Толстого»</p></MUSEUM>
  <SPIRUBR><p>С.Л. Толстой</p></SPIRUBR>
  <IDENTIF><p>ГМТ КП-4578</p></IDENTIF>
  <CREAT1><p>1878</p></CREAT1>
  <INV><p>0-2134</p></INV>
  <AUTHR><p>Автор: Ходасевич Ф.И.</p></AUTHR>
  <CREAT2><p>1878</p></CREAT2>
  <V_PAICODE><p>107</p></V_PAICODE>
  <GEGR><p>Тула</p></GEGR>
  <TYPOLOG><p>Фотофонд. Фото, фоно, кино, видео</p></TYPOLOG>
  <IMGFILS><p>/media/spf/8/328690/as001001.spf</p></IMGFILS>
  <COMPNAM><p>Погрудно, 3/4 вправо. В тёмной куртке с белым воротником.</p></COMPNAM>
  <ID><p>57</p></ID>
  <EVECOD><p>1011193</p></EVECOD>
  <CREAT><p>1878</p></CREAT>
</document>
```

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>#{PNAM}</title>
        <author>#{AUTHFUL}</author>
      </titleStmt>
      <editionStmt>
        <!-- Скопее всего, у нас отсутствует + это optional -->
      </editionStmt>
      <extent>
        <!-- Скопее всего, у нас отсутствует + у нас не текст + это optional -->
      </extent>
      <publicationStmt>
        <publisher>#{VLAD1}</publisher>
        <authority>#{VLAD1}</authority>
      </publicationStmt>
      <seriesStmt>
        <!-- Скопее всего, у нас отсутствует + это optional -->
      </seriesStmt>
      <notesStmt>
        <!-- Скопее всего, у нас отсутствует + это optional -->
      </notesStmt>
      <sourceDesc>
        <!-- Что-то из трёх -->
        <date when="#{roa1}">#{CREAT}</date>
        <date notBefore="#{roa1}">#{CREAT1}</date>
        <date notAfter="#{roa1}">#{CREAT2}</date>
        <p>Фотография</p>
        <p>#{DESCM}</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
</TEI>
```


Что уже успели решить

Структура выгрузки из КАМИС сложная:

Папка 1: XML-описание

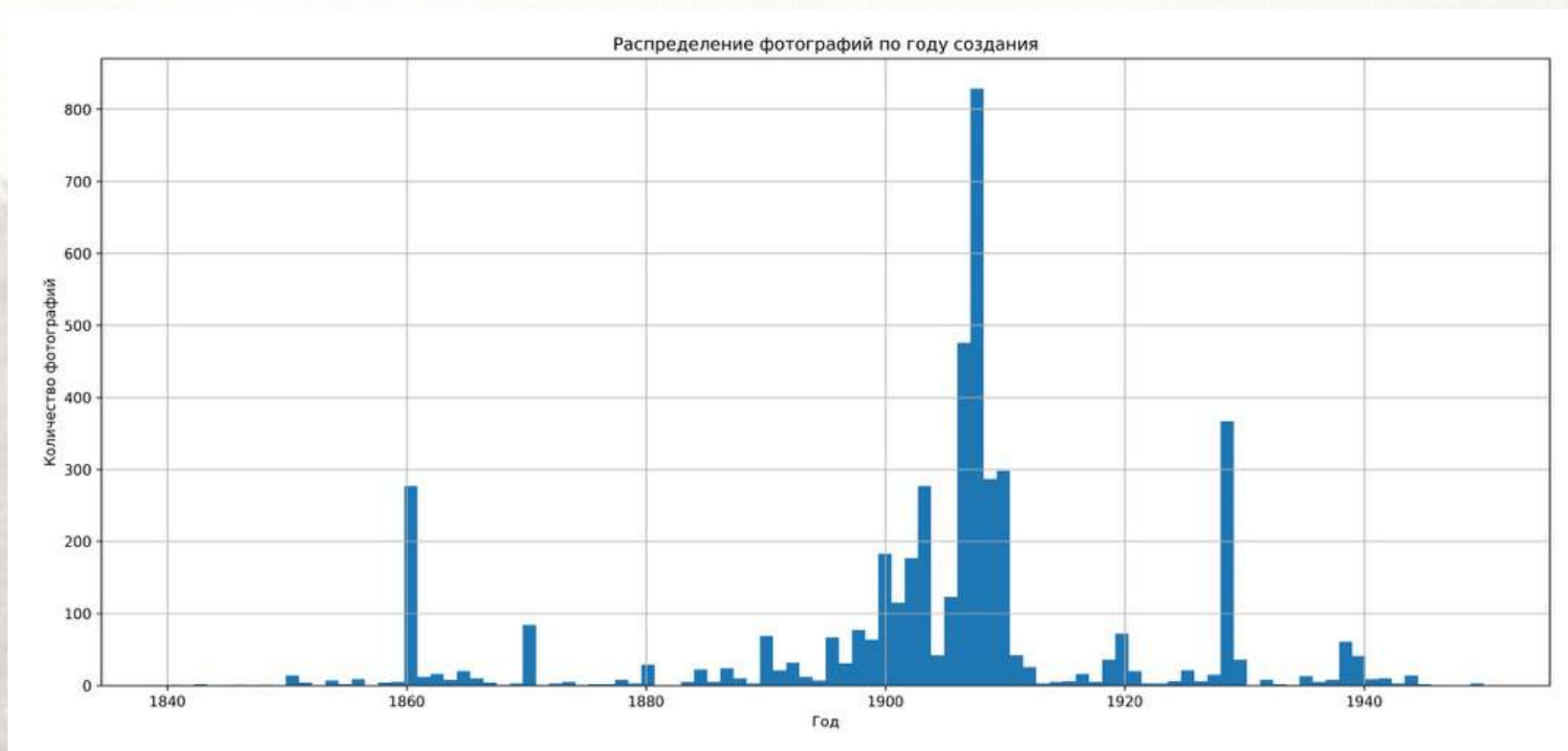
Папка 2: миниатюра фотографии (200 пикс по большей стороне) + сама фотография

=> сделали свою таблицу с метаданными по всем фотографиям с указанием всех путей.

Теперь можно проводить какие-то количественные исследования, даже не открывая сами фотографии!

Исх. № 18314/19

Что узнали из метаданных



Итого: 19324/19

Что узнали из метаданных

Рубрика	Кол-во фотографий
1900-е годы	325
Театральные постановки и экранизации произведений Толстого	217
Л.Н. Толстой в кругу друзей и знакомых	199
Интерьеры	157
Разные лица (персоналии)	156
М.Н. Толстая и её семья	112
Дома и др. строения	111
Л.Н. Толстой и С.А. Толстая	109
Парк и сады (аллеи, лес, пруды)	107

Что осталось сделать

- Доделать всю конвертацию в TEI
- Выложить датасеты в открытый доступ (coming soon)
- Прикрутить компьютерное зрение и достать с его помощью больше информации про содержимое фотографий
- Стать популярными :)



Спасибо за внимание!

Инв. № 18374/19