

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»  
АССОЦИАЦИЯ «ИСТОРИЯ И КОМПЬЮТЕР»

# **ЦИФРОВАЯ ГУМАНИСТИКА: РЕСУРСЫ, МЕТОДЫ, ИССЛЕДОВАНИЯ**

Материалы  
Международной научной конференции  
(г. Пермь, 16–18 мая 2017 г.)

**Часть 1**



Пермь 2017

УДК 004:3  
ББК 32.97+6/8  
Ц75

**Цифровая** гуманитаристика: ресурсы, методы, исследования: материалы Междунар. науч. конф. (г. Пермь, 16–18 мая 2017 г.): в 2 ч. / Перм. гос. нац. исслед. ун-т. – Пермь, 2017. – Ч. 1. – 175 с.  
ISBN 978-5-7944-2908-4 (ч. 1)  
ISBN 978-5-7944-2907-7

Сборник содержит материалы докладов, представленных на конференции, посвященной состоянию и основным проблемам развития Digital Humanities. В первой части рассматриваются социальные, институциональные, глобальные, языковые и мультикультурные аспекты цифровых гуманитарных наук, вопросы использования информационных технологий в области текстологии, текстометрии, лингвистики, компьютерного источниковедения и математических методов обработки информации.

**УДК 004:3**  
**ББК 32.97+6/8**

*Издается по решению ученого совета  
историко-политологического факультета и ученого совета  
механико-математического факультета ПГНИУ*

#### **Редакционный совет**

д-р ист. наук С.И. Корниенко, д-р филол. наук В.А. Баранов,  
д-р филол. наук К.И. Белоусов, канд. филос. наук Е.М. Березина,  
канд. ист. наук Д.Б. Вершинина, канд. пед. наук Д.А. Гагарина,  
д-р филол. наук Е.В. Ерофеева, д-р филол. наук С.Л. Мишланова,  
канд. ист. наук Н.Г. Поврозник, А.Р. Харисова

**Ответственный редактор**  
С.И. Корниенко

*Издание осуществлено при поддержке гранта РФФИ, проект № 17-06-20071*

ISBN 978-5-7944-2908-4 (ч. 1)  
ISBN 978-5-7944-2907-7

© ПГНИУ, 2017

## ОГЛАВЛЕНИЕ

### СОЦИАЛЬНЫЕ, ИНСТИТУЦИОНАЛЬНЫЕ И ГЛОБАЛЬНЫЕ АСПЕКТЫ ЦИФРОВЫХ ГУМАНИТАРНЫХ НАУК

<i>Антопольский А.Б.</i> Об организации мониторинга информационных ресурсов академических организаций социально-гуманитарного профиля .....	6
<i>Багрецова В.А., Кижнер И.А.</i> Проблема разнообразия интеллектуальной традиции в сообществе Digital Humanities .....	9
<i>Баканов С.А.</i> «Цифровая история»: вызовы и риски для профессионального сообщества историков .....	12
<i>Бородкин Л.И.</i> Digital History и историческая информатика: конвергенция или дивергенция? .....	15
<i>Владимиров В.Н.</i> Историческая информатика: прошлое, настоящее и будущее .....	19
<i>Мазур Л.Н.</i> Цифровая революция как фактор визуализации исторической науки .....	22
<i>Мохов В.П.</i> Социальное и институциональное в цифровой гуманитаристике .....	27
<i>Румянцев М.В.</i> Формирование Российской ассоциации цифровых гуманитарных наук .....	30
<i>Семченкова М.Н.</i> Проблемы правовой защиты результатов исследований цифровой гуманитаристики .....	33
<i>Ярков А.П., Пушкаревич О.А.</i> Информационные технологии экстремистских сюжетов. опыт анализа и противодействия .....	36

### КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА, ТЕКСТОЛОГИЯ И ТЕКСТОМЕТРИЯ

<i>Баранов В.А.</i> Количественный и статистический анализ средневековых славянских текстов: инструментарий корпуса «Манускрипт» и методика его использования .....	40
<i>Белоусов К.И., Дусакова Г.Ф., Леонов Д.В.</i> Анализатор ритмической структуры ритмической структуры текста: атрибуция текстов на основе ритмических паттернов .....	49
<i>Блинова О.В.</i> Транскрипция устного диалогического дискурса: форматы, инструменты и опыт их использования. ....	52
<i>Григоршин С.В.</i> Древние корпусы текстов в цифровую эпоху: разработка инновационного ресурса по шумерской клинописи .....	55
<i>Гулевская А.Е., Мишланова С.Л.</i> Мультимодальный словарь фитонимов нового типа .....	58
<i>Ерофеева Е.В.</i> Ментальный лексикон: сеть или поле? .....	62
<i>Крылов С.А., Фролова О.Е.</i> Использование специального корпуса в лингвистических целях .....	65

<i>Ланин В.В., Шаляева И.М., Скурихина А.Ю.</i> Мониторинг глобальных процессов на основе данных из интернет-новостей.....	67
<i>Москин Н.Д., Лебедев А.А., Варфоломеев А.Г.</i> Применение метода главных компонент для анализа поэтических и фольклорных текстов.....	70
<i>Нитко Я.М.</i> Графосемантическое моделирование пространственно-распределенных эргонимов.....	74
<i>Орехов Б.В.</i> Объяснимы ли ошибки В. А. Жуковского с помощью анализа данных?.....	77
<i>Русинова И.И., Гранова М.А.</i> К вопросу о принципах подачи в электронном словаре неоднословных единиц, называющих мифологических персонажей.....	80
<i>Филипсон С.К., Ланин В.В.</i> Онтология маркеров академического стиля.....	83
<i>Худякова Е.С.</i> Машинное обучение в социолингвистических задачах ..	85
<i>Чуприна С.И., Костарева Т.А.</i> Интеграция подходов на базе машинного обучения и методов онтологического инжиниринга для решения задач класса АОТ.....	89
<i>Ягунова Е.В., Крутченко О.В.</i> Моделирование процедур контекстной предсказуемости в зависимости от жанровых и стилевых характеристик текста .....	93
<i>Яруллин Д.В.</i> Применение методов машинного обучения в анализе трансмедийных объектов.....	97
<i>Ovchinnikova I., Ermakova L., Mothe J.</i> Do Frequent media words worsen query expansion?.....	100

## **КОМПЬЮТЕРНОЕ ИСТОЧНИКОВЕДЕНИЕ И МАТЕМАТИЧЕСКИЕ МЕТОДЫ**

<i>Баранова Е.В.</i> Формирование советской городской идентичности: контент-анализ видовых открыток Калининграда 1950–1980-е гг. ...	103
<i>Брюханова Е.А.</i> Материалы переписи 1897 года как источники по социальной топографии сибирских городов.....	106
<i>Будеркина Е.Ю.</i> Методика обработки печатных СМИ для последующего контент-анализа при условии плохой сохранности материала (на примере СМИ Калининградской области в послевоенное время).....	109
<i>Володин А.Ю.</i> Оцифровка как информационное моделирование: освоение многослойности электронных форматов в исторических исследованиях .....	112
<i>Гилязева С.А.</i> Использование объектно-ориентированного подхода при проектировании источник-ориентированных систем.....	116
<i>Иванков Э.В.</i> «Бумажные войны в 1917 г.»: опыт контент-анализа агитационных листовок кадетов и эсеров.....	119

<i>Кирьянов И.К., Сотник А.В.</i> Возможности применения информационных технологий при изучении парламентской элиты позднеимперской России .....	123
<i>Колдаков Д.В., Сарафанов Д.Е.</i> Информационные технологии в изучении метрических книг Барнаула второй половины XVIII в. (к методике кодирования и идентификации записей об участниках церковных обрядов) .....	127
<i>Kornienko S.I., Gagarina D.A., Kharisova A.R.</i> Meta description as a kind of distant reading of provincial periodicals collections .....	131
<i>Мельниченко М.А., Тышкевич Н.Б.</i> «Прожито» от рукописи до корпуса: сбор, разметка, анализ дневниковых текстов.....	134
<i>Поврозник Н.Г.</i> Указатели к стенографическим отчетам заседаний Государственной думы начала XX в.: анализ изменения структуры источника на основе XML .....	137
<i>Поврозник Н.Г., Сметанин А.В.</i> Социально-сетевой анализ гласных губернских земских собраний: к постановке проблемы .....	140
<i>Рожнева Ж.А., Акашева А.А.</i> Электронная почтовая переписка как потенциальный исторический источник .....	143
<i>Саенко А.В.</i> Политические процессы в Польше на основе контент-анализа «Калининградской правды» .....	147
<i>Станкевич Ю.Д., Кижнер И.А.</i> Методика сбора данных для цифрового научного издания с визуальными материалами .....	151
<i>Турсунова Э.Т.</i> Источниковедческие проблемы использования интернет-ресурсов в изучении политической истории Кыргызстане на современном этапе .....	154
<i>Uspenskiy V.S.</i> Social network analysis as a tool for the social history of the Sovereign's court in 16th century.....	158
<i>Чечулин В.Л.</i> Негэнтропийный подход к описанию социальных революций (на примере России и СССР) .....	161
<i>Яценко Ю.В., Гагарина Д.А.</i> Репрезентация Японии в Пермской губернской периодике в годы Первой мировой войны .....	163
<i>Brenner S., Sablatnig R.</i> Reading between the lines – multispectral imaging of ancient manuscripts .....	166
<i>Goderle W.</i> Space and the representation of space digitization and analysis of the Central European Census 1869–1910 .....	169
<i>Thorvaldsen G.</i> Access to census microdata and aggregates .....	172

# СОЦИАЛЬНЫЕ, ИНСТИТУЦИОНАЛЬНЫЕ И ГЛОБАЛЬНЫЕ АСПЕКТЫ ЦИФРОВЫХ ГУМАНИТАРНЫХ НАУК

УДК 002.2.001.5:009

## ОБ ОРГАНИЗАЦИИ МОНИТОРИНГА ИНФОРМАЦИОННЫХ РЕСУРСОВ АКАДЕМИЧЕСКИХ ОРГАНИЗАЦИЙ СОЦИАЛЬНО-ГУМАНИТАРНОГО ПРОФИЛЯ

*Антопольский А.Б.*

Институт научной информации по общественным наукам РАН,  
Россия, Москва, Нахимовский просп., 51/21, ale5695@yandex.ru

Описывается мониторинг информационных ресурсов академических организаций социального и гуманитарного профиля. Излагаются цели и задачи исследования, очерчивается сфера мониторинга, обсуждаются методические вопросы его проведения, приводятся краткие результаты.

*Ключевые слова:* мониторинг, информационные ресурсы, инфосфера, социальные и гуманитарные науки.

Недавнее трагическое событие – пожар в ИНИОНе – послужило стимулом к активизации дискуссий о дальнейшем развитии института как одного из центральных элементов инфосферы общественных наук. Очевидно, что происходит чрезвычайно динамичное развитие этой инфосферы, благодаря Интернету и другим информационно-телекоммуникационным технологиям появляется много новых каналов коммуникации, а многие традиционные устаревают. Это приводит к необходимости проведения специальных исследований состояния информационных ресурсов (ИР) и тенденций их изменений.

Подобное исследование информационных ресурсов академических организаций социально-гуманитарного профиля начато в ИНИОНе. Его задачами являются:

- получение исходных данных для дальнейшего скоординированного развития информационных систем подведомственных ФАНО учреждений социально-гуманитарного профиля;
- организация эффективной навигации в информационном пространстве социально-гуманитарных наук подведомственных ФАНО учреждений;
- определение параметров перспективной программы оцифровки библиотечных, архивных и музейных фондов учреждений РАН;

- определение параметров перспективной программы обеспечения сохранности (архивного хранения) ценной научной электронной информации подведомственных ФАНО учреждений;

- разработка предложений по условиям и перспективам развития информационного пространства в сфере общественных и гуманитарных наук и определение роли ИНИОНа РАН по координации развития информационного пространства.

В ходе проведения исследования предполагалось организовать мониторинг ИР академических организаций социально-гуманитарного профиля, описать и классифицировать ИР, создать соответствующий каталог и разместить его в Интернете.

В качестве источников информации использовались:

- результаты опроса академических организаций, подведомственных ФАНО,

- данные с сайтов учреждений,

- данные информационных систем и каталогов, статистической отчетности и др.

В ходе исследований решались многочисленные методические вопросы, связанные с границами инфосферы общественных и гуманитарных наук, типологией информационных ресурсов, несопоставимостью данных из различных источников.

Прежде всего, это касалось определения организационной сферы исследования. Всего были проанализированы ресурсы около 200 учреждений ФАНО, в т. ч. учреждений отделений социогуманитарного профиля, универсальных ресурсов библиотек, архивов и музеев, научных центров и учреждений, входящих в состав НЦ.

Принципиальной особенностью мониторинга был интегрированный подход к ИР – цифровым, традиционным, первичным, вторичным и др. В ходе отработки методики мониторинга решались многочисленные методические вопросы. Вот некоторые из них:

- многие ИР, например, библиотечные фонды или коллекции научных отчетов, могут быть универсальными по тематике.

- не совсем ясно, к какому профилю можно отнести множество различных ИР по истории естественных и точных наук, прежде всего архивные и музейные фонды, а также массивы документов научных центров и региональных отделений РАН;

- многие социогуманитарные по тематике ИР, например, лингвистические, создаются и поддерживаются учреждениями, не входящими в профильные;

- ИР, реально имеющиеся в учреждениях и на их сайтах, не всегда однозначно относятся к выделенным типам ИР;

- тематика ИР во многих случаях не может быть определена по имеющимся данным, поэтому на данном этапе она была исключена;
- количество признаков, описывающих ресурсы, было минимизировано для обозримости результатов;
- некоторые виды ИР, например, музейные фонды, практически не описаны.

В результате были получены следующие данные по отдельным типам ресурсов

1. Архивные фонды (св. 5 тыс. фондов 44 учреждений).
2. Фонды традиционных библиотек – св. 70 библиотек гуманитарного или универсального профиля.
3. Музейные фонды – св. 100 музеев РАН.
4. Сайты учреждений, сайты-сателлиты и аккаунты – около 300.
5. Электронные библиотеки, другие информационные системы – около 400.
6. Лингвистические ИР (словари, корпуса, процессоры) – около 100.
7. Библиотечные традиционные и электронные каталоги, библиографические и реферативные БД – св. 300.
8. Библиографические и аннотированные списки, перечни и указатели – св. 200.
9. Ресурсы по неопубликованным документам (отчеты, диссертации, описания проектов, грантов и пр.), в т. ч. отчетов в ЕГИСУ, – 1.5 тыс., диссоветов – 78.
10. Конференции, семинары, другие научные события – около 130 перечней.
11. Персональные ресурсы – св. 100 перечней.
12. Справочно-энциклопедические ресурсы – около 90.
13. Периодические, продолжающиеся и сериальные издания – св. 400.
14. Фото, аудио и видео ресурсы – менее 100.

Приведенные цифры являются приблизительными или обобщающими. Так, например, под персональными ресурсами понимается наличие на сайте учреждения перечня персон (сотрудников, академиков, лауреатов и др.) с некоторыми сведениями о них, обычно с биографиями и списками трудов. Количество персон в таких списках установить сложно.

Имеется определенное число пересечений, например, разница между электронными каталогами и библиографическими БД неочевидна.

Особую важность и сложность представляет анализ периодики. Будучи пока еще важнейшим инструментом научной коммуникации,

сфера научных журналов отличается крайней неупорядоченностью, дублированием электронной и традиционной форм, разнообразием способов архивного хранения и доступа и другими свойствами, затрудняющими анализ.

Главным выводом исследования является то, что мониторинг ИР нужно вести постоянно, углубленно и желательнo коллективными усилиями.

Более точные и полные результаты мониторинга будут представлены в виде специального ресурса на сайте ИНИОНа до конца года.

### **Библиографический список**

1. *Васютин О.* Калининград-Кёнигсберг: архитектура советского и постсоветского периодов (конец первой половины XX – начало XXI вв.) URL: [http://sg39.ru/upload/iblock/cec/oleg-vasyutin.-\\_kaliningrad\\_-\\_kyenigsberg.-arkhitektura-sovetskogo-i-postsovetskogo-periodov.pdf](http://sg39.ru/upload/iblock/cec/oleg-vasyutin.-_kaliningrad_-_kyenigsberg.-arkhitektura-sovetskogo-i-postsovetskogo-periodov.pdf) (дата обращения: 30.03.2017).
2. *Тагрин М.* Мир в открытке. М., 1978.

### **ORGANIZATION OF MONITORING OF SOCIAL-HUMANITARIAN ACADEMIC ORGANIZATIONS' INFORMATION RESOURCES**

*Antopolskiy A.B.*

Institute of Scientific Information on Social Sciences of the Russian Academy of Sciences, 51/21, Nahimovskiy ave., Moscow, Russia, ale5695@yandex.ru

The paper analyses the monitoring of social and humanitarian academic organizations' information resources. The author describes the goals of the study, the scope of monitoring, discusses the methodological issues, and summarizes the results.

*Key words:* monitoring, information resources, infosphere, Social Sciences and Humanities.

УДК 303.064

### **ПРОБЛЕМА РАЗНООБРАЗИЯ ИНТЕЛЛЕКТУАЛЬНОЙ ТРАДИЦИИ В СООБЩЕСТВЕ DIGITAL HUMANITIES**

*Багрецова В.А., Кижнер И.А.*

Сибирский федеральный университет, Россия, Красноярск, просп. Свободный, 79, bagvlada@yandex.ru

Цифровые гуманитарные науки – междисциплинарная область, которая испытывает влияние многих академических традиций. В докладе будут показаны основные источники разнообразия (единства) традиции в этой области, представлены результаты анализа университетских центров, публикаций и проек-

тов, которые были отмечены в ответах респондентов, ученых, активно работающих в цифровых гуманитарных науках.

*Ключевые слова:* Digital Humanities, ДН, цифровые гуманитарные науки, гуманитарная информатика, интеллектуальная традиция, источники влияния.

Рассматривая историю развития цифровых гуманитарных наук (ДН), мы видим, что эта междисциплинарная область объединяет разнообразие подходов, научных направлений и интеллектуальных традиций. Проводятся исследования относительно того, какие географические регионы и демографические группы представлены среди специалистов в этой области [1]. Однако важно получить представление не только о демографическом и географическом разнообразии в области цифровых гуманитарных наук, но и об интеллектуальном разнообразии [2, 4].

Источники влияния – один из важнейших факторов формирования научной традиции, анализ которых позволит выявить наиболее значимые элементы, из которых она складывается. Многие авторы обращают внимание на то, что англо-американские центры имеют большое влияние на развитие цифровых гуманитарных наук [5, 8]. Большая часть цитируемых работ в этой области публикуется учеными из США, Великобритании, Германии. Метод, который учитывает только центрально-ориентированные исследования ДН, создает неверную картину мира [5].

Задача этой работы – выявить ключевые особенности англо-американской традиции и традиций других стран. Это поможет обнаружить пересечения между методологиями различных стран и выявить те важные параметры, которые объединяют научные традиции. Целью исследования является выявление основных каналов трансфера знаний и определение того, какое место они занимают на интеллектуальном поле.

Для определения источников влияния интеллектуальной традиции в ДН мы провели анкетирование участников конференций Digital Humanities 2013 (Небраска-Линкольн, США) и Digital Humanities 2014 (Лозанна, Швейцария). Эта ежегодная конференция Альянса цифровых гуманитарных наук [2] дает представление о том, как мир видит ДН [3]. Прежде чем проводить анкетирование, мы собрали информацию об участниках конференций, чтобы проанализировать состав участников по ряду параметров, представленных в исследовании Террас [7]. Затем были собраны данные о каждом участнике конференции.

На данном этапе проекта мы собрали информацию о 1463 участниках конференции. Как и ожидалось, ученые из США (38 %), Германии (12 %), Британии (11 %), Канады (9 %), а также из Франции, Голландии, Швейцарии, Японии, Италии и Ирландии составляют большую часть сообщества цифровых гуманитарных наук и представляют доклады, связанные с академическими традициями этих стран. Кафедры, представившие максимальное количество докладов, –

кафедры филологии, истории и кафедры (лаборатории) цифровых гуманитарных наук.

В докладе на конференции будет подробно рассказано о результатах анализа состава участников конференции и анализа ответов респондентов на вопросы анкеты. Будут приведены примеры научных публикаций и проектов в области цифровых гуманитарных наук.

Значимость проекта заключается в выявлении источников влияния на формирование интеллектуального поля, определение их места на иерархическом дереве цифровых гуманитарных наук, создании базы данных источников влияния и интеллектуальных традиций в цифровых гуманитарных науках.

### **Библиографический список**

1. *Fiormonte, D.* Digital Humanities from a global perspective // Laboratorio dell'ISPF. 2014. URL: [http://www.ispf-lab.cnr.it/2014\\_203.pdf](http://www.ispf-lab.cnr.it/2014_203.pdf) (дата обращения: 10.04.2017).
2. *ADHO.* URL: <https://adho.org/conference> (дата обращения: 10.04.2017).
3. *Eichmann N., Jorgensen J., B. Weingart S.* Representation at Digital Humanities Conferences (2000-2015) // The scottbot irregular. Data are everywhen. 2016. URL: <http://scottbot.net/representation-at-digital-humanities-conferences-2000-2015/#return-note-41565-1> (дата обращения: 10.04.2017).
4. *EADH: Diversity and Inclusivity.* URL: <https://eadh.org/about/diversity-and-inclusivity> (дата обращения: 10.04.2017).
5. *Ortega E.* Notes towards moving ideas around the DH world // Readers of Fiction. 2014. URL: <http://lectoresdeficcion.blogs.cultureplex.ca/2014/02/03/notes-towards-moving-ideas-around-the-dh-world/> (дата обращения: 10.04.2017).
6. *Risam R.* Other Worlds, Other DHs: Notes towards a DH Accent // Digital Scholarship in the Humanities Advance Access. 2015. URL: [http://www.academia.edu/1932310/Towards\\_a\\_Cultural\\_Critique\\_of\\_Digital\\_Humanities](http://www.academia.edu/1932310/Towards_a_Cultural_Critique_of_Digital_Humanities) (дата обращения: 10.04.2017).
7. *Terras M.* Disciplined: Using Educational Studies to Analyse “Humanistes Computing” // Defining Digital Humanities/ Ashgate Publishing Company, 2013. С. 67–96.
8. *Terras M.* Quantifying Digital Humanities // UCL Center for Digital Humanities. 2011. URL: <http://www.ucl.ac.uk/infostudies/melissa-terras/DigitalHumanitiesInfographic.pdf> (дата обращения: 10.04.2017).

### **INTELLECTUAL DIVERSITY IN DIGITAL HUMANITIES**

*Bagretsova V.A., Kizhner I.A.*

Siberian Federal University, Russia, Krasnoyarsk, Svobodny ave., 79  
[bagvlada@yandex.ru](mailto:bagvlada@yandex.ru)

Intellectual diversity is an important part of the concept of diversity in digital humanities, a field that is interested in studying culture with its various forms and attitudes. We demonstrate some of the sources for intellectual diversity in the field using the results of the survey that was sent to the attendees of DH2013 and DH2014, annual ADHO conferences

*Key words:* DH, Digital Humanities.

## «ЦИФРОВАЯ ИСТОРИЯ»: ВЫЗОВЫ И РИСКИ ДЛЯ ПРОФЕССИОНАЛЬНОГО СООБЩЕСТВА ИСТОРИКОВ

Баканов С.А.

Челябинский государственный университет,  
Россия, Челябинск, ул. Братьев Кашириных, 129, [bakanov-s@mail.ru](mailto:bakanov-s@mail.ru)

Рассматривается преемственность между лингвистическим и цифровым вызовами, стоящими перед современной исторической наукой. Подчеркивается, что наступление цифровой эпохи способствовало дальнейшему углублению кризиса научного исторического знания. Цифровые гуманитарные науки, частью которых является цифровая история, представляют угрозу для существования истории в качестве научной дисциплины, потому что важнейшие характеристики цифровых гуманитарных наук, такие как открытость, доступность, демократичность, ориентация на массового пользователя и т. д. противоречат самому смыслу существования истории как науки, изучающей объективную историческую реальность. Они препятствуют критике исторических источников, историческому и историографическому анализу и нарушают принцип научной объективности.

*Ключевые слова:* цифровая история, цифровые гуманитарные науки, цифровой вызов, лингвистический вызов, научная объективность исторического знания.

Более четырех десятилетий историческая наука испытывает постоянное и все нарастающее воздействие внешней среды, ставящее под сомнение сам объект исторического знания – историческую реальность. Первым тяжелым ударом для самоопределения и идентичности историков стал лингвистический вызов 1970-х – 1980-х гг., приведший к осознанию того, что прошлая реальность может оказаться не объектом познания субъекта-историка, но всего лишь дискурсивной практикой этого исследователя, конструирующей совершенно искусственную реальность. Это поставило перед историческим сообществом ряд вопросов: о критериях объективности и грани между фактом и вымыслом, о способах контроля со стороны исследователя над собственной творческой деятельностью, наконец, о самой познаваемости прошлого [1]. Вместо одной объективной исторической реальности стала допускаться множественность реальностей субъективных. Сомнения в научности истории звучали и раньше, но именно с этого момента представление о том, что историю следует относить скорее к *Arts*, а не к *Sciences* стали разделяться и значительной частью представителей самого исторического цеха. Это неизбежно отразилось на авторитете исторических исследований сначала в академической, а затем и в публичной среде. У современной цивилизации возникло определенное недоверие

к профессиональным историкам, что в свою очередь привело к росту влияния любителей истории, чьи исторические реконструкции стали пользоваться активным спросом. С появлением Интернета этот спрос стал обеспечиваться ростом технических возможностей новых медиа, заинтересованных в контенте, ориентированном на вкусы разнородной аудитории. Популярная история, по сравнению с исторической наукой, оказалась более доступной и более востребованной публикой.

Наступление цифровой эпохи способствовало дальнейшему углублению кризиса научного исторического знания и стало для него новым вызовом. Основные положения этого цифрового вызова изложены в двух манифестах Digital Humanities (DH), опубликованных Альянсом цифровых гуманитарных наук в 2009 и 2010 гг. В одном из них говорится: «Цифровая революция» современного общества видоизменяет и ставит под вопрос традиционные формы создания и распространения знаний. Ученый-гуманитарий более не является единственным производителем, слугой и распространителем гуманитарного знания» [2], а в другом декларируется «реконфигурация иерархических отношений между учителем и учеником; экспертом и не-экспертом, академическим и обывательским сообществом <...> DH – это царство открытых ресурсов с открытым исходным кодом. Его экономика основана на изобилии, а не на редкости. Поэтому DH оценивает копию выше, чем оригинал <...> Отныне Процесс, а не результат или продукт является новым Богом» [3]. Главными целями нового движения провозглашаются открытый и свободный доступ к данным и демократизация знания, понимаемая как переход от экспертных систем знания к открытым. Такой подход применительно к истории предполагает тотальную оцифровку прошлого, то есть создание электронной копии всего массива данных об исторической реальности и обеспечение свободного доступа любого пользователя к этой копии. Ответвлением от направления Digital Humanities стала цифровая история, которая и была ориентирована на решение подобных задач. Сфера ее интересов широка: от формирования коллекций электронных копий исторических источников, виртуальных энциклопедий и тематических электронных ресурсов до визуализации исторических данных с помощью ГИС и 3D-технологий, а также элементов дополненной реальности. С одной стороны, цифровые технологии открыли новые горизонты для исторических исследований, но с другой – вызвали и появление серьезных рисков.

Во-первых, между исследователем и историческим источником начала образовываться новая преграда в виде электронной копии документа, которая существенно затрудняет возможности научной критики данного источника и его археографическое описание. Тезис о том, что копия важнее оригинала, подрывает основы источниковедения.

Во-вторых, применительно к электронным коллекциям документов справедливым является вопрос о репрезентативности выборки источников, принципах этого отбора и личных пристрастиях создателей таких коллекций и тематических ресурсов. Тезис о демократизации знания и свободном доступе не только к ресурсу, но и к его исходному коду открывает бесконечные возможности конструирования новых субъективных исторических реальностей, убивая принцип научной объективности в истории.

В-третьих, в жертву открытому и быстрому доступу к исторической информации приносится сам исторический анализ. Тезис о том, что результат исследования не важен, а важен только процесс, уничтожает в исторических исследованиях проблематику. Ведь если результат не важен, то проблему можно не решать, а только бесконечно ее описывать, используя все новые технические возможности.

В-четвертых, меняется форма научных дискуссий. Из пространства научной периодики они перемещаются в новые медиа, что становится препятствием для полноценного историографического анализа. Более того, в мире множественных исторических реальностей историография оказывается просто невостребованной.

В итоге цифровой вызов для исторической науки может стать не просто продолжением постмодернистского шока, но финальным аккордом ее существования в качестве науки.

### **Библиографический список**

1. *Ретина Л.П.* Вызов постмодернизма и перспективы новой культурной и интеллектуальной истории // *Одиссей: человек в истории.* М., Наука, 1996. С. 26.
2. *Manifesto for Digital Humanities*, 2010. URL: [http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf) (дата обращения: 11.03.2017).
3. *The Digital Humanities Manifesto 2.0*, 2009. URL: <http://tcp.hypotheses.org/411> (дата обращения: 11.03.2017).

### **“DIGITAL HISTORY”: CHALLENGES AND RISKS FOR THE HISTORIANS’ PROFESSIONAL COMMUNITY**

*Bakanov S.A.*

Chelyabinsk State University,  
129, Bratiev Kashirinykh st., Chelyabinsk, Russia, [bakanov-s@mail.ru](mailto:bakanov-s@mail.ru)

The paper examines the continuity between the linguistic and digital challenges faced by contemporary historical science. It highlights that the beginning of the digital age contributed to a further deepening of the crisis of historical knowledge. Digital Humanities, of which the Digital History is a part, pose a threat to the existence of history as an academic discipline. The most important ideas of Digital Humanities, such as openness, availability, democratic nature, orientation to the mass user, etc., contradict the very meaning of the existence of history as a science aimed at studying the so-called objective historical reality. They prevent critical analysis of

historical sources, as well as historical and historiographic analysis and violate the principle of scientific objectivity.

*Key words:* Digital History, Digital Humanities, digital challenge, linguistic challenge, scientific objectivity of historical knowledge.

УДК 004.9+93

## **DIGITAL HISTORY И ИСТОРИЧЕСКАЯ ИНФОРМАТИКА: КОНВЕРГЕНЦИЯ ИЛИ ДИВЕРГЕНЦИЯ?**

*Бородкин Л.И.*

Московский государственный университет им. М.В. Ломоносова,  
Россия, Москва, Ломоносовский просп., 27-4, Шуваловский корпус,  
borodkin@hist.msu.ru

В данной работе продолжена дискуссия по вопросам соотношения цифровых гуманитарных наук, цифровой истории и исторической информатики. Анализируются различия между исторической информатикой и цифровой историей, которые проявились за последние годы более четко. Рассмотрение этих различий представляет интерес не только в терминологическом плане, но и с точки зрения понимания тех задач и вызовов, которые обозначились в ходе «цифрового поворота» в исторической науке. Историческая информатика с самого начала позиционировалась как направление, развивающееся в рамках исторической науки, а цифровая история рассматривается сегодня как раздел цифровых гуманитарных наук. Тенденции последних лет выявляют доминирование процессов дивергенции этих направлений.

*Ключевые слова:* цифровая история, историческая информатика, цифровые гуманитарные науки, цифровой поворот, квантитативная история.

Прошло пять лет со времени выпуска в свет первого номера журнала «Историческая информатика», в котором была открыта дискуссия по вопросам соотношения цифровых гуманитарных наук (*Digital Humanities – DHum*), цифровой истории (*Digital History – DHist*) и исторической информатики (*Historical Information Science*) [1, 2]. Это соотношение было тогда не очень ясным, поскольку границы между указанными направлениями были достаточно размытыми. Так, список определений *Digital Humanities* насчитывал более 250 вариантов, дефиниции *Digital History* также включали немало противоречивших друг другу элементов. В течение пяти лет содержание этих направлений немного уточнилось, хотя вопросов здесь осталось больше, чем ответов. Тем не менее, различия между *DHist* и исторической информатикой проявились за эти годы более четко. Рассмотрение этих различий представляет интерес не только в терминологическом плане, но и с точки зрения понимания тех задач и вызовов, которые обозначи-

лись в ходе «цифрового поворота» в исторической науке; они обсуждались и в качестве одной из главных тем на XXII Международном конгрессе исторических наук, состоявшемся в Китае в 2015 г. (автору довелось принять участие в этих дискуссиях; одним из главных был вопрос о том, сводится ли цифровой поворот к совершенствованию инфраструктуры исторических исследований или он способен еще и повлиять на методологию исторической науки) [3]).

Обратимся к определениям. Еще четверть века назад, когда в русском языке появился термин «*историческая информатика*», это направление было определено как междисциплинарная область исследований на стыке исторической науки и информатики, в основе которой лежит совокупность теоретических и прикладных знаний, необходимых для создания, обработки и анализа электронных версий исторических источников всех видов [4, с. 31]. Это направление сформировалось (как за рубежом, так и в России) на факультетах истории, в исследовательских институтах исторического профиля и с самого начала позиционировалось в рамках исторической науки. Как и в других странах, в России это направление корнями уходило в количественную историю и большинство членов нового сообщества имело опыт активного применения ЭВМ именно при решении задач обработки статистических источников. Конечно, в ходе «микрокомпьютерной революции» 1980–1990-х гг. сфера использования историками информационно-коммуникационных технологий резко расширилась: пришла эра баз данных, компьютерного анализа исторических текстов, визуальных источников; затем (в середине 1990-х гг.) наступила эра Интернета, сопровождавшаяся дебатами на конференциях Ассоциации «История и компьютер» (АИК) того времени; в эти годы в трудах АИК появились публикации исследований с использованием ГИС, 3D моделей и т. д. В центре внимания сообщества АИК всегда был главный вопрос: что дают новые методы и технологии историческим исследованиям, какое приращение знания они приносят в ту или иную область исторической науки. Термин «цифровая история» практически не был известен членам международного профессионального сообщества *History and Computing* вплоть до начала 2000-х гг. Поэтому «ветераны движения» не без удивления узнают из многочисленных недавних историографических реконструкций, что цифровая история впервые заявила в 1960–1970-х гг. на ниве количественной истории (особенно любопытно это звучит по отношению к клиометрике – новой экономической истории, чьи достижения были увенчаны в 1993 г. Нобелевской премией).

Что касается *цифровой истории*, то ее многочисленные определения постепенно сузились, и сегодня в большинстве случаев она рассматривается как раздел цифровых гуманитарных наук, изучающий использование компьютерных технологий и цифровых медиа для ис-

торического анализа, исследований и визуализации данных. Большое внимание здесь уделяется вопросам оцифровки исторических источников (подробнее об этом см. [5]). Два основных направления *DHist* – знакомство интернет-аудитории с цифровыми архивами, интерактивными картами, хрониками событий (т. е. углубление исторических знаний пользователей Сети) и создание новых исследовательских инструментов для учёных-историков. Отметим, что создание нового программного обеспечения – удел небольших групп разработчиков, состоящих в основном из программистов. А вот первое из указанных двух направлений *DHist* – действительно сравнительно новое поле, которое было на периферии внимания сообщества специалистов по исторической информатике, ориентированного на решение (с помощью новых методов и ИКТ) исследовательских задач исторической науки.

Обобщая различные подходы к определению цифровой истории, англоязычная статья в «Википедии» отмечает, что это отрасль цифровых гуманитарных наук, реализующаяся, как правило, в качестве цифровой публичной истории и связанная в первую очередь с вовлечением онлайн-аудитории в исторический контент или с использованием цифровых методов исследований. Цифровая история, таким образом, ориентируется на создание цифровых архивов, интерактивных презентаций и карт, линий времени, аудиофайлов и виртуальных миров, чтобы сделать историю более доступной для пользователя. Впрочем, ясности в вопросе о структуре *DHist* до сих пор нет, некоторые авторы включают в нее, например, корпусную лингвистику или даже рассматривают *DHist* как расширение квантитативной истории, клиометрики (?!).

Формат данной публикации не позволяет провести более детальное сравнение двух рассматриваемых направлений. Резюмируя, отметим, что если историческая информатика позиционируется как междисциплинарное направление *исторических* исследований, имеющее сбалансированное соотношение прикладной (ресурсной) и аналитической компонент, то цифровая история относит себя к мультидисциплинарной области цифровых гуманитарных наук (в которой доминируют филологические науки). В этом контексте следует упомянуть о специфике задач цифровой истории, которая чувствует себя не очень уютно под «тентом» *DHum*. Так, С. Робертсон отмечает, что в центре внимания *DHist* находятся вопросы создания коллекций оцифрованных данных, их презентации, диссеминации в онлайн, работа с цифровыми картами, в то время как цифровые филологические науки включают в качестве доминирующих практик интеллектуальный информационный поиск и анализ текста, а также его семантическую разметку [6].

Таким образом, прошедшее пятилетие выявило более четко различия между исторической информатикой и цифровой историей. Первая развивает на новом витке те подходы к созданию тематических электронных ресурсов и их компьютеризованному анализу, которые сложились в течение четверти века, уделяя основное внимание получению содержательно значимых результатов на основе применения апробированных технологий анализа исторических источников разных видов, статистических методов и компьютерных моделей; этот арсенал методов и технологий постепенно расширяется. Цифровая история в большей степени ориентирована на формирование цифровой инфраструктуры и развитие онлайн-интерактивных коммуникаций с широким кругом Интернет-пользователей. Степень востребованности этой деятельности весьма высока.

Можно предположить, что дальнейшее развитие тенденций «цифрового поворота» усилит процессы дивергенции, обозначит более четко функции двух рассматриваемых направлений, имеющих заметные различия и в то же время дополняющих друг друга. Критерием успешности их развития будет уровень полученных практических результатов.

### **Библиографический список**

1. *Таллер М.* Дискуссии вокруг Digital Humanities // Историческая информатика. 2012. №1. С. 5–13.
2. *Бородкин Л.И.* Digital history: применение цифровых медиа в сохранении историко-культурного наследия? // Историческая информатика. 2012. №1. С. 14–21.
3. *Бородкин Л. И.* Цифровой поворот в дискуссиях на XXII Международном конгрессе исторических наук (Китай, 2015 г.) // Историческая информатика. 2015. № 3–4. С. 56–67.
4. *Историческая информатика: учебное пособие.* / под ред. Л.И. Бородкина и И.М. Гарсковой. М.: Мосгорархив, 1996.
5. *Володин А. Ю.* «Цифровая история»: ремесло историка в цифровую эпоху // Электронный научно-образовательный журнал «История», 2015. Т. 6, вып. 8(41).
6. *Robertson S.* The Differences between Digital Humanities and Digital History // Debates in the Digital Humanities 2016. University of Minnesota Press, 2016.

### **DIGITAL HISTORY AND HISTORICAL INFORMATION SCIENCE: CONVERGENCE OR DIVERGENCE?**

*Borodkin L.I.*

Lomonosov Moscow State University,  
27-4, Lomonosovsky ave., Moscow, Russia, borodkin@hist.msu.ru

In this paper, the discussion continues on the relationship of the digital humanities, digital history and historical information science. The differences between historical information science and digital history are analyzed, which have manifested themselves more clearly in recent years. Consideration of these differences is of interest

not only in the sense of terminology, but also from the point of view of understanding the tasks and challenges that have emerged in the process of the “digital turn” in historical science. Historical information science from the very beginning was positioned within the framework of historical science, and the digital history is considered today as a section of the digital humanities. Trends in recent years reveal the dominance of divergence processes in these approaches development.

*Key words:* digital history, historical information science, digital humanities, digital turn, quantitative history.

УДК 94:004

## **ИСТОРИЧЕСКАЯ ИНФОРМАТИКА: ПРОШЛОЕ, НАСТОЯЩЕЕ И БУДУЩЕЕ<sup>1</sup>**

*Владимиров В.Н.*

Алтайский государственный университет,  
Россия, Барнаул, просп. Ленина, 61, vlad@hist.asu.ru

Рассматриваются некоторые итоги и перспективы развития исторической информатики в России. Отмечается, что историческая информатика оказывает влияние на традиционные исторические исследования, в частности, на источниковедческие процедуры. Развитие исторической информатики обусловило появление в России «новой исторической демографии», базирующейся на первичных источниках и современных информационных технологиях. В настоящее время историческая информатика отвечает на новые вызовы, связанные с широким распространением мобильных устройств и ростом потока информации. Делается вывод, что адекватной реакцией на этот вызов может стать повышение эффективности использования исследовательских средств, оптимизация информационных потоков и постоянный рост профессионального уровня специалистов. Следует избегать опасности размывания границ между профессиональным научным историческим знанием и его популяризированным вариантом.

*Ключевые слова:* историческая информатика, источниковедение, информационные технологии, историческое знание, историческая демография.

Историческая информатика как сложившаяся научная дисциплина существует уже четверть века. Пройден этап первоначального внедрения компьютера в исторические исследования и образование, когда надо было доказывать саму необходимость этого процесса. Пройден период превращения информационных технологий в полноценный инструмент исторического исследования и один из важнейших компонентов лаборатории историка. В основном сгладился антагонизм между историками, применяющими компьютер, и историками, его отвергающими: последних становится все меньше; нынешнее состоя-

---

© Владимиров В.Н., 2017

<sup>1</sup> Публикация подготовлена при финансовой поддержке РГНФ (грант № 15-01-00207а).

ние этого противостояния можно охарактеризовать как дружественный нейтралитет. Наконец, в широком доступе появились электронные версии исторических источников всех видов, создано огромное число интернет-сайтов, посвященных истории, некоторые из них являются высокопрофессиональными и по праву занимают свое место в информационной базе исторической науки.

Влияние исторической информатики на историческую науку в целом нельзя оценивать только количеством опубликованных работ и числом их цитирований, хотя и то, и другое на сегодняшний день достаточно велико, и вряд ли лет двадцать тому назад на это надеялись даже самые закоренелые оптимисты и энтузиасты, работающие в рамках исторической информатики. Важно и то, что мы сегодня видим реальное влияние этой дисциплины на общее развитие научных исторических исследований. История все больше становится наукой в традиционном и полном смысле этого слова, в исторических исследованиях все больше внимания обращается на строгость исследовательских процедур и верифицируемость полученных результатов, все больше исторических работ становятся системными и структурированными, с четко определенным предметом, методологией и методикой исследования.

Важнейшим моментом представляется «обратное» влияние применения математических методов и компьютерных технологий на традиционные процедуры исторического исследования. Думается, что многие исследователи, использующие информационные технологии, обратили внимание на тот факт, что их отношение к обычной и ставшей уже рутинной внешней и внутренней критике исторического источника становится строже, ученый в этом отношении становится более требовательным к самому себе и своим оппонентам.

Неоценимо влияние исторической информатики и на другие специальные исторические дисциплины. Как уже отмечалось, особенно рельефно это проявляется в развитии исторической демографии [1, с. 23]. Фактически на наших глазах сформировалась «новая историческая демография», отличительной чертой которой является обращение к первичным номинативным источникам и широкое применение современных информационных технологий [2]. Ее особенностью является также то, что она существует как бы параллельно с продолжающей развиваться на базе советской историографической традиции демографической историей.

Наконец, историческая информатика в какой-то степени «порождает» принципиально новые для отечественной исторической науки предметные области. В качестве примера можно назвать историческое профессиоведение, вырастающее из синтеза исторической информатики с исследованиями по истории труда и профессий, перво-

димыми на базе международного исторического стандарта классификации профессий HISCO [3].

В настоящее время историческая информатика, как и вся историческая наука, столкнулась с новыми вызовами, обусловленными «мобильной информационной революцией», когда средства обработки информации, равные по мощности компьютерам десятилетней давности, лежат в кармане практически у каждого гражданина информационного общества, а сам поток информации стал настолько интенсивным, что специалистам – профессионалам в той или иной области науки, в том числе и гуманитарной, – становится трудно не только детально изучать новое в своей области, но и даже просто следить за его появлением.

Возможных путей выхода из этой ситуации может быть, по видимому, несколько, но основной, на мой взгляд, связан с дальнейшей интенсификацией труда гуманитария, повышением эффективности использования исследовательских средств, оптимизацией информационных потоков и постоянным ростом профессионального уровня специалистов. Для историка это прежде всего повышение информационной отдачи источника как за счет оптимизации доступа к документам и материалам, так и путем развития способов и приемов извлечения новой информации. При этом речь идет, конечно, о повышении профессионального уровня специалистов и дальнейшем подъеме «планки» исторических исследований.

Разумеется, можно пойти и другими путями, связанными с более экстенсивными способами освоения исторического пространства. Одно из таких направлений, реально существующее в течение уже примерно десятилетия, переносит центр тяжести на проблемы оцифровки и репрезентации исторических ресурсов. Однако, учитывая то, что полноценно этим занимаются лишь достаточно крупные объединения специалистов, абсолютное большинство которых составляют не профессиональные историки, интегрировать это направление в исследовательский тренд исторической науки можно только в качестве специального и не основное.

Еще один вариант реакции на информационное перенасыщение – доведение междисциплинарного подхода до полного снятия границ между науками, переход к нетрадиционным, менее трудоемким способам изучения источников. Возможно, это принесет какие-то текущие результаты, но, думается, не окажет серьезного влияния на саму суть исторического исследования, не обеспечит должного «приращения» исторического знания. Очень хотелось бы избежать опасности «размывания» границы между строго профессиональным научным историческим знанием и его популяризированными вариантами, ориентированными на широкий круг интересующихся историей.

## Библиографический список

1. *Владимиров В.Н.* Историческая геоинформатика вчера и сегодня // Информационный бюллетень Ассоциации «История и компьютер». № 43. Материалы международной научной конференции «Геоинформационные системы и технологии в исторических исследованиях». Барнаул, 2015. С. 21–24.
2. *Владимиров В.Н., Сарафанов Д.Е., Щетинина А.С.* «Новая историческая демография» в России: эволюция или скачок в развитии? // Известия Уральского федерального университета. Серия 2. Гуманитарные науки. 2016. № 3. С. 29–53.
3. *Брюханова Е.А., Владимиров В.Н.* Кодирование исторических профессий. Барнаул, 2015. 220 с.

### HISTORICAL INFORMATION SCIENCE: THE PAST, THE PRESENT, AND THE FUTURE

*Vladimirov V.N.*

Altai State University, 61, Lenina ave., Barnaul, Russia, vlad@hist.asu.ru

The article summarizes some results and prospects of the development of Historical Information Science in Russia. It states that Historical Information Science influences traditional historical studies, source studying procedures in particular. Its development was a driving force for the rise of Historical Demography in Russia and for the advent of a “new historical demography” as its part based on nominative sources and modern information technologies. Today, Historical Information Science faces new challenges related to the widespread use of mobile devices and intense flow of professional information. The article concludes that the adequate reactions to the challenges are the enhanced efficiency of research means, optimization of information flows, as well as the constant growth of researchers’ proficiency. It is important to avoid blurring the boundaries between professional historical knowledge and its popular variants.

*Key words:* Historical Information Science, Historical Source Study, Information Technology, Historical Knowledge, Historical Demography.

УДК 930.2

### ЦИФРОВАЯ РЕВОЛЮЦИЯ КАК ФАКТОР ВИЗУАЛИЗАЦИИ ИСТОРИЧЕСКОЙ НАУКИ<sup>1</sup>

*Мазур Л.Н.*

Уральский федеральный университет им. первого президента России Б.Н. Ельцина, Россия, Екатеринбург, просп. Ленина, 51, Lmazur@mail.ru  
В статье анализируется влияние цифровой революции и визуального поворота на историческую науку. Цифровая революция меняет информационную среду, технологии и сам характер исторических исследований – они приобретают

---

© Мазур Л.Н., 2017

<sup>1</sup> Тема поддержана грантом РФФИ № 15-06-05611 «Типология раннесоветской семьи по материалам Партийных переписей 1920-х гг.».

ярко выраженный междисциплинарный характер. Новым направлением исторических исследований стали виртуальные исторические реконструкции. Они не только иллюстрируют возможности информатизации исторической науки, но и связаны с другим глобальным процессом, меняющим профиль исторических исследований, – визуализацией истории. Визуальная история характеризуется следующими чертами: она опирается преимущественно на визуальные источники (чертежи, литографии, гравюры и проч.); использует технологии визуализации исторических фактов; содержит визуализированный результат исследования, созданный с использованием мультимедийных технологий. Новая культурологическая ситуация, порождаемая визуальным поворотом, ставит перед историками новые исследовательские задачи, решить которые возможно только опираясь на новые информационные технологии.

*Ключевые слова:* виртуальные реконструкции, цифровая революция, история, историческая реконструкция, визуальный поворот.

Историческая наука за последние сто лет пережила не один «поворот», среди которых особо выделяются технологические – это информатизация (цифровая революция) и визуализация (визуальный поворот). И тот и другой претендуют на статус революции, способной преобразить историческую науку, обеспечить ее переход на другой качественный уровень, уже получивший название Digital History [1]. Цифровая революция принципиально меняет информационную среду и методический арсенал исторических исследований [2] и разворачивается в следующих направлениях: 1) оцифровка архивных и библиотечных фондов; 2) создание Big Data – информационных ресурсов нового поколения; 3) использование новых способов презентации исторического знания в электронной форме с использованием гипертекстовых, мультимедийных и ГИС-технологий. Каждое из этих направлений нуждается в серьезном изучении. В рамках данного сообщения остановимся на одном из них – создании виртуальных реконструкций памятников и событий прошлого. В них с наибольшей полнотой реализуются не только «плоды» информатизации, но и результаты визуального поворота.

На сайте исторического факультета МГУ создана библиотека виртуальных исторических реконструкций [3], где представлены десятки проектов, связанных с решением задач виртуальной реконструкции различных исторических объектов – городской инфраструктуры, промышленных комплексов и фортификационных объектов; дворцовых, храмовых, монастырских комплексов и усадеб; а также интерьеров, быта, одежды, вооружения, археологических находок. В отдельный раздел выделена реконструкция исторических событий, технологий. Как видим, технологии виртуальных реконструкций уже сегодня охватили разные типы исторических задач – от презентации материальных объектов до визуализации событий и процессов. Одним из

успешных примеров использования 3D-технологий является проект реконструкции Страстного монастыря, реализованный под руководством чл.-корр. РАН Л.И. Бородкина в Московском государственном университете [4, 5]. Активные исследования в этом направлении проводятся в Сибирском федеральном университете (г. Красноярск) [6], Тамбовском государственном университете и других научных центрах.

Проекты виртуальных реконструкций не только являются иллюстрацией возможностей информатизации исторической науки, они связаны с другим глобальным процессом, меняющим профиль исторических исследований – визуализацией истории. Для нее характерны следующие черты: во-первых, опора преимущественно на визуальные источники (чертежи, литографии, гравюры и проч.); во-вторых, использование технологий визуального анализа источников; в-третьих, представление результата исследования в визуализированной форме с использованием мультимедийных технологий, в том числе 3D-инсталляций, видеозаписей и проч. И здесь возникает вопрос о характере взаимодействия этих двух процессов – цифровой революции и визуального поворота, т. е. что является причиной/следствием, что первично в этой гонке технологических новаций?

Визуальный поворот в этом споре выглядит как более масштабное явление, связанное со сменой типов культуры, если не учитывать, что основу культуры составляет коммуникационная среда и формирующие ее технологии работы с информацией. С учетом этого фактора выделяются три типа культуры, в исторической перспективе последовательно сменяющие друг друга: дописьменная (устные технологии); письменная (письменные технологии); визуальная культура (визуальные технологии). Смена культурных типов непосредственно соотносится с периодизацией исторической науки, где выделяются<sup>2</sup>:

– *классический период* (античность, Средние века), когда исторические сочинения рассматривались, прежде всего, как *искусство рассказать* историю. Этот этап вполне можно назвать «устно-историческим», поскольку устные свидетельства выступали информационной основой исторического сочинения, устным был и способ презентации исторических знаний, а в качестве базового принципа исторического искусства было соблюдение правил ораторского искусства;

– *рационалистический период* (XVII–XIX вв.), основной чертой которого стало утверждение в историческом исследовании научных принципов, опирающихся на критику источников, верификацию используемых фактов и результатов их аналитико-синтетической обработки. Информационно-технологическую основу истории составили

---

<sup>2</sup> Первые два этапа периодизации были описаны А.С. Лаппо-Данилевским [7, с. 23]

письменные технологии, которые распространялись на все этапы исследования: от изучения преимущественно письменных источников до текстовой репрезентации результатов исследования. В этот период история оформилась как научный *текст*, *описывающий* события и факты.

– *период модернизма* (первая половина XX в.), когда историческая наука испытывает на себе влияние факторов, связанных с формированием новой коммуникационной среды, в том числе с развитием кинематографа, фотографии, телевидения, радио, превративших историю в факт массового сознания. В XX в. складывается постнеклассическая модель исторической науки, которая сохранилась до настоящего времени. Ее особенностью является становление исторической науки как *способа интерпретации прошлого*;

– *период постмодернизма* (вторая половина XX – начало XXI в.)<sup>3</sup>. Историческая наука входит в эпоху многочисленных поворотов (лингвистический, антропологический, визуальный и проч.), расширяющих ее инструментарий. Визуальный поворот, отчетливо проявившийся в последние 15–20 лет, занял особое место, затронув не только источниковую базу исследований, но и методы исторического познания, тематику и язык исторической науки, в центре внимания которой оказалась *несобытийная история*, интерпретируемая в контексте понятия ОБРАЗ. По выражению З. Чеканцевой, «образы делают историю видимой» [9].

Исследование образа эпохи/события/пространства/личности выдвигает перед историком особые требования:

1) к используемым историческим источникам (аудиовизуальные источники, художественные тексты, памятники материальной культуры). В результате история выходит за границы изучения только письменных документов, расширяя возможности в изучении прошлого;

2) к методике исследования, которая должна соответствовать задачам визуальных реконструкций, включающей несколько этапов: 1 – деконструкцию образа (на уровне источника), 2 – его реконструкцию (на уровне исследования) и 3 – презентацию. Презентация образа может быть виртуальной, игровой, иллюстративной, текстовой, т. е. опирается на все возможные информационные технологии: от устной до цифровой.

Таким образом, новая культурологическая ситуация, порождаемая визуальным поворотом и сменой культурных парадигм, ставит перед историками новые исследовательские задачи, решить которые

---

<sup>3</sup> Теория постмодернизма была создана на основе концепции Жака Деррида, который ввел понятие текста и сформулировал его определение: «мир – это текст», «текст – единственно возможная модель реальности» [8].

возможно только опираясь на новые информационные технологии. Цифровая революция меняет не только информационную среду исторической науки, но и саму сущность исторических исследований, они приобретают ярко выраженный междисциплинарный характер.

### Библиографический список

1. *Seefeldt D., Thomas W.G.* What Is Digital History? // Perspectives on History. URL: <https://www.historians.org/publications-and-directories/perspectives-on-history/may-2009/intersections-history-and-new-media/what-is-digital-history> (дата обращения: 07.04.2017).
2. *Можжаева Г.В.* Digital Humanities: цифровой поворот в гуманитарных науках // Гуманитарная информатика. 2015. Вып. 9. С. 8–23. URL: <http://cyberleninka.ru/article/n/digital-humanities-tsirovoy-povorot-v-gumanitarnyh-naukah> (дата обращения: 07.04.2017).
3. Библиотека виртуальных исторических реконструкций. Исторический факультет МГУ. URL: <http://hist.msu.ru/Labs/HisLab/3D/3DLibrary-1.htm> (дата обращения: 07.04.2017).
4. *Бородкин Л. И., Жеребятьев Д. И., Моор В. В., Ким О. Г.* Виртуальная реконструкция московского Страстного монастыря (XVII–XX вв.): синтез источников, сочетание 3D-инструментов // Информ. бюллетень ассоциации «История и компьютер». Калининград: БФУ им. И.Канта, 2014, № 4. С. 8–12.
5. *Мироненко М. С.* Современные подходы к 3d-реконструкции объектов культурного наследия: проблемы визуализации и восприятия (на примере Московского Страстного монастыря и Чудова монастыря Московского Кремля) // Электр. науч.-образов. журн. «История». 2015. Т.6, вып. 8 (41). URL: <https://history.jes.su/s207987840001270-6-1> (дата обращения: 07.04.2017).
6. *Румянцев М., Смолен А., Барышев Р., Рудов И., Пиков Н.* Виртуальная реконструкция объектов историко-культурного наследия // Прикладная информатика. 2011. № 6 (36). С. 62–76.
7. *Лаппо-Данилевский А. С.* Методология истории. М.: Изд. дом «Территория будущего», 2006.
8. *Деррида Ж.* О грамматологии. М.: Ad Marginem, 2000.
9. *Чеканцева З. А.* Эпистемология исторического образа на рубеже XX–XXI вв. // Электр. науч.-образов. журн. «История», 2013. Т. 4, вып. 2(18). URL: <http://history.jes.su/s207987840000488-5-1> (дата обращения: 24.05.2015).

### DIGITAL REVOLUTION AS A FACTOR OF VISUALIZATION OF HISTORICAL SCIENCE

*Mazur L.N.*

Ural Federal University named after the first President of Russia  
B.N. Yeltsin, 51, Lenina ave., Yekaterinburg, Russia, [Lmaz@mail.ru](mailto:Lmaz@mail.ru)

The article analyzes the influence of the digital revolution and the visual turn on historical science. The digital revolution changes the informational environment, technologies and the very nature of historical research that acquires a pronounced interdisciplinary character. Virtual historical reconstructions have become a new direction of historical research. They not only illustrate the possibilities of informatization of historical science, but also are linked to visualization of history as

another global process that changes the profile of historical studies. Visual history is characterized by the following features: it relies primarily on visual sources (drawings, lithographs, engravings, etc.); it uses technology to visualize historical facts; it contains visualized research results created using multimedia technologies. A new cultural situation generated by the visual turn puts historians in front of new research tasks that can be solved only by relying on new informational technologies.

*Key words:* virtual reconstruction, digital revolution, history, historical reconstruction, visual turn.

УДК 316.42

## СОЦИАЛЬНОЕ И ИНСТИТУЦИОНАЛЬНОЕ В ЦИФРОВОЙ ГУМАНИТАРИСТИКЕ

*Мохов В.П.*

Пермский национальный исследовательский политехнический университет, Россия, Пермь, Комсомольский просп., 29, mvp@perm.ru

В статье отмечается, что цифровая гуманитаристика была основана на серьезных материальных предпосылках, основу которых составили компьютер, Интернет, гаджеты. Утверждается, что технологические и технические изменения привели к важным социальным и духовным изменениям, которые, по своей сути, носили институциональный характер. Одним из следствий развития цифровой гуманитаристики стало то, что доступ к гуманитарному знанию все меньше зависит от положения человека в системе социальной стратификации. Цифровая гуманитаристика привела к фундаментальным изменениям социального характера, включая массовое распространение Интернета и включение в него социально активной части населения или всего населения, особенно молодежи; возникновение специализированных социальных групп, занимающихся обслуживанием «цифры», изменились социальные практики производства, распространения, потребления культурных ценностей, знания, информации. Выдвинуто предположение, что мощное развитие цифровой гуманитаристики привело к возникновению значимых социальных и культурных угроз.

*Ключевые слова:* цифровая гуманитаристика, Digital humanities, социальные изменения, глобализация.

Цифровая гуманитаристика стала фактом нашей повседневности. Гуманитарные науки превратились в цифровые с помощью цифровых инструментов и ресурсов [1, с. 8] Мы не всегда осознаем, какой рубеж мы перешли за последние 10–15 лет. Фактически, мы живем в новой эпохе производства, распространения, потребления гуманитарного знания и культуры в целом.

Очевидно, что новая эпоха не была бы возможной без ряда институциональных предпосылок, связанных с развитием материального

производства, социальных изменений в обществе, изменения духовного мира жителей страны.

Материальную основу цифровой гуманитаристике создали три великих материальных прорыва: компьютер, интернет, гаджеты. Они сделали доступ к гуманитарному знанию более простым, мобильным, предсказуемым.

Вслед за материальными прорывами последовали изменения в общественном и индивидуальном сознании: люди из совершенно различных социальных групп, с значительно различающимися уровнями дохода, социальными статусами оказались примерно в одинаковом отношении к процессу потребления и распространения культуры в целом и гуманитарного знания в частности, что впервые создало уникальную ситуацию: доступ к гуманитарному знанию все меньше зависит от положения человека в системе социальной стратификации.

«Цифра» привела к двум противоположным следствиям: с одной стороны, началась унификация гуманитарного знания: ввиду общедоступности знаний в цифровой форме корпорации, заинтересованные в продвижении своих ценностей, стали диктовать моду на мировых рынках знаний. С другой стороны, в интернет-пространстве стали формироваться анклавы узкоспециализированного знания, адепты которого образуют интеллектуальные колонии наднационального характера.

Одновременно шли фундаментальные изменения социального характера. Во-первых, уровень информатизации населения и вовлеченности в интернет-общение и интернет-потребление информации выросли, достигнув в развитых странах почти полного охвата населения (Япония – 93 %, Германия – 87, Великобритания – 92, Корея – 90, Россия – 73 и т. д., в целом по планете – 48 %. Молодое поколение в развитых странах мира почти полностью охвачено сетью Интернет.

Во-вторых, создана инфраструктура, производящая специализированное оборудование и программы для использования в гуманитарной сфере (см., напр. [2]). Вместе с ними возникли социальные группы во всех видах культуры, гуманитарного знания, занятые обслуживанием соответствующих видов гуманитарной деятельности на основе использования медийных устройств, Интернета, программного обеспечения.

В-третьих, созданы новые формы презентации культурных феноменов в цифровой форме: виртуальные выставки, виртуальные музеи и электронные библиотеки, архивы. Газеты и журналы стали размещать на своих сайтах архивы номеров газет и журналов. Интернет-трансляции концертов, культурных и научных форумов стали обыденностью. Получение образования через интернет-доступ стало одной из форм обучения (см., напр. [3]). Сложилась новая социальная реальность, основанная на широком доступе к бывшим эксклюзивным благам и ценностям.

В-четвертых, изменилась сама культурная и научная коммуникация. Электронное общение ускорило обмен идеями и данными, культурными достижениями. Множественность выбора [4, с. 336–337] создала иллюзию культурной свободы, отражая при этом множественность социальных миров и меньшинств.

В-пятых, принципиально иным стало производство гуманитарного знания и культуры в целом: сократилось время на поиск исходных данных, доступ к цифровым копиям произведений культуры, гуманитарным достижениям ускорил усвоение гуманитарной информации. Как писал М. Маклюэн, «моментальный мир электроинформационных средств включает нас целиком и сразу», «"время" прекратилось, "пространство" исчезло» [5, с. 345].

Вместе с тем, распространение цифровой гуманитаристики сформировало очевидные угрозы общемировой культуре.

Во-первых, стала явной угроза мировой «образованщины», освоение культуры стало подменяться извлечением информации из массовых источников без глубокой внутренней переработки и освоения. При этом стоит отметить, что широкое распространение цифровой гуманитаристики повышает минимальный уровень доступности культуры, но понижает средний уровень освоения культурных достижений. Знание подменяется информацией, культура – ярлыками культуры.

Во-вторых, выросла угроза манипуляции культурным уровнем человека с помощью заранее сформированного контента и организованного доступа к нему.

В-третьих, возникает угроза «подмены прошлого» через манипуляцию архивными и литературными документами, размещенными в открытом доступе.

В-четвертых, резко выросла угроза идеологических (религиозных, культурных, общественных) «эпидемий», которые могут охватить сотни миллионов человек, что делает мир менее предсказуемым.

В-пятых, цифровая гуманитаристика стала инструментом глобализации [6]. Специфика глобальной игры заключается в том, что «игроки соревнуются по нелинейным правилам, меняющимся после (или во время) каждого хода» [7, с. 529]. Цифровая гуманитаристика объективно работает против национального государства, она взламывает культурные, информационные, поведенческие границы и разрушает национальный суверенитет в гуманитарной сфере. Глобальные игроки начинают диктовать правила в сфере культуры, науки и т. д.

### **Библиографический список**

1. *Можсаева Г.В.* Digital humanities: цифровой поворот в гуманитарных науках // Гуманитарная информатика. 2015. № 9. С. 8–23.

2. *Бородкин Л.И.* Дигитализация, визуализация, репрезентация. Аналитика? // Информационный бюллетень Ассоциации «История и компьютер». 2015. № 43. С. 3–8.
3. *Пенн Марк Дж., Залеси Кинни Э.* Микротенденции. М., 2009.
4. *Нейсбит Д.* Мегатренды. М., 2003.
5. *Маклюэн М.* Рождение сети // Информационное общество: сборник. М., 2004.
6. *Можсаева Г.В., Можсаева-Ренья П.Н., Сербин В.А.* Цифровая гуманитаристика: организационные формы и инфраструктура исследований // Вестник Томского государственного университета. 2014. № 389. С. 73–81.
7. *Тоффлер Э., Тоффлер Х.* Революционное богатство. М., 2008.

## **SOCIAL AND INSTITUTIONAL IN DIGITAL HUMANITIES**

*Mokhov V.P.*

Perm National Research Polytechnic University,  
29, Komsomolsky ave., Perm, Russia, mvp@perm.ru

Digital Humanities as a discipline has been based on serious material prerequisites, connected with computer, Internet, and gadgets. Technological improvements has led to considerable social and moral changes of institutional character. One of the effects of Digital Humanities development is that the access to humanitarian knowledge depends less on the person's position in the system of social stratification. Digital Humanities has led to fundamental changes of social character including mass distribution of Internet and the involvement of socially active proportion of population or the whole population, especially young people, in it. Moreover, specialized social groups being occupied with "digit" service appeared. Social practices of business, distribution, consumption of cultural value, knowledge and information have changed as well. At the same time, rapid development of Digital Humanities resulted in meaningful social and cultural threat.

*Key words:* Digital Humanities, social changes, globalization.

УДК 008

## **ФОРМИРОВАНИЕ РОССИЙСКОЙ АССОЦИАЦИИ ЦИФРОВЫХ ГУМАНИТАРНЫХ НАУК**

*Румянцев М.В.*

Сибирский федеральный университет,  
Россия, Красноярск, просп. Свободный, 79, m-rumyantsev@yandex.ru

Представлены промежуточные результаты институционализации отечественной научно-исследовательской сети Digital Humanities (цифровых гуманитарных наук). Создание Российской ассоциации цифровых гуманитарных наук рассматривается как этап интеграции российской цифровой гуманитаристики в мировое научно-образовательное пространство.

*Ключевые слова:* Digital Humanities, цифровые гуманитарные науки, цифровая гуманитаристика, Russian Association for Digital Humanities.

Становление Digital Humanities (ДН, цифровых гуманитарных наук, цифровой гуманитаристики), которые сегодня объединяют ученых, использующих информационные технологии в гуманитарных исследованиях, пришлось на 2000-е годы. Область исследований Digital Humanities генетически связана с деятельностью таких международных профессиональных объединений, как Ассоциация «Гуманитарные науки и компьютер» (Association for Computers and the Humanities, основана в 1978 г.) и Ассоциация компьютерных методов в литературе и лингвистике (Association for Literary and Linguistic Computing, основана в 1973 г.), на базе которых в 2005 году был сформирован Альянс организаций цифровых гуманитарных наук (Alliance of Digital Humanities Organizations, ADHO). Таким образом Альянс институционально оформил право на существование Digital Humanities.

Сегодня членами ADHO являются несколько тысяч исследователей. Под эгидой ADHO ежегодно проводится крупная конференция, в которой принимают участие ученые со всего мира [1]. ADHO также курирует выпуск нескольких научных изданий, связанных с указанной проблематикой [2].

В России тематика исследований на стыке гуманитарных наук и информационных технологий хорошо представлена в работах научных коллективов ряда ведущих университетов<sup>1</sup>. Регулярно проводятся конференции профессиональных сообществ: Ассоциации «История и компьютер» (основана в 1992 г.), Некоммерческого партнерства «Автоматизация деятельности музеев и информационные технологии» (основано в 1996 г.), специалистов в области компьютерной лингвистики (конференция «Диалог» проводится с 1995 г.). Очевидно, что российская наука, связанная с гуманитарной информатикой, имеет такую же авторитетную историю и не менее существенные достижения, как и наука зарубежная. Тем не менее, до недавнего времени в

---

<sup>1</sup> Кафедры и центры цифровой гуманитаристики сегодня функционируют в Балтийском федеральном университете им. И. Канта, Санкт-Петербургском государственном университете, Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики, Московском государственном университете имени М.В. Ломоносова, Национальном исследовательском университете «Высшая школа экономики», Российском государственном гуманитарном университете, Тамбовском государственном университете имени Г.Р. Державина, Нижегородском государственном университете имени Н.И. Лобачевского, Пермском государственном национальном исследовательском университете, Уральском федеральном университете имени первого президента Б.Н. Ельцина, Алтайском государственном университете, Национальном исследовательском Томском государственном университете, Новосибирском государственном университете, Сибирском федеральном университете.

составе участников ежегодной конференции Альянса организаций цифровых гуманитарных наук представителей от России не было, а с 2014 года доля российских участников составляет не более 1 %. Несмотря на то, что, как уже было сказано, в стране стабильно функционирует более 10 научно-образовательных центров цифровой гуманитаристики, на сайте члена ADHO centerNet, который представляет информацию о мировой сети центров ДН, данные о российских центрах отсутствуют.

По нашему мнению, подобная изолированность российских ученых, вероятно связанная с медленным освоением мировых стандартов научных публикаций [1], лишает их возможности включаться в актуальный мировой контекст формирующейся области знаний.

В 2015 году на Международной конференции «Информационные технологии в гуманитарных науках» (Сибирский федеральный университет, Красноярск) участниками было принято решение обратиться в Европейскую ассоциацию цифровых гуманитарных наук (European Association for Digital Humanities, EADH) с ходатайством о вступлении. EADH является одним из членом ADHO, наряду с другими образующими Альянс организациями: Canadian Society for Digital Humanities / Société canadienne des humanités numériques (CSDH/SCHN); centerNet; Australasian Association for Digital Humanities (aaDH); Japanese Association for Digital Humanities (JADH); Humanistica, L'association francophone des humanités numériques/digitales (Humanistica). Намерение российских центров ДН фиксировал Меморандум, а сама организация стала сетевой<sup>2</sup>. Президиум EADH одобрил эту инициативу и предложил варианты вхождения в ассоциацию.

На сегодняшний момент организация Российской ассоциации цифровых гуманитарных наук находится в заключительной фазе: разработан устав организации, подготовлены соответствующие документы для ее официальной регистрации. Предполагается, что на очередной конференции в Красноярске в сентябре 2017 года пройдет презентация Российской ассоциации. Официальное признание российской научной сети цифровых гуманитарных наук в качестве ассоциированного члена EADH не только позволит российским ученым войти в мировую научную сеть, выстраивать партнерские отношения с зарубежными коллегами, но и воспользоваться преимуществами, которые дает членство в ассоциации: подпиской на журналы, стипендиальной и грантовой поддержкой исследований, предпочтениями при участии в конференциях и др.

---

<sup>2</sup> Сайт Российской научной сети цифровых гуманитарных наук. URL: <http://dhrussia.ru/>

## Библиографический список

1. *Кижнер И.А.* Конференция DigitalHumanities 2016: перспективы и вызовы для развития сообщества гуманитарной информатики / И.А. Кижнер, М.В. Румянцев, И.А. Рудов, Р.Г. Хлебопрос // Электронный научно-образовательный журнал «История». Т. 7. В. 7 (51). URL: <https://history.jes.su/s207987840001676-2-1> (дата обращения: 20.03.2017).
2. Официальный сайт Альянса организаций цифровых гуманитарных наук. Раздел «Publications». URL: <http://adho.org/publications> (дата обращения: 20.03.2017).

### **BUILDING RUSSIAN ASSOCIATION FOR DIGITAL HUMANITIES**

*Rumyantsev M.V.*

Siberian FederalUniversity,

79, Svobodny ave., Krasnoyarsk, Russia, m-rumyantsev@yandex.ru

The paper presents the intermediate results for the institutional development of the national Russian Digital Humanities network. Establishing the Russian Association for Digital Humanities is seen as an essential stage for the involvement of the Russian Digital Humanities in the international research environment.

*Key words:* Digital Humanities, Computers and the Humanities, Russian Association for Digital Humanities.

УДК 34

### **ПРОБЛЕМЫ ПРАВОВОЙ ЗАЩИТЫ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ ЦИФРОВОЙ ГУМАНИТАРИСТИКИ**

*Семченкова М.Н.*

Балтийский федеральный университет им. Иммануила Канта,  
Россия, Калининград, ул. Фрунзе, 6, semchenkova.m@mail.ru

Статья посвящена рассмотрению правовых проблем, возникающих в процессе использования информационных технологий в гуманитарных науках. Результатом исследований цифровой гуманитаристики являются объекты интеллектуальной собственности, защита прав на которые осуществляется посредством авторского и патентного права. В связи с этим, в статье проанализированы правовые проблемы, с которыми сталкивается исследователь в процессе применения цифровых методов к традиционным гуманитарным ресурсам. Рассмотрены наиболее эффективные способы защиты прав на объекты интеллектуальной собственности, созданные в процессе исследовательской деятельности. Разработан механизм правовой защиты при создании цифровых архивов с использованием цифровых документов личного происхождения. Подробно описан способ определения статуса объекта интеллектуальной собственности в процессе оцифровки традиционных источников гуманитарных наук. Сделан

вывод о необходимости взаимодействия междисциплинарного научного направления цифровой гуманитаристики с правом.

*Ключевые слова:* цифровая гуманитаристика, объекты интеллектуальной собственности, результат исследования, авторское право, патентное право.

В сфере цифровой гуманитаристики проводится большой спектр гуманитарных исследований, в которых широко применяются цифровые инструменты получения знания. Результатом таких исследований являются объекты интеллектуальной собственности, защита прав на которые осуществляется посредством как авторского, так и патентного права. Выбор эффективного способа защиты права на объект интеллектуальной собственности, являющегося результатом исследований в сфере цифровой гуманитаристики, становится одной из основных проблем, стоящих перед исследователем. Решение данной проблемы позволит защитить интеллектуальные права исследователя на результаты интеллектуальной деятельности, полученные в процессе проведенных исследований в сфере цифровой гуманитаристики. Это обеспечит и перспективность использования результатов исследования, как объектов интеллектуальной собственности в будущем, позволит избежать правовых проблем связанных с незаконным заимствованием, копированием и использованием результатов исследований другими лицами.

Особенностью цифровых гуманитарных наук является придание традиционным источникам гуманитаристики новой цифровой формы. Активно используя потенциал информационных технологий, применяемым на этапе обработки традиционного материала, гуманитарная наука делает их самостоятельным объектом исследования. К примеру, применение к нетекстовым ресурсам гуманитарных наук информационных технологий в виде оцифровки больших коллекций изображений. Подобное использование цифровых методов исследования в гуманитарных науках приводит к возникновению правовых проблем. Так как процесс оцифровки связан с использованием некоего объекта интеллектуальной собственности, являющегося по отношению к оцифрованному объекту оригиналом и созданным изначально автором в иной материальной форме, возникает проблема определения правового статуса оцифрованного объекта. У первоначально созданного объекта интеллектуальной собственности всегда есть автор или лицо, обладающее исключительными правами на такой объект, что ставит перед исследователем необходимость обладания соответствующим правом на использование цифровых методов в отношении такого объекта. Для определения правового статуса оцифрованного объекта необходимо обратиться к положениям гражданского законодательства, которое относит к объектам авторских прав производные произведения. Производные произведения представляют собой переработку другого произ-

ведения. В процессе оцифровки объект, изначально выраженный только в материальной форме, приобретает качественно новую цифровую форму. Следовательно, такой объект будет являться производным производением, которое защищается авторским правом. Однако это не единственный объект интеллектуальной собственности, который может возникнуть при использовании цифровых методов в гуманитарных исследованиях. Правовая защита отдельного объекта исследования в виде оцифрованного ресурса необходима для дальнейшего его использования в больших и основательных проектах, которые позволят создавать коллекции источников, пригодные для многозадачных исследований.

Для создания цифровых архивов в качестве исторических источников могут использоваться цифровые документы, личного происхождения. Для использования в исследованиях документов, принадлежащих какому-либо лицу, необходимо понимать, что какую бы общественную ценность не представлял такой документ, его использование без согласия автора или лица, обладающего исключительным правом на такой документ, будет нарушением авторского права. В случаях, если лицо добровольно готово отдать, принадлежащий ему объект интеллектуальной собственности, это не значит, что такое лицо передает свои интеллектуальные права на данный объект интеллектуальной собственности. В связи с этим, необходимо принять правовые способы для защиты прав исследователя на результат его исследования, в котором используются такой объект, что позволит избежать неблагоприятных последствий в будущем. Наиболее распространенным способом защиты права в данном случае является заключение договора с лицом, которое передает объект интеллектуальной собственности для использования его в исследовании.

Цифровая гуманитаристика – динамично развивающееся научное направление, которое открывает новые возможности и перспективы для получения инновационных научных знаний. Взаимодействие цифровой гуманитаристики с правом позволит защитить результаты исследований, проекты, научные знания, создаваемые в рамках этого междисциплинарного направления. Охрана интеллектуальной собственности необходимый этап в применении цифровых технологий в гуманитарных исследованиях, в их развитии, а также использовании в будущем.

### **Библиографический список**

1. *Гражданский кодекс Российской Федерации. Ч. 4 // Доступ из справочно-правовой системы «КонсультантПлюс», 2006.*
2. *Гуманитарная информатика: сб. статей / под ред. Г.В. Можяевой. Томск: Изд-во Томск. ун-та, 2015.*
3. *Динсман М. Цифровая гуманитаристика: интервью с Дэвидом Голомбиа // ГЕФТЕР: интернет-журн. 2016.*

## **PROBLEMS OF LEGAL PROTECTION OF RESEARCH RESULTS IN DIGITAL HUMANITIES**

*Semchenkova M.N.*

Baltic Federal University named after Immanuel Kant,  
6, Frunze st., Kaliningrad, Russia, semchenkova.m@mail.ru

The paper is devoted to the legal problems arising in the process of using information technologies in the Humanities. The results of Digital Humanities research are objects of intellectual property protected by copyright and patent law. In this connection, the paper analyzes the legal problems faced by the researcher in the process of applying digital methods to traditional humanitarian resources. The most effective ways of protecting the rights to intellectual property objects created as a result of research activity are considered. The mechanism of legal protection when creating digital archives with the use of digital documents of personal origin is developed. A method for determining the status of an object of intellectual property arising in the process of digitizing traditional sources of the Humanities is described in detail. The author states the necessity of interaction between Digital Humanities and Law Studies.

*Key words:* Digital Humanities, objects of intellectual property, research result, copyright, patent law.

УДК 364.043

## **ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ЭКСТРЕМИСТСКИХ СЮЖЕТОВ. ОПЫТ АНАЛИЗА И ПРОТИВОДЕЙСТВИЯ**

*Ярков А.П.*

Тюменский государственный университет,  
Россия, Тюмень, ул. Володарского, 6, ayarkov@rambler.ru

*Пушкаревич О.А.*

Свободный исследователь, Россия, г. Тюмень

Статья основана на опыте работы экспертного сообщества в г. Тюмени. Предметом исследования стал анализ одного из материалов, представленных для выявления содержания экстремистской направленности. В результате комплексной экспертизы установлены использованные авторами сюжета средства, направленные на дисгармонизацию межконфессиональных отношений. При этом расчёт только на электронные средства выявления экстремистских текстов несостоятелен. Для развенчания искажающих смысл священных для представителей разных конфессий текстов необходимо объединение представителей разных направлений светской науки и богословия.

*Ключевые слова:* противодействие экстремизму, комплексная экспертиза, приемы и средства.

В поле внимания экспертов – сложно построенные психолингвистические и ментально-культурологические материалы. Поэтому для анализа 9-минутного фрагмента сюжета DVD-диска (не указываем авторов, название) привлечены: психолингвист, религиовед, раввин, имам, переводчики четырех языков, специалист по компьютерным технологиям.

Предшествующий сюжет настраивал (спокойным аудио- и видеоритмом) на восприятие материала, утверждающего ценности ислама. Внимание же привлек сюжет о переводе (с фарси) текста, начиная с воспроизведения с 3 мин. 31 с.: *«Это слова состоит из 4 буквенных символов (знаков): "мим", "ха", "мим", "дол". Теперь прочитаем эти слова, но без изменений – в оригинальном варианте: "Мухаммад". Это имя мусульманского пророка. Без огласовки фатха читается как «Махмад». Согласно иврито-английскому словарю правильное произношение этого слова – "Мухаммад", а не "Махмад". Как нам поверить, что это слова читается как "Мухаммад"? Выход один – надо показать этот материал (откровение) иудейскому раввину, чтобы он прочитал. Слушайте чтение газели Сулеймана 5 глава стих 16, которая читается с еврейского сайта».*

Экспертами обращено внимание, что в оригинале это цитата из «Шир а-Ширим» («Песни Песней»), приписываемая царю Соломону (30-я часть Танаха, 4-я книга Ктувим), т. е. каноническая книга Ветхого Завета, где нет упоминаний о Мухаммаде. Там говорится о Боге («Всевышнем») в интерпретации иудаизма, что он – «*куломахамадим*». Для малосведующего совпадение может стать «открытием», убеждающим в «предназначенном» в Ветхом Завете. Арабский и иврит близки. Но это не свидетельство идентичности религии. К тому же в иврите есть корень, состоящий из букв «хет», «мем», «далет», которые образуют слово «хамед». Это в переводе означает «милый», «приятный» и др., но никак не указывает на имя пророка Мухаммеда. В последующих фрагментах это имя начинает повторяться рефреном, закрепляя утверждение, что «Мухаммед» упомянут в «Шир а-Ширим».

Очевидно влияние такой техники на психопатологическую личность, а также на личность с неустойчивой психикой. Их поведение и реакция непредсказуемы: могут совершаться действия, когда личность находится в плену ощущений, мотивов, галлюцинаций. Гипнотическим ключом к управлению психикой представителя этого психотипа является жесткое директивное внушение.

Далее сказано на фарси: *«Пожалуйста, будьте внимательны к "Ийм" – это форма множественного числа. Оно используется как уважительная обращение: "Светом озаряем (уничтожим) темноту, нету бога кроме бога, и Мухаммад пророк его. Пускай мир услышит". Здесь дан знаменитый сайт переводов, где можно перевести слово в*

слово, букву в букву. Слова на иврите, которые обсуждаем и копируем онлайн (прямо) из газеты Сулеймана. Надо упомянуть, что эти слова взяты с того же сайта, где мы взяли "правильное чтение". Нужные слова копируем и ставим для перевода. Теперь будьте внимательны: переводит ли этот веб-сайт имя собственное? Переводчики Евангелия перевели эти слова как "похваленный, возлюбленный". Теперь сами посмотрите: "Мухаммад!"». При этом с 7 мин. 27 с. меняется темпоритм – звуковой фон становится жестким, учащенным.

Следует текст: «Перевод с этого сайта похож на имя Мухаммад». В кадре появляется «Доктор Ахмад Дидат – один из исследователей Евангелия». Его прямая речь, произносимая с эмоциональным напором, с категоричным безапелляционным утверждением: «Вы не имейте права перевести собственные имена!!! ...Например, если в тексте написано "мистер Блек", то и в переводе должны перевести "мистер Блек!". Но евреи и христиане сделали так, что имя Мухаммадин перевели как "личность", "возлюбленный". Но имя Мухаммада дано здесь в оригинальном тексте на еврейском (иврите) языке».

Следующий сюжет с наложенным текстом на фарси: «На этом видео обратите внимания на слова молодого еврея: "Ты и твой Иисус уходите ... Вы не должны сюда приходить, это не ваш дом. Мы убили Иисуса, и гордимся этим. Слышал, что сказал? Мы убили Иисуса, и гордимся этим"». В оригинальном тексте, произносимом иудеем на английском, нет упоминаний о Христе вообще. Далее в сюжете прием манипуляционной психотехники с использованием сублиминальных раздражителей. К таковым относятся визуальные психокоррекционные техники, основанные на алгоритмах предварительной обработки информации. Внушаемая таким образом видеoinформация принципиально не может быть воспринята на осознаваемом уровне, но, тем не менее, она попадает в мозг и там с помощью нейропсихических механизмов встраивается в семантические поля памяти, усваивается, изменяя психологические установки человека.

Эксперты пришли к выводу: видеосюжет предназначен для сознательного информационного и психологического нагнетания напряженности, внушения информации с помощью трансовых техник.

### **Библиографический список**

1. Корнилов Г.А., Имаков Т.З. Экспертные исследования и судебные экспертизы в теории и практике выявления и раскрытия преступлений экстремистской направленности: научно-практическое пособие. М.: Изд.дом И.И. Шумиловой, 2014.
2. Копнина Г.А. Речевое манипулирование. М.: Флинта: Наука, 2010.
3. Красных В.В. Виртуальная реальность или реальная виртуальность? (Человек. Сознание. Коммуникация). М.: Диалог-МГУ, 1998.

4. *Рикер П.* Конфликт интерпретаций. Очерки о герменевтике. М.: Academia-Центр, 1995.
5. *Секераж Т.Н.* Методологические проблемы исследования текстов по делам об экстремизме // Психология и право. 2011. № 2.
6. *Социокультурное исследование* / под ред. Ф.И. Минюшева. М.: Изд-во МГУ, 1994.

## **INFORMATION TECHNOLOGIES OF EXTREMIST PLOTS: EXPERIENCE OF ANALYSIS AND COUNTERMEASURES**

*Iarkov A.P.*

Tyumen State University, 6, Volodarskogo st., Tyumen, Russia,  
ayarkov@rambler.ru

*Pushkarevich O.A.*

Independent researcher, Russia, Tyumen

The paper is based on the experience of the expert community in Tyumen. It analyses one of the submissions to identify extremist content. As a result of the complex expertise, the means used by the authors of the plot and aimed at disharmony in interreligious relations were set up. The author notes that the reliance only on electronic means of revealing extremist texts is untenable. Technologies used by the authors of such programming for psychological creating "an enemy" cultivate medieval stereotypes. For debunking the distorting meaning of sacred texts, it is necessary to unite representatives of different directions of secular science and theology.

*Key words:* counteraction to extremism, comprehensive expertise, receptions and means.

# КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА, ТЕКСТОЛОГИЯ И ТЕКСТОМЕТРИЯ

УДК 801.8:004

## КОЛИЧЕСТВЕННЫЙ И СТАТИСТИЧЕСКИЙ АНАЛИЗ СРЕДНЕВЕКОВЫХ СЛАВЯНСКИХ ТЕКСТОВ: ИНСТРУМЕНТАРИЙ КОРПУСА «МАНУСКРИПТ» И МЕТОДИКА ЕГО ИСПОЛЬЗОВАНИЯ<sup>1</sup>

*Баранов В.А.*

Ижевский государственный технический университет  
им. М.Т. Калашникова. Россия, Ижевск, ул. Студенческая, 7,  
victor.a.baranov@gmail.com

Описываются возможности модулей исторического корпуса «Манускрипт» (manuscripts.ru), предназначенных для количественного и статистического анализа лингвистических данных. На примере прилагательных в двух рукописях Апостола – в Христинопольском (Львовский исторический музей, ОР, № 37, XII в., 291 л.) и Толстовском (Российская национальная библиотека, Q.п.1.5., XIV в., 93 л.) – демонстрируются основные параметры запросных форм модуля n-грамм и модуля статистики, а также методика их применения. Показаны способы поиска различий в частоте использования членных и нечленных форм с суффиксами *-ьск-* и *-ьн-*, приемы выявления семантически и грамматически устойчивых сочетаний существительных с прилагательными, а также инструменты визуализации разночтений в соответствующих друг другу контекстах списков. Делается вывод о разной активности форм с суффиксами *-ьск-* и *-ьн-* в рукописях, о последовательном соответствии форм с суффиксом *-ьн-* в кодексе XIV в. формам с *-ьск-* в списке XII в.

*Ключевые слова:* исторический текстовый корпус, корпусные, количественные, статистические методы, нечленные, членные прилагательные, коллокации, разночтения.

*1. Исторический корпус «Манускрипт»* (manuscripts.ru), помимо основных средств поиска и демонстрации лингвистических данных в виде одноктекстовой и многотекстовой форм запроса и вывода, предоставляет пользователям и другие инструменты выборки и демонстрации материала – модуль n-грамм и модуль статистики. В работах [1, 2, 3, 4] уже были описаны назначение и параметры запросных форм обоих модулей, в частности на примере корпуса М.В. Ломоносова. Это позволяет дать их самую общую характеристику.

---

© Баранов В.А., 2017

<sup>1</sup> Статья написана при поддержке Российского государственного научного фонда (РГНФ) в рамках проекта «Древнерусские списки Апостола XII–XIV вв.: лингвистическое исследование и электронная публикация» (грант № 15-04-00213).

Модуль n-грамм ([http://manuscripts.ru/mns/cred\\_ngr.stat](http://manuscripts.ru/mns/cred_ngr.stat)) обеспечивает возможность выявить наиболее частотные или статистически значимые сочетания словоформ в кодах корпуса. Запросная форма включает количественные и статистические (меры ассоциации) параметры оценки сочетаний, структурные (количество компонентов сочетания, расстояние между ними, учет порядка следования и нек. др.) и лингвистические (исключение служебных слов, учет границ конструкций, грамматические признаки словоформ и др.) характеристики сочетаний. Упорядоченные по количеству или величине статистической меры перечни позволяют делать вывод о степени активности или устойчивости сочетаний.

Модуль статистики (<http://manuscripts.ru/mns/cred.stat>) предназначен для демонстрации количественного распределения символов, словоформ, сочетаний словоформ в пределах рукописи или нескольких рукописей. Запросная форма дает возможность указать образец поиска, единицу шага подсчета (символы, страницы, листы, фрагменты) и его длину. Результат выводится на экран в виде графика абсолютной или относительной частоты встречаемости искомой единицы в каждом следующем шаге рукописи.

Рассмотрим методику количественного анализа прилагательных в двух рукописях Апостола – в Христинопольском (Львовский исторический музей, ОР, № 37, XII в., 291 л.<sup>2</sup>) (далее – ХА) и Толстовском (РНБ, Q.п.1.5., XIV в., 93 л.<sup>3</sup>) (далее – ТА) апостолах<sup>4</sup>. Анализ направлен на обнаружение различий в функционировании двух словообразовательных групп прилагательных (с суффиксами **-ьск-** и **-ьн-**), которые (различия) – как мы ожидаем – должны проявляться в том числе и в частоте использования соотносимых и альтернативных форм.

## *2. Объем подкорпуса*

Общее количество словоформ в транскрипциях может быть установлено с помощью количественного указателя словоформ, построенного в многотекстовой запросной форме (маска словоформы – %): ХА – 65981, ТА – 23838.

---

<sup>2</sup> Апостол толковый, с толкованиями на полях.

<sup>3</sup> Апостол последовательный (четий).

<sup>4</sup> Адрес подкорпуса – <http://manuscripts.ru/mns/portal.main?p1=68>. Работа над транскрипциями списков в настоящее время продолжается. Поэтому эта публикация демонстрирует лишь методику анализа лингвистических данных, а ее выводы относительно функционирования прилагательных в списках являются предварительными.

### 3. Членные и нечленные формы

С помощью многотекстовой запросной формы и масок на основе регулярных выражений<sup>5</sup> определяем количество нечленных и членных форм (табл. 1).

Таблица 1. Количество нечленных и членных форм

Текст	-ьск-		-ьн-		Всего
	нечленные	членные	нечленные	членные	
ХА	87	<b>149</b>	<b>514</b>	393	1143
ТА	<b>28</b>	12	<b>103</b>	<b>114</b>	257

То же в относительных величинах (в процентах к общему числу форм транскрипции) представлено в табл. 2.

Таблица 2. Относительное количество нечленных и членных форм

Текст	-ьск-		-ьн-		Всего
	нечленные	членные	нечленные	членные	
ХА	0,132	<b>0,226</b>	<b>0,779</b>	0,596	1,732
ТА	<b>0,118</b>	0,050	<b>0,432</b>	<b>0,478</b>	1,078

Соотношения показывают существенные различия в количестве форм в группах:

-ьск-: преобладание членных в ХА и преобладание нечленных форм в ТА;

-ьн-: преобладание нечленных в ХА и примерное равенство нечленных и членных форм в ТА (выделено полужирным);

ХА vs. ТА: относительное количество прилагательных в ХА значительно больше, чем в ТА, в первую очередь за счет членных форм с суффиксом **-ьск-** и нечленных с суффиксом **-ьн-** (выделено курсивом).

### 4. Пре- и постпозиция форм

С помощью модуля n-грамм устанавливаем соотношение нечленных и членных форм в пре- и постпозиции по отношению к существительным<sup>6</sup>; используем регулярные выражения и ищем только те

<sup>5</sup> Нечленные формы с суффиксом **-ьск-**: `.*(ск)[^взлнпрстхушн]$,` членные формы: `.*(ск)[^взлнпрстхушн]{2,4}$,` нечленные формы с суффиксом **-ьн-**: `.*(ьн)[^взклнпрстхцшщнѐ]$,` членные формы: `.*(ьн)[^взклнпрстхцшщнѐ]{2,4}$,` где ряд после символа `^` позволяет устранить из выборки словоформы с финалиями, не являющимися падежными окончаниями.

<sup>6</sup> Грамматическая разметка текста осуществлена лемматизатором корпуса. В ХА приведены к леммам и получили морфологические пометы 81,7 % словоформ, в ТА – 78,6 %.

сочетания, между компонентами которых нет «пунктуационных» знаков (табл. 3).

Таблица 3. Соотношение нечленных и членных форм с суффиксами *-ьск-* и *-ьн-* в пре- и постпозициях по отношению к существительным

Текст	<i>-ьск-</i>				<i>-ьн-</i>			
	нечленные		членные		нечленные		членные	
	до	после	до	после	до	после	до	после
ХА	14	40	47	55	69	96	88	89
ТА	6	16	3	1	4	16	17	27

Данные показывают, что в ХА преобладают сочетания с постпозицией рассматриваемых форм по отношению к существительным, это свойственно нечленным формам с суффиксом *-ьн-*, но особенно формам с суффиксом *-ьск-*; в ТА примеры единичны, поэтому непоказательны.

#### 5. Устойчивые сочетания

Поиск на основе лемм биграмм, состоящих из прилагательных и существительных, и оценка этих сочетаний с помощью меры ассоциации *T-score*<sup>7</sup> позволяет установить в ХА (таблицы а) и в ТА (таблицы б) семантически и грамматически устойчивые сочетания и сравнить их<sup>8</sup>.

Таблицы 4а, 4б. Коллокации с препозицией форм с суффиксом *-ьск-* в ХА и ТА<sup>9</sup>

$F^{l0}$	$F_{w1} \dots F_{w2}^{11}$	$T$ - $score$	$F$	$F_{w1} \dots F_{w2}$	$T$ - $score$
7	корнифнскъ (14) кпнстолнѧ (165)	2.633	1	жндовьскъ (6) преданнѧ (2)	0.999
4	ѡфьсьскъ (8) кпнстолнѧ (165)	1.990	1	жндовьскъ (6) басьнѧ (3)	0.999

<sup>7</sup>  $T - score = \frac{F(w_1, w_2) - F(w_1) \times F(w_2)}{\sqrt{F(w_1) \times F(w_2)}}$ , где  $F(w_1)$  – частота первого компонента;

$F(w_2)$  – частота второго компонента;  $F(w_1, w_2)$  – частота сочетания  $w_1 w_2$ ,  $N$  – общее число биграмм в корпусе.

<sup>8</sup> Мера оценивается как одна из лучших для извлечения коллокаций [5, разд. 4.3], позволяет найти сочетания, являющиеся характерными для подкорпуса [6, с. 608].

<sup>9</sup> Леммы даны в нормализованной форме электронного грамматического словаря корпуса «Манускрипт».

<sup>10</sup> Количество биграмм в транскрипции.

<sup>11</sup> Компоненты биграммы и количество каждого из компонентов в транскрипции.

$F^{10}$	$F_{w1} \dots F_{w2}^{11}$	$T$ - $score$	$F$	$F_{w1} \dots F_{w2}$	$T$ - $score$
4	ЕВРЕНСКЪ (10) КНИСТОЛНА (165)	1.988	1	БОЖЬСКЪ (2) ЖИТНЕ (11)	0.999
4	РИМЬСКЪ (30) КНИСТОЛНА (165)	1.963	1	ЛАСДНИКНИСКЪ (1) ЦРЬКЪ (27)	0.999
2	КОЛАСАНСКЪ (4) КНИСТОЛНА (165)	1.407	1	ОУЧЕНТЕЛЪСКЪ (1) КЪНИГА (29)	0.999
2	ЕВРЕНСКЪ (10) ЯЗЪКЪ (146)	1.399	1	ЖЕНЬСКЪ (2) ВЕЩЬ (20)	0.998

Таблицы 5а, 5б. Коллокации с постпозицией форм с суффиксом *-ьск-* в ХА и ТА

$F$	$F_{w1} \dots F_{w2}$	$T$ - $score$	$F$	$F_{w1} \dots F_{w2}$	$T$ - $score$
2	СЪНЬМНШТЕ (11) НЮДЕНСКЪ (12)	1.413	2	СЛОУЖЬБА (36) КОУМНРЬСКЪ (5)	1.409
2	ГЛАВНЪНА (42) КОРНИФНИСКЪ (14)	1.408	1	СЛОУШАНИЕ (1) КОРНИФНИСКЪ (1)	0.999
2	МОУДРОСТЬ (13) ПЛЪТЬСКЪ (55)	1.407	1	ПРАВЪ (5) ЕПИСКОУПЬСКЪ (1)	0.999
1	НИКОМЪ (1) ГРАДЬСКЪ (2)	1.000	1	СЛОУЖИТЕЛЪ (1) КОУМНРЬСКЪ (5)	0.999
1	ПРЕДЪЛЪ (3) СОУРЬСКЪ (1)	1.000	1	РИЗА (3) МОУЖЬСКЪ (2)	0.999
1	ВЛЪНИКНИ (1) МОУРЬСКЪ (6)	1.000	1	ЦРЬКЪ (27) АНИСКЪ (2)	0.999

Таблица 6. Коллокации со свободным положением форм с суффиксом *-ыск-* и существительного в ХА

<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>	<i>Коллокации, адрес в ХА</i>
3	ПОХОТЬ (59) ПЛЪТЬСКЪ (55)	1.704	ПЛЪТЬСКЪИХЪ ПОХОТНИ 56.2-16 <sup>12</sup> , ПЛЪТЬСКЪИНА ПОХОТНИ 66.1.1-18.19, ПОХОТНИ ПЛЪТЬСКЪИНА 206.2.1-4.23
2	СКВЪРНА (10) ПЛЪТЬСКЪ (55)	1.408	ПЛЪТЬСКЪИИ СКВЪРНИ 58.2.1-16.1, СКВЪРНИ ПЛЪТЬСКЪИНА 66.1.1-18.12
2	ЗАКОНЪ (422) ПЛЪТЬСКЪ (55)	1.166	ПЛЪТЬСКАГО ЗАКОНА 233.1.1-7.10, ЗАКОНУ ПЛЪТЬСКОУ 256.2.1-11.1

В ТА свободные сочетания с формами с суффиксом *-ыск-* отсутствуют.

Таблицы 7а, 7б. Коллокации с препозицией форм с суффиксом *-и-* в ХА и ТА

<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>	<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>
3	ИЗИИШНИ (10) ВРЕМА (49)	1.728	2	ИЗИИШНИ (3) ВРЕМА (17)	1.413
3	ПОСЛАДНИ (17) ДНИ (111)	1.716	2	ОУГОДНИ (4) БОГЪ (533)	1.351
3	ДЪЛЖНИ (27) ЛЮБИ (234)	1.677	1	ИИИЗГЛАГОЛАНИ (1) ВЪЗДИХАНИ (2)	0.999
2	ПОСЛАДНИ (17) ГОДИНА (6)	1.413	1	ВЪЩИНИ (4) ПРЕБИВАНИ (1)	0.999
2	ЗАКОННИ (14) ОСОУЖЕНИ (10)	1.413	1	ЗЕМЛИНИ (2) ХРАМИНА (2)	0.999
2	ИСТИНИНИ (35) ВЪДОВИЦА (10)	1.411	1	ГЛАНИНИ (1) СЪСОУДЪ (5)	0.999
2	ВЪРНИ (53) СЛОУЖИТЕЛ (11)	1.408	1	ПОСЛАДНИ (3) ТРОУБА (2)	0.999
2	ВЪРНИ (8) СЛОВО (288)	1.390	1	ИЗИИШНИ (3) ЧАСЪ (4)	0.999
2	СОУКЪТНИ (5)	1.386	1	ПОКОРНИ (2)	0.999

<sup>12</sup> 56 – номер листа, .2 – номер страницы на листе, 16 – номер строки.

<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>	<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>
	<b>ВЪРА</b> (522)			<b>ИСПОВѢДАННѢ</b> (6)	
2	<b>ПРАВЪДЪНЪ</b> (36) <b>СОУДЪ</b> (88)	1.380	1	<b>ПРЕЖЪНЪ</b> (1) <b>АПОСТОЛЪ</b> (13)	0.999
2	<b>СНЛЪНЪ</b> (13) <b>БОГЪ</b> (1062)	1.266	1	<b>ЗЕМЛЪНЪ</b> (2) <b>ГЛАГОЛЪ</b> (8)	0.999

Таблицы 8а, 8б. Коллокации с постпозицией форм с суффиксом **-ьн-** в ХА и ТА

<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>	<i>F</i>	<i>F<sub>w1</sub> ... F<sub>w2</sub></i>	<i>T-score</i>
9	<b>ЖИЗНЪ</b> (24) <b>ВЪУЪНЪ</b> (34)	<b>2.996</b>	3	<b>ЖИЗНЪ</b> (18) <b>ВЪУЪНЪ</b> (11)	<b>1.727</b>
6	<b>ЖНЕОТЪ</b> (75) <b>ВЪУЪНЪ</b> (34)	<b>2.434</b>	2	<b>ВРЕМѦ</b> (17) <b>ПРНѦТЪНЪ</b> (3)	<b>1.413</b>
4	<b>ОУМЪ</b> (27) <b>ГОСПОДЪНЪ</b> (10)	<b>1.998</b>	2	<b>УЛОВЕКЪ</b> (77) <b>ГРЕШЪНЪ</b> (20)	1.369
3	<b>ДЪНЪ</b> (111) <b>ГОСПОДЪНЪ</b> (10)	1.722	1	<b>ГЛАВНА</b> (1) <b>ОГНЪНЪ</b> (1)	0.999
3	<b>ЗАКОНЪ</b> (422) <b>ГРЕХОВЪНЪ</b> (10)	1.695	1	<b>ДОУХЪ</b> (2) <b>ЗЪЛОБЪНЪ</b> (1)	0.999
2	<b>ЦВѢТЪ</b> (4) <b>ТРАВЪНЪ</b> (2)	1.414	1	<b>ЗАПРЕЩЕННѢ</b> (2) <b>СТРАШЪНЪ</b> (2)	0.999
2	<b>ЖРЪТВА</b> (23) <b>ПРНѦТЪНЪ</b> (6)	1.413	1	<b>ЛЪСТЪЦЪ</b> (1) <b>ИСТНЪНЪ</b> (4)	0.999
2	<b>МЪЗДА</b> (34) <b>НЕПРАВЪДЪНЪ</b> (8)	1.411	1	<b>СЪСОУДЪ</b> (5) <b>ГНѢВЪНЪ</b> (1)	0.999
2	<b>ВРЕМѦ</b> (49) <b>ПРНѦТЪНЪ</b> (6)	1.411	1	<b>ХОУДОЖЕСТВО</b> (2) <b>ЛЪСТЪНЪ</b> (3)	0.999
2	<b>РАЗЛНУННѢ</b> (8) <b>БРАШЪНО</b> (39)	1.411	1	<b>БОИѦ</b> (2) <b>ЖИЗНЪНЪ</b> (4)	0.999
2	<b>ДЪНЪ</b> (111)	1.411	1	<b>СТРАСТЪ</b> (3)	0.999

F	$F_{w1} \dots F_{w2}$	T-score	F	$F_{w1} \dots F_{w2}$	T-score
	СОУДЪНЪ (3)			НЪЗНЪШЪНЪ (3)	
3	БОГЪ (1062) НСТННЪНЪ (35)	1.407	1	КОУМНРЪ (13) БЕЗГЛАСЪНЪ (1)	0.999

Свободные сочетания с формами с суффиксом **-ЪН-** в ХА и ТА отсутствуют.

Наибольшие величины меры ассоциации для сочетаний с препозицией прилагательных с суффиксом **-ЪСК-** и постпозицией прилагательных с **-ЪН-** (величины в табл. 1а, 1б, 7а, 8а, 8б выделены полужирным) позволяют предположить существование связи между устойчивостью сочетания прилагательного с конкретным суффиксом, его позицией по отношению к существительному и – предполагаем – временем создания списка.

#### б. Разночтения в параллельных контекстах

Модуль статистики дает возможность построить графики распределения в рукописях анализируемых форм и выявить в параллельных контекстах ХА и ТА соответствующие друг другу словоформы, например:

– **СОУЩЕН БО ПО ПЛЪ/ТН·ПЛЪТЬСКАЯ МОУДРЪСТВОУЮТЬ** · ХА 116.1-13 vs. **СУЩЕН БО ПО ПЛОТН ПЛОТЪНОЕ БО СМЪШЛАЮТЬ** ТА 6.2-9 *Rom. 8:5*,

– **МОУДРОСТЬ БО ПЛЪТЬСКАЯ СЪМЪРТЬ** · 116.1-16 vs. **СЪМЪШЛЕННЕ БО ПЛОТЪНОЕ СМЪРТЬ** ТА 6.2-11 *Rom. 8:6*,

– **НЕ ВЪСТЕ АН ЯКО АГГЕЛОМЪ СОУДИМЪ · НЪ ТЪКМО ЖЕ ЖНТННСКЪ/МЪ · / ЖНТНСКАЯ ЖЕ |·| ОУЕО СОУ/ДНЩА** ХА 146.1-3 vs. **НЕ ВЪСТЕ АН / ЯКО АНГЛЪМЪ СЪДНМЪ · А НЕ ТО / УЪЮ ЖНЗНЪНЪМЪ · / ЖНЗНЪНАЯ / ЖЕ ОУЕО СЪДНЩА** · 13.1-23 *1 Cor. 6:3*,

– **ЯКО ПЪРЪТЪНЪНЪНЪ · ТАКО/ВНН Н ПЪРЪТЪННН · Н ЯКО НЕЪ/СКЪНЪНЪ · ТАКОВНН Н НЕЪ/СКНЪНЪ** · ХА 168.1-1 vs. **ЯКОЖЕ ПЪРЪТЪННН · ТАЦН / ЖЕ Н ПЪРЪТННН · Н ЯКОЖЕ НЕЪННН / ТАЦН ЖЕ Н НЕЪННН** · 18.2-1 *1 Cor. 15:48*.

Регулярность соответствия форм с **-ЪСК-** в ХА формам с **-ЪН-** в ТА объясняет, в частности, значительное относительное преобладание в ТА прилагательных с суффиксом **-ЪН-** над формами с **-ЪСК-** (см. разд. 3 «Членные и нечленные формы»).

Выявленная различная количественная активность нечленных и членных прилагательных с суффиксами **-ЪН-** и **-ЪСК-**, их позиционные и комбинаторные характеристики в двух древнерусских списках Апостола определяют возможность поставить вопрос об условиях и при-

чинах существования различий, а на следующем этапе анализа ответить на него с помощью лингвистической интерпретации полученных количественных данных с привлечением лингвотекстологических (лингвистический анализ с учетом состава списков) и историко-лингвистических (анализ реализации лингвистической особенности в кодексах разного времени) методов.

### Библиографический список

1. Баранов В.А., Дубовцев С.В. Модуль статистики информационно-аналитической системы «Манускрипт»: функции и демонстрация данных // Информационные технологии и письменное наследие: материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября 2012 г.) / отв. ред. В.А. Баранов, А.Г. Варфоломеев. Петрозаводск; Ижевск, 2012. С. 23–26.
2. Баранов В.А. Организация поиска и демонстрации коллокаций в корпусе «Манускрипт» // Проблемы истории, филологии, культуры. М.; Магнитогорск; Новосибирск. 2014. 3 (45). С. 275–277.
3. Баранов В.А. Модуль n-грамм исторического корпуса «Манускрипт»: структурные и лингвистические параметры // Научное наследие В. А. Богородицкого и современный вектор исследований Казанской лингвистической школы: тр. и материалы междунар. конф. (Казань, 31 окт. – 3 нояб. 2016 г.): в 2 т. / под общ. ред. К.Р. Галиуллина, Е.А. Горобец, Г.А. Николаева. Казань: Изд-во Казан. ун-та, 2016. Т. 1. С. 50–61. URL: <https://drive.google.com/file/d/0BwBejXХryRcROhc5T3YxTENhLUk/view?usp=sharing> (дата обращения: 31.03.2017).
4. Баранов В.А. Опыт создания модуля n-грамм системы «Манускрипт» и оценки эффективности его использования для поиска коллокаций в корпусе М. В. Ломоносова // Интеллектуальные системы в производстве. 2016. № 4 (31). С. 124–131. URL: <https://drive.google.com/file/d/0BwBejXХryRcRV3ZpemZKcC1CMTQ/view?usp=sharing> (дата обращения: 31.03.2017).
5. Evert S. Association Measures // Computational Approaches to Collocations. URL: <http://collocations.de/AM/index.html> (дата обращения: 31.03.2017).
6. Пивоварова Л.М., Ягунова Е.В. От коллокаций к конструкциям // Acta Linguistica Petropolitana. Труды ин-та лингв. исслед. 2014. Т. 10, № 2. С. 568–617. URL: [http://elibrary.ru/download/elibrary\\_23195921\\_30041482.pdf](http://elibrary.ru/download/elibrary_23195921_30041482.pdf) (дата обращения: 31.03.2017).

### QUANTITATIVE AND STATISTIC ANALYSIS OF MEDIEVAL SLAVONIC TEXTS: TOOLS OF CORPUS «MANUSCRIPT» AND METHODOLOGY OF THEIR USE

*Baranov V.A.*

Kalashnikov Izhevsk State Technical University,  
7, Studencheskaya st., Izhevsk, Russia, victor.a.baranov@gmail.com

The article gives the description of historical corpus “Manuscript” (manuscripts.ru) modules possibilities for a quantitative and statistic analysis of linguistic data. The main parameters of the query forms of the n-gram module and the statistic module and the methodology of their use are demonstrated on the example of adjectives in two Apostle’s manuscripts: Christinopolitanus (the Lviv Historical Museum, MS 37,

XII cent., 291 ff.) and Tolstovsky (the National Library of Russia, Q.p.1.5, XIV cent., 93 ff.). Two ways of search of differences in the frequency of independent and inseparable adjectives with the suffixes -ьск- and -ьн- usage, the techniques of revelation of semantically and grammatically stable combinations of nouns with adjectives and also the tools of visualization of variant readings in the matching contexts of the copies are shown. The conclusion of the various degrees of the forms with the suffixes -ьск- and -ьн- activity in the manuscripts and the successive replacement of the first ones by the second ones in the later codex is made.

*Key words:* historical text corpus; corpus, quantitative, statistic methods; inseparable, independent adjectives; collocations; variant readings.

УДК 801.63

**АНАЛИЗАТОР РИТМИЧЕСКОЙ СТРУКТУРЫ ТЕКСТА:  
АТРИБУЦИЯ ТЕКСТОВ НА ОСНОВЕ  
РИТМИЧЕСКИХ ПАТТЕРНОВ**

*Белоусов К.И., Дусакова Г.Ф.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, belousovki@gmail.com

*Леонов Д.В.*

ООО «Би.Си.Си.», Россия, Санкт-Петербург,  
Лесной просп., 64, лит. А, dvleonov@gmail.com

Исследование посвящено разработке нового метода атрибуции текстов, основанного на анализе ритмики размеров предложений в тексте. Программная реализация метода осуществлена в приложении «Анализатор ритмической структуры текста». Описывается обработка ритмики текста, выявление последовательности размеров, следующих друг за другом предложений, вычисление матрицы переходов типов предложений текста и корреляции между матрицами переходов размеров предложений анализируемого текста с эталонными авторскими матрицами.

*Ключевые слова:* текст, размер предложения, ритм предложений, АРСТ, атрибуция текстов.

Размер (длина) предложения – один из первичных текстовых параметров, в котором специфически преломляется много факторов: семантическая и синтаксическая сложность, установка на простоту/затрудненность восприятия и пр. Длина предложения интересна тем, что обнажает глубинный ритм целого текста, особенно наглядно, если учесть характер распределения единиц разных размеров от начала текста к концу. Средние величины предложения служат фоном, на котором выявляется композиционная роль «экспрессивных» единиц, то есть, сверхкратких и сверхдлинных предложений [1].

Ритм позволяет отдельным разрозненным элементам преодолеть неупорядоченность во времени и пространстве и образовать единый целостный объект. Нужно, чтобы части системы функционировали в едином темпе и подчинялись единой ритмической организации, что и обеспечивает когерентное (согласованное) взаимодействие этих частей. Как отмечают Е.Н. Князева и С.П. Курдюмов, «... для построения сложной организации необходимо когерентно соединить структуры внутри нее, синхронизировать темп их эволюции. В результате объединения структуры попадают в один темпомир, <...> начинают «жить» в одном темпе» [2, с. 247].

Традиционно исследователи соотносят средний размер предложения с индивидуально-авторским стилем или целым жанром [3; 4].

Нами предлагается метод изучения распределения размеров предложений в тексте, позволяющий исследовать единичные тексты и выборочные совокупности. *В основе метода лежит положение о необходимости рассмотрения размеров предложений только в составе целого текста.* Говорить о коротких, средних и больших размерах предложения можно только относительно того целого, в котором эти предложения *содержатся*. В этом смысле априорные шкалы, структуры и стандарты при исследовании ритмов размеров предложений в тексте становятся ненадежными.

В нашем исследовании рассматривается проблема ритма размеров предложения как параметр атрибуции текста. **Цель** работы – выявление возможности установления авторства художественного текста на основе ритмических паттернов. **Методом** исследования является анализ ритма размера предложений в художественных текстах. Метод реализован в программном средстве АРСТ (Анализатор ритмической структуры текста). **Материалом** послужили более 300 рассказов 20 русских прозаиков.

Алгоритм обработки данных.

1. Текст загружается в программное средство и представляется в виде цепочки чисел, отражающих следующие друг за другом размеры предложений в тексте. Например, рассказ А.П. Чехова «Смерть чиновника» записывается в следующем виде: 23 – 8 – 2 – 7 – 6 – 17 – 3 – 6 – 11 – 2 – 20 – 5 – 21 – 12 – 5 – 7 – 2 – 18 – 2 – 3 – 5 – 3 – 2 – 9 – 8 – 4 – 18 – 1 – 5 – 2 – 16 – 13 – 4 – 20 – 6 – 8 – 20 – 7 – 9 – 4 – 6 – 6 – 5 – 13 – 20 – 9 – 15 – 1 – 2 – 3 – 8 – 6 – 2 – 5 – 3 – 20 – 13 – 5 – 7 – 10 – 6 – 5 – 5 – 11 – 3 – 7 – 3 – 5 – 5 – 8 – 7 – 24 – 13 – 4 – 14 – 8 – 7 – 6 – 6 – 15 – 12.

2. Абсолютные значения размеров предложений переводятся в значения порядковой шкалы в рамках двух моделей: квартилей и децилей. На основе полученной цепочки размеров предложений, относящихся к тем или иным квартилям/децилям, генерируются соответствующие матрицы переходов для данного текста.

3. Полученные матрицы переходов для каждого текста с помощью метода ранговой корреляции сопоставляются с эталонными матрицами писателей и получают коэффициент корреляции с каждой из «писательских» матриц.

4. Создается таблица  $N \times M$ , где  $N$  – количество отобранных для анализа произведений,  $M$  – количество анализируемых авторов; в ячейках располагаются коэффициенты ранговой корреляции для пары: Текст<sub>1</sub> – Автор<sub>1</sub>.

5. Анализируется полученная таблица; делаются выводы о вероятностной атрибуции текстов с помощью параметра ритма размеров предложений текстов.

### **Библиографический список**

1. Белоусов К.И. Деятельностно-онтологическая концепция формообразования текста: дис. ... д-ра филол. н. Оренбург, 2006. 374 с.
- Князева Е.Н., Курдюмов С.П. Антропный принцип в синергетике // Вопросы философии. 1997. № 3. С. 62–79.
2. Лесский Г.А. О зависимости между размером предложения и характером текста // Вопросы языкознания. 1963. № 3. С. 92–112.
3. Yule G.U. On sentence length as a statistical characteristic of style in prose // Biometrika. 1939. XXX. P. 363–390.

### **ANALYZER OF THE RHYTHMIC STRUCTURE OF THE TEXT: ATTRIBUTION OF TEXTS BASED ON RHYTHMICAL PATTERNS**

*Belousov K.I., Dusakova G.F.*

Perm State University, 15, Bukireva st., Perm, Russia,  
belousovki@gmail.com

*Leonov D.V.*

ООО “BCC”, 64 A, Lesnoy ave., St.-Petersburg, Russia,  
dvleonov@gmail.com

The study is devoted to the development of a new method of attribution of texts, based on the analysis of the rhythm of the size of sentences in the text. The software implementation of the method is implemented in the application "Analyzer of the rhythmic text structure". It describes the processing of the rhythm of the text, the identification of a sequence of sizes, successive sentences, the calculation of the matrix of transitions of the types of text sentences, and the correlation between the matrices of the transitions of the sizes of sentences of the analyzed text with reference author matrices.

*Key words:* text, sentence size, rhythm of sentences, ARST, attribution of texts.

## ТРАНСКРИПЦИЯ УСТНОГО ДИАЛОГИЧЕСКОГО ДИСКУРСА: ФОРМАТЫ, ИНСТРУМЕНТЫ И ОПЫТ ИХ ИСПОЛЬЗОВАНИЯ

*Блинова О.В.*

Санкт-Петербургский государственный университет, Россия,  
Санкт-Петербург, Университетская наб., 7/9, o.blinova@spbu.ru

В докладе рассматриваются форматы и конвенции дискурсивной транскрипции. Анализируются некоторые различия в макроструктуре транскриптов, обусловленные использованием определённых инструментов расшифровки и аннотирования (типы нотации, способы расположения аннотации). Рассматриваются параметры отличий между конвенциями дискурсивной транскрипции (resp. различия в микроструктуре транскриптов). Перечисляются и кратко характеризуются инструменты расшифровки и аннотирования. Затрагивается проблема обмена данными и совместимости (interoperability) разных форматов.

*Ключевые слова:* дискурсивная транскрипция, инструменты расшифровки и аннотирования, форматы транскрипции, конвенции транскрипции.

«Дискурсивная транскрипция» (ДТ) – «создание письменной репрезентации событий речи» [1]. Существует множество форматов дискурсивной транскрипции, см. [2], [3], [4], [5]. Предпочтение того или иного формата, тех или иных конвенций транскрипции обусловлено характером коммуникации (диалог или монолог др.), теоретической рамкой исследования и его задачами [6]. При анализе форматов и конвенций ДТ вслед за [2] полезно проводить различие между микроструктурой и макроструктурой транскрипта. Микроструктура – это собственно контент транскрипта. Макроструктура – это способы визуализации текста расшифровки, его привязки к временной оси, к звуку и осциллограмме, а также способы расположения и хранения сопутствующей информации (т. е. как расположены реплики или другие единицы членения, где расположена аннотация, где хранятся метадаанные, какие именно метадаанные и др.).

Формат во многом предопределён возможностями, которые предоставляет инструмент расшифровки и аннотирования. Программа, предназначенная для расшифровки, определяет способ линейной организации транскрипта (способ размещения реплик разных говорящих, нотацию) и «место», где расположена аннотация, исследовательский комментарий к содержимому транскрипта. Транскрипт может быть организован вертикально (вертикальная, или «драматургическая» нотация), по аналогии с музыкальной партитурой (партитурная нотация),

или по столбцам (когда реплики говорящего А всегда в столбце слева, а реплики говорящего Б всегда в столбце справа).

Аннотация может располагаться в строках (слоях, уровнях) под строками расшифровки (это интерлинейарный или многослойный, multilayer, формат аннотирования). Кроме того, она может быть внедрена в сам транскрипт (это так называемый вкрапленный, interspersed формат аннотации, ср. пример, где размечены части речи: *мам\_S смотри\_V / кошка\_S пришла\_V //*). Существует и формат столбцов (column-based format), в котором разные типы информации разнесены по разным столбцам.

Параметры отличия в микроструктуре транскриптов следующие:

1. Использование стандартной vs модифицированной орфографии.

2. Введение разных единиц членения дискурса («юнитов») – интонационных единиц, реплик, элементарных дискурсивных единств, интонационных фраз, минимальных дискурсивных единиц и др. В вертикальной нотации определённый юнит обычно совпадает со строкой.

3. Использование разных знаков для маркирования пауз различной длительности, пауз незаполненных (без вокализации) и заполненных, пауз с разными функциями.

4. Передача разнообразной просодической информации (информации о длительности, темпе произнесения, интенсивности, громкости, частоте основного тона фрагментов речевой цепи).

5. Отражение информации, связанной с turn-taking (системой передачи очерёдности говорения от одного участника коммуникации другому), а именно: пауз между комплексами реплик разных говорящих, перебивов, наложений речи.

6. Использование разнообразных знаков для маркирования неречевых аудиособытий.

7. Способы маркирования непонятных для расшифровщика мест, введения альтернативной расшифровки.

Инструменты расшифровки и аннотирования многочисленны (ELAN, CLAN, ANVIL, EXMARaLDA, Folker, VoiceWalker, LIPP, SALT, Transcriber, CoNLL, Blitzscribe2, MTRANS и др.). В каждом используются свои форматы файлов, что затрудняет обмен данными между исследователями, работающими с разным инструментарием. Для того, чтобы сделать обмен данными возможным, создатели программ предусматривают возможности экспорта в другие форматы. Так, например, транскрипт в формате CHAT (CLAN) может быть, по утверждению [3], автоматически преобразован в форматы Praat, Phon, ELAN, CoNLL, ANVIL EXMARaLDA, LIPP, SALT, LENA, Transcriber и ANNIS.

ELAN допускает, в частности, обмен данными с Toolbox, FLEx, Praat, CLAN [7]. В контексте работы с дискурсивной (не с фонетической) транскрипцией в рамках подходов разговорного анализа, функциональной прагматики и др. актуальны прежде всего возможности экспорта из ELAN в формат CHAT (CLAN) (resp. экспорт файла \*.eaf в формат \*.cha) и в формат FLEx.<sup>1</sup>

Наконец, необходимо упомянуть отдельные инструменты, предназначенные для преобразования формата файлов, например, Transformer [8], для конвертации файлов в единый формат [9] или для одновременной работы с разными форматами, например, NITE XML Toolkit [10].

### Библиографический список

1. *Du Bois J. et al.* Santa Barbara Papers in Linguistics. Vol. 4. Discourse transcription. Santa Barbara: Dep. of linguistics UCSB, 1992.
2. *Schmidt Th.* A TEI-based Approach to Standardising Spoken Language Transcription // Journal of the Text Encoding Initiative [Online]. № 1. June 2011. URL <http://jtei.revues.org/142> (дата обращения: 20.04.17).
3. *MacWhinney B.* The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
4. *Рассказы о сновидениях: Корпусное исследование устного русского дискурса / под ред. А.А. Кибрика и В.И. Подлеской.* М.: Языки славянских культур, 2009.
5. *Bird S., Liberman M.* A formal framework for linguistic annotation // Speech Communication. 2001. № 33. P. 23-60.
6. *Edwards J.A.* The Transcription of Discourse // Schiffrin D., Tannen D., Hamilton H.E. (eds.) The Handbook of Discourse Analysis. Malden, MA: Blackwell Publishers Inc., 2001. P. 321-348.
7. *Sloetjes H. et al.* ELAN – aspects of interoperability and functionality // Proceedings of Interspeech 2011. P. 3249-3252.
8. *Transformer.* URL <http://www.oliverehmer.de/transformer/> (дата обращения: 20.04.17).
9. *TEI Drop.* URL <http://exmaralda.org/en/tei-drop-en/> (дата обращения: 20.04.17).
10. *NITE XML Toolkit.* URL <http://groups.inf.ed.ac.uk/nxt/index.shtml> (дата обращения: 20.04.17).

### **SPOKEN DIALOGUE TRANSCRIPTION: FORMATS, TOOLS, AND SOME EXPERIENCE FROM THEIR USE**

*Blinova O.V.*

Saint Petersburg State University, 7/9, Universitetskaya nab.,  
St. Petersburg, Russia, o.blinova@spbu.ru

The paper discusses formats and conventions of discourse transcription. Some differences in transcripts macrostructure and microstructure; the parameters that differ

---

<sup>1</sup> Именно экспорт из ELAN обсуждается в докладе подробно.

transcripts content; some tools for transcription and annotation, and the problem of converting between formats and interoperability are in question.

*Key words:* discourse transcription, tools for transcription and annotation, transcription formats, transcription conventions.

УДК 930.272:004

## **ДРЕВНИЕ КОРПУСЫ ТЕКСТОВ В ЦИФРОВУЮ ЭПОХУ: РАЗРАБОТКА ИННОВАЦИОННОГО РЕСУРСА ПО ШУМЕРСКОЙ КЛИНОПИСИ**

*Григоришин С.В.*

Тюменский государственный университет, Тюмень, ул. Ленина, 23,  
s.v.grigorishin@gmail.com

Статья посвящена описанию проекта *Древние корпуса текстов в цифровую эпоху*, который с октября 2016 г. проводится научно-исследовательской группой постдока ТюмГУ на базе Института истории и политических наук Тюменского государственного университета. Проблема, на решение которой направлены усилия участников проекта, заключается в слабой поддержке базовых ресурсов по ассириологии клинописной системой символов. Приоритет остается за отражением шумерских и аккадских текстов в научной транскрипции, между тем как развивающиеся с невероятной скоростью цифровые технологии уже давно способны обрабатывать информацию на идеографических языках. Самым ярким подтверждением этого факта служит динамика развития китайского Интернета, который основан на распознавании символики идеографической системы письма. Таким образом, академическая ассириология уже способна выйти на новый уровень работы с первоисточниками – с помощью программы распознавания и набора текстов в старовавилонской и новоассирийской клинописи, разработанной на основе стандартного юникода. Главным результатом такой работы должна стать разработка, внедрение и поддержка компьютерной программы по набору символов шумерской клинописи по стандарту действующего юникода. В статье также дан перечень мировых достижений в этой области и основанные научные центры, где были достигнуты позитивные результаты по использованию электронных ресурсов в сфере ассириологии. Вместе с тем описаны наиболее актуальные проблемы, требующие детальной проработки и быстрого решения.

*Ключевые слова:* шумерская клинопись, ассириология, электронные ресурсы, цифровая эпоха, корпус текстов, цифровая библиотека, лексические списки.

Идея тюменского проекта *Древние корпуса текстов в цифровую эпоху* (*The Ancient Corpora in Digital Age*) возникла в связи с осмыслением роли цифровых технологий в области истории и филологии, а также быстрым прогрессом в мировой ассириологии. Для понимания того, чем занимается научная группа проекта, нужно

кратко ознакомиться с контекстом развития электронных ресурсов в ассириологии.

В 1997 году ассириологи из Восточного института Оксфордского университета разработали онлайн-проект *Электронный корпус текстов шумерской литературы* (*The Electronic Text Corpus of Sumerian Literature*) [1]. Уже в следующем году благодаря совместным усилиям Калифорнийского университета в Лос-Анджелесе и берлинским Институтом истории науки Макса Планка возник *Проект клинописной цифровой библиотеки* (*The Cuneiform Digital Library Initiative*) [2]. Начиная с 2000 года существует мультиинституциональный проект *Электронные сервисы и Древние ближневосточные архивы* (*Electronic Tools and Ancient Neareastern Archives*) [3], он объединяет три базы данных: Abzu, Core Texts и eTact. В 2002 году обрел новую жизнь Пенсильванский шумерский словарь. Отныне он называется *Электронный Пенсильванский шумерский словарь* (*The Electronic Pennsylvania Sumerian Dictionary*) [4]. С 2003 года проект *Цифровой корпус клинописных лексических текстов* (*Digital Corpus of Cuneiform Lexical Texts*) [5] – под руководством главного специалиста в этой области Ника Вельхейса из Калифорнийского университета в Беркли. При наличии цифровой библиотеки шумерской литературы ETCSL резонно возникал вопрос: когда появится библиотека аккадской литературы? Восполнили этот пробел двое ученых – Михаэль Штрек из Института Древнего Востока Лейпцигского университета и Натан Вассерман из Института археологии Еврейского университета в Иерусалиме. Проект получил название *Источники ранней аккадской литературы* (*Sources of Early Akkadian Literature*) [6]. Так в общих чертах выглядит цифровая ипостась современной ассириологии. Проекты ETCSL, CDLI, ETANA, ePSD, DCCLT и SEAL охватывают все темы и ресурсы мировой ассириологии. В то же время, они продолжают совершенствоваться в техническом и содержательном плане. Такая потребность оптимизирует появление новых проектов по цифровым технологиям в ассириологии.

Для этих целей научной группой постдока ТьюмГУ С. Григоришина создан проект ACDA при кафедре археологии, истории Древнего мира и Средних веков Тюменского государственного университета [7]. Проект *The Ancient Corpora in Digital Age* имеет конкретную цель – дополнить существующий в транскрипции ETCSL клинописным эквивалентом. Эта цель достигается путем решения следующих задач: 1) выбор печатных изданий текстов шумерской литературы, скопированных от руки и представленных на сайте ETANA в разделе Core Texts; 2) ознакомление и работа с колляциями – сличение и проверка корректности клинописных копий из печатных изданий; 3) поиск и отбор фотографий табличек с текстами шумерской литературы, работа с которыми способствует верификации отдельных кли-

нописных знаков; 4) набор клинописного текста вручную с помощью современного юникода (U+12000–U+1237F Cuneiform); 5) проверка и сличение с оригинальными клинописными табличками набранного текста, а также отправка текстов ведущим ассириологам с целью ознакомления с полученным результатом; 6) на финальном этапе публикация электронного клинописного корпуса текстов на собственном сайте с открытым доступом.

Инновационность проекта ACDA заключается в том, что впервые в сфере цифровых технологий клинописные тексты будут представлены не в статичном, а динамичном виде. Другими словами, статичный клинописный текст «переходит» в динамичный клинописный текст. Рядом с фотографиями отдельных клинописных табличек, печатными изданиями текстов в клинописи и их отсканированными электронными копиями появится клинописный электронный контент. Клавиатурный набор клинописных знаков позволит в краткие сроки осуществлять легкий и удобный поиск необходимых первоисточников. К настоящему времени определен круг печатных источников, с которых будет производиться клавиатурный набор текста. Первые результаты проекта ACDA будут представлены на материале шумерского мифа *Энлиль и Нинлиль*, которым открывается ETC SL.

#### **Библиографический список**

1. The Electronic Text Corpus of Sumerian Literature. General information. URL: (дата обращения: 27.03.2017).
2. The Cuneiform Digital Library Initiative Wiki: Submission Guidelines. URL: [http://cdli.ox.ac.uk/wiki/doku.php?id=submission\\_guidelines](http://cdli.ox.ac.uk/wiki/doku.php?id=submission_guidelines) (дата обращения: 27.03.2017).
3. Electronic Tools and Ancient Neareastern Archives/ About the ETANA Project. URL: <http://www.etana.org/about> (дата обращения: 27.03.2017).
4. The Electronic Pennsylvania Sumerian Dictionary. About the PSD. (дата обращения: 27.03.2017).
5. Digital Corpus of Cuneiform Lexical (дата обращения: 27.03.2017).
6. Sources of Early Akkadian Literature. Project Description. (дата обращения: 27.03.2017).
7. Постдок ТюмГУ изучает древние письменаэ (дата обращения: 27.03.2017).

#### **THE ANCIENT CORPORA IN DIGITAL AGE: AN INNOVACION RESOURSC E DEVELOPING FOR SUMERIAN CUNEIFROM TEXTS**

*Grigorishin S.V.*

Tyumen State University, 23, Lenina st., Tyumen, Russia,  
[s.v.grigorishin@gmail.com](mailto:s.v.grigorishin@gmail.com)

The paper deals with the description of the project “Ancient Textual Corpora in the Digital Age” which has been conducted by a postdoc-led research team at the Institute of History and Political Sciences of the University of Tyumen (Tyumen, Rus-

sia). The project is targeted at the problem of the existing electronic tools on Assyriology being poorly supported by the system of cuneiform symbols. The existing online resources, like ETCSL, give preference to the transcribed academic versions of Sumerian and Akkadian texts, whereas the increasingly developed technology has long allowed to process data in ideographic languages, and the ballooning Chinese web-content based on the recognition of ideographic symbols is the best confirmation of this fact. Thus, academic Assyriology is now able to move to the next level in its work with the original sources by means of a recognition programme as well as of a text layout in Unicode-based symbols of Old Babylonian and New Assyrian cuneiform. The main result to achieve by the research group will be the development, implementation, and support of a software programme aimed at producing a layout of Sumerian cuneiform symbols encoded within the standard Unicode allocation. The paper also provides a list of major achievements in this research area and a list of major research centers in which the use of electronic data has been successfully used as applied to Assyriology.

*Key words:* Sumerian cuneiform, Assyriology, electronic resources, digital case, digital library, vocabulary lists.

УДК 81'374

## **МУЛЬТИМОДАЛЬНЫЙ СЛОВАРЬ ФИТОНИМОВ НОВОГО ТИПА**

*Гулевская А.Е., Мишланова С.Л.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, aekhokhlova@mail.ru

Излагается краткая история возникновения такой своеобразной знаковой системы как язык цветов, и создания словарей фитонимов начиная с XVIII в. и заканчивая нашим временем, а также перспективы создания современного мультимодального словаря фитонимов нового типа с использованием облачного хранилища Google. Описывается процесс составления и структура словаря, а также объясняется необходимость разработки словаря нового типа, перечисляются достоинства использования облачных хранилищ для размещения подобных словарей и работы над ними. Рассматриваются возможности словаря в литературоведческих, лингвокультурологических, лингвистических, лексикографических, семиотических исследованиях, а также исследованиях в области изучения метафоры. Словарь создан на материале оригинальных произведений английского поэта Дж. Г. Байрона и словаря флористических символов Генри Филиппса "Floral Emblems".

*Ключевые слова:* фитоним, облачное хранилище, словарь, язык цветов.

Словари цветов возникли в период развития ботаники как науки. Составление ее пришлось на XVIII в., – эпоху стремительного развития транспортного сообщения между странами. Это дало европейцам возможность привозить экзотические виды растений со всего

мира и разводить их у себя на родине. Ботаника вызвала интерес высшего образованного сословия и быстро вошла в моду.

В XIX в. ботаникой стали увлекаться женщины. Для них это была одна из немногих возможностей заниматься наукой. Тем не менее в то время все еще бытовало мнение, что женский интеллект значительно уступает мужскому, поэтому ботаника для женщин принимала особые формы. Одной из них был язык цветов. Он являл собой своеобразную систему знаков, роль слов в которой исполняли растения и плоды.

В языке цветов, как в полноценной знаковой системе существовало такое явление как полисемия. Фитонимы могли обладать множественством значений, и со временем количество этих значений могло изменяться. Значение того или иного знака в этой системе зависело от периода, географического положения и многих других факторов. Со временем складывались разные традиции языка цветов, значения растений в которых могли отличаться вплоть до противоположности.

За всю историю существования языка цветов появилось великое множество словарей, большая часть из них к сожалению, недоступна, но упоминания встречаются в литературе, письмах и записках.

Несмотря на развитие информационных технологий и увеличение скорости передачи информации, доступных, понятных и современных словарей фитонимов, отвечающих критериям качества и научности, до сих пор нет. В сети Интернет можно найти несколько сайтов, предлагающих так называемые словари языка цветов, источники в которых не указаны либо недостоверны. Наша цель – это создание с помощью информационных технологий современного, удобного, доступного и универсального мультимодального словаря фитонимов, который мог бы быть применен во множестве сфер.

При разработке нашего словаря мы используем облачное хранилище Google. Возможности этой системы позволяют создавать и редактировать электронные таблицы, документы, презентации и т. д., а также разместить и хранить их в так называемом облачном хранилище. Это значит, что, имея электронную ссылку на этот словарь, любой заинтересованный пользователь сможет редактировать, исправлять, дополнять и видоизменять словарь, благодаря чему работать над ним сможет огромное количество человек, внося в него изменения в реальном времени. Отсюда следует, что, в отличие от бумажных словарей, на издание которых требуется значительное время, такой словарь будет актуальным всегда, поскольку обновления, внесенные в него создателями, будут появляться мгновенно.

В качестве исходного материала нами был выбран словарь флористических символов Генри Филиппа “Floral Emblems” (Цветочные эмблемы), а также произведения английского поэта Джорджа Гордона Байрона, в которых мы и искали фитонимы, содержащиеся в филипп-

совском словаре. Всего нами найдено более ста словоупотреблений фитонимов, содержащихся в словаре.

Словарь составлен с помощью технологии «Google Таблицы». Он имеет форму электронной таблицы с несколькими заполняемыми полями:

1. Название фитонима на русском языке;
2. Название фитонима на английском языке;
3. Изображение растения, его плодов и листьев;
4. Значение растения из словаря Г.Филлипса;
5. Значения из дополнительных источников: энциклопедического, ботанического, медицинского, толкового, этимологического словарей;
6. Название литературного произведения;
7. Дата создания произведения;
8. Цитата, содержащая фитоним (литературный контекст);
9. Контекстуальное значение фитонима в произведении;
10. Совпадение/несовпадение со значением из словаря Г. Филлипса;
11. Идентификация метафорического значения.

Работа над подобными словарями необходима, поскольку она поможет систематизировать знания о такой специфической знаковой системе как «язык цветов», которая зародилась на востоке и активно распространилась по всей Европе, оказывая значительное влияние на языки и культуры на протяжении долгого времени.

Использование современного электронного мультимодального словаря значительно облегчит исследовательскую работу во многих областях и направлениях, например:

1. в литературоведческих и лингвокультурологических исследованиях;
2. интерпретации текста литературных произведений;
3. исследованиях, посвященных метафоре (благодаря его сложной структуре, т. е. наличию в словаре статей из энциклопедических и ботанических словарей, отрывков из литературных произведений, исторических справок о возникновении и использовании растений, возможно будет обнаружить неочевидные мотивирующие признаки метафоры, которые стерлись с течением времени);
4. лексикографических и семиотических исследованиях – работа с литературными контекстами поможет зафиксировать новые неизвестные значения слов.

Такой словарь по своей сути является «метасловарем», в котором собраны статьи из различных источников, что позволяет в полной мере изучить семантику языковых знаков.

### **Библиографический список**

1. *Ненарокова М.П.* Необыкновенная энциклопедия: язык цветов в справочной литературе // *Современные тенденции развития науки и технологий.* 2016. №. 3–4. С. 71–79.
2. *Ненарокова М.П.* Язык цветов: между литературой и ботаникой // *Проблемы национальной литературы художественные поиски второй половины XX в. и современность.* Новосибирск: Наука, 2015. С. 106–114.
3. Phillips H. *Floral Emblems.* 1825. URL: [https://books.google.ru/books/about/Floral\\_Embles.html?id=H11IAAAAYAAJ&redir\\_esc=y](https://books.google.ru/books/about/Floral_Embles.html?id=H11IAAAAYAAJ&redir_esc=y) (дата обращения 10.04.2017).
4. *The Collected Poems of Lord Byron.* 1832. URL: <https://books.google.ru/books?id=y9HrRPW1SpYC&printsec=frontcover&dq=george+gordon+byron&hl=ru&sa=X&ved=0ahUKewiexo3hzJbRAhUJiywKHYEMDioQ6AEIZjAH#v=onepage&q=george%20gordon%20byron&f=false> (дата обращения 10.04.2017)
5. *The Poetical works of Lord Byron.* 1855. URL: <https://books.google.ru/books?id=8N0cUz0Ze3oC&printsec=frontcover&dq=george+gordon+byron&hl=ru&sa=X&ved=0ahUKewiexo3hzJbRAhUJiywKHYEMDioQ6AEIUDAE#v=onepage&q=george%20gordon%20byron&f=false> (дата обращения 10.04.2017)
6. *The Works of Lord Byron: With His Letters and Journals,* 1832. URL: <https://books.google.ru/books?id=tccNAAAAQAAJ&pg=PA188&dq=george+gordon+byron&hl=ru&sa=X&ved=0ahUKewiexo3hzJbRAhUJiywKHYEMDioQ6AEIPjAC#v=onepage&q=george%20gordon%20byron&f=false> (дата обращения 10.04.2017)
7. *The Works of the Right Hon. Lord Byron,* 1825. URL: [https://books.google.ru/books?id=Fak\\_AAAAYAAJ&printsec=frontcover&dq=george+gordon+byron&hl=ru&sa=X&ved=0ahUKewiexo3hzJbRAhUJiywKHYEMDioQ6AEINjAB#v=onepage&q=george%20gordon%20byron&f=false](https://books.google.ru/books?id=Fak_AAAAYAAJ&printsec=frontcover&dq=george+gordon+byron&hl=ru&sa=X&ved=0ahUKewiexo3hzJbRAhUJiywKHYEMDioQ6AEINjAB#v=onepage&q=george%20gordon%20byron&f=false) (дата обращения 10.04.2017)

## **NEW TYPE MULTIMODAL DICTIONARY OF PHYTONYMS**

*Gulevskaya A.E., Mishlanova S.L.*

Perm State University, 15, Bukireva st., Perm, Russia,  
aekhokhlova@mail.ru

The article presents a brief history of origin and development of “the language of flowers” as a special kind of a sign system, the preconditions of creation of phytonyms dictionaries, from the 18th century to the present time, as well as the prospects for creating a modern multimodal dictionary of phytonyms of a new type using Google’s cloud storage. The authors describe the process of creating and structuring the dictionary, and explain the necessity of developing a new type of dictionary. The article lists the advantages of using cloud technology for storage and processing of information. The article deals with the possibilities of using the dictionary in literary, linguoculturological, linguistic, lexicographic, semiotic studies, as well as researches in the field of metaphor studies. The dictionary is based on the original works of English poet George Gordon Byron and the dictionary of floral symbols “Floral Emblems” written by Henry Phillips.

*Key words:* phytonym, cloud storage, dictionary, language of flowers.

## МЕНТАЛЬНЫЙ ЛЕКСИКОН: СЕТЬ ИЛИ ПОЛЕ?<sup>1</sup>

*Ерофеева Е.В.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, elenerofee@gmail.com

*Лещенко Ю.Е.*

Пермский государственный гуманитарно-педагогический университет, Россия, Пермь, ул. Сибирская, 24, naps536@gmail.com

Обсуждаются основные подходы к моделированию ментального лексикона: модулярный, сетевой (коннекционистский), интегральный. С точки зрения модулярного подхода ментальный лексикон репрезентируется в виде относительно самостоятельных модулей; сетевые модели лексикона представляют его в виде сложной нейронной сети; комплексные интегральные модели совмещают в себе модульную и сетевую архитектуру. Структура ментального лексикона характеризуется наличием множественных внутриуровневых и межуровневых связей, формирующих подструктуры лексикона разного объема. Эти подструктуры имеют полевой характер и обеспечивают параллельный доступ к разнотипной информации, хранящейся в ментальном лексиконе. Ментальный лексикон является динамической системой: его связи могут перестраиваться в зависимости от ситуации и активировать разные внутриуровневые и межуровневые поля. В докладе приводятся данные экспериментальных исследований, демонстрирующие различные связи и поля в ментальном лексиконе.

*Ключевые слова:* ментальный лексикон, моделирование, нейронная сеть, полевая структура лексикона, внутриуровневые связи, межуровневые связи.

Исследования ментального лексикона – части языковой способности человека, обеспечивающей связь когниции и языка [1], – ведутся уже более полувека, однако единства мнений относительно его сущности и структуры до сих пор нет. Существует несколько подходов к моделированию ментального лексикона, внутри каждого из которых также наблюдается вариативность мнений относительно устройства лексикона и способов его моделирования. Среди основных подходов можно назвать модулярный, разрабатываемый в рамках генеративистской традиции, сетевой, развивающийся в рамках коннекционизма, а также различные интегральные подходы, совмещающие представления о ментальном лексиконе как сети и модулях внутри сети.

Собственно модулярные подходы основываются на представлениях Дж. Фодора о модулярности когнитивной архитектуры, т. е. о структуре когнитивной системы как множестве относительно авто-

номных модулей, каждый из которых обеспечивает специфические процессы, в том числе языковые [2].

Коннекционистские модели когниции опираются на представление о мозге как о единой нейронной сети, в которой происходит постоянная (параллельная и одновременная) переработка входящей информации. Сетевые модели воспроизводят структуру и/или функциональные аспекты биологических нейронных сетей, в которых закрепляется любой когнитивный опыт индивида, в том числе опыт, связанный с усвоением языка [3, 4].

Сложность строения изучаемого объекта такова, что экспериментальное подтверждение находят и те, и другие типы моделей, что привело исследователей к созданию комплексных интегральных моделей.

Интегральные модели ментального лексикона представляют его как сеть, включающую совокупность подсетей (фонологическая/орфографическая, семантическая и др.), каждая из которых выполняет определенные функции [5, 6]. Подсети являются компонентами (подсистемами) ментального лексикона и могут рассматриваться как функциональные аналоги модулей.

В работах А.А. Залевской интегральная модель представляет ментальный лексикон как «сложную систему многоярусных, многократно пересекающихся полей, с помощью которых упорядочивается и хранится... разносторонняя информация о предметах и явлениях окружающего мира... и о лингвистических особенностях обозначающих их вербальных единиц» [7: 73].

Таким образом, ментальный лексикон можно представить как структуру с множественными связями: внутри- и межуровневыми. Во-первых, связи элементов лексикона образуют модули (фонетико-фонологический, грамматический, концептуальный и др.). Во-вторых, существуют внутриуровневые связи, формирующие подсети в каждом из модулей. Организация данных подсетей имеет полевой характер: в них выделяются ядро и периферия, разные связи при активации имеют разную силу, связи могут пересекаться между собой и т. п. В-третьих, существуют межуровневые (межмодульные) связи, обеспечивающие интеграцию всех подсистем (компонентов) в единое целое. Межуровневые связи, в свою очередь, также формируют поля, обеспечивающие параллельный доступ к разнотипной информации, хранящейся в ментальном лексиконе.

Ментальный лексикон является динамической системой: связи могут перестраиваться в зависимости от ситуации и активировать разные внутриуровневые и межуровневые поля.

В докладе приводятся данные экспериментальных исследований, демонстрирующие различные связи и поля в ментальном лексиконе.

## Библиографический список

1. Кубрякова Е.С. Язык и знание: На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира. М.: Языки слав. культуры, 2004. 560 с.
2. Fodor J. The Modularity of Mind: An Essay on Faculty Psychology. Cambridge: MIT Press, 1983. 158 p.
3. Rumelhart D.E., Hinton G.E., McClelland J.L. A General Framework for Parallel Distributed Processing // Rumelhart D.E., & McClelland J.L. and the PDP Research Group (Eds). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MA: MIT Press. 1986. Vol. 1: Foundations.
4. Ritter H, Kohonen T. Self-Organizing Semantic Maps // Biological cybernetics. 1989. № 61. P. 241–254.
5. Li P., Farkas I. A self-organizing connectionist model of bilingual processing // R. Heredia & J. Altarriba (Eds.). Bilingual Sentence Processing. 2002. P. 59–85.
6. Залевская А.А. Лексикон человека как функциональная самоорганизующаяся система // Проблемы организации внутреннего лексикона человека. Калинин, 1977. С. 3–74.

### MENTAL LEXICON: A NETWORK OR A FIELD?

*Erofeeva E.V.*

Perm State University, 15, Bukireva st., Perm, 614990, Russia,  
elenerofee@gmail.com

*Leshchenko Y.Y.*

Perm State Humanitarian-Pedagogical University, Sibirskaya st. 24,  
Perm, Russia, naps536@gmail.com

The paper discusses major approaches to mental lexicon modeling: the modular, the network (connectionist) and the integral ones. From the point of view of modular approach mental lexicon is represented as a set of relatively independent modules; network mental lexicon models visualize it as a complex neural network; integrated models combine the modular and network architecture. Mental lexicon structure is characterized by a multitude of intra-level and inter-level connections which form substructures of different size. These substructures are shaped as field formations and ensure parallel access to diverse information stored in mental lexicon. Mental lexicon is a dynamic system: its connections can reorganize depending on the situation and activate various intra-level and inter-level fields. In the paper the data demonstrating diverse connections and fields in mental lexicon are presented.

*Key words:* mental lexicon; modeling; neural network; field lexicon structure; intra-level connections; inter-level connections.

## ИСПОЛЬЗОВАНИЕ СПЕЦИАЛЬНОГО КОРПУСА В ЛИНГВИСТИЧЕСКИХ ЦЕЛЯХ<sup>1</sup>

*Крылов С.А.*

Институт Востоковедения РАН,  
Россия, Москва, ул. Рождественка, 12, krylov-58@mail.ru

*Фролова О.Е.*

Московский государственный университет имени М.В. Ломоносова,  
Россия, Москва, ГСП-1, Ленинские горы, 1, стр. 51,  
olga@frolova@list.ru

Рассматривается взаимодействие общелитературного и профессионального, юридического, языков в сложившейся системе функционирования документов в «бумажном» и электронном виде. Монополия юристов как создателей и интерпретаторов юридической документации оставляет непрофессионального носителя русского языка на периферии важнейшей сферы правовой коммуникации – общения государства и гражданина. Потребность создания текстов в юридической сфере в продуктивных и рецептивных видах речевой деятельности: в письменных жанрах заявления, претензии, иска, ходатайства, а также в устных жанрах ответов на вопросы в судебном заседании, необходимость понимания и интерпретации законодательных и правовых письменных документов, речи прокурора и адвоката актуализируют знание юридической терминологии неюристами. В своих попытках стать участником правовой коммуникации неюрист сталкивается с несовпадением объемов значения, полисемией и омонимией слов общелитературного языка и юридических терминов. Специальные корпуса и юридические словари онлайн позволяют снять остроту коммуникативных проблем между юристами и неюристами.

*Ключевые слова:* общелитературный язык, профессиональный язык, термин, терминосистема, поисковая система «Консультант Плюс», язык права, Национальный корпус русского языка, подкорпус.

В настоящую эпоху актуальным становится взаимодействие общелитературного и специального, профессионального языков. Необходимость понимания, адекватной интерпретации юридических документов побуждает не-юристов обращаться за информацией к специальным источникам, одними из которых являются справочно-поисковая система «Консультант Плюс» (далее – КП) [1] и специальные словари, размещенные в сети Интернет [2, 3].

---

© Крылов С.А., Фролова О.Е., 2017

<sup>1</sup> Исследование выполнено при поддержке РФФИ, грант 14-29-05028 «Исследование методов и разработка моделей и средств оценки научных текстов на основе их когнитивных структур».

Соответственно, предметом изучения лингвиста в подобной социокультурной ситуации становятся различия в семантике между словами как терминами (словами как членами терминосистем) и словами общелитературного языка. В описании нуждаются следующие расхождения в специальном языке законодательства и общелитературном языке: а) несовпадение родовых и видовых понятий (*правонарушение* и *преступление*); б) нетождественность интерпретации семантики слова общелитературного языка, имеющего статус юридического термина (*оскорбление*); в) совпадение неактуальной правовой номинации правонарушителя и употребление слова в общелитературном языке (*спекулянт*); г) несовпадение объема понятия в общелитературном и специальном языке (*иждивенец*); д) несовпадение коннотации слова общелитературного языка и его нейтрального статуса в юридической терминосистеме (*иждивенец, инвалид, пенсионер*); е) непрозрачность ситуации преступления в валентностях юридических терминов в общелитературном языке (*коррупцированный*).

В современном НКРЯ представлены несколько подкорпусов: основной, синтаксический, газетный, параллельный, обучающий, диалектный, поэтический, устный, акцентологический, мультимедийный, исторический, включающий старорусский [4]. Как видим, в НКРЯ отсутствует официально-деловой подкорпус, который включал бы в себя законодательные документы и позволял наблюдать за актуализацией единицы в специальных и неспециальных текстах.

Разрыв, существующий в современном Национальном корпусе русского языка и в юридическом терминологическом узусе, вызывает необходимость обращения к двум источникам: в КП и специальных словарях онлайн исследователь получает толкование семантики единицы и примеры профессионального узуса, в НКРЯ – наблюдает поведение юридического термина в непрофессиональной речи.

### **Библиографический список**

1. *КонсультантПлюс*. URL: <http://www.consultant.ru> (дата обращения: 12.01.2017).
2. Большой юридический словарь. URL: [http://petroleks.ru/dictionaries/dict\\_big\\_law.php](http://petroleks.ru/dictionaries/dict_big_law.php) (дата обращения: 12.01.2017).
3. Юридический словарь. URL: <http://www.klerk.ru/slovar/law/> (дата обращения: 12.01.2017).
4. *Национальный корпус русского языка*. URL: <http://www.ruscorpora.ru/> (дата обращения: 12.01.2017).

## USE OF THE SPECIAL CORPUS FOR LINGUISTIC PURPOSES

*Krylov S.A.*

Institute of Oriental Studies Russian Academy of Sciences,  
12, Rozhdestvenka st., Moscow, Russia, krylov-58@mail.ru

*Frolova O.E.*

Moscow State University, GSP-1, 1-51 Leninskie Gory, Moscow,  
Humanities Building, Russia, olga\_frolova@list.ru

The interaction of general literary and professional, juristical, languages is considered in the current system of functioning of documents in “paper” and electronic form. The monopoly of lawyers as creators and interlopers of law documentation leaves an unprofessional Russian speaker on the periphery of the most important sphere of law communication – communication between the state and the citizen. The need to create texts in the law field in productive and receptive types of speech activity: in written genres of the application, claim, lawsuit, petition, as well as in oral genres of answers to questions in the court session, the need to understand and interpret legislative and legal written documents, the speech of the prosecutor and Lawyer actualize the knowledge of legal terminology by non-jurists. In his attempts to become a participant in law communication, the non-jurist is faced with a discrepancy in the volume of meaning, polysemy and homonymy of the words of the literary language and juristical terms. Special Internet corpora and online legal dictionaries allow to remove the severity of communicative problems between lawyers and non-jurists.

*Key words:* the literary language, professional language, term, terminology, search system “Consultant Plus”, the language of law, the National Corpus of the Russian language, the subcorpus.

УДК 004.9

### МОНИТОРИНГ ГЛОБАЛЬНЫХ ПРОЦЕССОВ НА ОСНОВЕ ДАННЫХ ИЗ ИНТЕРНЕТ-НОВОСТЕЙ

*Ланин В.В., Шаляева И.М., Скурихина А.Ю.*

Национальный исследовательский университет  
«Высшая школа экономики», Россия, Пермь, ул. Студенческая, 38,  
alena.skurikhina@mail.ru

Представлен результат исследовательского проекта, цель которого – объединить онтологический подход к обработке текста и инструменты интеллектуального анализа процессов. Онтологии предметных областей и источников данных используются для поиска новостей в Интернете и последующего извлечения из их текстов фактов. Инструменты Process Mining позволяют находить закономерности, связи между отдельными событиями или типами событий для построения формальных моделей процессов, которые могут быть использованы для последующего анализа экспертами. Результаты апробации

---

© Ланин В.В., Шаляева И.М., Скурихина А.Ю., 2017

показывают практическую значимость подхода: были построены онтологии нескольких предметных областей, извлечённая на их основе информация, обработанная с помощью средств углублённого анализа процессов, позволила выявить существующие между событиями причинно-следственные связи.

*Ключевые слова:* онтология, поиск информации, анализ текста, извлечение фактов, анализ событий, Process Mining.

В данной работе предлагается подход к формированию максимально понятного для человеческого восприятия отображения информации о значимых для пользователя событиях, их последствиях и предпосылках, представленных в новостных сообщениях сети Интернет. Поскольку человек наиболее восприимчив к визуальному представлению данных, в качестве результата предлагаются графические модели процессов, данные о которых представлены в Интернет. Границы процессов определяются областью интересов пользователя.

Выбор в качестве источника данных новостных сообщений обусловлен тем, что любое значимое событие получает очень быстрый отклик в Интернет-новостях. Новостные отчеты предоставляют статус событий практически в реальном времени, описывая все аспекты случившегося на естественном языке, являясь одним из наиболее доступных источников информации о современном обществе. Кроме того, новостные тексты характеризуются наличием большого числа фактографических данных, обеспечивая тем самым базу для существующих методов обработки текстов. Новости не специализируются на определенных классах событий, а значит, могут быть источником данных для любых заданных пользователем областей (например: экология, здравоохранение, экономика, политика и пр.). В связи с чем новости являются неисчерпаемым источником данных для различных систем анализа [1, 2, 3, 4], однако данные работы не фокусируют внимание на возможности формирования и визуализации цепочек взаимосвязанных событий, принадлежащих различным предметным областям (глобальных процессов).

Для пользователей, которым важно анализировать информационную реакцию, причины и последствия определённых событий, проектируемая система является аналитическим инструментом, который обеспечит единую точку доступа к заданным интернет-источникам. Система настраивается на предметные области, определяемые каждым пользователем; она способна накапливать исторические данные и самостоятельно запрашивать актуальную информацию. Система позволит экономить время, затрачиваемое как на поиск интересующих пользователя данных в сети, так и на их анализ. В отличие от действующих сейчас агрегаторов новостей представленный подход предлагает авто-

матически выявлять зависимости между событиями и предоставлять отчёты в форме визуальных моделей процессов.

В качестве структуры для описания и хранения данных, представленных в Интернет в виде текстов на естественном языке, была выбрана многоуровневая онтология, представляющая собой описание источников информации, онтологические словари предметных областей и описание требуемых форматов выходных данных – журналов событий (XES). Данные из Интернет-новостей извлекаются в соответствии с представленными в пользовательских онтологических словарях экземплярами событий и фактов и их лексико-семантическим описанием.

В работе [6] представлена общая схема реализации данного подхода с применением различных средств обработки информации, в том числе инструмента анализа журнала событий Process Mining (ProM). Предварительно вся найденная информация о событиях преобразуется в формат, пригодный для обработки данными средствами. Результатом применения ProM являются графические модели процессов в удобной для анализа форме.

### **Библиографический список**

1. *Chaney D.* Text mining newspapers and news content: new trends and research methodologies // 79th IFLA General Conference and Assembly. Singapore, 2013.
2. *Bing Z., Vogel S.* Adaptive parallel sentences mining from web bilingual news collection // In 2002 IEEE International Conference on Data Mining. Maebashi City, Japan, 2002. P. 745–748.
3. *Mittermayer M.-A., Knolmayer G.* Text Mining Systems for Predicting the Market Response to News: A Survey // Working Paper No. 184, Institute of Information Systems, Univ. of Bern. Bern, 2006.
4. *Bautin M., Vijayarenu L., Skiema S.* International sentiment analysis for news and blogs // In Second Int. Conf. on Weblogs and Social Media (ICWSM 2008). 2008.
5. *van der Aalst, W. M. P., Adriansyah A., de Medeiros A.K.* Process Mining Manifesto // BPM 2011 Workshops, Part I. T. 99. Springer-Verlag, 2012. P. 169–194.
6. *Шалыева И.М., Ланин В.В., Лядова Л.Н.* О проекте разработки системы мониторинга глобальных процессов на основе Интернет-новостей // Сб. тр. VII Межд. науч.-техн. конф. «Технологии разработки информационных систем» (ТРИС-2016). Таганрог. 2016. С.166–170.

### **GLOBAL PROCESS MONITORING ON THE BASIS OF INTERNET NEWS**

*Lanin V.V., Shalyaeva I.M., Skurikhina A.Y.*

National Research University Higher School of Economics,  
38, Studencheskaia st., Perm, Russia, alena.skurikhina@mail.ru

The result of the research project, which is aimed to combine ontology information retrieval technology and process mining tools, is presented. The domain ontologies and ontologies of data sources are used to search news in the Internet and to extract

facts from texts of news. Process Mining tools allows finding regularities, relations between single events or event types to construct formal models of processes which can be used for the next ensuing analysis by experts. An applicability of the approach is studied with example of some domains: ontologies were described, information, extracted from the Internet and handled with Process Mining tools, allowed to identify relationships between events.

*Key words:* ontology, information retrieval, text mining, fact extraction, event analysis, Process Mining.

УДК 81'33

## **ПРИМЕНЕНИЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ ДЛЯ АНАЛИЗА ПОЭТИЧЕСКИХ И ФОЛЬКЛОРНЫХ ТЕКСТОВ**

*Москин Н.Д., Лебедев А.А., Варфоломеев А.Г.*

Петрозаводский государственный университет,  
Россия, г. Петрозаводск, просп. Ленина, 33, moskin@petsu.ru

В работе представлены результаты факторного анализа структуры текстов авторских стихотворений и фольклорных песен (на примере стихотворений Н.А. Клюева и бесёдных песен Заонежья XIX – начала XX вв.), выполненного с применением теоретико-графовых моделей. На основе этих моделей были вычислены числовые характеристики, которые были обработаны с помощью процедуры факторного анализа (метод главных компонент). В результате исследования обеих коллекций выявились две основные компоненты. Их собственные значения, составляющие больше единицы, в сумме объясняют более 75 % дисперсии. В обоих случаях первая компонента связана с размерностью графа (и соответственно текста). Анализ второй компоненты показывает, что веса признаков для разных коллекций принимают сопоставимые значения, но с противоположными знаками, что говорит о возможности дифференциации текстов по данному фактору.

*Ключевые слова:* теоретико-графовая модель, стихотворения, Н.А. Клюев, фольклорные песни, метод главных компонент.

Фольклорная и литературная традиции тесно связаны между собой; многие авторы, как поэты, так и прозаики, искали вдохновение непосредственно в фольклоре, стремясь привнести в собственные произведения и образы, характерные для народных произведений, и мудрость народа как таковую. При этом проблема, связанная с разграничением фольклорных текстов и литературных произведений, по-прежнему остается не до конца решенной, когда речь идет об анализе текста стихотворного произведения и событий, которые в этом тексте представлены.

Для исследования были выбраны бесёдные песни Заонежья XIX – начала XX вв. (50 текстов) и стихотворения Н.А. Клюева (30 тек-

стов), в чьем творчестве «народные» элементы и близость к русской культуре традиционно привлекали внимание как читателей, так и исследователей стихотворного текста. Клюева называли народным поэтом и «провозвестником народной культуры», «духом добра и красоты, исходящим от потаенной народной России» [1, с. 12–13]. В поэзии Клюева сближается фольклорная и литературная традиции – проявляется это в особой образности, смысловой насыщенности стихотворений автора, а также в использовании Н.А. Клюевым слов и выражений, присущих живой народной речи.

В ходе исследования были построены теоретико-графовые модели текстов, отражающие их семантическую структуру (описание моделей приведено в [2]). Отметим, что ранее авторами (на примере стихотворений П.А. Вяземского, И. Бродского, Э. По и др.) были изучены и другие виды теоретико-графовых моделей, которые отражают синтаксическую структуру текстов [3, 4]. Далее тексты загружались в информационную систему «Фольклор» [5], с помощью которой были вычислены следующие числовые характеристики:

- 1) число слов в тексте (*word*);
- 2) число строк в тексте (*string*);
- 3) число вершин в графе (*vertex*);
- 4) число ребер в графе (*edge*);
- 5) максимальная степень вершины графа (*max*) – это максимальное количество ребер, инцидентных вершине (соединенных с вершиной);

- 6) параметр связности графа (*link*). Он находится как отношение числа пар вершин, связанных между собой, к числу ребер в соответствующем полном графе с таким же количеством вершин [2, с. 58];

- 7) коэффициенты гиперболической регрессии *a* и *b*. Объекты в фольклорных песнях и стихотворениях Н. Клюева неравнозначны между собой. В сюжете часто доминируют два основных персонажа (вершины с максимальной степенью), а остальные объекты являются второстепенными, они встречаются в тексте реже, большинство одна–два раза. Если сопоставить каждой вершине графа ее степень и отсортировать их в порядке убывания, то получатся графики, которые можно интерполировать гиперболической кривой вида  $y=a/x+b$ , предварительно пронормировав исходный массив [2, с. 64];

- 8) процент глобальных связей (*global*), которые отражают парадигматические связи в тексте, от общего числа связей.

Далее с помощью статистического пакета SPSS была выполнена процедура факторного анализа (методом главных компонент) [6]. Задача факторного анализа – перейти от множества исходных признаков к существенно меньшему числу новых переменных (факторов), которые интерпретируются как причина структурной взаимосвязи не-

скольких исходных признаков. В результате исследования стихотворений Н. А. Клюева выявились два фактора (компоненты). Их собственные значения составляют больше единицы и в сумме объясняют 78,313 % дисперсии. После применения метода вращения «варимакс» с нормализацией Кайзера была получена следующая таблица.

Таблица 1. Матрица повернутых компонент стихотворения Н.А.Клюева

	Компонента	
	1	2
<i>vertex</i>	0,944	-0,114
<i>edge</i>	0,941	0,048
<i>word</i>	0,877	0,27
<i>link</i>	-0,874	0,218
<i>string</i>	0,819	0,244
<i>a</i>	-0,717	0,677
<i>b</i>	0,233	-0,92
<i>max</i>	0,179	0,862
<i>global</i>	-0,109	-0,464

Как видно из таблицы, первую компоненту можно условно назвать «размерность графа». Чем выше ее уровень, тем больше число вершин (вес 0,944), число ребер (вес 0,941), число слов (вес 0,877) и число строк (вес 0,819) в тексте. При этом уменьшаются значения параметра связности (вес -0,874) и параметра *a* (вес -0,717). Согласно значительному факторному весу вторую компоненту можно назвать «максимальная степень вершины графа» (вес 0,862). Чем выше ее уровень, тем меньше значения параметра *b* (вес -0,92) и параметра *global* (вес -0,464). С другой стороны, увеличивается значение параметра *a* (вес 0,677).

В результате исследования фольклорных песен также выявились две компоненты. Их собственные значения составляют больше единицы и в сумме они объясняют 75,397 % дисперсии. После применения аналогичного метода была получена следующая таблица.

Таблица 2. Матрица повернутых компонент бесёдные песни Заонежья

	Компонента	
	1	2
<i>edge</i>	0,918	0,251
<i>word</i>	0,917	0,068
<i>string</i>	0,897	0,064
<i>vertex</i>	0,864	0,363
<i>max</i>	0,788	-0,38
<i>link</i>	-0,72	-0,185

	Компонента	
	1	2
<i>a</i>	-0,327	-0,924
<i>b</i>	-0,254	0,884
<i>global</i>	0,19	0,43

Как видно из второй таблицы, первая компонента также связана с размерностью графа. Чем выше ее уровень, тем больше число вершин (вес 0,864), число ребер (вес 0,918), число слов (вес 0,917), число строк (вес 0,897), а также максимальная степень вершины (вес 0,788). При этом уменьшается значение параметра связности (вес -0,72). Что касается второй компоненты, то чем выше ее уровень, тем меньше значения параметра *a* (вес -0,924). С другой стороны, увеличиваются значения параметра *b* (вес 0,884) и параметра *global* (вес 0,43).

Таким образом, в обоих случаях первая компонента связана с размерностью графа (и соответственно текста). Анализ второй компоненты показывает, что веса признаков *a*, *b*, *max*, *link* и *global* для разных коллекций принимают сопоставимые значения, но с противоположными знаками (что отчетливо видно на графике компонент). Это говорит о возможности дифференциации текстов по данному фактору.

### Библиографический список

1. Ключев Н.А. Сердце единорога. Стихотворения и поэмы. СПб.: Изд-во РХГИ, 1999.
2. Москин Н.Д. Теоретико-графовые модели фольклорных текстов и методы их анализа. Петрозаводск: Изд-во ПетрГУ, 2013.
3. Варфоломеев А.Г., Кузнецов Д.В., Лебедев А.А., Москин Н.Д. Применение теоретико-графовых моделей для анализа переводов англоязычных поэтических текстов // Филологические науки. Вопросы теории и практики. Тамбов, 2014. № 12. (42). Ч. 1. С. 42–47.
4. Кузнецов Д.В., Лебедев А.А., Москин Н.Д., Варфоломеев А.Г. Теоретико-графовые модели с упорядоченной иерархической структурой и их использование в анализе синтаксиса поэтических текстов // Ученые записки Петрозаводского государственного университета. Серия «Естественные и технические науки». Петрозаводск, 2013. №6 (135). С. 113–118.
5. Москин Н.Д. Инструменты исследования текстовых коллекций на основе теоретико-графовых моделей в информационной системе «Фольклор» // Прикладная информатика. М., 2010. № 4 (28). С. 48–62.
6. Наследов А.Д. SPSS 19: профессиональный статистический анализ данных. СПб.: Питер, 2011.

## **APPLICATION OF THE PRINCIPAL COMPONENTS METHOD FOR THE ANALYSIS OF POETIC AND FOLK TEXTS**

*Moskin N. D., Lebedev A. A., Varfolomeyev A. G.*

Petrozavodsk State University, 33, Lenina ave.,  
Petrozavodsk, Russia, moskin@petsu.ru

The paper presents the results of factor analysis of author's poems and folk songs structure (on the example of N.A. Klyuev's poems and Karelia folk songs of the XIX – early XX century), made with the use of graph-theoretic models. Several numerical characteristics of graphs were calculated, after that they were processed using the procedure of factor analysis (principal components method). The analysis of both collections revealed two main components. Their eigenvalues are greater than 1 and together they explain more than 75 % of the variance. In both cases the first component is associated with the dimension of the graph (and the text). The analysis of the second component shows that the characteristics weight for different collections take comparable values, but with opposite the signs. This suggests the possibility of the texts differentiation by means of that factor

*Key words:* graph-theoretic model, poems, N.A. Klyuev, folk songs, principal components method.

УДК 81'27

## **ГРАФОСЕМАНТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННО-РАСПРЕДЕЛЕННЫХ ЭРГОНИМОВ**

Нитко Я.М.

Пермский государственный национальный исследовательский  
университет, Россия, Пермь, ул. Букирева, 15,  
renee1618harmony@mail.ru

Статья посвящена моделированию семантического пространства эргонимов, распределенных в локализованном городском топосе. Впервые строится семантическая модель эргонимикона локализованного городского пространства в «привязке» к географическим координатам городских объектов, номинируемых анализируемыми эргонимами. Семантизация фрагментов городского пространства позволяет рассматривать образ адресата (представителя целевой аудитории), характерного для того или иного городского локуса. В работе высказывается предположение о том, что разные фрагменты городского пространства могут быть «настроены» на разные типы потребителей. В результате исследования была сформирована база наименований языковой среды городского локуса. Были выявлены наиболее значимые особенности в гендерном соотношении, социокультурные и экономические особенности, переходное состояние самоиндификации российского общества на региональном и национальном уровнях, отсутствие потенциала у русской мифологии к коммерциализации. Выявленные особенностей позволяет говорить о специфике типа потребителя для данного локализованного городского топоса: преобладание формирующихся черт «западного» типа над «региональным» и «национально-ориентированным» типами, что, в свою очередь, дает возможность раскрыть

закономерные особенности создания ономастикона современного города и прогнозировать формирование его имиджа, регулировать процесс экономического и культурного развития современного общества.

*Ключевые слова:* эргоним, эргонимикон, графосемантическое моделирование, семантический граф, семантическая карта города, тип потребителя.

Данное исследование посвящено изучению семантики и прагматики городского пространства, создаваемого эргонимами. Основное отличие от аналогичных исследований состоит в рассмотрении семантики эргонимов в зависимости от их географического расположения. Впервые строится семантическая модель эргонимикона локализованного городского пространства в «привязке» к географическим координатам городских объектов, номинируемых анализируемыми эргонимами. Семантизация фрагментов городского пространства позволяет рассматривать образ адресата (представителя целевой аудитории), характерного для того или иного городского локуса. В работе высказывается предположение о том, что разные фрагменты городского пространства могут быть «настроены» на разные типы потребителей.

**Материалом** для исследования послужили 137 эргонимов, обозначающих организации сферы услуг и государственные учреждения, располагающиеся на улице Сибирской г. Перми (состояние на 2013 г.).

**Метод анализа:** графосемантическое моделирование [1; 2]. Графосемантическое моделирование (ГСМ) представляет собой метод графической экспликации структурных связей между семантическими компонентами одного множества. Метод ГСМ позволяет представить набор данных (выборку и/или целостность) в виде системы, в которой каждый из компонентов имеет иерархическую и топологическую определённость по отношению к другим компонентам и всей системе в целом. Эта структурная контекстуальность, в свою очередь, дает возможность интерпретировать каждый компонент системы в отношении к возможным причинам его появления и вариантам его дальнейшего развития.

Данный метод включает в себя следующие этапы:

- 1) проведение компонентного анализа выделенных контекстов;
- 2) проведение полевого анализа (на основании выделенных на первом этапе компонентов);
- 3) обнаружение основных связей между полями в пределах всей выборки;
- 4) графическая экспликация результатов анализа;
- 5) интерпретация полученной модели.

В данном исследовании топологическая связь эргонима с его адресом была зафиксирована в метаданных каждого контекста, что

позволяет представить общую картину особенностей семантизации исследуемого городского пространства.

Определяющим для заключения выводов исследования является интерпретация созданных графических моделей – семантического графа и фрагмента семантической карты города.

Наиболее значимые особенности были выявлены в гендерном соотношении: тяготение к социокультурному пространству западных стран, усиленное стремление к интеграции с западной экономикой, переходное состояние самоидентификации российского общества на региональном уровне и национально-этническом уровне, отсутствие потенциала у русской мифологии к коммерциализации, что, вероятно, является закономерным явлением в стране на современном этапе ее развития.

Выявленные особенности позволяют говорить о специфике типа потребителя для данного локализованного городского топоса: преобладание формирующихся черт «западного» типа над «региональным» и «национально-ориентированным» типами. Отмечается формирование черт потребителя, ориентированных на культуру данного региона («регионального» типа потребителя). Для организации экономического пространства локализованного топоса характерны ориентация на личные имена и обращение к образу пространства, что предполагает возможность самоидентификации для адресата и формирования особого типа потребителя.

### **Библиографический список**

1. Система графосемантического моделирования [электр. издание] / Д.А. Баранов, К.И. Белоусов, И.В. Влацкая, Н.Л. Зелянская. Свидетельство о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011. М., 2011.
2. Белоусов К.И., Зелянская Н.Л. Применение метода графосемантического моделирования в лингвомаркетологических исследованиях // Вестник Оренбургского государственного университета. 2005. № 8. С. 40–46.

### **GRAPHOSEMANTIC MODELING OF SPATIALLY DISTRIBUTED ERGONYMS**

*Nitko Y.M*

Perm State University, 15, Bukireva st., Perm, Russia,  
renee1618harmony@mail.ru

This article is devoted to modeling the semantic space of ergonyms distributed in localized urban topos. The semantic model of ergonymikon of localized urban space is created in "binding" to the geographical coordinates of urban objects, which are nominated by the analyzed ergonyms. The semantization of urban space fragments enables to consider the image of the typical for a particular urban locus addressee

(the target audience). It is assumed, that different urban space fragments can be "tuned" to different types of consumers. As a result of the study a collection of urban locus language environment was formed. Besides, the most important features in gender correlation, socio-cultural and economic features, increased aspiration for integration with western economies, transitional condition of the Russian society self-identification at the regional and national-ethnic level, deficiency of Russian mythology potential to commercialization have been identified. These peculiarities permit to speak about the specific type of the customer for the localized urban topos: the predominance of the "western" type features over "regional" and "national-oriented" types. This makes it possible to discover natural specific features of modern urban onomasticon, to predict formation of its image and to regulate the process of economic and cultural development of modern society.

*Key words:* ergonym, ergonymikon, graphosemantic modeling, semantic graph, semantic map of the city, type of consumer.

УДК 801.6

## **ОБЪЯСНИМЫ ЛИ ОШИБКИ В.А. ЖУКОВСКОГО С ПОМОЩЬЮ АНАЛИЗА ДАННЫХ?<sup>1</sup>**

*Орехов Б.В.*

Национальный исследовательский университет «Высшая школа экономики», Россия, Москва, ул. Мясницкая, 20,  
nevmenandr@gmail.com

В статье речь идёт о метрических отступлениях, которые мы наблюдаем в переводе «Одиссеи», выполненном В.А. Жуковским. Русский гекзаметр, конвенционально принятый для передачи античного метра, подразумевает обязательные 6 ударений в строке (соответствующих 6 долгим слогам оригинала). Но в 49 случаях (на 12 с лишним строк поэмы) Жуковский ошибается и делает 5 или 7 ударений в строке. В статье изложены результаты поиска закономерностей между этими ошибками и другими параметрами текста: сюжетом, лексикой, некоторыми особенностями синтаксиса. Общий вывод неоднозначен. С одной стороны, мы действительно можем установить закономерности между песнями, в которых есть такие ошибки, и общим движением сюжета. Песни с ошибками приходятся на ту часть поэмы, в которой Телемах находится в отъезде с Итаки. С другой стороны, более точной обусловленности, которая показывала бы, чем именно строка с ошибкой отличается от всех остальных строк, не обнаруживается.

*Ключевые слова:* поэтический перевод, В.А. Жуковский, анализ данных, метрические аномалии, русский гекзаметр.

---

© Орехов Б.В., 2017

<sup>1</sup> В данной научной работе использованы результаты проекта «Точки роста в языковой системе: корпусные исследования и моделирование», выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2017 году.

В представленном В.А. Жуковским варианте перевода поэмы Гомера есть 49 случаев особого рода, которые мы будем далее называть метрическими аномалиями. Речь идет в основном о тех же дактилических строках, длина которых, однако, отличается от гекзаметра на одну стопу как в сторону увеличения, так и в сторону уменьшения:

Многих он в помощь себе приведёт из песчаного Пилоса, многих (II, 326)

Как видно, в этих примерах Жуковский использует семистопный дактиль вместо привычного (и требуемого) шестистопного. Таких строк во всем переводе насчитывается 30.

В этой земле не знаком мне; скажи, где дорога (VI, 177)

Здесь строка оказывается на стопу короче, чем остальные, меньше, они исчерпываются 19 стихами.

Можно ли выявить внутритекстовые причины метрических аномалий, используя методы анализа данных? Даже не перечисляя конкретные методы, следует сказать, что это серьезный вызов. Анализ данных служит для обнаружения тенденций, характерных для больших массивов информации. В нашем случае ситуация обратная: мы пытаемся выявить механизм появления не тенденций, а исключений.

Метрические аномалии распределены по песням неравномерно. Больше всего их (по 6) в IX и XII песнях (обе переведены в 1844 г.). В песнях I, III, XVI, XIX–XXIV аномалий нет совсем. При этом песни с наибольшим числом аномалий представляют нам аномалии разного типа. В упоминавшихся IX и XII песнях число в 6 аномалий складывается из 3 семистопных и 3 пятистопных строк. В то же время V песнь, насчитывающая 5 аномалий, имеет совсем другое соотношение их типов: здесь 4 семистопные строки и только 1 пятистопная.

Как кажется, оперирование контекстом песни предполагает уместность подхода, при котором мы разделили все песни на две неравные группы. В первую попали те, в которых встречаются метрически аномальные строки, во второй оказались те, в которых таких строк не наблюдается. Таким образом, в первой группе будут II, IV–XV, XVII, XVIII песни; во второй – I, III, XVI, XIX–XXIV. Легко заметить, что первая группа захватывает, в основном, середину поэмы, а во второй оказываются песни «сюжетной рамки» начала и финала гомеровского текста.

Сначала попытаемся провести анализ с помощью традиционно герменевтического инструментария. Существуют ли видимые глазу отличия песен первой группы от песен второй группы на каком-либо из возможных уровней (прежде всего, конечно, сюжетном)?

В принципе, наиболее значимый сюжетный фактор, который делит весь текст «Одиссеи» на две части, – это состояние главного героя: половину повествования он находится на суше (Огигия, Итака), а

вторую половину – в морском путешествии (реальном или данном через нарратив). При этом рассказ о путешествиях Одиссея приходится именно на середину поэмы. Однако Одиссей покидает Огигию в середине V песни (метрические аномалии появляются уже во II и в IV) и прибывает на Итаку в XIII, в то время как метрические аномалии встречаются и в XIV, XV, XVII и XVIII песнях (хотя и единичны в двух последних). Таким образом, 6 песен из 15 нарушают предложенную модель распределения «аномальных» и «неаномальных» песен поэмы.

Существует несколько метрик, позволяющих оценить тематическую близость текстов. В этом исследовании мы будем использовать самую распространенную, построенную на основе вычисления TF-IDF. После того, как мы произвели все нужные подсчеты, мы приступили к собственно автоматической классификации, инструментом которой выступил самый распространенный алгоритм кластеризации, т. е. разбиения имеющегося множества объектов на группы, а именно алгоритм k-средних. Мы осуществили кластеризацию. В первый кластер попали песни, в которых мы ожидали бы увидеть метрические аномалии, во второй — песни, в которых аномалий в идеальном случае не было бы. Если посчитать меру согласия (каппа Коэна), то можно сделать следующий вывод: опыт кластеризации на основе распределения частотности слов в песнях (которое в конечном счете имеет в виду автоматическое определение тематики текста) показал, что кластеры хотя и не вполне строго, но в целом соответствуют делению на песни с метрическими аномалиями и без них.

Для анализа строк был выбран ряд текстовых характеристик, которые можно легко определить автоматически: 1) позиция в песне (влияет ли то, в начале, в середине или в конце песни находится аномальная строка; все песни разделены на 5 частей); 2) количество слов в строке; 3) набор ударных гласных в строке и т. д.

Лучший результат, который показал анализ атрибутов, нельзя назвать удовлетворительным, его мера качества оценивается алгоритмом в 0,05481. Эта цифра означает, что если потенциально и существует какой-то фактор, на 100 % объясняющий появление метрических аномалий, то выявленный набор атрибутов справляется с этой задачей только на 5,5 %. Фактически таким результатом можно пренебречь.

## **CAN WE EXPLAIN V.A. ZHUKOVSKY'S MISTAKES BY MEANS OF DATA ANALYSIS?**

*Orekhov B. V.*

National Research University Higher School of Economics, 20,  
Myasnitckaya st., Moscow, Russia, nevmenandr@gmail.com

In this article we are talking about metric deviations that we observe in the translation of the "Odyssey" made by V.A. Zhukovsky. Russian hexameter, the standard

adopted for transmission antique meter, implies 6 accents in the string (simply speaking, the respective 6 long syllables). But in 49 cases (more than 12 lines of a poem) Zhukovsky is wrong and makes 5 or 7 beats in a line. The article presents the results of a search of regularities between these errors and other settings of the text: plot, language, some peculiarities of the syntax. The overall conclusion is ambiguous. On the one hand, we can establish patterns between songs that have such errors, and the overall movement of the plot. Songs have mistakes on the part of the poem, in which Telemachus is departing from Ithaca. On the other hand, the more accurate determination, that would show what makes the line with an error different from all other lines, is not detected.

*Key words:* Poetic translations, Zhukovsky, data analysis, metric anomalies, Russian hexameter.

УДК 81

## **К ВОПРОСУ О ПРИНЦИПАХ ПОДАЧИ В ЭЛЕКТРОННОМ СЛОВАРЕ НЕОДНОСЛОВНЫХ ЕДИНИЦ, НАЗЫВАЮЩИХ МИФОЛОГИЧЕСКИХ ПЕРСОНАЖЕЙ<sup>1</sup>**

*Русинова И.И., Гранова М.А.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, irusinova@mail.ru

В работе обосновывается актуальность создания электронного «Словаря мифологических рассказов Пермского края». Наибольшую сложность при лексикографировании представляют однословные номинации духов-«хозяев» различных локусов (домового, дворового, банника, лешего и под.). Авторами приводится тематическая классификация таких единиц и ставится вопрос о принципах их подачи в указанном словаре. Анализ исследовательских работ и существующих словарей демонологической лексики позволяет сделать вывод об отсутствии четких принципов теоретического рассмотрения однословных номинаций мифологических персонажей и их словарного описания в современной науке. По нашему мнению, такие единицы должны включаться в словарь. Критериями для их отбора являются, во-первых, регулярная встречаемость во множестве текстов, а во-вторых, отражение в их семантике релевантных признаков персонажа.

*Ключевые слова:* мифологические рассказы Пермского края, демонологическая лексика, электронный словарь.

Фольклористами и этнолингвистами пермских вузов собран обширный корпус мифологических текстов. Оптимальной формой систематизации лексики, отражающей мифологические представления, может стать словарь.

---

© Русинова И.И., Гранова М.А., 2017

<sup>1</sup> Публикация подготовлена при финансовой поддержке РФФИ (грант № 17-14-59009а(р)) и гранта, выполняемого в рамках тематических планов по заданию Министерства образования РФ, № 34.1505.2017/ПЧ.

Однако в условиях развития информационного общества бумажные словари перестают соответствовать современным запросам. К «недостаткам» печатных словарей можно отнести следующее: относительно небольшой объем описываемого материала, невозможность регулярного обновления словарных данных, доступность относительно узкому кругу читателей. Решить указанные проблемы способно создание электронных словарей, которые за счет гипертекстовой организации могут включать неограниченный объем данных и постоянно пополняться. Обеспечить доступность такого словаря для пользователей из любой точки мира помогает Интернет.

Учитывая сказанное, нам представляется актуальным создание электронного «Словаря мифологических рассказов Пермского края».

Материалом для него станут полевые экспедиционные записи, собранные в 1980–2010-х гг. студентами и преподавателями ПГНИУ, и тексты, опубликованные в сборниках сказочной прозы Пермского края.

Одной из групп лексики, которая войдет в электронный словарь, станут номинации духов-«хозяев» различных локусов (домового, дворового, банника, овинника, водяного, русалки, болотного, лешего, полудницы).

В пермских говорах встречаются как однословные (*лесной, лешак, суседко*), так и неоднословные номинации духов-«хозяев». В рамках настоящего исследования мы рассмотрим вопрос о лексикографировании номинативных сочетаний.

В пермских материалах находим широкий спектр таких единиц. Это:

1) номинации по месту обитания (*лесная баба, лесной дьявол, лесной человек*);

2) номинации, указывающие на атрибуты персонажа (*дедушко с кузовом, дядька с грибами*);

3) номинации, указывающие на одежду персонажа (*маленькая бабка в сарафане и рубаше*);

4) номинации, включающие цветообозначения (*баба с белыми волосами, в белом платье в горошек; чёрный дядька; красенький телёночек*);

5) номинации, указывающие на рост/размер персонажа (*здоровая баба; мужичок с метр*).

Возникает вопрос: как представлять подобные номинации в электронном словаре? Здесь возможны два варианта:

а) считать неоднословное сочетание целостной номинативной единицей и подавать как вокабулу самостоятельной словарной статьи;

б) считать названием лишь главное слово в сочетании, а обозначения характеристик и атрибутов (типа *необъятный размер* или *крас-*

ные сапоги) подавать отдельно от номинации в разделах «Характеристики» и «Атрибуты» соответственно.

Рассмотрением неоднословных номинаций мифологических персонажей в теоретическом аспекте занимались Е.Е. Левкиевская и О.И. Васнева.

Е.Е. Левкиевская называет сочетания, включающие указание на внешний облик персонажа, актуальными именами, которые являются одним из типов референции объекта при коммуникации [1]. Таким образом, по ее мнению, эти единицы имеют не языковую, а речевую природу. Следовательно, они не должны быть объектом описания в словаре.

С другой стороны, в этой же работе указывается, что, если такие сочетания выступают как постоянное обозначение данного мифологического явления во множестве текстов в рамках какой-либо традиции, они конвенционализируются, превращаются в имя персонажа. Следовательно, они должны входить в словарь.

О.И. Васнева представляет неоднословные сочетания как тип номинаций мифологического персонажа [2]. Она считает такие сочетания языковыми единицами, и поэтому их необходимо включать в словарь.

Анализ исследований по славянской мифологии и русских словарей демонологической лексики [3, 4] позволяет сделать вывод об отсутствии четких принципов теоретического рассмотрения неоднословных номинаций мифологических персонажей и их словарного описания.

Мы считаем, что неоднословные сочетания, называющие демонов через их облик и характеристики, должны подаваться в описанном электронном словаре. Но при этом нужно выработать четкие критерии отбора единиц. Для нас значимыми выступают два критерия: 1) единицы должны присутствовать во множестве текстов; 2) их семантика должна отражать релевантные признаки персонажа. Таким образом, в лексикографируемых названиях пермских демонов должны отражаться только те признаки, которые значимы для народной духовной культуры и являются персонажеобразующими: рост выше или ниже человеческого; черный, белый и красный цвета и т. д. Именно такие семиотически значимые признаки переводят «обычную» номинацию в сферу мифологической: например, именно **красные сапоги** позволяют нам опознать *мужика в красных сапогах* как лешего. Поэтому характеристики и атрибуты, на наш взгляд, не могут быть просто «выброшены» из состава номинаций. Однако на практике произвести отбор единиц для лексикографирования довольно сложно, поскольку их структура обладает очень широкой вариативностью.

## Библиографический список

1. *Левкиевская Е.Е.* Восточнославянский мифологический текст: семантика, диалектология, прагматика: дис. ... д-ра филол. наук. М., 2007. 634 с.
2. *Васнева О.И.* Типы и способы номинации демонологической лексики (на материале пермских говоров) // Филологические науки. Вопросы теории и практики. 2013. № 4 (22), ч. 1. С. 37–39.
3. *Новичкова Т.А.* Русский демонологический словарь. СПб.: Петербургский писатель, 1995. 640 с.
4. *Власова М.Н.* Новая АБЕВЕГА русских суеверий. СПб.: Северо-Запад, 1995. 383 с.

## ABOUT THE PRINCIPLES OF PRESENTATION THE MYTHOLOGICAL CHARACTERS MULTIWORD NOMINATIONS IN THE ELECTRONIC DICTIONARY

*Rusinova I.I., Granova M.A.*

Perm State University, 15, Bukireva st., Perm, Russia, irusinova@mail.ru

The authors substantiate the relevance of the electronic “Dictionary of the Perm Krai mythological stories” creation. The greatest lexicographical difficulty is the multiword nominations of the spirits-“owners” of the various loci (brownie, yard spirit, bath spirit, goblin, etc.). In the article a thematic classification of these units is provided and the question of the principles of these units presentation in the dictionary is raised. Analysis of the research papers and the existing dictionaries of the demonological vocabulary allow drawing a conclusion about the absence of clear principles of theoretical consideration of the mythological characters multiword nominations and their dictionary descriptions in modern science. In our opinion, such units should be included in the dictionary. Criteria for their selection are, firstly, a regular occurrence in many texts, and secondly, reflection of the character’s relevant features in their semantics.

*Key words:* mythological stories of the Perm Krai, demonological vocabulary, electronic dictionary.

УДК 005

## ОНТОЛОГИЯ МАРКЕРОВ АКАДЕМИЧЕСКОГО СТИЛЯ<sup>1</sup>

*Филипсон С.К., Ланин В.В.*

Национальный исследовательский университет «Высшая школа экономики», Россия, Пермь, ул. Студенческая, 38, vlanin@hse.ru

В докладе описывается подход к разработке онтологии, предназначенной для систематизации стилевых маркеров академического английского языка. Онто-

---

© Филипсон С.К., Ланин В.В., 2017

<sup>1</sup> Статья подготовлена в ходе проведения исследования № 17-05-0020 в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2017 г. и в рамках государственной поддержки ведущих университетов Российской Федерации "5-100".

логия содержит описание общелингвистических понятий, на основе которых описываются конкретные стиливые конструкции. Представление онтологии в человекочитаемом и машиночитаемом форматах позволяет использовать ее как в процессе изучения предметной области, так и для автоматической обработки языкового материала.

*Ключевые слова:* маркер стиля, научная публикация, онтология.

Публикация результатов исследований в научных изданиях является одним из существенных показателей результативности ученых и исследовательских коллективов. Английский язык значительно расширяет аудиторию читателей, но для авторов, не являющихся носителями языка, подготовка таких материалов довольно часто вызывает сложность, т. к. текст должен удовлетворять нормам «академического английского». На данный момент экспертами разработан значительный объем методических материалов по академическому письму, проводятся специализированные учебные курсы. Анализ литературы по данной тематике показал, что приведенные рекомендации не систематизированы, а в некоторых случаях даже имеют явные внутренние противоречия. Следует также учесть, что многие из изданий имеют свои стилистические «традиции», которые необходимо учитывать при подготовке материалов. В настоящее время ведутся исследования текстов с применением компьютерных технологий, благодаря которым доступна обработка объемных корпусов текстов. Корпусные исследования дают эмпирический материал, на основании которого могут быть построены эталонные языковые модели и исследованы языковые закономерности, описаны языковые явления, свойственные той или иной области языка, т. е. выявлены так называемые «маркеры стиля».

На протяжении нескольких лет сотрудниками и студентами департамента иностранных языков и кафедры информационных технологий в бизнесе Пермского филиала «Высшей школы экономики» проводятся исследования стиля научной речи на английском языке с помощью специализированного программного обеспечения [1]. К настоящему времени выделенные маркеры стиля представлены в виде несистематизированного набора. Появилась острая необходимость систематизации маркеров стиля, притом используемый способ должен предусматривать возможности расширения и адаптации, так как язык является динамической и постоянно развивающейся системой. В качестве способа классификации набора маркеров было решено создать онтологический ресурс [2, 3], описывающий основные понятия лингвистики и выявленные маркеры стиля. Онтология не только позволит решить широкий спектр задач, но и сможет стать отличным инструментом для исследователя, изучающего академический английский, и основой автоматизированного поиска маркеров в тексте.

## Библиографический список

1. *Strinyuk S. A., Shuchalova Y., Lanin V.* Academic Papers Evaluation Software, in: Application of Information and Communication Technologies (AICT), 2015 9th International Conference on, 14–16 Oct. 2015. Rostov-on-Don: IEEE, 2015. P. 506–510.
2. *Загоруйко Ю.А.* Построение порталов научных знаний на основе онтологий // Вычислительные технологии. 2007. Т. 12. Спецвып. 2. С. 169–177.
3. *Kent D. Bimson, Richard D. Hull* Unnatural Language Processing: Characterizing the Challenges in Translating Natural Language Semantics into Ontology Semantics // Semantic Web 2015. P. 119–135.

## ACADEMIC STYLE MARKER ONTOLOGY

*Philipson .K., Lanin V.V.*

National Research University Higher School of Economics, 38,  
Studencheskaya st., Perm, Russia, vlanin@hse.ru

The report describes an approach to an ontology development. The ontology is intended to organize academic English style markers. Ontology contains two main parts: the first part is general linguistic concepts description and the second part is specific style constructions. The ontology representation in human readable and machine readable formats allows to use it as a data domain study recourse and as a basis for automatic language processing.

*Key words:* style marker, scientific paper, ontology.

УДК 81'33

## МАШИННОЕ ОБУЧЕНИЕ В СОЦИОЛИНГВИСТИЧЕСКИХ ЗАДАЧАХ

*Худякова Е.С.*

Пермский государственный национальный исследовательский  
университет, Россия, Пермь, ул. Букирева, 15,  
khudiakova.es@gmail.com

Проблемой исследования является оценка эффективности методов машинного обучения при решении социолингвистических задач. Конкретно на основе метода машинного обучения с учителем (в учебной выборке тексты сведены до лексем, каждой автоматически присвоены частеречные характеристики) проведена попытка определения этничности автора текста. Результаты эксперимента показали, что только 1 группа текстов («русские») устойчиво определяются машиной. Тексты «татар» определяются, но с низкой вероятностью, тексты «коми-пермяков» отнесены в результате переобучения то к группе «русских», то «татар». Эффективность атрибуции этничности автора текста объяснима малым количеством параметров обучения, а также лингвистическим эффектом дисперсии языковых средств, обусловленным внутренними свойствами самого языка.

*Ключевые слова:* машинное обучение, социолингвистика, этничность, нейронная сеть, алгоритм обучения с учителем.

Применение математических методов в социолингвистике является стандартом. Однако перепоручение машине задач выявления факторов социального варьирования текстов не слишком частотно. В социолингвистике возможен индуктивный подход к лингвистическому материалу, когда анализ лингвистических единиц приводит к выводам о социальных характеристиках информанта, а возможен дедуктивный подход – когда исследователь идет от известных социальных характеристик информанта и в тексте пытается найти единицы, маркирующие эти характеристики. По-видимому, первый подход хорошо соотносится с задачами классификации документов на основе машинного обучения: существует выборка текстов с потенциально неограниченным набором лингвистических единиц (их отбирает исследователь), и целью исследования является обнаружение зависимости между набором единиц и присваиваемыми текстам классами (социальными характеристиками).

Проблема определения социальных характеристик автора текста (в нашем случае – этничности автора) является более общей по отношению к предметной области атрибуции авторства текста, в рамках которой машинное обучение уже применялось с достаточно устойчивыми результатами [4, с. 82].

В широком смысле машинное обучение (МО) предполагает обнаружение скрытых в массивах данных закономерностей на основе предъявления машине этих размеченных данных [3]. Данную область можно разделить на основе применяемых методов на две группы: МО на основе статистики (кластеризация) [6] и на основе нейронных сетей. Вторая группа успешно применяется в компьютерной лингвистике для кластеризации текстов по теме [1] и по тональности (область *sentiment analysis*) [2].

В данном исследовании ставится задача определения социальных характеристик автора текста (конкретно – этничности: русской, татарской и коми-пермяцкой) с применением метода машинного обучения с учителем.

В учебную выборку вошло 30 текстов на русском языке, по 10 текстов от русских, татарских и коми-пермяцких информантов. Все тексты входят в базу устной спонтанной речи жителей Пермского края, собранной кафедрой теоретического и прикладного языкознания ПГНИУ. Объем русской выборки – 3908 словоформ, татарской – 3998, коми-пермяцкой – 4657.

Все учебные тексты вручную обработаны (удалены знаки разметки), каждая словоформа с применением приложения *Mystem* от Яндекс сведена до лексемы (результаты проверены вручную), каждой лексеме в том же приложении присвоена частеречная характеристика. Таким образом, для обучения использовано два лингвистических па-

раметра: абсолютное вхождение лексемы (связан с семантикой текста) и часть речи (показывает малоконтролируемую автором характеристику текста). Для каждого документа определена относительная частота встречаемости каждой части речи (14 по числу выделяемых *Mystem*).

Каждой лексеме присвоен целочисленный индекс для последующего построения вектора каждого документа. Из каждого документа брались по 500 лексем ( $N$ ), если лексема совпадала с полученным ранее списком, ей присваивался целочисленный идентификатор, если не совпадала – 0.

В качестве тестовых текстов методом случайной выборки из базы отобрано 3 текста, препроцессинг тестовых документов проводился аналогичным образом.

Для создания сети использовалась библиотека PyBrain, выбран алгоритм обучения с учителем [4]. Использовался метод обратного распространения ошибки. Входной слой включал  $N+14$  (количество лексем и количество частей речи) нейронов. Скрытый слой по умолчанию задавался сигмоидной функцией, выходной слой содержал 3 нейрона («русский», «татарский», «коми-пермяцкий»), функция активации *softmax*.

Обучение сети показало вероятность определения этничности только для «русских» текстов (вероятность серии обучений – 0,7 – 0,83), татарский текст отнесен к коллекции «татарских» текстов, но с вероятностью 0,27 – 0,3. Коми-пермяцкий текст в серии экспериментов отнесен то к русской, то к татарской выборке.

Невысокая эффективность атрибуции этничности автора текста может быть объяснена, во-первых, малым количеством лингвистических параметров для обучения, во-вторых, лингвистической дисперсией: тексты созданы на русском языке и некоторые закономерности объясняются свойствами самого языка, а не индивидуальными особенностями авторов текста. При этом важно заметить соотносимость результатов с данными, полученными ранее автором данной работы: частеречное наполнение коми-пермяцких текстов практически не отличается от русских, наиболее частотные лексемы в текстах русских и коми-пермяков также статистически не различаются; существенными различиями текстов выступают лишь жанр текста и семантические категории, реализованные в них [7, с. 39–40]. Однако семантика текста пока, к сожалению, практически не поддается машинному анализу.

### **Библиографический список**

1. *Grozin V. A., Dobrenko N. V., Gusarova N. F., Ning T.* The Application of Machine Learning Methods for Analysis of Text Forums for Creating Learning Objects // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 27–

- 30 мая 2015 г.). Вып. 14 (21): в 2 т. Т. 1: Основная программа конференции. – М.: Изд-во РГГУ, 2015. С. 202–213
2. *Kutuzov A., Andreev I.* Texts in, Meaning out: Neural Language Models in Semantic Similarity Tasks for Russian // Компьютерная лингвистика и интеллектуальные технологии По материалам ежегодной Международной конференции «Диалог» (2015) Вып. 14: в 2 т. Т. 2: Доклады специальных секций. – М.: Изд-во РГГУ, 2015. С. 133-144
3. *Manning Ch.D., Raghavan P., Schütze H.* An Introduction to Information Retrieval. – Cambridge: Cambridge University Press, 2009.
4. Schaul T., Bayer J., Wierstra D., Sun Y., Felder M., Sehnke F., Rückstieß Th., Schmidhuber J. PyBrain // Journal of Machine Learning Research, 2010. P. 743–746.
5. *Батура Т. В.* Формальные методы определения авторства текстов // Вестник ИГУ. Сер.: Информационные технологии. 2012. Т. 10. № 4. С. 81–94.
6. *Романенко А.А.* Кластеризация коллекции текстов // Машинное обучение и анализ данных. 2012. Т. 1. №3. С. 305–310.
7. *Худякова Е.С.* Особенности структуры устных спонтанных текстов монолингвов и билингвов (на материале монологов русских, татар и коми-пермяков) // Вестник Пермского университета. Российская и зарубежная филология. 2013. Вып. 4 (24). С. 33–46.

## MACHINE LEARNING IN SOCIOLINGUISTIC PROBLEMS

*Khudiakova E.S.*

Perm State University, 15, Bukireva st., Perm, Russia,  
Khudiakova.es@gmail.com

The problem of the study is evaluation of the effectiveness of machine learning methods in solving sociolinguistic problems. Specifically, it was made an attempt to determine the ethnicity of the author of the text based on the method of supervised learning (in the training sample the texts are brought to the lexeme and part of speech). The results of the experiment showed that only 1 group of texts (“Russians”) are stably determined by the machine. The texts of the “Tatars” are defined, but with a low probability, the texts of the “Komi-Permyaks” are attributed to the group of “Russians” or “Tatars” as a result of retraining. The effectiveness of attribution of the text’s author ethnicity explained by a small number of learning parameters, as well as by the linguistic effect of the linguistic dispersion, due to the internal properties of the language itself.

*Key words:* machine learning, sociolinguistics, ethnicity, neural network, supervised learning.

## ИНТЕГРАЦИЯ ПОДХОДОВ НА БАЗЕ МАШИННОГО ОБУЧЕНИЯ И МЕТОДОВ ОНТОЛОГИЧЕСКОГО ИНЖИНИРИНГА ДЛЯ РЕШЕНИЯ ЗАДАЧ КЛАССА АОТ<sup>1</sup>

*Чуприна С.И., Костарева Т.А.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, chuprinas@inbox.ru

Освещаются проблемы современных систем автоматической обработки текстов (АОТ) и демонстрируется возможность нивелирования одного из недостатков методов машинного обучения, связанного с отсутствием объяснительных способностей у разрабатываемых на их основе программных систем. Статья описывает авторский подход к интеграции методов онтологического инжиниринга и машинного обучения на примере решения задачи кластеризации коллекции текстовых документов методом *k*-средних. Для автоматического построения онтологических профилей документов, попавших в один и тот же кластер, используется подход на основе лексико-синтаксических шаблонов. Визуализация онтологических профилей позволяет наглядно представить основные понятия и взаимосвязи, описывающие то общее, что характеризует смысловое содержание документов кластера.

*Ключевые слова:* автоматическая обработка текста, машинное обучение, автоматизация построения онтологий, лексико-синтаксические шаблоны.

Несмотря на имеющееся многообразие систем автоматической обработки текстов, до сих пор остается масса нерешенных проблем, которые могут быть, на наш взгляд, отнесены к одной из двух основных категорий.

Первая категория связана со сложностью систематизации и классификации самих систем класса АОТ. Это затрудняет выбор наиболее адекватного решения, учитывающего как специфику конкретной задачи обработки текстов, так и специфические особенности самих текстов (разного рода неструктурированной документации, научных статей, историй болезни, отзывов из социальных сетей о том или ином событии, объекте и т. п.). Кроме того, растет число пользователей, которым нужны инструменты не только для решения таких, ставших уже традиционными, задач как фильтрация спама, проверка орфографии и стиля написания статьи, автоматический перевод выделенных фрагментов текста, но и относительно новых задач, требующих обработки больших массивов текстовых данных. Это пертинент-

---

© Чуприна С.И., Костарева Т.А., 2017

<sup>1</sup> Исследование выполнено в рамках государственного задания ПГНИУ на 2017–2019 гг. при финансовой поддержке Минобрнауки России (проект № 34.1505.2017/ПЧ).

ный поиск информации в ответ на запросы на естественном языке (ЕЯ-запросы), рубрицирование больших текстовых коллекций, выдача рекомендаций, анализ мнений, отзывов, оценок и многие другие. Важными отличительными особенностями задач обработки больших массивов неструктурированных данных являются их сложность, отсутствие формализации, разнородность.

Вторая категория проблем связана с необходимостью совершенствования самих методов АОТ, а также способов их интеграции для нивелирования выявленных недостатков различных подходов. В настоящее время еще нет полноценного набора решений для унифицированной обработки структурированных, полу-структурированных и неструктурированных данных, а средства визуальной аналитики массивных данных не имеют простых и доступных обычному пользователю средств адаптации под их персональные предпочтения. Даже в такой популярной и быстро развивающейся области применения методов машинного обучения как кластеризация текстовых документов до сих пор не решена так называемая проблема «explainability», когда система не способна продемонстрировать объяснение полученному результату.

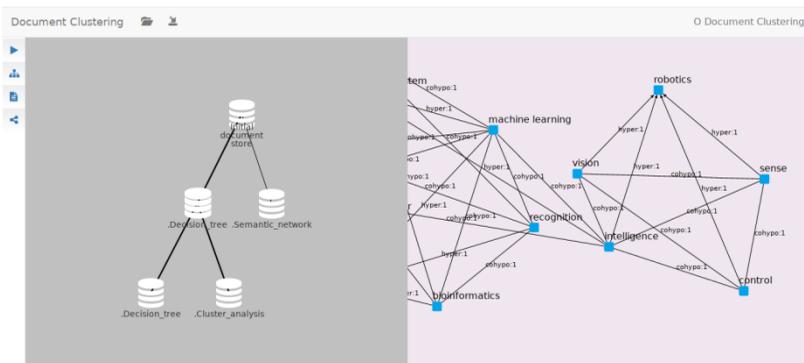
Мы считаем, что для решения указанной проблемы необходима интеграция современных подходов машинного обучения и инженерии знаний, в частности, методов онтологического инжиниринга. Объяснительные возможности экспертных систем хорошо известны и являются одним из преимуществ инженерии знаний. Однако трудоемкость построения и отладки баз знаний сдерживает процесс интеграции методов инженерии знаний и машинного обучения. Онтологии, совмещая в себе преимущества графовых моделей представления знаний (семантических сетей, фреймов) и логических (модели на основе дескриптивных логик), позволяют приблизиться к решению этой проблемы.

Активно развивающиеся в последнее время методы так называемого «онтологического обучения» (Ontology Learning) [1], позволяют автоматически строить онтологии не только на базе корпуса текстов, но и на основе отдельных научных статей, коротких текстов, извлеченных из различных форумов и социальных сетей, и т. п. Важное место здесь уделяется исследованию метрик семантической близости понятий [2], выявленных в результате парсинга неструктурированных текстовых документов, а также вопросам визуализации полученных результатов.

В данной работе демонстрируется один из результатов наших исследований по автоматической генерации онтологических профилей неструктурированных текстовых документов применительно к задачам кластеризации. Онтологические профили представляют собой основные понятия и взаимосвязи, отражающие смысловое содержание тек-

стовых документов. В разработанном нами приложении они генерируются для каждого из автоматически построенных кластеров с целью визуального представления понятий и взаимосвязей, описывающих то общее, что характеризует смысловое содержание документов конкретного кластера. Это может считаться одним из подходов к автоматической генерации объяснений для результатов машинного обучения.

Эксперименты проводились на коллекции текстов статей Википедии, посвященных методам искусственного интеллекта. Кластеризация выполнялась на основе широко известного метода k-средних (k-means), который, несмотря на имеющиеся недостатки, по-прежнему остается одним из наиболее широко используемых на практике алгоритмов кластеризации [3]. На рисунке представлен скриншот демонстрационного прототипа нашего приложения. В левой части экрана в виде иерархической структуры визуализируется результат автоматической кластеризации 20 документов из указанной выше коллекции, в правой – представлены онтологические профили текстовых документов, автоматически сгенерированные для текстов из соответствующего кластера. Такое визуальное представление отражает семантику документов кластера (на рисунке выделен кластер с именем «Decision\_tree» на втором уровне иерархии) и является своего рода объяснительной компонентой и наглядной иллюстрацией полученных результатов кластеризации.



### Фрагмент визуального представления онтологического профиля документов кластера

Кроме того, визуализация онтологических профилей документов может помочь программистам и прикладным лингвистам обнаружить ошибки в разрабатываемом программном обеспечении в ходе реализации новых методов кластерного анализа.

Для качественного построения онтологических профилей произвольных текстовых документов мы используем хорошо зарекомен-

довавшие себя на практике традиционные подходы из области АОТ. В разработанной нами системе TAISim [4] компоненты, отвечающие за реализацию основных этапов автоматической обработки текста, унаследованы от системы Serelex [2]: этапы графематического и морфологического анализа реализованы традиционными методами, а этап синтаксического и семантического анализа заменяется анализом на базе лексико-синтаксических шаблонов [5, 6]. Мы расширили набор поддерживаемых шаблонов и используем результаты анализа текстов на их основе для автоматического построения онтологических профилей документов. В настоящее время описанные в данной работе механизмы совершенствуются с целью их использования для решения задач анализа поведения пользователей социальных сетей в рамках проекта № 34.1505.2017/ПЧ – государственного задания ПГНИУ на 2017–2019.

### **Библиографический список**

1. *Perspectives On Ontology Learning* / J. Lehmann, and J. Voelker (Eds.). Studies in the Semantic Web AKA. IOS Press, 2014.
2. *Panchenko A., Morozova O., Naets H. A Semantic Similarity Measure Based on Lexico-Syntactic Patterns. In Proceedings of KONVENS, 2012. P. 174–178.*
3. *Коэльо Л.П., Ричард В. Построение систем машинного обучения на языке Python. 2-е издание. ДМК Пресс, 2016.*
4. *Chuprina S., Kostareva T., Nam A. Using ontology-driven methods to develop frameworks for tackling NLP problems // In Proceedings of AIST-2016 (Supplement), CEUR-WS.org, 2016.Vol. 1710. P. 102–113.*
5. *Klaussner C., Shekova D. Lexico-Syntactic Patterns for automatic Ontology Building // Proceedings of the Student Research Workshop associated with RANLP, 2011. P. 109-114.*
6. *Sabirova K., Lukanin A. Automatic Extraction of Hypernyms and Hyponyms from Russian Texts // Supplementary Proceedings of the 3rd International Conference on Analysis of Images, Social Networks and Texts (AIST'2014), 2014. Vol. 1197. P. 35–40.*

### **ONTOLOGY ENGINEERING METHODS MEET MACHINE LEARNING TO TACKLE NLP PROBLEMS**

*Chuprina S.I., Kostareva T.A.*

Perm State University, 15, Bukireva st., Perm, Russia, chuprinas@inbox.ru

The paper highlights the problems of modern Natural Language Processing and demonstrates how ontology engineering methods help eliminate Machine Learning problems related to explainability. Using ontology profiles visualization to tackle explainability problems within implementation of k-means clustering algorithm for text documents analysis is presented. Pattern-based approach for ontology profile construction is discussed.

*Key words:* Natural Language Processing, Machine Learning, Ontology Learning, Lexico-Syntactic Patterns.

## МОДЕЛИРОВАНИЕ ПРОЦЕДУР КОНТЕКСТНОЙ ПРЕДСКАЗУЕМОСТИ В ЗАВИСИМОСТИ ОТ ЖАНРОВЫХ И СТИЛЕВЫХ ХАРАКТЕРИСТИК ТЕКСТА

*Ягунова Е.В., Крутченко О.В.*

Санкт-Петербургский государственный университет, Россия,  
Санкт-Петербург, Университетская наб., 11,  
iagounova.elena@gmail.com, krutchenko.olga@gmail.com

Формализуются процедуры информационной избыточности и контекстной предсказуемости с помощью различных вычислительных метрик контекстной предсказуемости, определяются наиболее адекватные метрики и функции для верификации модели оценки и сравнения моделей в зависимости от жанра и типа текста.

*Ключевые слова:* контекстная предсказуемость, биграмма, научные тексты, художественные тексты, Dice, Surprisal, условная вероятность, информационная энтропия, вычислительный эксперимент, cloze-тест.

Избыточность является неотъемлемым свойством естественного языка и текста на естественном языке в частности, необходимым для восприятия и понимания. Избыточность присуща всем текстам без исключения, однако она не является постоянной величиной и зависит от многих параметров, одним из которых является функциональный стиль текста [1, 2].

Общее количество информации, содержащейся в тексте, называется информационной насыщенностью текста. Информационная насыщенность является абсолютным показателем качества текста (в отличие от информативности, которая зависит от степени новизны темы для читателя, и следовательно является относительным показателем качества). По степени информационной насыщенности пять основных функциональных стилей можно расположить следующим образом в порядке возрастания: разговорный, художественный, публицистический, научный, официально-деловой [1, 6].

Для исследования контекстной предсказуемости были выбраны для сравнения два функциональных стиля: научный и художественный, существенно отличающихся избыточностью текстов. Научные тексты были отобраны в два подкорпуса, каждый из которых принадлежит одной предметной области и однороден по жанру и теме. Основным ориентиром в исследовании послужили разработки моделей контекстной предсказуемости в информатике и смежных дисциплинах. Чаще всего такого рода модели опираются на скрытые марковские

процессы. Такие модели позволяют рассматривать текст как совокупность процессов перехода из одного состояния в другое.

Данная работа связана с квантитативной типологией стилей и жанров, активно развивающейся в наше время; так, была предположена зависимость процедур контекстной предсказуемости и модели текста. Работы, связанные с изучением избыточности, велись и в нашей стране уже в 60-е годы (например, исследования Н.Н. Леонтьевой, Р.Г. Пиотровского, Т.Н. Никитиной, М.И. Откупщиковой, специально посвященные этой теме [3]). Особенно полно данный вопрос рассматривается в статьях Р.Г. Пиотровского [4] и [5].

Материал:

- миникорпус художественных текстов 782 300 словоупотреблений (с/у);
- миникорпус научных текстов (37 796 с/у) делится на 2 подкорпуса: научные статьи по корпусной лингвистике (15 093 с/у) и когнитивной психологии (22 703 с/у).

На данном этапе подготовки к вычислительному анализу контекстной предсказуемости на основе двух корпусов текстов можно предположить, что значение контекстной предсказуемости для научных текстов будет намного выше по сравнению с художественными ввиду большей информационной насыщенности.

В результате вычислительного эксперимента были получены модели текстов. Вычисленная информация о контекстной предсказуемости представлена в виде результирующих таблиц (матриц признаков) на основе простейших лексико-морфологических характеристик (средняя длина слова, доля частотных слов, коэффициент лексического разнообразия, доля глаголов, доля существительных, доля служебных слов и т. д.), коллокаций, характеризующихся максимальной связностью, интегральных информационных характеристик (меры удобочитаемости, энтропии и удивительности (surprisal)).

Стоит отметить, что значение среднего арифметического энтропии и метрики удивительность не являются взаимобратными, поскольку для вычисления энтропии использовалась вероятность появления в тексте одиночного словоупотребления, в то время как расчет значения метрики удивительность производился с учетом контекста (на основе условной вероятности).

На данном этапе исследования можно сделать вывод о том, что значения выбранных признаков непосредственно зависят не только от свойств контекстной предсказуемости текста, но также и от объема исследуемого текста. Это говорит о необходимости пополнения корпусов текстов для проведения дальнейшего исследования. Несмотря на это, на основе полученных результатов уже можно сделать вывод, что гипотеза об ожидаемом более высоком показателе контекстной пред-

сказуемости для научных текстов по сравнению с художественными текстами подтверждается.

Верификация модели текста и его обработки, проведенная в ходе вычислительного эксперимента, включала проведение cloze-теста. Для проведения теста были выбраны 4 фрагмента текстов (2 фрагмента художественного стиля, 2 – научного), принадлежащих различным произведениям. Каждый из фрагментов по объему составляет от 100 до 120 слов. В каждом фрагменте пропущено по 10 слов, которые предлагается восстановить информантам.

Выбор того, какие слова фрагмента будут пропущены, осуществлялся на основе метрики surprisal (в последнее время она является основной и наиболее часто используемой при изучении контекстной предсказуемости). В рамках проведения эксперимента с информантами предполагалось проверить работоспособность модели текстов, построенной на основе вычислительного эксперимента. Поэтому в каждом фрагменте текста в качестве пропущенных слов были выбраны слова с высоким (8–11), средним (4–8), низким (0–3) значением метрики surprisal. Предполагается, что более высокое значение меры говорит о том, что данное слово хуже восстанавливается из контекста. Результаты эксперимента сопоставлялись и со значением энтропийных характеристик.

Основные предположения эксперимента подтвердились – слова с низким значением меры surprisal восстанавливаются информантами однозначно или с использованием синонимов в 85–100 % случаев.

По результатам эксперимента была выделена группа слов (10 % от общего количества пропущенных), которые однозначно восстанавливаются информантами, однако имеют высокое значение меры surprisal (от 8 до 11). Изначально предполагалось, что данная группа слов будет менее восстанавливаема. Значения же энтропийных характеристик для данной группы слов незначительно превышают среднюю энтропию текста.

Данный результат можно объяснить тем, что для расчета меры surprisal использовалось в качестве контекста одно предшествующее слово, а информант, заполняя пропуски, ориентировался на контекст большей длины. Факт использования более широкого контекста подтверждается тем, что слова этой группы входят в выделенные цепочки (чанки), образованные на основании метрики Dice с учетом более широкого контекста (пять слов из левого контекста и два слова из правого).

Результатом работы является модель-прототип (на языке Python), позволяющая построить модель текста, реализующую различные признаки контекстной предсказуемости, и оценка работы модели-прототипа, в том числе с помощью cloze-теста.

Эксперимент с информантами подтвердил эффективность и работоспособность построенной модели текстов, а также наметил возможные пути для дальнейшего развития исследования и совершенствования написанной программы.

По результатам построения моделей текстов подтвердились гипотезы о различии вариантов моделей, приспособленных для художественных или научных текстов. Так, для научных текстов более высок обобщенный показатель контекстной предсказуемости, для научного и художественного текстов различаются иерархии значимости признаков (метрик) и средняя длина чанков.

### **Библиографический список**

1. Ягунова Е.В. Основы теоретической, вычислительной и экспериментальной лингвистики, или Размышления о месте лингвиста в компьютерной лингвистике // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. М.: МИЭМ, 2011.
2. Ягунова Е.В. Исследование избыточности русского звучащего текста // Избыточность в грамматическом строе языка / Отв. ред. М. Д. Воейкова. СПб.: Наука, 2010. 462 с.
3. Бердичевский А. Языковая сложность (Language complexity) // Вопросы языкознания. 2012. № 5.
4. Пиотровский П. Г. Лингвистический автомат (в исследовании и непрерывном обучении). СПб., 1999.
5. Пиотровский П. Г. Информационные измерения языка. Л., 1968.
6. Бабайлова А.Э. Текст как продукт, средство и объект коммуникации при обучении неродному языку. // Изд. Саратовского университета, 1987.

### **MODELING OF CONTEXT PREDICTION PROCEDURES OF THE TEXT DEPENDING ON GENRE AND STYLE CHARACTERISTICS OF THE TEXT**

*Yagunova E. V., Krutchenko O. V.*

St. Petersburg State University, 11, Universitetskaya nab., St. Petersburg,  
Russia, yagounova.elena@gmail.com, krutchenko.olga@gmail.com

The paper formalizes the procedures of information redundancy and contextual predictability with the help of various computational contextual predictability metrics, identifies the most adequate metrics and features for further verification in the process of the constructing a model of the text, the significant part of the paper are the model evaluation and comparison of models depending on the genre and type of text.

*Key words:* contextual predictability, bigram, scientific texts, fiction texts, Dice, Surprisal, conditional probability, information entropy, computational experiment, cloze-test.

## ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ ТРАНСМЕДИЙНЫХ ОБЪЕКТОВ

*Яруллин Д.В.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, d.v.yarullin@gmail.com

В работе описываются явление трансмедийности и существующие подходы к его изучению. Дается рабочее определение трансмедийности, подчеркивается ее интертекстуальный аспект. Трансмедийный образ трактуется как когнитивный продукт, складывающийся в сознании отдельного субъекта. Ввиду того что трансмедийные образы существуют в интертексте, их субъективное восприятие в большинстве случаев оказывается типичным. Это дает возможность рассматривать их в рамках теории фреймов. С опорой на фреймовую модель предлагается принцип формализации трансмедийного образа для его последующего автоматического анализа. Автоматический анализ осуществляется при помощи технологий машинного обучения, обучающая выборка содержит пары «элемент нарратива – актуализированный признак». На вход искусственной нейронной сети подается ряд элементов нарратива для классификации его как содержащего конкретный трансмедийный образ в его типичном восприятии или не содержащего. Продукт может найти свое применение в межавторских коллективах, работающих над единой трансмедийной вселенной (франшизой), и анализе т. н. фейковых новостей.

*Ключевые слова:* трансмедийность, интертекст, фрейм, трансмедийный образ, формализация, машинное обучение, искусственная нейронная сеть.

В последнее время феномену трансмедийности посвящается все больше теоретических работ, описывающих его в терминах различных дисциплин, включая семиотику, нарратологию и даже маркетинг [1, 2]. Отдельно взятые методы могут опираться на техники сравнительных исследований, нарративный анализ или непосредственное изучение источников [3].

Ввиду многообразия подходов к изучению трансмедийности нет устоявшегося понимания этого явления. Для описания феномена трансмедиа мы воспользуемся определением, данным К. Сколари: это «особого рода нарративная структура, границы которой расширяются одновременно за счет использования разных знаковых систем (вербальной, иконической и пр.) и разных медиумов (кино, комиксов, телевидения, видеоигр и др.). Разные знаковые системы и медиумы вместе создают трансмедийную вселенную смыслов» [4]. Важное замечание делает Г. Дженкинс: «Трансмедийная история рассказывается посредством различных платформ, и каждый новый текст вносит весомые изменения в общий нарратив» [5]. Также нам важен ас-

пект интертекстуальности трансмедийного повествования, выделенный М. Киннер [6].

Исходя из этого, мы можем заметить, что в трансмедийных системах целостные образы не формируются в рамках одного медиума, но являются продуктом сопоставления и анализа нескольких отдельных образов, реализованных в каждом конкретном медиа. Следовательно, трансмедийный образ – когнитивный продукт, созданный в сознании аудитории. При этом, опираясь на работы классиков постмодернизма [7, 8], мы соглашаемся с пониманием человеческой культуры как единого «интертекста», который, в свою очередь, служит как бы предтекстом любого вновь появляющегося текста. Важным следствием было растворение суверенной субъективности человека в «текстах-сознаниях», составляющих «великий интертекст» культурной традиции [9, с. 101]. Так, субъективное восприятие ряда образов, складывающееся в сознании в единый трансмедийный образ, чаще всего оказывается типичным, поскольку рождено в интертексте культурной традиции.

Это позволяет нам рассмотреть трансмедийный образ в рамках теории фреймов М. Минского [10], развитой Ч. Филлмором [11]. Фрейм определяется исследователями как когнитивная модель типичной ситуации, сценария, объекта и т. п., у которой есть формальная структура – слоты и их значения. Представление трансмедийного образа как фрейма открывает путь к формализации его значимых признаков, отвечающих за целостность его восприятия. Это дает нам возможность автоматического анализа нарратива в разных медиумах на предмет считывания отдельно взятого трансмедийного образа.

Обучив искусственную нейронную сеть (обучающая выборка содержит пары «элемент нарратива – актуализированный признак»), мы можем последовательно подавать на ее вход ряд элементов нарратива для классификации его как содержащего конкретный трансмедийный образ в его типичном восприятии или не содержащего. Подобный продукт может найти свое применение как в межавторских коллективах, работающих над единой трансмедийной вселенной (франшизой), для облегчения задачи проверки «каноничности» создаваемого в новом произведении образа, так и, что нам видится более перспективным, в анализе т. н. фейковых новостей, поскольку образы и персоналии, описываемые в СМИ, также трансмедийны по своей сути.

### Библиографический список

1. *Jenkins H.* Transmedia storytelling and entertainment: An annotated syllabus. *Continuum: Journal of Media & Cultural Studies*. 2010. 24, 6. P. 943–958.
2. *Saldre M., Torop P.* Transmedia space I. Ibrus, C. A. Scolari (eds.), *Crossmedia Innovations: Texts, Markets, Institutions*. Frankfurt: Peter Lang, 2012. P. 25–44.

3. *Gambarato R.R.* Transmedia Project Design: Theoretical and Analytical Considerations // *Baltic Screen Media Review*. 2013. №.1. P. 80–100.
4. *Scolari C. A.* Transmedia storytelling: Implicit consumers, narrative worlds, and branding in contemporary media production // *International Journal of Communication*. 2009. 3. P. 586–606.
5. *Jenkins H.* *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press, 2006.
6. *Kinder M.* *Playing with Power in Movies, Television, and Video Games: From Muppet Babies to Teenage Mutant Ninja Turtles*. Berkeley: University of California Press, 1991.
7. *Barthes R.* *The Pleasure of the Text – Le Plaisir du texte*. Paris; Seuil, 1973. URL: <https://emberilmu.files.wordpress.com/2011/08/roland-barthes-the-pleasure-of-the-text.pdf> (дата обращения: 30.03.2017).
8. *Kristeva J.* *La révolution du langage poétique*. Paris: Seuil, 1974.  
*Ильин И.П.* Постмодернизм: словарь терминов. М.: ИНИОН РАН; INTRADA, 2001. 384 с.
9. *Minsky M.* A framework for representing knowledge. MIT AI Laboratory Memo 306, June, 1974. URL: <http://web.media.mit.edu/~minsky/papers/Frames/frames.html> (дата обращения: 30.03.2017).
10. *Fillmore C. J.* Frame semantics and the nature of language// *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*. Vol. 280. P. 20–32. URL: <http://www.icsi.berkeley.edu/pubs/ai/frame semantics76.pdf> (дата обращения: 30.03.2017).

## MACHINE LEARNING METHODS IN ANALYSIS OF TRANSMEDIA OBJECTS

*Yarullin D.V.*

Perm State University, 15, Bukireva st., Perm, Russia,  
d.v.yarullin@gmail.com

The article describes the phenomenon of transmedia and existing approaches to its study. A definition of transmedia is given with focus on intertextual aspect of the phenomenon. Transmedia image is defined as cognitive product of one's consciousness. As transmedia images exist in the intertext, the subjective perception of them is typical for most cases. It allows applying to them the frame theory. Frame theory gives a basis for a method of transmedia image formalization. The goal of such formalization is automatic analysis using technologies of machine learning. Training dataset contains pairs «narrative element — part of transmedia image». When elements of a particular narrative are used as an input for artificial neural network, the system classifies this narrative as containing or not containing the transmedia image. The product may be used for collectives working on a transmedia universe (franchise) and for fake news analysis.

*Key words:* transmedia, intertext, frame, transmedia image, formalization, machine learning, neural network.

## DO FREQUENT MEDIA WORDS WORSEN QUERY EXPANSION?

*Ovchinnikova I.*

Perm State University, 15, Bukireva st., Perm, Russia,  
ira.ovchi@gmail.com

*Ermakova L.*

LISIS – Université de Lorraine, 5, boulevard Descartes, 77454,  
Champs-sur-Marne, France, liana87@mail.ru

*Mothe J.*

ESPE, Université de Toulouse, Institut de Recherche en Informatique  
de Toulouse, UMR5505 CNRS, 118 Route de Narbonne,  
F-31062 Toulouse Cedex 9, France, josiane.mothe@irit.fr

This paper offers a linguistic approach to the study of the potency of query expansion while retrieving information from the web. The expansion allows enhancing the results; however, some queries show lower effectiveness after expansion. The objective of the study is to analyze linguistic features of initial query (IQ) as predictors for the expansion potency by different systems. The IQ is considered as a ‘bag of words’ with their linguistic descriptions, frequency first of all. The interdependence of different linguistic features of a query term determines the term value and its validity for the expansion. Analyzing two sets of terms from IQ (from queries that failed and from queries that were improved after expansion), we found out the negative impact of frequent terms from media on query expansion. This effect reflects the semantic variety of the frequent term connections in texts of different genres

*Ключевые слова:* initial query, information retrieval, query expansion, word frequency, pragmatics, query analysis.

1. IQ processing by a search engine results in extraction of documents with the query terms as frequent words. The small number of query terms (more than 90 % of queries are 3 words or less long) generates a problem of ambiguity of users’ queries. To solve this problem, some systems diversify their response by expanding the query with terms either being extracted from external resources such as WordNet or co-occurring in the documents containing initial query terms. In the latter case, the query expansion (QE) can use pseudo-relevance feedback which considers the first retrieved documents using the IQ to extract the terms to expand it. Sometimes pseudo-relevance feedback lowers results, demonstrating poorer relevance to users’ interests than the IQ. The poor results can be caused by non-relevant expansion of IQ with terms from non-relevant documents.

The objective of this paper is to reconsider the basis for prognosis of the necessity and productivity of the IQ expansion. We argue that the rele-

vance of the expansion depends on the linguistic parameters of the IQ terms. Since IR systems deal with words and word chunks, their frequencies in texts of different genres and styles should have an impact. In this paper, we examine a possibility to recognize by virtue of frequency, whether the expansion will improve or impair the results.

2. We consider an IQ as a “bag of words” with their linguistic characteristics. We examine impact of the frequency in interaction with pragmatics on the effectiveness of IQ processing. Word frequency correlates with semantics and pragmatics on the basis of different rates of word occurrence in the various discourses.

In our previous publication [1] we analyze the effectiveness of QE using two different QE systems. Based on these results, we chose two sets of query terms. The first set consists of the terms from queries which failed after expansion by the systems (*Low* set). IQs for the control set (*Imp* set) were randomly chosen from the queries which were improved after expanding. Our research question is “Are there cues to decide whether a frequent term in an IQ need to be expanded or not?” To answer the question, we test the distributions of the word frequencies<sup>1</sup> for two sets of query terms. Our analysis is based on WT10G TREC collection<sup>2</sup>. TREC WT10G is a subset of the Internet archive making 10GB, containing more than 1.6 million documents and 98 requests with judgments of relevance.

3. The data are represented in the table. Word frequencies in the *Low* set vary in wider limits than ones in the *Imp*.

**Frequency of terms in WordAndPhrase for *Low* and *Imp***

Set	Average Total F	Me Total F	Average F newsp.	Me F newsp.	Average F magaz.	Me F magaz.
<b>Low</b>	51796	27922	10309	6896	12151	7313
<b>Imp</b>	39922	11103	5589	2022	6926	3081

Frequency of words in *Low* is essentially higher in media (F newspaper plus F magazine), as well as in all communicative spheres (Total F). Terms from *Low* occur in the corpora almost twice often than terms from *Imp*. We found out that the most frequent words from media tend to fail as a query terms after being expanded by their collocations.

---

<sup>1</sup> <http://www.wordandphrase.info/frequencyList.asp> WordAndPhrase resource, based on COCA, represents 1,500,000 frequent English words in 60,000 lists 25 words each.

<sup>2</sup> [trec.nist.gov](http://trec.nist.gov).

## References

1. *Ermakova, L., Mothe, J., & Ovchinnikova, I.* (2014). Query expansion in information retrieval: What can we learn from a deep analysis of queries? In: International Conference on Computational Linguistics-Dialogue 2014 (Vol. 20, No. 13, pp. 152–162).

### **ЧАСТОТНЫЕ В МЕДИА СЛОВА КАК ТЕРМИНЫ ПОИСКОВОГО ЗАПРОСА**

*Овчинникова И.Г.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, ira.ovchi@gmail.com

*Ермакова Л.*

Университет Лотарингии, 5, boulevard Descartes, 77454  
Champs-sur-Marne, France, liana87@mail.ru

*Мот Ж.*

Институт исследований информационных технологий, Тулуза, 118  
Route de Narbonne, F-31062 Toulouse Cedex 9, France,  
josiane.mothe@irit.fr

В статье исследуется продуктивность автоматического расширения первичного запроса на основе анализа частотности и сферы использования его терминов. Расширение осуществляют различные системы извлечения информации. Нередко расширение запроса оказывается неудачным, ухудшает результативность поиска. Цель нашего исследования – выявить прогностические свойства терминов первичного запроса, которые позволили бы предсказать необходимость его расширения любой из систем и, если таковое необходимо, эффективность расширенного запроса. Первичный запрос представлен как набор слов с некоторыми лингвистическими параметрами. Основным параметром слова для ИП является его частотность. Частотность слова в определенном дискурсе предопределяет его ценность для запроса и пригодность для расширения за счет коллокаций, синонимов и ассоциаций. В результате анализа двух наборов из терминов первичных запросов (ухудшенного (Low) и улучшенного (Imp) при расширении различными системами) мы обнаружили негативное влияние частотных в медиа слов на результаты поиска посредством расширенного запроса. Это негативное воздействие обусловлено разнообразием связей таких слов в текстах различных жанров и тематики.

*Key words:* Первичный запрос, извлечение информации, расширение запроса, частотность, прагматика, дискурс, анализ запроса.

# КОМПЬЮТЕРНОЕ ИСТОЧНИКОВЕДЕНИЕ И МАТЕМАТИЧЕСКИЕ МЕТОДЫ

УДК 947(084)

## ФОРМИРОВАНИЕ СОВЕТСКОЙ ГОРОДСКОЙ ИДЕНТИЧНОСТИ: КОНТЕНТ-АНАЛИЗ ВИДОВЫХ ОТКРЫТОК КАЛИНИНГРАДА 1950-1980-Е ГГ.

*Баранова Е.В.*

Балтийский федеральный университет им. Иммануила Канта,  
Россия, Калининград, ул. Невского, 14, EBaranova@kantiana.ru

Актуализируется изучение истории советской идентичности в контексте специфики так называемых «трофейных» городов. Особенности архитектурно-градостроительной «колониализации», в ходе которой рождался образ советского города, осложнялся в процессе формирования новой культурной идентичности горожан. Официальные визуальные образы видовой фотографии были одним из средств интеграции «трофейных» городов в советское пространство как в социокультурном, так и в буквальном смысле, через тиражирование на почтовых открытках. Изучение формирования образа советского Калининграда даст возможность проанализировать пропагандистские установки власти с одной стороны, а с другой – раскрыть систему общественных стереотипов и штампов в восприятии советской действительности. Междисциплинарный анализ визуальной семиотики городских ландшафтов, запечатленных на почтовых открытках 1950–1980-х гг. позволит выявить роль зрительных образов в формировании советской и городской идентичности жителей изучаемых городов.

*Ключевые слова:* визуальная история, идентичность, образ жизни, контент-анализ, видовые открытки.

Основной научной проблемой исследования стало изучение особенностей формирования советской идентичности в т. н. «трофейных» городах (на примере г. Калининграда). Источником стала коллекция видовых открыток, изданных с 1956 по 1990 г. в количестве 303 штук. Планировалось рассмотреть, как закрепленные в формате городской видовой открытки визуальные образы города отразили процессы создания общесоветской (на уровне страны) и локальной (городской) советской идентичностей в условиях делегитимации немецкого прошлого.

На основе коллекции была разработана база данных. Она содержит характеристики на оборотной стороне открытки: тиражность, издательство, комментарий – описание открытки, литование (цензура), год издания, цена. Но наиболее интересными представляются распре-

деленные по категориям образы на лицевой стороне (объект, назначение, наличие людей, объекты коммунального хозяйства).

Приступая к техническому анализу, стоит отметить, что открытки подразделяются на маркированные и не маркированные. Маркированные почтовые открытки (цельные вещи) – это открытки с типографским изображением почтовой марки. Издание маркированных открыток было в компетенции Министерства связи СССР. Наша коллекция содержит в большинстве именно маркированные открытки, издававшиеся с 1968 по 1991 г. Отдельные экземпляры с видами Калининграда и области выходили каждый год. Средний тираж составлял от 20 000 до 100 000 экземпляров по каждой открытке.

Помимо маркированных открыток, большое значение придавалось выпуску сувенирных видовых открыток. Первые видовые открытки в Калининграде появились в 50-х гг. План издания утверждался Обкомом КПСС. К концу 60-х – началу 70-х гг. печать открыток постепенно переходит к центральным издательствам. Всего в Советском государстве существовало 6 больших издательств, выпускающих в том числе открытки: «Филателин ГФК», «Плакат», «Правда», «Турист», «Советская Россия», «Планета». В нашей коллекции представлены все издательства [2, с. 76].

Объекты, нашедшие отражение на калининградских видовых открытках (см. табл.), можно разделить на три типа – архитектурные объекты (54,2 %), монументальная пластика (27,4 %) и природные виды (18,3 %).

Категория	Время				
	50-е	60-е	70-е	80-е	Итого
<b>Архитектура</b>	<b>5</b>	<b>8</b>	<b>11</b>	<b>59</b>	<b>83</b>
Немецкая архитектура	4	4	1	32	41
Советская архитектура	1	4	10	27	42
<b>Памятники</b>	<b>5</b>	<b>4</b>	<b>2</b>	<b>32</b>	<b>43</b>
Памятники немецкие	–	1	–	4	5
Памятники военные	3	1	1	13	18
Памятники культурные	2	2	1	15	20
<b>Природные виды</b>	<b>8</b>	<b>1</b>	<b>5</b>	<b>14</b>	<b>28</b>
<b>Итого</b>	<b>18</b>	<b>13</b>	<b>18</b>	<b>105</b>	<b>154</b>

Обращаясь к архитектурным объектам, стоит сказать, что изображались туристические доминанты, объекты сферы услуг для гостей (гостиницы, театры, рестораны) и административные здания.

Второй многочисленной группой, нашедшей свое отражение в видовых открытках, стали памятники. Калининградская монументальная пластика развивалась под влиянием Ленинского плана монументальной

тальной пропаганды. Особенность региона предопределила преобладание военных памятников.

Важным представляется вопрос о наличии объектов немецкого наследия на открытках. Так, на первых открытках немецкая архитектура преобладает, и только в 70-е гг. ситуация выравнивается и количество немецких и советских объектов уравнивается. В 1971 г. происходит важное событие для закрепления дальнейшей судьбы Калининграда: принимается Хельсинская конвенция о неизменности послевоенных границ европейских государств. В результате происходит окончательное подтверждение статуса Калининграда, и архитектурно-градостроительная деятельность в городе принимает более интенсивный и уверенный характер. Начало 80-х гг. ознаменовано явлением реконструкции-реставрации довоенных построек. Этому явлению мы обязаны восстановлением, сохранением и приспособлением под новые функции чудом уцелевших на тот момент исторических зданий старого города [1].

Интересной стала задача сравнения объектов, изображенных на открытках и комментариев на оборотной стороне. Несмотря на присутствие немецких объектов в кадре, они абсолютно игнорируются, поэтому для определения верных акцентов использовался текст-описание на обратной стороне. Так, открытка, где в центре четко видны руины главной доминанты Кёнигсберга – Королевского замка, имеет название «Калининград. Дом профсоюзов». Внимание зрителя переносится на советское строение. Переориентация зрителя на советскую действительность стала заметна после проведения частотного анализа текстовых комментариев к открыткам. Преобладают термины советской послевоенной эпохи: победа, память, штурм, героизм, мемориал, воин.

В нашем исследовании Калининград стал объектом «положительной» советской пропаганды (т. е. формирующей «правильные» образы и символы) как образец нового советского города, в котором постепенно стиралась память о многовековой довоенной немецкой истории. Анализ визуальных характеристик идеального образа советского Калининграда, соотнесение их с текстовыми способствует более глубокому изучению реалий повседневности.

### **Библиографический список**

1. *Васютин О.* Калининград-Кёнигсберг: архитектура советского и постсоветского периодов (конец первой половины XX – начало XXI вв.) URL: [http://sg39.ru/upload/iblock/cec/oleg-vasyutin.-\\_kaliningrad-\\_-kyenigsberg.-arkhitektura-sovetskogo-i-postsovetskogo-periodov.pdf](http://sg39.ru/upload/iblock/cec/oleg-vasyutin.-_kaliningrad-_-kyenigsberg.-arkhitektura-sovetskogo-i-postsovetskogo-periodov.pdf) (дата обращения: 30.03.2017).
2. *Таггин М.* Мир в открытке. М., 1978.

# **FORMATION OF THE SOVIET URBAN IDENTITY: CONTENT ANALYSIS OF CITY VIEW POSTCARDS OF KALININGRAD, 1950S–1980S**

*Baranova E. V.*

Immanuel Kant Baltic Federal University,  
Kaliningrad, Nevskogo st., 14, Russia, EBaranova@kantiana.ru

The article actualizes the study of the history of the Soviet identity in the context of the specificity of the so-called “captured” cities. The features of architectural and urban “colonization”, during which the image of the Soviet city was born, was complicated by the processes of formation of a new cultural identity of citizens. Official visual images of the city view photography were one of the means of integrating the “captured” cities in the Soviet space both in sociocultural and literal sense, through replicating on postcards. The study of the formation of the Soviet Kaliningrad image will give an opportunity to analyze the propagandistic attitudes of the authorities, on the one hand, and to reveal the system of public stereotypes and cliches in the perception of the Soviet reality, on the other hand. Interdisciplinary analysis of visual semiotics of urban landscapes, imprinted in the postcards of the 1950s-1980s, will reveal the role of visual images in the formation of the Soviet and urban identity of the inhabitants of the studied cities.

*Key words:* visual history, identity, lifestyle, content analysis, postcards.

УДК 94(571)(093)

## **МАТЕРИАЛЫ ПЕРЕПИСИ 1897 ГОДА КАК ИСТОЧНИКИ ПО СОЦИАЛЬНОЙ ТОПОГРАФИИ СИБИРСКИХ ГОРОДОВ<sup>1</sup>**

*Брюханова Е.А.*

Алтайский государственный университет,  
Россия, Барнаул, просп. Ленина, 61, elena@hist.asu.ru

Социальная топография предусматривает изучение демографических, социальных, профессиональных, экономических и других характеристик населения на уровне городского пространства, т. е. в пространстве городских районов, улиц, домов, отдельных квартир и строений. Социальная топография позволяет выявить городские социальные сообщества, внутригрупповые связи и другие тенденции урбанизации, а социально-топографическое зонирование – оценить влияние «особых» факторов сибирских городов на формирование городской среды. В тезисах особое внимание уделяется источникам для реконструкции пространственной структуры городов, приводятся примеры отечественных исследований по исторической топографии городов XIX–XX вв., рассматривается состав данных переписных листов переписи 1897 г., оценивается возможность их применения как источников «пространственных» данных для

---

© Брюханова Е.А., 2017

<sup>1</sup> Публикация подготовлена при финансовой поддержке РФФИ (грант № 17-06-00496).

городского населения. Делается вывод о высокой репрезентативности первичных материалов переписи 1897 г. для изучения социальной топографии сибирских городов.

*Ключевые слова:* перепись 1897 г., топография города, Сибирь, Тобольская губерния.

В конце XIX – начале XX вв. происходит интенсификация процессов урбанизации, в том числе на территории Сибири. Несмотря на преобладание общих тенденций, сибирские города отличались рядом специфических особенностей, таких как влияние этнических групп (г. Березов, Якутск), принудительных миграций (каторжные тюрьмы г. Тобольска), железнодорожного строительства (г. Тюмень, Курган), добровольных переселений (г. Тара, Тюкалинск), присутствие военного контингента (г. Благовещенск, Владивосток) и т. д. Рассматривая ключевые вопросы урбанизации – социально-демографические аспекты, профессиональный состав, промышленные, культурные объекты, благоустройство и т. п., исследователи практически не обращались к анализу внутреннего пространства городов. При этом изучение социальной топографии города, предусматривающей соединение социальных, экономических и пространственных аспектов, может быть направлено, с одной стороны, на выявление городских социальных сообществ, внутригрупповых связей и других тенденций урбанизации (например, маргинальных слоев), с другой – возможности социально-топографического зонирования позволят оценить влияние «особых» факторов сибирских городов на формирование городской пространственной среды.

Определенные сложности изучения социальной топографии городов обусловлены наличием и составом сохранившегося корпуса источников. Для таких целей необходим комплекс массовых источников, содержащих социально-демографическую информацию о жителях с точными пространственными «координатами» (дом, улица, район города), позволяющими соотнести их с картографическим материалом. Так, А.А. Акашева при реконструкции пространственной структуры Нижнего Новгорода конца XIX в. использовала подробные характеристики домовладений и городских участков, отраженных в ежегодных подворных полицейских отчетах [1, с. 6]. В.Ю. Козлова для изучения этнической топографии города Пермь применяла «Списки владельцев недвижимых имуществ», опубликованные в исследуемом городе в 1898 [2, с. 88–100].

Для изучения социальной топографии сибирских городов особую ценность представляют первичные переписные листы переписи 1897 г. Наибольшую репрезентативность имеют материалы Государственного архива города Тобольска, в котором сохранились переписи-

ные листы по городам Тобольск, Тюмень, Туринск, Тюкалинск, Ишим, Курган, Ялуторовск (Ф. И417. Оп. 2). Следует отметить, что материалы более информативных вторых и третьих страниц переписных листов уже использовались для исследования демографических [3], социальных [4] и этнических [5] аспектов как городского, так и сельского населения Тобольской губернии. При этом сведения первой страницы переписного листа, сообщающие не только пространственные характеристики домохозяйства (улица, дом, квартира), сведения о собственнике объекта недвижимости, но и описание всех имеющихся строений (из чего построены, чем крыты?), оставались вне поля зрения исследователей. Набор сведений немного видоизменялся в зависимости от формы переписного листа (А, Б, В, перечневые ведомости). Например, для формы В (для городских квартирохозяев) вводились дополнительные вопросы о количестве незаселенных квартир, а для Военской перечневой ведомости – вопрос о месте расквартирования воинской части (в казарме, в частном доме, «на квартире у обывателя»?). Аккумуляция подобных данных и сопоставление их с картографическим материалом позволит провести подробную реконструкцию социального пространства ряда сибирских городов.

Таким образом, материалы переписи 1897 г. могут стать уникальным источником не только для статистического, но и пространственного анализа демографических и социальных аспектов городского населения Сибири на рубеже XIX–XX вв.

### Библиографический список

1. *Акашева А.А.* Нижний Новгород в 1860–1890-х гг. Методика реконструкции социокультурного пространства города: автореф. дисс. ... канд. ист. наук. Нижний Новгород, 2006.
2. *Козлова В.Ю.* Город и городское население Пермского Прикамья второй половины XIX в. в социальном, культурном и этническом измерениях: дисс. ... канд. ист. наук. Пермь, 2011.
3. *Лискевич Н.А., Машаринова А.Х.* Формирование и расселение групп коми на территории Ялуторовского уезда Тобольской губернии в XIX – начале XX в. // Вестник археологии, антропологии и этнографии. 2012. № 2 (17). С. 113–119.
4. *Зверев В.А.* Семейно-брачный строй в деревнях Западной Сибири (по материалам Всероссийской переписи населения 1897 г.)// Вестник Томского государственного университета. Серия «История». 2009. № 4(8). С. 63–70.
5. *Брюханова Е.А., Владимиров В.Н., Стрекалова Н.В.* Сословие, профессия, социальная страта: к проблеме исследования социальной структуры провинциального российского губернского города в конце XIX – начале XX в. (на материалах Тамбова и Тобольска)// Вестник Томского государственного университета. 2016. № 411. С. 38–46.

## THE 1897 CENSUS DATA AS SOURCES ON SOCIAL TOPOGRAPHY OF SIBERIAN CITIES

*Bryukhanova E.A.*

Altai State University, 61, Lenina ave., Barnaul, Russia, elena@hist.asu.ru

Social topography provides the study of demographic, social, occupational, economical and other population's characteristics at the level of urban space that is the space of city districts, streets, households, apartments, and other buildings. Social topography makes it possible to identify urban social communities, intra-group links and other trends of urbanization. Socio-topographical zoning gives the possibility to estimate the influence of "special" factors of Siberian cities on the evolution of urban environment. Special attention is paid to the sources of the city's spatial structure reconstruction; the examples of Russian researches in the field of historical topography of the cities of XIX-XX centuries are given; the composition of the 1897 census questionnaires is considered; and the possibility of their application as sources of "spatial" data for the urban population is estimated. It is concluded that the primary materials of the 1897 census are highly representative for the social topography of Siberian cities.

*Key words:* census 1897, topography of the city, Siberia, Tobolsk province.

УДК 087.6.82-92

## МЕТОДИКА ОБРАБОТКИ ПЕЧАТНЫХ СМИ ДЛЯ ПОСЛЕДУЮЩЕГО КОНТЕНТ-АНАЛИЗА ПРИ УСЛОВИИ ПЛОХОЙ СОХРАННОСТИ МАТЕРИАЛА (НА ПРИМЕРЕ СМИ КАЛИНИНГРАДСКОЙ ОБЛАСТИ В ПОСЛЕВОЕННОЕ ВРЕМЯ)

*Будеркина Е.Ю.*

Балтийский федеральный университет имени Иммануила Канта,  
Россия, Калининград, ул. А. Невского, 14, ebuderkina@mail.ru

Рассматривается разработка метода обработки печатных СМИ на примере газеты «Калининградская правда» с 1946 по 1950 г. для последующего проведения количественного и качественного контент-анализа. Необходимость проведения подобного исследования обоснована качеством сохранности газетных листов в подшивках и связанных с ним проблем приведения данных в машиночитаемый вид. Были выдвинуты два алгоритма действий при использовании указанного метода, призванные выделить наиболее оптимальный путь обработки данных и подготовки их к контент-анализу. По каждому из них была проделана обработка материала. В каждом из двух алгоритмов были выделены плюсы и минусы перевода информации в машиночитаемый вид. В результате проведенного исследования был выработан оптимальный алгоритм метода работы с оцифрованным материалом, в дальнейшем он был использован для обработки материала.

*Ключевые слова:* «Калининградская правда», контент-анализ, машиночитаемый текст, СМИ.

---

© Будеркина Е.Ю., 2017

Проведение как количественных, так и качественных исследований, требующих обработки большого массива источников, содержащих изначально не структурированную и не оцифрованную информацию, требует решения задач по оцифровке источника, приведения его в машиночитаемый вид с последующим анализом информации с помощью математических методов. Проведение контент-анализа с использованием современных пакетных программных комплексов позволяет значительно сократить время исследования и минимизировать трудозатраты исследователя. И.С. Винокурова отмечала, что обработка текстовой информации в итоге должна учитывать «тип исходной информации, конечный результат анализа и способ перехода от первичных данных к конечному результату» [1]. Она же приводит утверждение М. Майлса и А. Хубермана, в соответствии с которым при анализе информации существуют следующие задачи:

- сжатие информации;
- демонстрация данных через описание и представление материала в виде теоретических схем;
- формулирование итоговых заключений, их верификация [1].

Информационным источником в нашем случае является региональная газета «Калининградская правда», исследуемый период – с 1946 по 2016 г. Оцифровка издания проводилась с 2015 г. Центром социально-гуманитарной информатики БФУ им.И.Канта, по итогам работы которого были оцифрованы номера газеты с 1946 по 1991 г. Оцифровка велась с использованием планетарного сканера BookEye-3 в формате JPEG с максимальным разрешением 300 dpi. При этом страницы газет в ранние годы имеют механические повреждения, среди которых: смятия, пометки карандашом и ручкой, склейки клейкой лентой, разрывы. Эти обстоятельства затрудняют перевод полученных электронных данных в машиночитаемый вид и ставят необходимость определения наиболее оптимального пути приведения информации в требуемый вид и последующей подготовки ее к количественному и качественному контент-анализу.

Были проверены два алгоритма действий. В соответствии с первым оцифрованные источники в количестве 54 номеров анализировались с использованием программного обеспечения ABBYY FineReader 11 без последующей дополнительной редактуры со стороны исследователя. В этом случае был получен массив машиночитаемых данных, пригодных для проведения контент-анализа, при этом не требовались дополнительные временные и трудовые затраты. Однако в связи с неудовлетворительной сохранностью номеров газеты за исследуемый период программное обеспечение создает массив данных с количеством ошибок в среднем варьирующееся от 20 % до 40 %. Подобный материал не годится для проведения контент-анализа.

В рамках второго алгоритма авторы добавляли процедуру редакторской правки исследователем после обработки информации программой АБВУУ FineReader 11. В этом случае время обработки одного номера газеты увеличилось с одного часа до суток в зависимости от качества исходного материала. Полученные данные оказались предпочтительнее для использования в контент-анализе.

Поскольку для проведения исследования авторам необходима лишь некоторая часть от общего количества статей в газете, оптимальным методом был выбран следующий:

- выделение необходимых материалов;
- обработка с помощью специализированного программного обеспечения;
- проведение контент-анализа на основе полученных данных.

В ходе исследования был выделен наиболее оптимальный метод подготовки и обработки информации к последующему проведению количественного и качественного контент-анализа. Данный метод апробирован на архиве газеты «Калининградская правда» с 1946 по 1950 г. в рамках диссертационного исследования по выявлению образа региона.

### **Библиографический список**

1. Винокурова И.С. Применение компьютера для контент-анализа социологической информации. URL: <http://www.sociologos.ru/upload/File/Vinokurova.pdf> (дата обращения: 20.03.2017).
2. Проект Л-2015-73540 Разработка учебного ресурса «Онлайн архив региональной городской прессы советского периода». URL: <http://mizar2.kantiana.ru/> (дата обращения: 15.03.2017).

### **METHOD OF PRINTED MEDIA PROCESSING FOR THE FUTURE CONTENT ANALYSIS IN CONDITION OF POOR MATERIAL PRESERVATION (ON EXAMPLE OF THE POST-WAR KALININGRAD MEDIA)**

*Buderkina E. Yu.*

Immanuel Kant Baltic Federal University, 14, A. Nevskogo st.,  
Kaliningrad, Russia, ebuderkina@mail.ru

The paper demonstrates a method of processing the printed mass media based on the “Kaliningradskaya Pravda” newspaper of 1946-1950 for the future quantity and quality content analysis. The need for such an analysis is based on the quality of preservation of printed paper version in folders and the following problems with machine adaptation made in such conditions. Two ways of doing that were developed in connection with such method in order to find the most optimal way of material processing and following preparation for content analysis. Based on the above mentioned ways, the processing of materials was made. In both ways, their pluses and minuses were pointed out in regard to machine adaptation process. As a result,

the optimal way of dealing with such material was identified and used for material review in later periods.

*Key words:* “Kaliningrad Pravda”, content analysis, machine adapted text, mass-media.

УДК 930.23

## **ОЦИФРОВКА КАК ИНФОРМАЦИОННОЕ МОДЕЛИРОВАНИЕ: ОСВОЕНИЕ МНОГОСЛОЙНОСТИ ЭЛЕКТРОННЫХ ФОРМАТОВ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ<sup>1</sup>**

*Володин А.Ю.*

Московский государственный университет им. М.В. Ломоносова,  
Россия, Москва, Ломоносовский просп., 27, корп. 4,  
volodin@hist.msu.ru

В статье рассматривается проблема информационного моделирования как важного исследовательского этапа в процессе создания оцифрованных архивов исторических источников в контексте «цифрового поворота» в исторической науке. Реконструктивность исторических знаний требует особого внимания к разработке принципов эвристики, критики и интерпретации оцифрованных архивов. Новой чертой цифровой эпохи становится подход к исследовательским материалам как к данным, требующим компьютерной обработки, визуализации и интерпретации. Особенность «цифрового поворота» можно заметить в том, что разные типы данных на практике чаще всего создают неожиданные гибриды форматов представления ретроспективной информации. Особое внимание уделяется рассмотрению важной составляющей «цифровой реальности» историка – многослойности современных электронных форматов, позволяющей существенно повысить информативную отдачу оцифрованных исторических источников. Слой как совокупность объектов цифрового файла, объединенная группой признаков, позволяет на практике реализовать давнюю идею сохранения всей истории изменений, исправлений, искажений любого исторического источника.

*Ключевые слова:* оцифровка, информационное моделирование, многослойность, электронные форматы.

Историческое исследование по сути является реконструкцией, соединяющей воедино три уровня знаний: энциклопедические, исследовательские и источниковые. Историческая реконструкция основывается на моделировании ключевых черт и связей, с одной стороны, исторических явлений и процессов, а с другой – исторических источников и их информационного потенциала. Исторические источники в процессе оцифровки становятся многослойными контейнерами, вклю-

---

© Володин А.Ю., 2017

<sup>1</sup> Публикация подготовлена при финансовой поддержке РФФИ (грант № 17-01-50134).

чающими факсимильную, дипломатическую и метаинформацию. Таким образом, историческое исследование, реконструктивное по своей сути, становится вариантом информационного моделирования как на уровне анализа, так и на уровне синтеза. Причем историческая реконструкционная модель основывается на ретроспективной информации, представительность которой требует проверки и обоснования.

Сегодня оцифровка стала обыденной практикой историков-исследователей [2]. По этой причине работа с цифровыми источниками (причем с практической точки зрения уже не так важно, идет речь о файле-оригинале или электронной копии аналогового источника) становится важным и актуальным вопросом методологии исторического исследования. Если сформулировать эту задачу в традициях источниковедческой школы акад. И.Д. Ковальченко, то «главная из прикладных задач источниковедения состоит во все более явственно обнаруживающейся потребности *повышения информативной отдачи источников*» [3, с. 120; курсив мой – А.В.]. Первые попытки обобщения возможностей и ограничений «цифрового поворота» уже предприняты в мировой историографии – Р. Розенцвейгом и Д. Коэном – по-английски [10], С. Нуаре – по-французски [6] и Г. Коллером – по-немецки [8].

Под информационной моделью можно понимать структурированную информацию, характеризующую существенные свойства и состояния конкретного объекта или процесса и его связи с другими объектами или процессами. Как справедливо отмечает основатель исторической информатики акад. Л.И. Бородкин, «когда говорят о применении моделирования в истории, имеют в виду, как правило, области социально-экономической или политической истории. Однако компьютерные модели успешно применяются и в историко-культурных исследованиях, на базе уже не количественных (статистических) данных, а текстовых и визуальных источников. Важность последних актуализируется и в контексте «визуального» и «цифрового» поворотов, переживаемых исторической наукой» [1, с. 6–7]. Понятие информационной модели может оказаться полезным прежде всего по той причине, что при оцифровке исторической информации возникает сложное соотношение разнотипной информации – вербальной, изобразительной, пространственной, потоковой.

Оцифровка существенно изменила ключевые практики историков-исследователей: эвристику, критику и интерпретацию. Эвристика приобрела черты не только поиска в местах хранения источников (архивах, библиотеках, музеях), но и поиска в сетях общего доступа. Критика источника требует выявления в цифровой среде таких новых черт, как аутентичность (свойство электронного документа, гарантирующее его идентичность заявленному), мультимедийность (сплете-

ние воедино текста, изображений, звука и видео) и интероперабельность (способность взаимодействовать и функционировать с разными программными продуктами). Интерпретация источников ставит новые задачи к «связности» знания, главной технической составляющей которой становится смысловая разметка электронных документов, позволяющая их соединять в том числе и на основе многомерной скрытой информации [5].

Условно можно выделить несколько типов цифрового представления исторической информации: текст, таблица и изображение, а вслед за ними и динамические потоковые форматы – аудио и видео, и далее – программные коды. Особенность «цифрового поворота» можно заметить в том, что эти типы на практике чаще всего создают неожиданные гибриды. Простой пример: оцифрованная средневековая рукопись первоначально становится изображением, а процедура установления текста, даже если в ней и принимают участие компьютерные технологии, например, алгоритмы распознавания или таблицы базы данных, в основном согласуется с классическим текстологическим подходом. Таблица придает структуру информации самого разного характера, а связанные таблицы уже могут превратиться в базу данных. Важным свойством, которое позволяет увидеть перспективы цифрового подхода к историческим источникам, является многослойность цифровых форматов.

Слой, по определению, – это совокупность объектов цифрового файла, объединенная каким-либо признаком или группой признаков. Метафора цифрового слоя стала классическим инструментом сначала в работе с компьютерной графикой (например, в программах Photoshop, 3dsMax), затем в географических информационных системах (слой однотипных пространственных объектов в пределах некоторой территории и в единой системе координат); теперь это понятие слоя становится залогом интерактивности (например, в археографической разметке при установлении текста для сохранения локационной связи с оригиналом). Многослойность информации, учитывая возможности установления связей как отдельных слоев, так и отдельных элементов слоев с другими произвольными объектами, можно считать самостоятельной собственно цифровой составляющей современных исследовательских практик, с одной стороны, усложняющей, но, с другой – и существенно обогащающей традиционные подходы за счет многомерности и интерактивности. Причем, метафора слоя позволяет реализовать давнюю идею сохранения всей истории исправлений или искажений исторического источника, когда с открытием файла можно обратиться к любой из его существовавших версий (например, так, как сохраняются версии исправлений статей в Википедии).

Необходимость сочетать структурированную и графическую информацию, слабоструктурированную или неструктурированную информацию обеспечивает основу для мультимедийного понимания современных электронных многослойных ресурсов. Многообразие форматов современной передачи информации во многом скрывает от взгляда существенные различия, необходимые для их осмысления. Например, один и тот же файл может сообщить разное количество информации в зависимости от той программы, в которой он открыт. Исследователь оказывается в прямой зависимости от функционала программного обеспечения, связанной с гибридизацией цифровой информации и созданием метамедийных форматов [9].

Информационные технологии и цифровые достижения, напрямую касающиеся профессиональных нужд историков, не должны оставаться закрытым самодостаточным полем исследований отдельных специалистов, не должны стать маргинальным направлением хотя бы по той причине, что цифровая эпоха касается самих основ методологии и методики исторического исследования [7]. Важно, чтобы «цифровой поворот» не становился бы препятствием на пути развития общества ученых-историков, а открывал новые исследовательские перспективы и решал актуальные научные задачи, в том числе за счёт технического потенциала многослойных оцифрованных исторических источников.

### **Библиографический список**

1. *Бородкин Л.И.* Моделирование исторических процессов: от реконструкции к анализу альтернатив. СПб., 2016.
2. *Володин А.Ю.* Цифровая история: ремесло историка в цифровую эпоху // ЭНОЖ «История». 2015. № 08. DOI: 10.18254/S0001228-9-1
3. *Ковальченко И.Д.* Методы исторического исследования. 2-е изд. М., 2003.
4. *Cohen D.J., Rosenzweig R.* Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web. University of Pennsylvania Press, 2005.
5. *McGann J.* Marking Texts of Many Dimensions // *A New Companion to Digital Humanities.* Wiley-Blackwell, 2016. P. 358–376.
6. *Noiret S.* La digital history: histoire et mémoire à la portée de tous // READ/WRITE BOOK 2. Une introduction aux humanités numériques. Marseille: OpenEdition Press, 2012. P. 151–177.
7. *History in the digital age / ed. T. Weller.* London; New York: Routledge, 2013.
8. *Koller G.* Geschichte digital. Historische Welten neu vermessen. Kohlhammer, 2016.
9. *Manovich L.* Software Takes Command. Bloomsbury Academic, 2013.
10. *Rosenzweig R., Grafton A.* Clio Wired: The Future of the Past in the Digital Age. Columbia University Press, 2011.

## **DIGITIZATION AS INFORMATION MODELING: DEVELOPING OF MULTI-LAYERED ELECTRONIC FORMATS IN HISTORICAL RESEARCH**

*Volodin A. Yu.*

Moscow State University, 27-4, Lomonosovsky ave., Moscow, Russia,  
volodin@hist.msu.ru

The article deals with the problem of information modeling as an important research phase in the process of creating digitized archives from primary historical sources in the context of the digital turn in History. Reconstruction is the core feature of historical knowledge which requires special attention to the development of the principles of heuristics, criticism and interpretation of the digitized archives. A new feature of the digital age is the approach to research materials as to data, requiring computer processing, visualization, and interpretation. Another feature of the digital turn in History is that, in research practice, different hybrids of formats for the presentation of retrospective information have been created. Special attention is paid to the multi-layered electronic formats that enhance the informative impact of digitized historical sources and become an important part of contemporary digital reality of a historian. Layer as a collection of digital file objects, united by a group of features, makes it possible to put into practice an old idea of preserving the entire history of changes, corrections, and distortions of any historical source.

*Key words:* digitization, information modeling, multi-layering, electronic formats.

УДК 004; 930.2

## **ИСПОЛЬЗОВАНИЕ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ПОДХОДА ПРИ ПРОЕКТИРОВАНИИ ИСТОЧНИКО- ОРИЕНТИРОВАННЫХ СИСТЕМ<sup>1</sup>**

*Гилязева С.А.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, sof.gilyazeva@yandex.ru

Описывается возможность применения объектно-ориентированного подхода при проектировании источникно-ориентированных информационных систем на примере разработки системы «Указатели к стенографическим отчетам». Обоснован выбор объектно-ориентированного подхода при описании структуры исторического источника – указателя. Названы причины использования технологии XML в качестве способа хранения информации в системе. Приведен пример выделения структурных единиц источника как объектов информационной системы. Описаны требования к информационной системе как инструменту хранения и обработки информации исторического источника.

*Ключевые слова:* информационная система, источникно-ориентированная система, технология XML, XSD.

---

© Гилязева С.А., 2017

<sup>1</sup> Исследование выполнено при поддержке РФФИ (грант № 16-06-00569, руководитель Кирьянов И.К.)

Для разработки источникно-ориентированной информационной системы можно использовать объектно-ориентированный подход.

Источникно-ориентированная информационная система основывается на структуре источника. Для описания структуры исторического источника и дальнейшего проектирования ИС можно применить объектно-ориентированный подход. Исторический источник – это физический объект, представляющий собой совокупность разнородных объектов с определенными свойствами и связями между ними. Результатом объектно-ориентированного проектирования структуры источника является концептуальная модель, на основе которой затем можно строить диаграмму классов с учетом технологий, используемых для реализации ИС [1, 2].

Указатели к стенографическим отчетам имеют строгую иерархическую структуру. Из массива указателей можно выделить типы указателей, состоящих из множества определенных объектов. Например, при моделировании указателей выделены такие объекты, как *Персона* (объект агрегирует в себе социокультурные характеристики), *Депутатская активность*, *Предмет*. Личный алфавитный указатель представляет собой перечень объектов *Персона*, а список членов также содержит информацию о персоналиях, но упорядоченную по территориальным единицам. Поэтому оба этих указателя состоят из одинаковых объектов (*Персона*), но данный объект находится на разных уровнях иерархии.

Объекты с четко прослеживаемой иерархической структурой удобнее хранить в формате XML (eXtensible Markup Language – расширяемый язык разметки), чем с помощью реляционных баз данных. Указатель представляет собой упорядоченный массив объектов – порядок следования объектов также легко отразить с помощью XML. Еще одним аргументом является обилие пустых или отсутствующих элементов на разных уровнях иерархии, которые в случае использования разметки можно опустить. Часть данных в ходе жесткой типизации при приспособлении модели к реляционным базам данных может потеряться или исказиться, поэтому предпочтительнее использовать XML, так как язык разметки позволяет не типизировать элементы вообще или типизировать некоторый набор [3].

XML-документы могут быть двух видов: корректно сформированные (well-formed) и действительные (valid). Корректно сформированные документы – это документы, которые удовлетворяют формальным требованиям языка. Действительные документы – это документы, которые, помимо формальных требований, удовлетворяют также пользовательским требованиям, описанным с помощью одного из способов: определение типа документа (DTD) и определение схемы XML (XSD). Определения представляют собой дополнительные фай-

лы, которые содержат описание используемой системы тегов, уровни вложенности, данные об обязательности элементов и количестве вхождений, используемых пользовательских типах, ограничениях и т. д. С помощью этих файлов можно проверять XML-документы на целостность, что может быть критичным при последующей автоматизированной обработке документов [4].

Источнико-ориентированная информационная система должна:

1. Обеспечивать извлечение и разметку информации с образа исторического источника и преобразовывать его в структуру, пригодную для дальнейшего хранения и использования в ИС.

2. Обеспечивать поиск информации в источнике, настройку представления результатов поиска, их визуализацию и верификацию, а также экспорт полученных данных.

Информационная система предполагает наличие двух ролей: оператор и пользователь.

Оператор с помощью модуля извлечения информации может обработать образы исторического источника, а именно:

- разметить текст образа в соответствии с разработанной системой тегов;
- исправить содержимое полученных тегов и при необходимости дополнить тегами, несущими информацию о физических особенностях исторического источника (качество, сохранность, повреждения);
- сохранить полученную разметку в файле в формате XML.

Итоговый файл перед сохранением проходит валидацию для того, чтобы при дальнейшем его использовании в информационной системе исключить возможные сбои и ошибки.

Пользователь с помощью ИС может:

- осуществлять разные виды поиска информации (поиск по заданным параметрам, полнотекстовый поиск);
- настраивать представление результатов поиска – выбирать те свойства объектов, значения которых будут присутствовать в результирующей выборке;
- экспортировать полученные результаты поиска в один из поддерживаемых форматов;
- просматривать файлы источника (как образ источника, так и XML-файл);
- верифицировать файл источника с образом.

### **Библиографический список**

1. *Гаспарян М.С., Лихачева Г.Н.* Информационные системы и технологии: учеб. пособие. М.: Евразийский открытый институт, 2011.
2. *Историко-ориентированные* информационные системы. URL: <http://digitalhistory.ru/> (дата обращения: 10.04.2017)

3. *Школы* Консорциума W3C. Семейство XML. URL: <http://xml.nsu.ru/> (дата обращения: 10.04.2017).

4. *Сайт Г. Бермана*. XML&QT. URL: <http://genberm.narod.ru/> (дата обращения: 10.04.2017).

## USE OF OBJECT-ORIENTED APPROACH IN DESIGNING SOURCE-ORIENTED SYSTEMS

*Gilyazeva S.A.*

Perm State University, 15, Bukireva st., Perm, Russia,  
[sof.gilyazeva@yandex.ru](mailto:sof.gilyazeva@yandex.ru)

The paper describes the possibility of application of the object-oriented approach in designing the source-oriented information systems. The use of approach is demonstrated by the example of the source-oriented system. The author substantiates the choice of the approach in describing the structure of such historical source as directories. The reasons of using XML technology as an information storage method are described. The example of allocation of source structural units as objects of the information system is given. The requirements to the information system as the instrument of storage and processing of historical source's information are described.

*Key words:* information system, source-oriented system, XML, XSD.

УДК 94(47).084.1

## «БУМАЖНЫЕ ВОЙНЫ В 1917 Г.»: ОПЫТ КОНТЕНТ-АНАЛИЗА АГИТАЦИОННЫХ ЛИСТОВОК КАДЕТОВ И ЭСЕРОВ

*Иванков Э.В.*

Балтийский федеральный университет им. Иммануила Канта, Россия,  
Калининград, ул. Чернышевского, 56 «а», [eduard.ivankov.2013@mail.ru](mailto:eduard.ivankov.2013@mail.ru)

Предвыборная политическая кампания 1917 г. как часть общего революционного процесса отличалась богатой вариативностью и разнообразием форм. Период с февраля по октябрь стал новым этапом в развитии печатной агитации и пропаганды. Цель настоящего исследования – контент-анализ агитационных листовок Партии народной свободы и Партии социалистов революционеров, хранящихся в коллекции «От февраля к октябрю» Российской Национальной библиотеки. В коллекции РНБ представлены уникальные листовки всех политических партий и общественных сил, принимавших участие в предвыборной борьбе за созыв Учредительного собрания.

*Ключевые слова:* Февральская революция, предвыборная агитация, листовки, избирательная кампания, контент-анализ, Октябрьский переворот, Учредительное собрание.

Основным информационным и агитационным источником для населения по-прежнему оставалась периодическая печать. Вместе с тем, мощным инструментом пропаганды стали агитационные листов-

ки. В доступной и концентрированной форме листовки доносили до населения основные лозунги, программные и тактические установки партий и, что немаловажно, содержали первые отклики о происходивших в стране событиях и их оценку.

Предвыборная агитация на всеобщих выборах 1917 г. изучена весьма слабо и поверхностно. Советские работы акцентировали главное внимание на организационно-политической деятельности РСДРП (б). Большевистские листовки достаточно широко введены в научный оборот, собраны в специальные каталоги [1].

Избирательная агитация остальных партий изучена в меньшей степени. В современных работах внимание уделяется составу и структуре партий, их программам, тактике, отчасти исследователи затрагивают взаимоотношения с избирателями, через призму пропаганды [2]. Специальных работ, посвященных агитационным листовкам 1917 г., фактически нет, за исключением исследования А.М. Кулегина [3].

Цель настоящего исследования состоит в контент-анализе кадетских и эсеровских листовок, хранящихся в коллекции «От февраля к октябрю» Российской Национальной библиотеки. В коллекции РНБ представлены листовки всех политических партий и общественных объединений, принимавших участие в предвыборной борьбе (2000 названий).

Предвыборная кампания как часть общего революционного процесса отличалась богатой вариативностью и разнообразием форм. Период с февраля по октябрь 1917 г. стал новым этапом в развитии печатной агитации и пропаганды. Россию буквально захлестнуло листовками, содержащими весь спектр политических проблем: вопросы войны и мира, власти и межпартийной борьбы на выборах [3, с. 10].

Наибольшую активность на выборах проявляли эсеры и кадеты, массово распространявшие листовки. Наличие финансов играло решающую роль в их агитационных возможностях. Среди социалистических партий эсеры были наиболее обеспеченными. Помимо доходов от издательства, источниками финансирования для партии были кружечные сборы и пожертвования, лотереи, митинги и другие массовые мероприятия. Эсеры получали солидную финансовую помощь от Американского Красного Креста [2, с. 102].

Тонкую предвыборную игру в 1917 г. вела ПНС. В отличие от остальных, кадетам не требовалось много времени для предвыборной мобилизации: у них был богатый опыт думских кампаний. Они быстро наладили эффективную агитационную сеть. Не было недостатка ни в средствах, ни в ораторах, ни в публицистах. В отличие от социалистов, кадетов не терзали внутренние конфликты, что позволяло им расчетливо тратить свои силы на агитацию [2, с. 113].

Условно все листовки можно разделить на 3 вида: 1) листовки-обращения, направленные к широким массам; 2) листовки-брошюры, содержащие политические программы; 3) листовки-плакаты, с помощью которых воздействовали на население визуальным образом. Самыми распространенными были листовки-обращения. Их материалы по тематике можно подразделить на 4 группы: 1. О земле, 2. О войне, 3. О выборах в местные органы земств и городских дум, 4. О выборах в Учредительное собрание.

Первая группа содержит предвыборные обращения ПНС и ПСР к крестьянству. В отличие от эсеров, кадетам было труднее найти подход к простому люду. Их предвыборные призывы носили декларативный характер: «Крестьяне! Если вы не хотите, чтобы у вас отобрали ваши надельные земли, голосуйте за список №5». Отвергая разом все социалистические варианты решения аграрного вопроса, кадеты предлагали обсудить этот вопрос в Учредительном собрании. Деревня осталась закрытой для кадетов [2, с. 114–115].

Ударным лозунгом эсеров была социализация земли, при этом земельный вопрос должен быть окончательно решен только Учредительным собранием. Впредь до этого все земли передавались в земельные комитеты, которые осуществляли ее распределение [4, с. 207–208].

Вторая группа листовок посвящена Первой мировой войне. Главным лозунгом кадетов был призыв: «Война до победного конца!». Их листовки со страстным патриотизмом призывали всех сплотиться перед врагом и отстоять свободу и независимость Родины. Внешнеполитическая позиция эсеров сводилась к скорейшему завершению войны, путем не сепаратного, а демократического мира, без аннексий и контрибуций. Главным лозунгом эсеров был «Демократический мир всему миру» [4, с. 207].

Третья группа листовок связана с выборами в городское самоуправление и уездные земства в мае-июне 1917 г. Для кадетского руководства они имели ключевое значение, поскольку противостоять власти Советов на местах должны были земства и городские думы. Подготовку к выборам ПНС провела с широким размахом, не давшим, впрочем, ожидаемого результата: в Петрограде они получили 21,9 % голосов, а в Москве еще меньше – 16,8 % [5, с. 139–140]. На их фоне эсеры выглядели более успешными: в 37 городах от них избрали 44 % гласных, а в 14 городах получили абсолютное большинство мест в думах, в остальных 29 городах обеспечили себе многочисленные думские фракции [5, с. 187–188].

Четвертая группа касается выборов в Учредительное собрание. После октября 1917 г. ответом на действия большевиков появилась лавина листовок других политических партий, осудивших Октябрьский переворот. Кадеты выпускали листовки, построенные в форме

диалога-спора («Спор большевика с кадетом»), использовали в агитации метод «от противного» («Долой кадетов»). Листовка «Враги народа и свободы» отличалась яркой антибольшевистской направленностью. Анализ текстов листовок четвертой группы позволяет сделать вывод, что между социалистами и кадетами не было общего антибольшевистского блока. В ходе предвыборной кампании в Учредительное собрание они подвергали друг друга, а не только большевиков острой критике.

С помощью контент-анализа были проанализированы 103 листовки ПНС и 102 листовок ПСР, изданных в марте-ноябре 1917 г. Большинство листовок освещают организационно-политическую работу партий в населении (56 %), выборы в органы местного самоуправления (28 %), выборы в Учредительное собрание (16 %).

Наиболее массовое распространение получили листовки-обращения, направленные либо ко всему населению («Свобода народу!»), либо к отдельной категории населения (у кадетов «Граждане-земледельцы», у эсеров «Товарищи-рабочие») или определенной социальной группе («Хозяйки, кухарки и горничные!» и т. п.).

#### **Библиографический список**

1. *Листовки* Великого Октября: Каталог из коллекции Музея революции. СССР. М., 1975.
2. *Протасов Л.Г.* Всероссийское Учредительное собрание: История рождения и гибели. М.: РОССПЭН, 1997.
3. *Кулегин А.М.* Политические листовки 1917 года в Петрограде как исторический источник: автореф. дис. ... кандидата исторических наук. СПб., 1993.
4. *Политические партии России: история и современность.* М., 2000.
5. *История политических партий России* / под ред. А.И. Зевелева. М.: Высшая школа, 1994.

#### **“THE PAPER WAR IN 1917”: EXPERIENCE OF CONTENT ANALYSIS OF THE CADETS AND THE SOCIALIST REVOLUTIONARIES LEAFLETS**

*Ivankov E.V.*

Immanuel Kant Baltic Federal University, 56, Chernyshevskogo st.,  
Kaliningrad, Russia, eduard.ivankov.2013@mail.ru

The election political campaign of 1917 as a part of the general revolutionary process differed in rich variability and a variety of forms. The period from February to October became a new stage in development of printing propaganda and promotion. The research is a content analysis of propaganda leaflets of the People's Freedom Party and of the Party of Socialist Revolutionaries from the «From February to October» collection of the National Library of Russia. The unique leaflets of all political parties and public forces taking part in pre-election fight for convocation of the Constituent assembly are represented in a collection.

*Key words:* February revolution, pre-election campaign, leaflets, election campaign, content analysis, October revolution, Constituent Assembly.

## **ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ ИЗУЧЕНИИ ПАРЛАМЕНТСКОЙ ЭЛИТЫ ПОЗДНЕИМПЕРСКОЙ РОССИИ<sup>1</sup>**

*Кирьянов И.К., Сотник А.В.*

Балтийский федеральный университет им. Иммануила Канта,  
Россия, Калининград, ул. Невского, 14, EBaranova@kantiana.ru

В позднеимперской России произошло усложнение политической элиты вследствие включения в ее состав депутатов Государственной Думы и выборных членов Государственного Совета. Основой для изучения когорты депутатов Государственной Думы четвертого созыва с повторяющимся парламентским статусом, образовавшей отдельную субэлитную группу, стал Интернет-портал «Парламентская история позднеимперской России». Исследование базируется на междисциплинарном подходе, в рамках которого применялись методы дискриптивной статистики, корреляционного анализа, фасетной классификации. Анализ полученных результатов показал, что когорта депутатов с повторяющимся парламентским статусом была более однородной по своим социокультурным характеристикам, чем когорта впервые избранных депутатов. Существенные различия в социокультурном облике этих групп определялись сословной принадлежностью, уровнем образования, допарламентскими профессиями, связанными с интеллектуальным трудом.

*Ключевые слова:* позднеимперская Россия, Государственная Дума, парламентская элита, корреляционный анализ, фасетная классификация.

Характерной чертой современной историографии парламентской истории позднеимперской России является расширяющееся применение междисциплинарных исследовательских практик. Исследователи, не удовлетворяясь применением только традиционных методов исторической науки, обращаются, в частности, к методам математической статистики (см., например: [1; 2; 3]). При этом следует учитывать то обстоятельство, что материалы парламентского делопроизводства по своей природе адаптированы к применению современных информационных технологий, что позволяет совершить в их отношении переход от традиционного источниковедения к компьютерному или e-источниковедению (см., например: [4; 5]). На развитие подобного междисциплинарного синтеза, характерного для такого научного направления, как Digital History, нацелена реализация проекта «Формирование парламентской субэлиты в позднеимперской России».

---

© Кирьянов И.К., Сотник А.В., 2017

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ, грант № 16-06-00569 «Формирование парламентской субэлиты в позднеимперской России» (руководитель Кирьянов И. К.)

Первый этап проекта связан с изучением социокультурных характеристик депутатов. В составе парламентской элиты поздней имперской России особую роль играла когорта депутатов Государственной Думы с повторяющимся парламентским статусом, образовавшая отдельную субэлитную группу. Представители данной группы отличались от других парламентариев степенью политического влияния и выделялись среди остальных депутатов, в частности четвертого созыва (1912–1917 гг.), своими социокультурными характеристиками, предполагавшими исполнение этой особой роли.

Информационной, инструментальной и технологической основой исследования выступает интернет-портал «Парламентская история поздней имперской России» ([www.parliament.psu.ru](http://www.parliament.psu.ru)). Один из его модулей – «Просопографическое исследование» – позволяет получать данные о социокультурных характеристиках депутатов для выявления признаков, в наибольшей степени определявших облик парламентской субэлиты.

Для удобства обработки данные были организованы на основе локальной реляционной базы данных и электронных таблиц, благодаря чему стало возможным применение как методов дескриптивной статистики, так и многомерного статистического анализа, в частности корреляционного анализа, фasetной классификации.

За время деятельности Думы четвертого созыва ее депутатами всего стали 462 человека, включая тех, кто замещал выбывших по тем или иным причинам депутатов, при этом 143 думца (31 %) обладали повторяющимся парламентским статусом – наивысший показатель в сравнении с предыдущими созывами. На состав субэлитной группы членов Государственной Думы четвертого созыва оказали влияние траектории получения повторяющегося парламентского статуса.

Корреляционный анализ позволил выявить социокультурные характеристики, в наибольшей степени связанные с повторным избранием. Например, показатели связи когорты депутатов с повторяющимся парламентским статусом, с такими признаками, как православное вероисповедание и принадлежность к великороссам, не имели принципиальных отличий от всего депутатского корпуса. Между тем в результате анализа полученных данных был сделан вывод о том, что среди субэлиты выделяются депутаты с иной этнической и конфессиональной принадлежностью – наиболее показателен в этом отношении сегмент поляков-католиков в субэлитной группе.

К отличительным характеристикам парламентской субэлиты следует отнести высокий уровень образования. Вполне естественно то, что уровень образования депутатов в целом значительно превышал средний для Российской империи, однако представителей субэлиты, как правило, отличал более высокий уровень образования среди пар-

ламентариев. Обращаясь к профессиональным додумским занятиям депутатов, выраженным через укрупненные признаки, можно выделить относительно устойчивую положительную связь между обладанием повторным парламентским статусом и занятием преимущественно интеллектуальным трудом (присяжные поверенные, преподаватели вузов, учителя, журналисты, врачи, агрономы, инженеры). Сословная принадлежность в большей степени, чем иные выделенные категории, была связана с переизбранием, в особенности принадлежность к дворянскому сословию. Несмотря на то, что преобладание дворян в Думе было следствием действовавшего избирательного закона, связь привилегированного социального положения и переизбрания говорит об особом значении этой характеристики для облика профессионального политика: высокое социальное положение представителей субэлиты связано не только с общей для парламента диспропорцией представительства социальных групп, но и с общественными ожиданиями, обеспечивающими повторное избрание в условиях высокой конкуренции.

На основе совокупности результатов корреляционного анализа можно сделать вывод о том, что отдельные социокультурные характеристики присущи в большей степени субэлите, чем остальной части депутатского корпуса, и обнаруживают связи с повторяющимся парламентским статусом. Поскольку субэлита оказалась достаточно неоднородной, о чем свидетельствуют сложные связи между рассматриваемыми признаками, продолжение исследования было связано с выделением социокультурных типов в составе как когорты депутатов с повторяющимся парламентским статусом, так и всего думского корпуса четвертого созыва, для чего был применен метод фасетной классификации.

В качестве фасетов выступили выделенные социокультурные признаки. Полученные результаты свидетельствуют о том, что когорта депутатов с повторяющимся парламентским статусом была более однородной по своим социокультурным характеристикам, чем когорта впервые избранных депутатов. Результаты фасетной классификации также свидетельствуют о том, что сословная принадлежность и уровень образования оказывались значимыми характеристиками для переизбрания в члены Государственной Думы поздней имперской России.

Продолжение данного междисциплинарного исследования предполагает выявление взаимосвязей между социокультурными характеристиками и моделями политического поведения парламентской элиты, определение степени влияния депутатов с повторяющимся статусом на принятие решений в отечественном парламенте начала XX в.

## Библиографический список

1. *Алескеров Ф.Т., Кравченко А.С.* Распределение влияния фракций в Государственных думах Российской империи, 1905–1917 гг.: препринт WP7/2005/03. М.: ГУ ВШЭ, 2005. 20 с.
2. *Кирьянов И.К., Корниенко С.И.* Количественные методы исследования истории парламентаризма в России начала XX в. // Проблемы методологии и источниковедения // Материалы III Научных чтений памяти акад. И.Д. Ковальченко / отв. ред. С.П. Карпов. М.: Изд-во МГУ; СПб.: Алетейя, 2006. С. 483–486.
3. *Кравченко А.С.* Структурный анализ III Государственной Думы Российской империи: препринт WP7/2006/02. М.: ГУ ВШЭ, 2006. 24 с.
4. *Селунская Н.Б., Бородкин Л.И., Григорьева Ю.Г., Петров А.Н.* Становление российского парламентаризма начала XX века. М.: Мосгорархив, 1996. 283 с.
5. *Кирьянов И.К., Корниенко С.И.* Стенографические отчеты Государственной Думы начала XX века: от традиционного к компьютерному источниковедению // Вестник Челябинского государственного университета. Сер. История. 2009. Вып. 30. № 6. С. 144–148.

### POSSIBILITIES OF APPLYING INFORMATION TECHNOLOGY IN RESEARCH ON PARLIAMENTARY ELITE IN LATE IMPERIAL RUSSIA

*Kiryanov I.K., Sotnik A.V.*

Perm State University, 15, Bukireva st., Perm, Russia,  
ikiryanov@yandex.ru

The political elite of late Imperial Russia became more complicated due to the inclusion of the State Duma deputies and the elected members of the State Council in its membership. Study of the Fourth Duma deputies with the repeated parliamentary status, who constituted a distinct sub-elite group, was conducted on the basis of the Internet-portal “Parliamentary history of Late Imperial Russia”. The research is based on the interdisciplinary approach, which suggested applying historical computing methods, like descriptive statistics, correlation analysis and faceted classification. The study showed that the cohort of deputies with the repeated parliamentary status was more homogeneous in terms of socio-cultural characteristics than the cohort of newly elected deputies. The major differences in the socio-cultural portrait of those groups were revealed in the estate their members belonged to, level of education and pre-parliamentary occupation.

*Key words:* Late Imperial Russia, State Duma, parliamentary elite, correlation analysis, faceted classification.

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИЗУЧЕНИИ  
МЕТРИЧЕСКИХ КНИГ БАРНАУЛА ВТОРОЙ ПОЛОВИНЫ  
XVIII В. (К МЕТОДИКЕ КОДИРОВАНИЯ И ИДЕНТИФИКАЦИИ  
ЗАПИСЕЙ ОБ УЧАСТНИКАХ ЦЕРКОВНЫХ ОБРЯДОВ)<sup>1</sup>**

*Колдаков Д.В., Сарафанов Д.Е.*

Алтайский государственный университет,  
Россия, Барнаул, просп. Ленина, 61, dima@hist.asu.ru,  
sarafanov@hist.asu.ru

Объединение разновременных данных на основе персональных сведений из исторических источников является одной из актуальных исследовательских задач в области исторической демографии. Для ее решения рассматриваются возможности использования русифицированной версии фонетического алгоритма Metaphone для кодирования данных участников церковных обрядов на примере приходов г. Барнаула второй половины XVIII столетия. Представлены особенности источников данного периода. Анализируются результаты кодирования фамилий, имен и отчеств. Показано, что применение кодированных значений, позволяет получить более представительную выборку из повторяющихся записей по сравнению с обычным поиском дубликатов вышеуказанных атрибутов, взятых как отдельно, так и в их сочетаниях. Предлагается комбинированная методика кодирования данных в условиях неполноты атрибутивной информации как подготовительный этап для дальнейшей идентификации персоналий и связывания записей из разных метрических книг.

*Ключевые слова:* историческая демография, метрические книги, фонетические алгоритмы, Metaphone, базы данных, идентификация персоналий, связывание записей.

Историческая демография в России сегодня является быстро развивающимся исследовательским направлением. В различных регионах сложились определенные историографические традиции в плане выбора проблематики, используемой методологии и методики исследования. Несмотря на то, что сегодня многими исследователями активно изучаются характеристики латентной информации метрических книг, тем не менее, большая часть показателей рассчитываются пока без привязки к «имени» человека.

В исследованиях, базирующихся на объединении разновременных данных, где основу составляют персональные сведения об одном и том же лице или прямых предках/потомках, существенными факторами на этапе установления дистанционных связей выступают наличие опре-

---

© Колдаков Д.В., Сарафанов Д.Е., 2017

<sup>1</sup> Исследование выполнено при финансовой поддержке РГНФ в рамках научного проекта №15-01-00207а «“Новая историческая демография”: становление, развитие, перспективы».

деленного (критического) объема записей в базе данных и «информационная полнота» источников (наличие дополнительных атрибутов – сведений о ФИО, возрасте, сословии, чине/звании, профессии и пр.).

Человек, участвовавший в каком-либо из церковных обрядов (крещение, венчание, отпевание), и зафиксированный в метрической книге, на протяжении последующих лет жизни мог фигурировать в источниках неоднократно. Записи о нем могут быть оформлены так, что поиск по шаблону и последующее сопоставление их в разных записях может оказаться недостаточно эффективным и результативным (например, при разном написании одних и тех же фамилий, имен, отчеств, в том числе при ошибках набора). Поэтому связывание записей более оптимально при идентификации персоны путем присвоения ей уникального кода.

В связи с этим подготовительным этапом для такой идентификации может стать кодирование соответствующих атрибутов персоны, с использованием, в частности, фонетических алгоритмов. Применение некоторых из них (например, Soundex) в практике связывания записей электронных версий исторических источников осуществлялось еще в начале 2000-х гг. [1]. Задачи сопоставления и идентификации объектов, относящихся к персональным данным, которые хранятся в информационных системах, весьма актуальны. Разрабатываются и предлагаются новые алгоритмы, в том числе те, в основе которых лежит известный алгоритм Metaphone [2, 3]. Русская адаптация его приводится в статье Петра Каньковски [4], представлена и программная реализация на VBA [5].

На материалах выборки базы данных «Население Барнаула второй половины XVIII – начала XX вв. по материалам метрических книг» мы протестировали этот алгоритм. Кодирование фамилий с его помощью дает достаточно хороший результат в случае различного их написания. Например, фамилиям «Кошкаргаев» и «Кошкоргаев» присваивается один код – «КАШКАРГА4», а однофамильцы разного пола кодируются практически одинаково – «Копылов» (КАПАЛ4) и «Копылова» (КАПАЛ9), что позволяет более эффективно осуществлять сортировку данных и искать с применением шаблонов.

Основу для анализа составили материалы метрических книг трех церквей г. Барнаула за вторую половину XVIII в. Источники сохранились за 1754, 1767, 1770, 1784, 1786, 1800 гг. Выбор раннего периода обусловлен «лучшими» условиями для выявления возможностей алгоритма Metaphone. В метрических книгах этого времени, на наш взгляд, чаще встречаются разночтения в написании фамилий духовенством, добавляются сложности прочтения скорописи второй половины XVIII в., что также может приводить к трансформации номинативных данных при занесении записей в БД.

Общее количество обрядовых записей составило 2 548. Однако количество персон, фиксируемых источниками, было больше. Так, в части «О родившихся» указывалась информация о четырех индивидах: новорожденном (ФИО), отце (ФИ), крестном (ФИО), крестной (ФИО). В таблице БД «Брачующиеся» встречаются данные о трех персонах: женихе (ФИО), невесте (ФИО), отце невесты (ФИ). В третьей части содержится информация об умершем (ФИО) и отце/муже умершего (ФИ). Отметим, что в источниках второй половины XVIII в. нет сведений о матери ребенка и поручителях жениха (2 чел.) и невесты (2 чел.), данные о которых начинают фиксироваться в XIX в.

Для реализации указанных выше процедур был выполнен ряд запросов на генерацию набора данных для всех десяти категорий персон. После чего произведено кодирование фамилий, при помощи алгоритма Metaphone, и последующая проверка корректности присвоения кодов. Была сформирована таблица, содержащая данные на 6 433 участников обрядов, после чего проведен поиск повторяющихся записей.

Повторяющиеся записи по коду фамилии дали результат в 5 252 записи, в то время как по полю, содержащему собственно фамилию – 5 167. Проведенная работа показала, что по Metaphone-коду находится большее число повторений от общей выборки данных (на 1,3 %). Алгоритм чаще «присваивает» один код для фамилий с различием гласных букв (Оболгина/Аболгина, Бурдычев/Бурдачев, Денисов/Дионисов). Однако это фиксируется не всегда. Например, фамилии Давыдов и Давидов имеют разные коды. Там, где встречаются вариации с согласной буквой, чаще сгенерированы разные коды (Бушков/Бушнов, Варнаков/Варнасов, Залуцкий/Залутцкий/Заруцкий).

Однако с именами и отчествами при кодировании их с использованием алгоритма Metaphone ситуация уже иная, и расхождение может быть весьма значительным, особенно, если сравнивать мужские имена и женские отчества с целью дальнейшего связывания записей для выявления отцов и дочерей и построения семейных карт: «Яков» (АКАФ) и «Яковлевна» (АКАВЛИВНА). В этом случае более оптимальным представляется кодирование имен и отчеств с помощью соответствующих таблиц-справочников, которые заполняются с учетом имеющихся в базе данных их написаний. Мы провели такое кодирование и осуществили поиск повторяющихся записей по комбинации нескольких полей. С Metaphone-кодом фамилии и кодом имени выявлено 2 055 записей, код фамилии/код имени/код отчества – 1 363 записи. Отметим, что при использовании собственно значений в связке фамилия/имя обнаруживается только 1 947 записей, в связке фамилия/имя/отчество – 1 282 записи.

Таким образом, кодирование с использованием фонетического алгоритма в сочетании с кодированием по словарям, несмотря на зна-

чительный объем «ручной» работы, позволило сделать более представительной выборку для дальнейшей идентификации участников церковных обрядов и связывания записей, которые станут следующим этапом нашего исследования.

### **Библиографический список**

1. *Варфоломеев А.Г., Сидоров Ю.В.* Об эффективности алгоритмов связывания записей // Новые информационные ресурсы и технологии в исторических исследованиях и образовании: сб. тезисов, докладов и сообщений Всероссийской конференции. М., 2000. С. 62–63.
2. *Гусятников В.Н., Палькин Е.А.* Эффективность алгоритмов сопоставления персональных данных // Программные продукты и системы. 2011. № 1. С. 100–103.
3. *Ставровиецкий В. В., Гапанюк Ю. Е., Галкин В. А.* Алгоритм нечеткого фонетического поиска на основе простых чисел // Молодежный научно-технический вестник № 07, июль 2012. URL: <http://sntbul.bmstu.ru/doc/457989.html> (дата обращения: 20.03.2017).
4. *Каньковски П.* «Как ваша фамилия», или Русский MetaPhone // Программист. 2002. № 8. С. 36–39.
5. *Форумы Aeroion.ru.* Нечеткий поиск информации. URL: <http://forum.aeroion.ru/topic461.html> (дата обращения: 20.03.2017).

### **INFORMATION TECHNOLOGY IN THE STUDY OF BARNAUL PARISH BOOKS OF THE SECOND HALF OF THE 18TH CENTURY (THE METHOD OF RECORDS CODING AND IDENTIFICATION OF PARTICIPANTS IN CHURCH CEREMONIES)**

*Koldakov D.V., Sarafanov D.E.*

Altay State University, 61, Lenina ave., Russia, [dima@hist.asu.ru](mailto:dima@hist.asu.ru),  
[sarafanov@hist.asu.ru](mailto:sarafanov@hist.asu.ru)

The combination of distant data on the basis of personal information from historical sources is one of the most topical research tasks in the field of historical demography. To solve this problem, researchers have studied the possibilities of using the Russian version of the phonetic algorithm Metaphone. The algorithm was used to encode the data of the church ceremony participants on the example of Barnaul parishes in the second half of the 18th century. The authors have come to a conclusion that the use of coded values makes it possible to obtain a more representative sample of duplicate records compared to the usual search for duplicates of the above attributes, taken both separately and in their combinations. A combined methodology for encoding data in the conditions of the incompleteness of attributive information is proposed as a preparatory stage for the further identification of personalities and the record linkage.

*Key words:* Historical Demography, Parish Books, Phonetic Algorithms, MetaPhone, Databases, Identification of Personalities, Record Linkage.

## META DESCRIPTION AS A KIND OF DISTANT READING OF PROVINCIAL PERIODICALS COLLECTIONS<sup>1</sup>

*Kornienko S.I., Gagarina D.A., Kharisova A.R.*

Perm State University, 15, Bukireva st., Perm, Russia,  
kornienko@psu.ru, dinara@psu.ru

Creation of information resources based on unique historical text sources is often faced with the impossibility of adequate OCR because of poor physical security, printing defects, fonts specificity and page structure. Appropriately, it is impossible to apply the full-text search to such documents or parts of them, which greatly reduces the possibility of effective search and analysis of information based on these resources. We can use images of PDFs of unrecognized sources, but they are available for “close reading” only. How to include all publications from the collection to “distant reading” process and whole text analytics? The method of extended meta description can successfully solve the problem of creating machine-readable versions of unrecognized items and to carry out information retrieval and analysis of such collections. The project is realized on the base of newspapers collections published in Perm province during the First World War, the 1917 Revolution and the Civil War in Russia.

*Key words:* newspapers, information system, meta data, distant reading, Russian history.

### **Formulation of the problem**

Creation of information resources based on unique historical text sources is often faced with the impossibility of adequate OCR because of poor physical security, printing defects, font’s specificity and page structure. Appropriately, it is impossible to apply the full-text search to such documents or parts of them, which greatly reduces the possibility of effective search and analysis of information based on these resources. We can use images of PDFs of unrecognized sources, but they are available for “close reading” only through the eyes on the monitor. The question here is how to include all publications from the collection to “distant reading” process and whole text analytics?

### **The project permnewspapers.ru**

Regional newspapers’ periodicals hold a firm place in the humanities. This kind of sources is especially significant at the level of micro-history and history of everyday life. At the same time, it is complicated to put provincial periodicals into scientific circulation due to problems of access, preservation, organization as well as effective search and analysis of information.

---

© Kornienko S.I., Gagarina D.A., Kharisova A.R., 2017

<sup>1</sup> The study is supported by Russian Foundation for Humanities (grant 16-11-59009). Исследование поддержано РГНФ, грант № 16-11-59009.

Center for Digital Humanities of Perm State University (Russia) in cooperation with Perm region museum have implemented a series of projects based on provincial newspapers of 19<sup>th</sup> – beginning of the 20<sup>th</sup> centuries. We select collections chronologically or thematically, digitize them and create information systems, which preserve and provide wide access to cultural heritage data as well as allow their scholarly usage (Gagarina 2016, Kornienko 2016, Kornienko 2015).

The source-oriented information system “Perm Provincial Periodicals” (<http://permnewspapers.ru/>) provides free access for scholars to images and full-texts of newspapers collections published in Perm province during the First World War, the 1917 Revolution and the Civil War in Russia. Now the system contains 10 newspapers collections, 2000 issues and 4000 publications of different types. The number of editions and publications is growing. We have digitized next part of sources that will be loaded to the information system soon.

The situation when we cannot OCR some issues or parts of them frequently occur with provincial newspaper periodicals published in Russia in the 19<sup>th</sup> – early 20<sup>th</sup> century. There are quite a large number of issues and publications in our newspapers collections, which cannot be recognized and presented in the information system in the form of electronic text in a machine-readable format according to the above reasons. The PDF-format with text below the image can be used to visualize such sources and this dual format creates the possibility to represent as much information of sources as possible source. In this case, the fragments that can be OCRed are included in the full-text search. Nevertheless, unrecognized fragments are not machine-readable and are not included in the full-text search and analysis across the entire collection. So the problem is not fully decided by PDF with text below the image.

### **An extended meta description as a solution**

Rather consistent solution of the problem can be achieved through the creation of an extended meta descriptions system of newspaper publications.

In this case, for the submission of the publication (or issue entirely) in addition to the PDF-format, we create a detailed metadescription, which includes a set of fields, allowing to attribute the publication on a content level. Thus, we create the information systems with electronic versions of publications shown in the image and text format as well as meta description with both formal and substantive attributes.

The structure of this publication meta description in addition to such formal elements as title, type, location in the issue, the author, etc. includes thematic rubrics and sub rubrics, keywords, personalities, place names and other meaningful parameters.

Subject-thematic rubrics indicate the relation between the publication and the main problems, events and processes that characterize the period of the sources and their local, regional or state level.

This method of creating thematic content metadescription of publications provides the possibility of distant reading with its approaches and methods, allows applying computerized methods of text processing.

#### **An example of thematic rubrics in the project and the conclusion**

As an example, there is a list of rubrics (and associated values of the attribute) for the publication in the period of the 1917 Revolution and the Civil War in Russia: *revolutionary movement, forms of social and political protest (information on strikes, demonstrations, rallies, etc.); the overthrow of the monarchy, the formation of the Provisional Government and the Petrograd Soviet; The Emperor and the imperial family; organization and activity of the Bolsheviks; Socialist parties in general; Constituent Assembly; the proclamation of independence of Ukraine. Central Rada; separation of Poland; separation of Finland; Russian conferences and congresses of the Soviets; the Central Executive Committee; organization and activities of the Mensheviks; organization and operation of the SRs; military operations and position at the front; attitude to the war of political parties; organization and operation of the Cadets; State Conference; organization and activities of the Provisional Government. Pre-Parliament; State power in the regions; the local Councils; the armed uprising in October 1917. October Revolution; the establishment of the Soviet regime; the first anti-Soviet organization and performance; organizing of anti-Soviet military units, the beginning of the Civil War; soviet coalition of Bolsheviks and left SRs; the creation of the Soviet state institutions; trade unions and public organizations.*

Thematic rubrics take into account the regional nature of the publication and provides rubricating associated with the local level. This lead to the rubrics, characterizing *the local government and self-government; relation to the center; the establishment of the Bolshevik regime in the region; peasant and working-class movement, the course of military operations in the region during the Civil War; red and white terror; the daily life of the province's population in the years of the Revolution and the Civil War.*

Thus, the method of extended meta description of newspaper articles can successfully solve the problem of creating machine-readable versions of unrecognized items and to carry out information retrieval and analysis of such collections.

#### **References**

1. Gagarina, Dinara, Kornienko, Sergey. 2016. Digital Editions of Russia: Provincial Periodicals for Scholarly Usage. In: Digital Editions: Academia, Cultural Heritage, Society. Cologne.
2. Kornienko, Sergey, Gagarina, Dinara. 2016. The First World War in Perm Provincial Periodicals. In: Digital Humanities 2016. Conference Abstracts. Kraków.

3. *Kornienko, Sergey, Gagarina, Dinara*. 2015. Information systems: new methods of Russian history sources study. In: International Multidisciplinary Scientific Conferences Social Sciences & Arts SGEM 2015. Sofia.

4. *Moretti Ф.* (Moretti, Franco). Дальнее чтение / пер. А. Вдовин, О. Собчук, А. Шеля; науч. ред. перевода: И. Кушнарева. М.: Издательство Института Гайдара, 2016.

### **МЕТАОПИСАНИЕ КАК МЕТОД «ДАЛЬНОГО ЧТЕНИЯ» КОЛЛЕКЦИЙ ГУБЕРНСКОЙ ПЕРИОДИЧЕСКОЙ ПЕЧАТИ**

*Корниенко С.И., Гагарина Д.А., Харисова А.Р.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, kornienko@psu.ru, dinara@psu.ru

Создание информационных ресурсов на основе исторических текстовых источников часто осложняется невозможностью качественного распознавания текста из-за плохой физической сохранности, дефектов печати, специфики шрифтов и структуры страниц. К таким документам невозможно применить полнотекстовый поиск, что значительно уменьшает возможности компьютеризированного анализа. Для работы с нераспознанными или плохо распознанными источниками можно использовать графические форматы или PDF-файлы, но в таком случае возможно только традиционное, «близкое» чтение текста. Расширенное метаописание отчасти решает проблему создания машиночитаемых версий трудно-распознаваемых источников, давая возможность выполнять поиск и другие виды анализа целых коллекций источников, тем самым являясь разновидностью «дального» чтения. Данный метод применяется в проекте, реализованном на базе коллекции газет, изданных в Пермской губернии во время Первой Мировой войны, Революции 1917 года и Гражданской войны в России.

*Ключевые слова:* периодическая печать, информационная система, метаданные, «дальнее чтение», Пермская губерния.

УДК 930.25

### **«ПРОЖИТО» ОТ РУКОПИСИ ДО КОРПУСА: СБОР, РАЗМЕТКА, АНАЛИЗ ДНЕВНИКОВЫХ ТЕКСТОВ**

*Мельниченко М.А.*

«Прожито», Россия, Москва, ул. Каретный ряд, 5/10,  
misha.melnichenko@gmail.com

*Тышкевич Н.Б.*

Национальный исследовательский университет «Высшая школа экономики», Москва, ул. Мясницкая, 20, natalie.tysh@gmail.com

Мы представляем «Прожито» (prozhito.org) – первый онлайн-корпус личных дневников, состоящий из 200 000 записей, имеющих привязку ко времени и

локации и представляющих личные нарративы XIX–XX вв. на русском и украинском языках. «Прожито» сочетает в себе технологии блоговых платформ и архивные традиции курирования личных архивов. Пользователь может работать не только с текстами определенных авторов, но и со всей коллекцией дневниковых записей, строя сложные поисковые запросы, указывая пол и возраст автора, типы дневников (например, военный, экспедиционный, детский и т. д.) и фильтруя результаты по точным датам и метоположению записей. Наш подход к публикации дневников делает Прожито с одной стороны профессиональным инструментом для современных ученых, а с другой – дает ключ к глобальной истории всем заинтересованным.

*Ключевые слова:* корпус, оцифровка, личные дневники, эго-тексты.

Несмотря на непреходящую актуальность исследований по личной и семейной истории источники личного происхождения оказываются обойдены вниманием исследователей. Ценнейшими из них являются дневники – эго-тексты, привязанные к хронологической линейке. Мы представляем первый онлайн корпус, построенный на материалах личных нарративов мужчин и женщин самых разных возрастов, культурных и социальных уровней, написанных в XIX–XX вв.

Задача волонтерского проекта «Прожито» – собрать в одной электронной библиотеке все личные дневники, как опубликованные, так и прежде неизвестные исследователям. Пользователи могут работать со всем массивом текстов, составляя сложные поисковые запросы по ключевым словам, и формировать произвольные выборки текстов с заданными внешними параметрами, выбирая привязку записи к месту и времени, пол и возраст автора, тематику дневника: например, тексты, написанные школьниками, экспедиционные дневники, фронтовые записки, документы из архива «Мемориала».

«Прожито» работает с двумя видами текстов – уже опубликованными и никогда прежде не публиковавшимися. При работе с опубликованными дневниками используются любые доступные публикации, как бумажные, так и доступные в сети их сканированные копии и распознанные тексты. Библиография опубликованных русскоязычных дневников, собранная участниками проекта, превышает 2 000 единиц. Из них в распоряжении «Прожито» уже есть около 900 текстов в виде бумажного издания, отсканированной книги, распознанного текста или издательского макета. На момент написания тезисов в систему загружено около 600 дневников общей совокупностью 200 000 подневных записей (25 млн словоупотреблений).

С момента открытия проекта в него было передано 60 рукописей разного объема (от одной до 80 тетрадей). Сбор опубликованных дневников, распознавание, разметка и публикация большого числа рукописных текстов – задача, которая может быть решена лишь с помощью привлечения большого количества участников. Вокруг «Про-

жито» мы успешно организовали волонтерское сообщество, в котором может принять участие любой человек, умеющий работать с текстовым редактором и готовый посвятить этому некоторое количество времени на регулярной основе.

Специально для привлечения непрофессионалов в «Прожито» разработан простой регламент работы с текстами, освоить который способен любой пользователь. Волонтеры набирают рукописные тексты, правят ошибки оцифровки, выставляют пропущенные даты. На следующем этапе редактор полуавтоматическим способом преобразует текст в XML-документ с указанием даты и места написания записи.

В случае работы с рукописями участники проекта снимают с них электронные копии и возвращают оригинальные документы владельцам. Однако на этом вклад наследников не заканчивается – они продолжают участвовать в работе с текстом на всех этапах и имеют право вычеркнуть из финальной версии любые фрагменты, публикацию которых по тем или иным причинам они считают невозможной. Сокращенный текст, утвержденный владельцем рукописи, загружается на сайт.

При работе с дневниками участники проекта не дают текстам свои оценки и не снабжают их комментариями. Команде «Прожито» интересна подлинность документа, а не изложенная в нем точка зрения. Задача проекта – создать максимально полную базу, которой бы могли пользоваться все желающие: ученые, журналисты, читатели.

Технически «Прожито» является мультиязычной платформой с возможностью поиска по разным языкам и орфографиям и последняя версия сайта помимо главного раздела с русскоязычными дневниковыми записями содержит тексты на украинском языке и один англоязычный дневник. В наши планы входит создание новых языковых разделов и автономных волонтерских сообществ для работы с ними.

«Прожито» имеет огромный потенциал для исследований на стыке «дальнего» и «близкого» чтения, поскольку представляет собой полноценный корпус текстов одного жанра, таким образом допуская применение как количественного, так и качественного анализа. Историки и культурологи используют «Прожито» для исследования повседневности и специфики эпох. Для лингвистов тексты «Прожито» ценны тем, что они представляют собой особый регистр, по дискурсивной структуре близкий к устной речи, а наличие информации о каждом авторе и датировки записей позволяет использовать их для социолингвистических и диахронических исследований.

## **“PROZHITO” FROM MANUSCRIPT TO CORPORA: COLLECTING, ANNOTATING AND ANALYSING DIARY TEXTS**

*Melnichenko M.A.*

“Prozhito”, 5/10, Karetny Ryad ave., Moscow, Russia,  
misha.melnichenko@gmail.com

*Tyshkevich N.B.*

National Research University Higher School of Economics, 20,  
Myasnitskaya st., Moscow, Russia, natalie.tysh@gmail.com

“Prozhito” (prozhito.org) is the first online database of 600 diverse private diaries (200 000 entries), tied to a chronological line and representing personal narratives from XIX–XX centuries in Russian and Ukrainian. “Prozhito” blends the structural experience of blog platforms and archival tradition of curating personal writings. User can work not only with particular texts but with the whole collection of diary entries, building complex search queries by specifying author’s gender and age, journal types (f.e. war, tourist, dream etc.) and filtering results by exact dates and places of records. Our publishing standard makes “Prozhito” both a professional instrument for modern scholars and a key to global history for amateurs.

*Key words:* corpora, digitizing, private diaries, ego-texts.

УДК 930:004

### **УКАЗАТЕЛИ К СТЕНОГРАФИЧЕСКИМ ОТЧЕТАМ ЗАСЕДАНИЙ ГОСУДАРСТВЕННОЙ ДУМЫ НАЧАЛА XX В.: АНАЛИЗ ИЗМЕНЕНИЯ СТРУКТУРЫ ИСТОЧНИКА НА ОСНОВЕ XML<sup>1</sup>**

*Поврозник Н.Г.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, [porvoznik.ng@gmail.com](mailto:porvoznik.ng@gmail.com)  
Изучение структуры указателей к стенографическим отчетам заседаний Государственной думы начала XX в. и ее изменения позволяет выявить возможности дальнейшего исследования источника, оценить его источниковый потенциал, представить степень подробного описания деятельностных характеристик депутатов, а также изменение в структуре представленных данных. Анализ указателей к стенографическим отчетам проводился на основе запросов к источникам, размеченным согласно разработанной схеме XML, отражающей их структуру. В статье сделан вывод о том, что указатели обладают различной полнотой данных о личностных характеристиках депутатов и их парламентской деятельности.

*Ключевые слова:* указатели к стенографическим отчетам, Государственная дума, XML, структура источника, деятельность депутатов.

---

© Поврозник Н.Г., 2017

<sup>1</sup> Исследование выполнено при поддержке РФФИ (грант № 16-06-00569, руководитель И.К. Кирьянов)

Значение указателей к стенографическим отчетам Государственной думы начала XX в. как исторического источника состоит в том, что его содержание в максимально структурированном виде отражает деятельность депутатов в парламенте, позволяет выявлять и анализировать качественные и количественные показатели их деятельности.

Анализ структуры указателей и ее изменения позволяет выявить возможности для дальнейшего изучения источника, оценить его источниковый потенциал, представить степень подробного описания деятельностных характеристик депутатов, а также изменение в структуре представленных данных.

Указатели к стенографическим отчетам заседаний Государственной думы относятся к высокоструктурированным источникам, в которых основные разделы сохранялись в течение всех четырех созывов. Анализ указателей к стенографическим отчетам проводился на основе запросов к источникам, размеченным согласно разработанной схеме XML, отражающей структуру источника. Всего было размечено и проанализировано 10 указателей, относящихся к заседаниям Государственной думы I–IV созывов.

Структура разметки отражает основные тематические разделы указателей и включает в себя следующие данные:

1) метаданные источника (созыв, заглавие, сессия, заседания, даты заседаний, издательство, год издания указателя);

2) личные характеристики депутата (ФИО, губерния избрания, признание правильности выборов, партия, возраст, социальный статус, сословие, курия избрания и некоторые другие сведения);

3) показатели, связанные с парламентской активностью депутата (подписал документы, говорит по теме, предлагает (в т. ч. формулу действий), вносит предложения, поправки, докладывает, избран в комиссию, добавляет к порядку дня, объясняется по личному вопросу, сообщает по партийной дисциплине (в т. ч. оглашает, объявляет от имени партии), отсутствует при вызове, отказывается от слова, предлагается на должность, прерывает ораторов, лишается слова, удаляется из зала заседаний (в т. ч. устраняется от заседаний), занимает должность, голосует, воздерживается (в т. ч. от голосования), заявляет, оглашает заявление от представителей территории (губернии, города избрания), оглашает внеочередное заявление, отказывается от членства в Думе, покидает заседание, информирует (новостное или сообщение для сведений), меняется порядком для выступления), в этой же категории выделены теги, обозначающие сообщение о депутате, замечания председателя депутату, судебное преследование, возглас с места, отпуск во время заседаний.

Перечисленные разделы повторялись в указателях на протяжении всех сессий и созывов, однако полнота данных в рамках разделов была различной. Так, в разделе с личными характеристиками депутата указатели I–II созывов содержат лишь краткие данные (ФИО, губерния избрания, признание правильности выборов и указание принадлежности к отделу). Наиболее полно личные сведения раскрывают указатели III и IV созывов, в которых перечисленные данные дополняются целой совокупностью характеристик (принадлежность к партии, сословие, звание (чин), национальность, дата рождения, семейное положение, образование (в т. ч. учебное заведение), членство в комиссиях и комитетах (внедумская деятельность), владение собственностью (с указанием типа собственности и размеров имущества). В указателях III созыва эта информация публикуется только в указателе за 1 сессию, тогда как в остальных она опускается. В указателях IV созыва эти данные фиксируются на протяжении всех сессий.

Степень полноты данных о деятельности депутатов в Государственной думе также различается. Так, наиболее детальные сведения изложены в указателях III созыва, тогда как наименее подробным является документ за 3 сессию IV созыва (в нем указаны данные лишь по 10 из 29 видов активности). Очевидно, что краткость данных в упомянутом указателе объясняется особенностями формирования документа, поскольку было проведено лишь 3 заседания в сессии. Это подтверждают данные стенографического отчета [1].

Интересна структура первых двух изданных указателей (сессии I–II созывов). В документах подробно определены виды деятельности депутатов, которые содержат также сведения о том, на какие позиции были предложены депутаты, как проходила баллотировка, сколько шаров получил парламентарий при избрании на должность. В остальных указателях эта ценная информация опускалась. Кроме того, в первых указателях более подробно изложены сведения о предложениях депутатов и внесенных поправках, тогда как в остальных указателях эти данные либо не публиковались совсем, либо указывались лишь отдельные случаи.

Отмечена также динамика в структуре данных в указателях одного созыва, в частности в указателях за 1–2 сессию III созыва не опубликованы сведения о депутатах, пропускаявших заседания, тогда как в документах за 3–5 сессию эти сведения присутствуют.

Таким образом, структура указателей к стенографическим отчетам претерпела существенные изменения за I–IV созывы, но в целом отражала общие, наиболее значимые показатели активности депутатов в Думе.

## Библиографический список

1. *Стенографический отчет*. Государственная Дума. Четвертый созыв. Сессия III. Заседания I-III. СПб., 1915.

### **INDEXES TO THE STENOGRAPHIC REPORTS OF THE STATE DUMA IN THE BEGINNING OF THE XX CENTURY: ANALYSIS OF CHANGES IN THE SOURCE STRUCTURE ON BASIS OF XML**

*Povroznik N.G.*

Perm State University, 15, Bukireva st., Perm, Russia,  
povroznik.ng@gmail.com

The study of the structure of indexes to the stenographic reports of the State Duma in the early twentieth century and its changes makes it possible to identify opportunities for further study of the source, to assess source potential, to present the degree of description details of deputies' activity characteristics and the change in the structure of the data submitted. The analysis of indexes to verbatim records was carried out on the basis of queries to sources marked up according to the developed XML markup scheme which reflects the structure of the source. The article concludes that indexes have different completeness of data on personal characteristics of deputies and their parliamentary activities.

*Key words:* Indexes to the stenographic reports, the State Duma, XML, the structure of the source, parliamentary activity of deputies.

УДК 930:004

### **СОЦИАЛЬНО-СЕТЕВОЙ АНАЛИЗ ГЛАСНЫХ ГУБЕРНСКИХ ЗЕМСКИХ СОБРАНИЙ: К ПОСТАНОВКЕ ПРОБЛЕМЫ<sup>1</sup>**

*Повроznик Н.Г., Сметанин А.В.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, povroznik.ng@gmail.com, smetanin.av@gmail.com

Статья посвящена возможностям использования социально-сетевого анализа для изучения гласных губернских земских собраний. В ходе анализа персональных сетей предполагается выявить наличие устойчивых связей депутатов, наиболее близких по общности взглядов гласных. На наш взгляд, использование методики построения семантических сетей с многообразием типов и уровней связей и различными отношениями между элементами позволит сохранить целостность изучаемого объекта, выделить структурные элементы и связи, не отрывая их от исторического контекста. Кроме того, выявление взаимосвязей элементов поведения и моделей в целом позволит судить и об их влиянии на эффективность деятельности губернских земских собраний в целом.

*Ключевые слова:* губернское земское собрание, социально-сетевой анализ, сеть гласных, персональная сеть, плотность сети, эффективность деятельности земств.

---

© Повроznик Н.Г., Сметанин А.В., 2017

<sup>1</sup> Исследование выполняется при поддержке РФФИ № грант № 17-06-00470.

Введение земского самоуправления в Российской Империи явилось важным звеном в цепи либеральных реформ Александра II. Значение земской реформы заключалось не только в том, что новый институт способствовал развитию хозяйства и инфраструктуры регионов, разрешению социальных проблем и противоречий, но и в том, что правительство признало необходимость привлечения общественности для рассмотрения этих вопросов и предоставило определенную самостоятельность в этом направлении. Министр финансов С.Ю. Витте в конфиденциальной записке отмечал значение земства как «школы представительных учреждений» [1, с. 92], указывая на его политическую роль. Изучение групп земских гласных на основе использования средств и методов сетевого анализа позволит прояснить внутреннюю фрагментацию земских собраний, выделить депутатские группы, понять основы их формирования и реконструировать систему взаимосвязей между депутатами и депутатскими группами для более глубокого понимания основ взаимодействия гласных, а в целом и развития земских собраний.

Губернские земские собрания значительно повлияли на развитие общественно-политической, экономической и социальной сфер тех губерний, в которых вводилось земское самоуправление. Решения, определяющие направления деятельности, хотя и ограниченные рамками Положения 1864 г., принимались именно в губернских земских собраниях и неизбежно зависели от корпуса гласных, а значит, и от установившихся депутатских групп, моделей поведения земских депутатов, их мнений, компетенций, личного опыта и других факторов.

Изучение процессов фрагментации представительных учреждений связано, в первую очередь, с анализом парламентских групп. Среди проанализированных – составы Конгресса США [2], Палаты представителей Нидерландов [3], итальянского парламента [4], Государственной думы Российской Империи [5] начала XX в. и некоторые другие. Особенность земства заключается в том, что оно имело больше условностей и ограничений, наложенных законодательством и обстоятельствами функционирования, прения по вопросам публиковались не полностью, журналы заседаний подлежали цензуре, – все это существенно осложняет изучение земского самоуправления по полной аналогии с существующими работами по парламентариям.

Несмотря на достаточно большой набор метрик для измерения параметров социальных сетей, нас больше всего интересует изучение персональных сетей гласных и их плотности. Анализ персональных сетей покажет наличие устойчивых связей депутатов, наиболее близких по общности взглядов гласных. Рассмотрение плотности персональной сети заключается в определении отношения реально существующих взаимосвязей к количеству максимально возможных. Про-

ведение такого анализа позволит увидеть: насколько единым было земство, как влияли отдельные земские группы на принятие решений. В этой связи сравнение данных показателей по очередным и чрезвычайным заседаниям позволит судить о монолитности групп в разных условиях – «повседневных» и экстренных по характеру сессий.

Установление взаимосвязей гласных в земствах требует выявления оснований для взаимодействия между депутатами, а также определения влияния этих факторов на эффективность деятельности губернских земских собраний в целом. Полагаем, что построение социальной сети, определение различных типов маркеров способно найти ответ на эти вопросы.

Таким образом, метод сетевого анализа, построение социальных и семантических сетей являются значимыми средствами изучения взаимодействия земских депутатов, образования группировок гласных на основе общего понимания проблем и схожести путей их решения, устойчивости этих групп анализа групп гласных, сформированных по антагонистическому принципу, а также для рассмотрения позиций депутатов по различным вопросам земской деятельности и их изменения во времени. Проведение сетевого анализа позволит выявить факторы, под воздействием которых мнения гласных могли трансформироваться тем или иным образом. Использование методики построения семантических сетей с многообразием типов и уровней связей и различными отношениями между элементами позволит сохранить целостность изучаемого объекта, выделить структурные элементы и связи, не отрывая их от исторического контекста. Выявление взаимосвязей элементов поведения и моделей в целом позволит судить и об их влиянии на эффективность деятельности губернских земских собраний в целом.

### Библиографический список

1. Самодержавие и земство. Конфиденциальная записка Министра финансов Статс-Секретаря С.Ю. Витте. (1899 г.). Изд. 2-е. Штутгарт, 1903. С. 92.
2. *Porter M.A., Mucha P.J., Newman M.E.J., Friend A.J.* Community structure in the United States House of Representatives // *Physica A*. 2007. № 386.
3. *Scott J.C.* Social Processes in Lobbyist Agenda Development: A Longitudinal Network Analysis of Interest Groups and Legislation // *The Policy Studies Journal*. 2013. Vol. 41, № 4.
4. *Moody J., Mucha P.J.* Portrait of political party polarization // *Network Science*. 2013. Vol. 1. Issue 1.
5. *Kirkland J.H., Gross J.H.* Measurement and theory in legislative networks: The evolving topology of Congressional collaboration // *Social Networks*. 2014. № 36.
6. *Suermondt. A.* Social Network Analysis of the Dutch House of Representatives: Bachelor Thesis Information Science. Amsterdam, 2010.

7. *Parigi P., Sartori L.* The political party as a network of cleavages: Disclosing the inner structure of Italian political parties in the seventies // *Social Networks*. 2014. № 36.

8. *Сметанин А.В.* Моделирование системы междепутатских связей в Государственной Думе Российской империи // Информационный бюллетень Ассоциации «История и компьютер»: матер. XII конф. Ассоциации «История и компьютер». Октябрь 2010. М., 2010. № 36. С. 92–94.

9. *Сметанин А.В.* Моделирование неиерархических организаций: случай парламента // *ArsAdministrandi*. 2009. № 1. С. 38–43.

## **SOCIAL-NETWORK ANALYSIS OF THE PROVINCIAL ZEMSTVO' DEPUTIES: STATEMENT OF THE PROBLEM**

*Povroznik N.G., Smetanin A.V.*

Perm State University, 15, Bukireva st., Perm, Russia,  
Povroznik.ng@gmail.com, smetanin.av@gmail.com

The article is devoted to the possibilities of social-network analysis for studying provincial zemstvo deputies. The analysis of personal networks shows the stable ties between deputies. Constructing semantic networks with a variety of types and levels of connections and different relations between elements allows us to preserve the integrity of the researched object, to define structural elements and connections without taking them away from the historical context. The identification of the relationships between the elements of the deputies' behavior in the zemstvos and network model analysis makes it possible to assess its influence on the effectiveness of the activities of the provincial zemstvo assemblies in general.

*Key words:* Provincial Zemstvo Assemblies, social network analysis, network of deputies, personal network, network density, efficiency of zemstvo.

УДК 930.2:004.773.3

## **ЭЛЕКТРОННАЯ ПОЧТОВАЯ ПЕРЕПИСКА КАК ПОТЕНЦИАЛЬНЫЙ ИСТОРИЧЕСКИЙ ИСТОЧНИК**

*Рожнева Ж.А.*

Национальный исследовательский Томский государственный университет, Россия, Томск, просп. Ленина, 36, zhar@ido.tsu.ru

*Акашева А.А.*

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, Россия, Нижний Новгород, ул. Гагарина, 23, annakasheva@yandex.ru

С развитием и распространением цифровых технологий в круг потенциальных исторических источников активно вторгаются изначально цифровые материалы, в том числе электронная почтовая переписка. Необходимость ее сохранения и вовлечения в научный оборот требует от историков, источниковедов,

архивистов и других специалистов выработки соответствующей методики источниковедческого анализа. В статье на основе проведенного исследования предлагаются основные направления внешней и внутренней критики электронной переписки: реконструкция «почтовой истории», выделение типовых черт и индивидуальных особенностей в организации электронных писем внутри почтового аккаунта; учет количественных характеристик содержимого электронного почтового ящика; выявление элементов внутренней структуры и оформления электронных писем, способствующих их атрибуции; определение взаимосвязей между отдельными письмами.

*Ключевые слова:* электронная почта, цифровые исторические источники, источниковедение электронных документов.

Электронные письма, наряду с другими формами электронных коммуникаций, практически вытеснили сегодня бумажную переписку, являвшуюся в недавнем прошлом непременным атрибутом деловых и личных взаимосвязей. По данным исследовательской компании Radicati Group в 2015 г. насчитывалось около 2,6 млрд пользователей электронной почты по всему миру. Объем общемирового почтового трафика составлял более 205 млрд электронных писем в день [1]. Электронная переписка, получившая столь широкое распространение как средство коммуникации между коммерческими организациями, государственными учреждениями, отдельными людьми, отражает их повседневную деятельность и, безусловно, содержит разнообразные сведения о современном периоде новейшей истории.

В отечественной историографии электронная почтовая переписка как потенциальный исторический источник частично рассматривается только в двух известных нам работах [2; 3], в которых отмечается, что в последние десятилетия электронные письма стали одним из главных средств обмена информацией, однако детального их анализа с источниковедческих позиций не проводится. В зарубежной исторической науке исследования электронной почтовой переписки также практически отсутствуют [4; 5].

Ставя целью выявить в первом приближении особенности электронных писем как потенциальных исторических источников, мы сочли уместным использовать такой подход, как «case-study», предполагающий изучение отдельных случаев. В этом качестве мы использовали собственную электронную почту, которая рассматривалась сквозь призму традиционного источниковедческого анализа.

При изучении электронной переписки базовыми объектами, на наш взгляд, становятся почтовые учетные записи (аккаунты), поскольку именно в них она создается и сохраняется. Люди и организации могут иметь любое количество электронных почтовых аккаунтов, состав которых может изменяться с течением времени. Реконструкцию

«почтовой истории» следует рассматривать как первый этап источниковедческого анализа.

Содержимое почтовых аккаунтов представляет собой сложный комплекс, который включает в себя не только полученные, но и отправленные сообщения, а также черновики и удаленные письма. В этом смысле электронная почта выгодно отличается от традиционной почтовой переписки, в рамках которой существование всех ее составляющих в виде единой совокупности практически невозможно.

Электронная переписка обладают типовой внутренней структурой, которая задается программной средой и представляет собой совокупность системных и пользовательских папок, содержащих отдельные письма. Обязательное наличие сведений об адресате и адресанте, теме письма, дате и времени получения или отправки значительно облегчает проведение внешней критики источника, но в то же время требует понимания механизмов формирования подобных метаданных.

Среди значимых количественных характеристик почтового аккаунта можно выделить: общее количество писем и количество писем по папкам; относительная доля разных групп сообщений; количество прочитанных сообщений; количество сообщений, на которые не был дан ответ, и др. Количественные данные позволяют составить представление о масштабе переписки и выдвинуть предположения об особенностях ее использования.

Электронные письма имеют свой формуляр, состоящий из двух частей, – стандартной, создаваемой по умолчанию почтовой программой, и пользовательской. Автор письма может оформлять его с помощью инструментов, аналогичных тем, что имеются в текстовых редакторах. В письмах также могут присутствовать рисованные эмодзи, так называемые «смайлики», призванные через графические образы передавать эмоциональное состояние автора. С точки зрения жанровых особенностей электронная переписка представляет собой гибридный жанр, в котором сочетаются признаки письменной и устной речи [6].

Электронные письма могут сопровождаться вложениями, которые представляют собой файлы разных форматов. В интерфейсе почтовой программы вложения выглядят как самостоятельные объекты, прикрепленные к письму, в действительности их содержание кодируется и пересылается как неотъемлемая часть электронного послания.

При выявлении информационного потенциала электронных писем следует принимать во внимание масштаб переписки и ее назначение, а также присутствие в электронной почте таких специфических по происхождению и содержанию писем, как спам-сообщения, новостные рассылки, письма от программ-роботов. Важной составляющей внутренней критики является определение композиционных и тематиче-

ских связей внутри комплекса электронных писем, что, исходя из технологии электронной почты, предполагает выявление так называемых «цепочек писем».

Поскольку электронные письма представляют собой изначально цифровые источники, то для их содержательного анализа можно использовать различные компьютерные методы обработки данных.

При всех преимуществах электронной почты, как средства коммуникации и потенциального исторического источника, основной проблемой при вовлечении ее в научный оборот является ее сохранность в долговременной перспективе. К сожалению, на данный момент отсутствуют устоявшиеся представления о необходимости и возможности долговременного сохранения электронных писем в институциональной архивной системе. В этой связи современные историки и архивисты должны изучать их особенности, не дожидаясь, когда они станут архивными документами, поскольку этого может и не случиться, если уже сегодня не искать подходы к их сохранению и введению в научный оборот как части историко-культурного наследия.

### Библиографический список

1. *Email Market 2015–2019* // The Radicati Group. A technology market research firm. URL: [http://www.radicati.com/wp/wp-content/uploads/2015/02/Email\\_Market\\_2015-2019\\_Executive\\_Summary.pdf](http://www.radicati.com/wp/wp-content/uploads/2015/02/Email_Market_2015-2019_Executive_Summary.pdf) (дата обращения: 10.03.2017).
2. *Злобин Е.В.* Источниковедческие особенности новейших информационных и коммуникационных технологий // *Технотронные документы – информационная база источниковедения и архивоведения*: сб. науч. ст. М.: РГТУ, 2011. С. 236–246.
3. *Корниенко С.И., Гагарина Д.А.* Интернет – среда и средство историографического исследования // *Интернет и современное общество*: тр. XI Всерос. объединенной конф. (28–30 октября 2008 г., Санкт-Петербург) СПб.: Фак. филологии и искусств СПбГУ, 2008. С. 62–63.
4. *Zalinger J., Freier N., Freire M., Shneiderman B.* Reading Ben Shneiderman's Email: Identifying Narrative Elements in Email Archives // *The Human-Computer Interaction lab*. URL: <http://hcil2.cs.umd.edu/trs/2009-31/2009-31.pdf> (дата обращения: 20.12.2016).
5. *Fostikov A., Mandić S.* New categories of historical sources: e-mail and forum. *Internet communication and history* // *Review of the National Center for Digitization*. 2007. Vol. 10. P. 40–46.
6. *Зализняк А., Микаэлян И.* Переписка по электронной почте как лингвистический объект // *Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной междунар. конф. «Диалог»* (31 мая – 4 июня 2006 г., Москва). URL: [www.dialog-21.ru/dialog2006/materials/html/Zaliziak.htm](http://www.dialog-21.ru/dialog2006/materials/html/Zaliziak.htm) (дата обращения: 10.03.2017).

## E-MAIL CORRESPONDENCE AS A POTENTIAL HISTORICAL SOURCE

*Rozhneva Zh.A.*

National Research Tomsk State University, 36, Lenin ave., Tomsk, Russia,  
zhar@ido.tsu.ru

*Akasheva A.A.*

Lobachevsky State University of Nizhni Novgorod, 23, Gagarin ave.,  
Nizhni Novgorod, Russia, annakasheva@yandex.ru

With the development and spreading of digital technologies, digital materials, including e-mail correspondence, has become a potential primary source for historical research. The necessity to preserve and introduce it to the scientific use makes historians, source studies experts, archivists and professionals in the related fields develop relevant methodology for source analysis. The authors propose such directions of internal and external criticism of e-mail correspondence as reconstruction of the “e-mailing history”; detection of typical and specific features in how e-mail letters are organized within an e-mail account; consideration of quantitative adjectives of the content of the e-mail account; identification of elements of the internal structure and formatting patterns of electronic letters; tracing interrelations between particular letters.

*Key words:* e-mail, digital historical sources, source study of electronic documents.

УДК 94(438).083

### ПОЛИТИЧЕСКИЕ ПРОЦЕССЫ В ПОЛЬШЕ НА ОСНОВЕ КОНТЕНТ-АНАЛИЗА «КАЛИНИНГРАДСКОЙ ПРАВДЫ»

*Саенко А.В.*

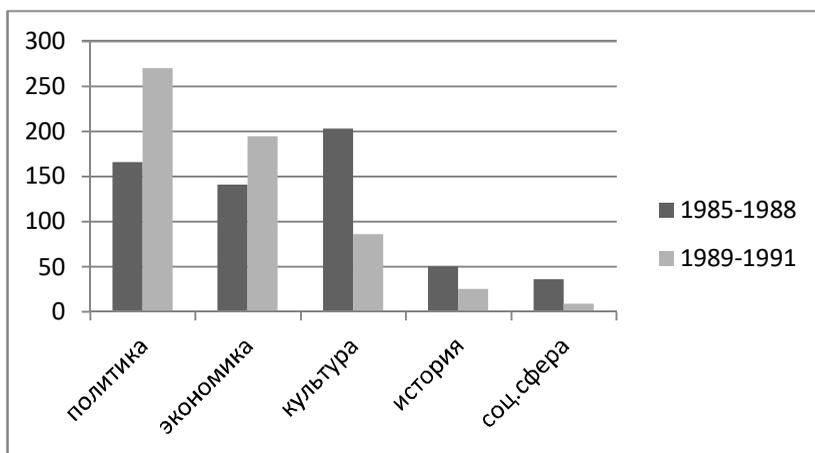
Балтийский федеральный университет им. Иммануила Канта,  
Россия, Калининград, ул. Невского, 14, angelinasaenko08@gmail.com

В данной работе предпринята попытка рассмотреть политические изменения, имевшие место в польском государстве в 1985–1991 гг. Известно, что как Польша важный геополитический сосед находилась под пристальным вниманием со стороны СССР. Проводившаяся в то время политика в отношении соседнего государства неизбежно отражалась в зеркале советской прессы, фиксируя как изменения, отклонения от осуществляемого курса, так и реалии советской системы и польско-советских взаимоотношений, что можно проследить на материалах «Калининградской правды». С помощью инструментов контент-анализа в газете и в политической жизни становится возможным проследить эволюцию общественно-политического устройства государства с выделением двух этапов преобразований: от последних лет существования коммунистического режима в 1985–1988 гг. к 1989–1991 гг., когда происходит трансформация системы внутренней и внешней политики государств.

*Ключевые слова:* Польша, СССР, контент-анализ, советская пресса, перестройка.

На протяжении всего XX в. отношения России и Польши проходили различные этапы своего развития. 1985–1991 гг. обозначили особый период в истории СССР: время перестройки. Так, если в 1987–1988 гг. перестройка включала в основном попытки «совершенствования социализма», то затем на первый план вышли реформа политической системы и курс на обновление идеологии [1, с. 40]. Период 1989–1991 гг. послужил точкой отсчета формирования новых отношений Польши и СССР, что связано с проведением радикальных экономических и политических преобразований в обеих странах.

Для анализа была использована генеральная совокупность материалов за 1985–1991 гг. Польской тематике в общей сложности посвящено 985 статей. Контент-анализ осуществлялся с помощью информационной системы «Время советское». Данный ресурс позволяет выполнять тематическое типирование и тегирование текстов, производить необходимую выборку статей, а также анализировать частотные характеристики по заданным темам.



#### Соотношение частоты появления категорий в двух периодах

В ходе исследования с помощью частотного контент-анализа были выделены следующие категории: Политика, Экономика, Культура, История и Социальная сфера. Во второй половине 1980-х гг. в Польше «продолжала действовать прежняя, сложившаяся еще в годы индустриализации система, жестко соединявшая политику с экономикой» [2, с. 834]. Начавшиеся процессы перестройки, затронув одну из сфер, необратимо вносили изменения в остальные. В 1985–1988 гг. наиболее актуальными являлись темы культуры и экономики, а интерес к политике, хотя и присутствовал, но не играл главной роли. Об-

ратная ситуация наблюдается в 1989–1991 гг., когда темы, связанные с политикой и экономикой, становятся самыми обсуждаемыми, а интерес к сфере культуры падает (см. ресурс).

Несмотря на то, что начавшиеся преобразования постепенно охватили все стороны жизни польского общества, более подробно стоит остановиться на категории Политика, так как она наиболее тесно связана с другими категориями, что выражается в преобладающем количестве статей. Неравномерное распределение обсуждаемых тем имело место и внутри категории.

Публикуемые в 1985–1988 гг. политические новости ограничивались краткой информацией о происходящем в Польше. Поддерживалась связь ПОРП и КПСС, выражалась поддержка проводимой политике «ускорения и перестройки», при этом зачастую Восточный блок противопоставлялся Западу. Как пример деятельности «диверсионных центров и разведывательных служб Запада» изображался профсоюз «Солидарность». Большая доля сообщений была связана с референдумом 1987 г. в ПНР, имевшего целью продемонстрировать поддержку общества экономической политике властей. Следует отметить, что большинство статей, содержащих информацию о внутренней политике польского правительства, имеют индикатор «положительно». Напротив, статьи с информацией о нелегальном профсоюзе «Солидарность» характеризуются индикатором «негативно». Интересны статьи с индикатором «жизненно», отражающие заинтересованный, иногда с долей удивления, взгляд соседей на происходящие процессы в обоих государствах.

К началу 1989 г. положение изменилось, пресса стала фиксировать новое отношение правящей партии к оппозиции, что выразилось в стремлении к диалогу и подготовке «круглого стола». Хотя в статьях все еще встречались сообщения об «экстремистских элементах, подрывающих систему», изображение деятелей оппозиции, в первую очередь Леха Валенсы, смещается от негативного к нейтральному. Образ Польши 1990 г. представляется весьма противоречивым, а ситуация в стране – напряженной. Приватизация, рост безработицы, увеличение количества забастовок, инфляция, банкротство множества предприятий – все эти моменты описывают сложности, происходившие в соседнем государстве. Заметно увеличивается число сообщений о демонстрациях «коммунистических» памятников, о повторяющихся актах вандализма и экстремизма. С долей сожаления описывается процесс отстранения от власти ПОРП и ее последующего самороспуска.

Образ Польши в годы перемен, представленный на страницах «Калининградской правды», отражает происходившие в это время политические, социальные и экономические преобразования, затронувшие все сферы жизни общества. По материалам газеты в польской

«перестройке» можно выделить два этапа: 1) 1985–1988 гг. – последние годы существования коммунистического режима с отсутствием возможности для легальной деятельности оппозиции. В это время почти все статьи о Польше были посвящены вопросам ее культурного развития и приграничного сотрудничества Калининградской области с соседними польскими воеводствами. 2) 1989–1991 гг. – «демократическая революция», когда после проведения досрочных выборов в сейм первое некоммунистическое правительство Польши приступило к проведению кардинальных экономических и политических реформ, которые, как тогда считали, были синхронны нашей перестройке. Все это не могло не привлечь внимание советской стороны, что нашло отражение в «Калининградской правде».

### **Библиографический список**

1. *Согрин В.В.* Политическая история современной России. 1985–2001: от Горбачева до Путина / Серия «Высшее образование». М.: Издательство «Весь Мир», 2001. 272 с.
2. *Матвеев Г.Ф., Носкова А.Ф. (отв.ред.), Лыкошина Л.С.* Польша в XX веке. Очерки политической истории: М., Индрик, 2012. 929 с.

### **POLITICAL PROCESSES IN POLAND BASED ON THE “KALININGRADSKAYA PRAVDA” CONTENT ANALYSIS**

*Saenko A.V.*

Immanuel Kant Baltic Federal University, 14, A. Nevskogo st.,  
Kaliningrad, Russia, angelinasaenko08@gmail.com

The paper considers the political changes taking place in Poland in 1985-1991. As an important geopolitical neighbor, Poland was in focus of the Soviet politics. The policy towards the neighboring state was inevitably reflected in the Soviet press, recording both changes, deviations from the political course, and the realities of the Soviet system and Polish-Soviet relationships. All this can be traced in the materials of the regional newspaper “Kaliningradskaya Pravda”. The content analysis tools give a possibility to see the evolution of the socio-political structure of the state. The results of analysis make it possible to distinguish the two stages: the last years of the Communist regime existence in 1985–1988 and the beginning of the transformation of internal and external policies of both states in 1989–1991.

*Key words:* Poland, the USSR, content analysis, the Soviet press, perestroika.

## МЕТОДИКА СБОРА ДАННЫХ ДЛЯ ЦИФРОВОГО НАУЧНОГО ИЗДАНИЯ С ВИЗУАЛЬНЫМИ МАТЕРИАЛАМИ

*Станкевич Ю.Д., Кижнер И.А.*

Сибирский федеральный университет, Россия, Красноярск,  
просп. Свободный, 79, superstd@ya.ru

Понимание необходимости открытого доступа к музейным изображениям и процедур предоставления изображений вызвано необходимостью повторного использования изображений для научных целей и образовательных целей и требует понимания реальной практики получения таких изображений для научных и образовательных проектов. Статья анализирует процедуры предоставления и правила использования изображений в образовательном проекте в Сибирском федеральном университете. Несмотря на то, что только двадцать процентов российских музеев в выборке из 184 музеев предоставили изображения для повторного использования, можно считать, что популяризация музеев и их коллекций могут служить стимулом открытого доступа для больших музеев в провинции.

*Ключевые слова:* музей, авторское право, открытый доступ, цифровые копии, публикация.

Цифровые издания научного характера, их создание, моделирование, технологические проблемы, связанные с их разработкой, давно находятся в фокусе международных научных исследований [3, 4, 5, 6]. Проблемы создания цифровых научных изданий визуального, а не текстового характера обсуждаются не столь интенсивно.

Цифровой документ научного и образовательного характера может и должен быть использован вне интерфейса, для которого он был создан. Возможность переноса данных из одного интерфейса в другой, позволит ответить на другие вопросы, интересующие исследователей [6].

Перенос визуальных и текстовых данных из одного интерфейса в другой осложняется юридическими ограничениями, связанными с правом на использование произведения путем его воспроизведения и распространения. Только автор произведения или его наследники имеют право давать согласие на воспроизведение, копирование и повторное использование произведения искусств. Правообладатель может потребовать от третьих лиц воздержаться от использования такого произведения.

Однако исключительное право действует не бессрочно. Оно ограничено жизнью автора и 70 годами, следующими за годами его смерти. По истечении срока действия исключительного права произ-

ведение искусство переходит в общественное достояние и может свободно использоваться любым лицом без согласия или разрешения и без выплаты авторского вознаграждения [1].

В тоже время в РФ существует закон о Музейном Фонде и музеях Российской Федерации. В соответствии со статьей 36 этого закона тиражировать изображение музейного предмета нельзя без согласия музейной администрации [2].

На практике происходит столкновение этих двух законов, которое приводит к множеству вопросов и препятствий. Публиковать цифровые копии любого произведения, перешедшего в общественное достояние, без письменного согласия музея запрещено.

Цель исследования – проверить, насколько выполняется закон «О Музейном Фонде» и ГК РФ в случае создания цифровых коллекций, которые будут использоваться в образовательных и научных целях или для цифрового издания академического характера.

Создание коллекции изображений избранных работ художественных и краеведческих музеев России на сайте Гуманитарного института Сибирского федерального университета позволит определить юридические и корпоративные ограничения, связанные со сбором данных для создания цифрового научного издания с визуальными материалами.

Для того, чтобы составить выборку музеев, которые участвуют в проекте, мы воспользовались веб-порталом Музеи России. Мы выбрали 184 исторических и художественных музея, расположенных в столицах, провинциальных городах, а также в небольших населенных пунктах. В первой группе рассматриваются музеи с числом посетителей в год более 50 000 человек. Другие две группы включены небольшие музеи (количество посетителей в год был меньше, чем 15 000 человек) и средние музеи с числом посетителей от 15 000 до 50 000 в год.

Мы направили в 184 музея официальные письма, адресованные на имя директора, которые приглашали участвовать в проекте, направленном на создание цифровых коллекций избранных произведений российских музеев. В нем также сообщалось, для какой цели будет создана такая коллекция, и какие преимущества получит музей, который согласится на участие в проекте.

Мы получили сорок положительных ответов из художественных и исторических музеев. Двадцать пять музеев предоставили нам цифровые изображения избранных предметов из их коллекций.

В докладе на конференции будут рассказаны условия, на которых музеи готовы предоставить изображения избранных работ, и показаны возможности оформления договора между музеем и университетом (гарантийное письмо, договор и так далее). Значимость проекта заключается в создании предпосылок для разработки модели данных,

системы управления коллекцией и интерфейса цифровой коллекции для популяризации произведений российских музеев среди студентов Сибирского федерального университета, а также предпосылок для дальнейшего развития модели и создания новых вариантов интерфейса в зависимости от целей и задач дальнейших исследований. Результаты исследования могут использоваться специалистами в области создания цифровых научных изданий.

### **Библиографический список**

1. *Гражданский* кодекс Российской Федерации. В 4 ч. 2. URL: <http://www.consultant.ru> (дата обращения: 17.04.2016).
2. *Федеральный закон* Российской Федерации «О музейном фонде Российской Федерации и музеях Российской Федерации». URL: <http://www.museum.ru/prof/mat/law> (дата обращения: 10.02.2016).
3. *Bielsten S.* Permissions, a Survival Guide: Blunt Talk about Art as Intellectual Property. 2006.
4. *Crews K. D.* Museum Policies and Art Images: Conflicting Objectives and Copyright Enrichment. 2012.
5. *Kelly K.* Images of Works of Art in Museum Collections: The Experience of Open Access. URL: <https://www.clir.org/pubs/reports/pub157/pub157.pdf> (дата обращения 17.04.2016).
6. *Robinson P.* Five desiderata for scholarly editions in digital form. URL: <http://dh2013.unl.edu/schedule-and-events/program/> (дата обращения 17.04.2016).

### **DIGITAL SCHOLARLY EDITIONS: ACCESS AND RE-USE OF DIGITAL IMAGES FROM RUSSIAN MUSEUMS**

*Stankevich Ju.D., Kizhner I.A.*

Siberian Federal University, Russia, Krasnoyarsk,  
Svobodny ave., 79, superstd@ya.ru

Digital scholarly editions require access to and re-use of museum images. The paper shows what could be done to obtain digital copies of images from Russian museum collections for an educational project in an academic setting. We discuss the results of an experiment and analyse the types of documents necessary to maintain the right to use images for a single project without moving them across interfaces.

*Key words:* museum, copyright, open access, digital copies, publication.

## **ИСТОЧНИКОВЕДЧЕСКИЕ ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ ИНТЕРНЕТ-РЕСУРСОВ В ИЗУЧЕНИИ ПОЛИТИЧЕСКОЙ ИСТОРИИ КЫРГЫЗСТАНЕ НА СОВРЕМЕННОМ ЭТАПЕ**

*Турсунова Э.Т.*

Кыргызский национальный университет имени Ж. Баласагына,  
Кыргызская Республика, г. Бишкек, ул. Фрунзе, 547, eiful@mail.ru

Статья посвящена источниковедческой проблеме использования интернет-ресурсов при изучении политической истории Кыргызстана в настоящее время. Описывается содержание состояния интернет-сайтов на территории Кыргызстана, степень отражения в них политической истории. Проанализированы особенности классификации, поиска и выявления исторической информации из интернет-ресурсов в Кыргызстане. А также было исследовано влияние процессов информатизации на современное общество Кыргызстана, выявлены особенности данных процессов. Электронные ресурсы во всемирной сети Интернет рассматриваются, как объективная реальность в современной отечественной истории, отражающая процессы информатизации. Даются методологические указания при поиске и выявления информации в Интернете.

*Ключевые слова:* Кыргызстан, процесс информатизации, информационное общество, электронные ресурсы, интернет-ресурсы, ИКТ, сайты.

На сегодняшний день переход к информационному обществу привел к коренному изменению общественно-политических и социально-экономических процессов в мире. Если информационная система в целом ранее служила средством связи и передачи информации в ходе эволюции человеческого общества, то она сейчас определяет вектор самого человеческого развития. Как и в других государствах, информатизация общества в Кыргызской Республике в процессе глобализации имеет свою структуру и стала его неотъемлемой частью.

Проблема использования интернет-ресурсов в качестве исторических источников, ставших сегодня закономерным явлением во всех сферах общества и имеющих уже определенный вес, характеризуется как важное направление в источниковедении Нового и Новейшего времени. Изучение истории эпохи информатизации не только раскрывает особенности вышеуказанного процесса и степень использования ИКТ в создании, хранении, переработке, передаче информации, но и радикально меняет сущность исторического источника и способствует возникновению его нового вида – электронных ресурсов. Обеспечение системы научных исторических фактов по современной истории Кыргызстана привело к проблеме поиска рациональных путей введения в

научный оборот электронных ресурсов, определения особенностей источниковедческого анализа и синтеза.

В этом плане электронные ресурсы являются одним из информационных источников, которые создаются и используются посредством ИКТ и расширяются переводом традиционных источников в электронный формат. Понятие «электронный ресурс» непосредственно связано с ИКТ, его совершенствованием и содержанием и исходя из особенностей этого постоянно будет расширяться [1, 2, 3, 4].

На сегодняшний день один из видов электронных ресурсов – интернет-ресурсы растут быстрыми темпами. В отличие от предыдущих периодов наблюдается не только количественный рост, но и повышается их качество, возрастает актуальность и востребованность. Рост интернет-ресурсов не ограничивается только размещением и публикацией информации, в то же время они превращаются в одно из направлений деятельности государственных органов и общественных организаций.

При изучении интернет-ресурсов как исторического источника следует обратить внимание на различие их происхождения: а) электронные ресурсы, ранее составленные и размещенные на веб-сайтах; б) документы, непосредственно составленные в интернет-среде, как результат неотъемлемой части современности.

По последним данным на территории Кыргызстана насчитывается около 1000 сайтов, из них 27 % составляют информационные сайты, остальные – рекламные и развлекательные. [5].

Сайты, отражающие новейшую историю Кыргызстана, можно разделить на три группы: государственные, общественные, индивидуальные. Если исследовать политическую историю Кыргызстана в постсоветское время, то интернет-ресурсы как источник информации выйдут на первое место. Если первоначально интернет-сайты выступали как средство отражения событий и процессов в постсоветском пространстве, доносили их до широких масс, то в последующем Интернет превратился в реальную силу, влияющую на ход и развитие событий. Эти выводы наглядно доказываются отражением в виртуальном мире мартовских дней 2005 г., апрельских 2010 г. и трагических событий в Оше и Джалал-Абаде, информация о которых, распространенная по всему миру, показала потенциальную силу интернет-ресурсов, как одного из факторов «информационной войны» [6].

Интернет-сайты вышеуказанных трех групп различаются между собой по информационному контенту, взглядам и способам доведения до общественности, которые исследователи должны обязательно учитывать при поиске необходимых источников в интернет-среде. Сегодня при изучении Новейшей истории Кыргызстана меняется корпус источников, в которых роль электронных, особенно интернет-

ресурсов, постоянно возрастает. Все это в свою очередь вызывает необходимость решать определенные исследовательские задачи для введения их в научный оборот.

При поиске информации в Интернете нужно помнить о ряде факторов. На сегодняшний день большое количество информации по всему миру связано с Интернетом, автоматически работающим с высокой скоростью, и есть ряд проблем, связанных с поиском адресатов, несмотря на поисковые системы, крайне затруднительно. Ибо сегодня отсутствуют единые требования и правила для устоявшихся определенных структур веб-сайтов и порталов, вследствие чего размещенные ресурсы носят бессистемный характер. Поэтому процессе поиска первоисточников появляются затруднения: во-первых, трудно разделить сайты по определенным типологиям; во-вторых, затруднен поиск источников из-за разных критериев и подходов цензуры в Интернете со стороны владельцев; в-третьих, поиск первоисточников осложняется практикой копирайтинга в Интернете, т. е. копирование и распространение владельцами интернет-сайтов популярной и уникальной информации.

При использовании интернет-ресурсов необходимо иметь в виду их отличие от других ресурсов и в условиях информатизации общества следует выделять два типа источников – устные и вещественные, как продукт деятельности человека в целом, которые можно перевести в электронный формат и, наоборот, из электронного – в традиционный, т. е. устные и вещественные.

При использовании интернет-ресурсов как средства познания прошлого, следует обратить внимание на особенности их природы, т. е. технической сущности. В условиях информационного общества работа с электронными историческими источниками занимает важное место в практике специалистов-историков. В связи с тем, что Интернет является свободной Всемирной сетью, он постоянно находится в распоряжении исследователей. Поэтому необходимо подчеркнуть, с одной стороны, равномерную доступность каждому исследователю интернет-ресурсов, с другой – возрастающую потребность в них все более широких слоев населения.

В информационном обществе электронизация и дигитализация в области научных исследований, т. е. перевод всех документов из традиционного формата в цифровой, становится в настоящее время одним из способов познания истории [6].

В результате изучения процесса информатизации в Кыргызстане, его влияния на формирование современного общества, источниковедческих проблем исторических источников, созданных с помощью ИКТ и являющихся частью реальности, особенностей объекта исследования исторической науки в информационном обществе и его

роли как источника по отечественной истории был сделан следующий вывод: проблемы, рассматриваемые в рамках диссертационной работы, по своей значимости и неотвратимости в дальнейшем будут претендовать на отдельное направление в исторической науке. В целом, интернет-ресурсы, фиксирующие основные события и процессы кыргызстанского общества, составляют основную группу источников новейшего времени по истории Кыргызстана, требующих тщательного, всестороннего источниковедческого анализа, ибо сегодня Интернет, как неотъемлемая часть общества не только носитель информации, но и результат деятельности человечества.

### **Библиографический список**

1. *Ханна М.* Может ли сохраняться созданный в цифровой форме документ? // Международный форум по информации. М., 2001. Т. 26, № 4. С. 15–18.
2. *McMichael A.* The Historian, the Internet, and the Web: A // Perspectives on History. 1998. № 36. URL: <http://www.historians.org/perspectives/issues/1998/9802/9802VIE2.CFM> (дата обращения: 03.04.2017).
3. *Sarantakes N.E.* So that a Tree May Live: What the World Wide Web Can and Cannot Do for Historians. A Reassessment // Perspectives: American Historical Association Newsletter. 1999. Vol. 37, № 2. URL: <http://www.historians.org/perspectives>. (дата обращения: 03.04.2017).
4. *Smith C.* Can You Do Serious History on the Web? // Perspectives: American Historical Association Newsletter. 1998. Vol.36, №2. P. 5–8. URL: <http://www.historians.org/perspectives/issues/1998/9802/9802COM.CFM> (дата обращения: 03.04.2017).
5. *Кыргыз* Республикасынын Транспорт жана коммуникация министрлигинин URL: [www.mtc.gov.kg](http://www.mtc.gov.kg) (дата обращения: 03.04.2017).
6. *Турсунова Э.Т.* Кыргызстан в условиях информатизации общества: источники и методы их изучения: дис. ... канд. ист. наук. Бишкек, 2014. 193 с.

### **SOURCE STUDIES PROBLEMS OF USING INTERNET RESOURCES IN RESEARCHING PRESENT POLITICAL HISTORY OF KYRGYZSTAN**

*Tursunova Je.T.*

Kyrgyz National University named after Zh. Balasagyn,  
547, Frunze st., Bishkek, Kyrgyz Republic, [eiful@mail.ru](mailto:eiful@mail.ru)

The article is devoted to the source studies problem of using internet resources in researching present political history of Kyrgyzstan. The current state of websites on the territory of Kyrgyzstan and the degree of reflection of political history on those sites are described. Features of classification, search and detection of historical information from online resources in Kyrgyzstan are analyzed. The author analyses the influence of informatization processes on the present society of Kyrgyzstan and investigates the peculiarities of those processes. Electronic resources are viewed as an objective reality in contemporary national history, reflecting the processes of

informatization. Methodological guidelines for searching and identifying information online are presented.

*Key words:* Kyrgyzstan, informatization process, information society, electronic resources, web resources, information and communication technologies, websites.

УДК 930:94.(47)"08/16"

## **SOCIAL NETWORK ANALYSIS AS A TOOL FOR THE SOCIAL HISTORY OF THE SOVEREIGN'S COURT IN 16<sup>TH</sup> CENTURY**

*Uspenskiy V.S.*

Higher School of Economics in St. Petersburg, Department of History,  
Russia, St. Petersburg, Promyshlennaia st., 17, vuspenskiy@hse.ru

The history and structure of Russian medieval and Early Modern nobility is traditionally understood by historians largely in an institutional manner. Different studies track how the official documents change the hierarchies and invent new ranks from the first attempts of stratification in the middle of the 16th century to the "classical" pre-Petrine structure of Russian nobility, based on a hierarchical system of ranks. This (over)institutionalized understanding of the social structure of the Early Modern society is driven by the composition of our written sources, produced by the government bureaucrats. Such a picture seems too 'modern' to be complete and thorough. Our expectations based on general understanding of the organization of the Late Medieval (and Early Modern) societies force us to look for certain kinds of 'informal' structures based on person-to-person relations and, more specifically, patronage. This paper will present an attempt to determine informal relations within the Sovereign's court in the 3d quarter of 16th century. My observations were made as a result of the analysis of a database of joint service appointments stated in Deployment Registers (*Razryadnye knigi*) in 1550s-1580s. I apply the method of SNA to complement our institutional understanding of hierarchies with a view of personal ties.

*Ключевые слова:* Ivan the Terrible, Royal Court, Muscovy, Oprichnina and Zemshchina.

In the middle of the 16th century, the structure and organization of power in Russia was undergoing substantial changes (on the most recent studies of the Russian state-building in the beginning of the 16th century see Krom 2010). The transformation process of bureaucratic hierarchies, the folding structure Tsar's court, is well studied (Pavlov 1992; Stanislavskiy 2004) But some of the processes that take place around the organization of the elite of the Russian state in the 16th century remain almost unknown.

The history and structure of Russian medieval and Early Modern nobility is understood by historians largely in an 'institutional' manner. Our studies follow the hierarchy changes in official documents, from the first attempts of stratification in the middle of the 16th century to the "classical"

structure of Russian nobility: the complex system of ‘*chiny moskovskiy i gorodovye*’.

This (over)institutionalized understanding of the social structure of the Early Modern society is driven by the composition of our written sources, produced by the government bureaucrats. The 16th century government was trying to control and systematize its nobility, and we, historians, follow the government blindly. The history of social strata is therefore limited to a page in the history of the state building<sup>1</sup>.

Such a picture seems too “modern” to be complete. Our expectations based on general understanding of how the Late Medieval and Early Modern societies work force us to look for certain kinds of ‘informal’ structure based on person-to-person relations, patronage etc.

This paper will try to present a new methodological approach to studying the personal connections and career making at the tsar’s court of Ivan the Terrible in the third quarter of the 16th century. I will try to apply Social Network Analysis to visualize certain type of personal ties within the Sovereign’s Court. The deployment registers (‘*razryadnye knigi*’) – provide us with the sources of situational ties within the Tsar’s court<sup>2</sup>. Razryad was a record of services executed by the members of the Sovereign’s Court. Some of the tasks (military or administrative) required more than one person, in those cases, several nobles were given temporary appointments, as a rule, such a team included a high-born noble as its leader and several nobles of lesser status as his officers. I will use Social Network Analysis to deal with the following question: how were the minor “officers” appointed with the senior noble? Were they appointed randomly or is there a pattern of joint appointments that indicates that they were somehow connected to their commander?

The social network analysis clearly shows that the official appointments were not made at random, but were usually following patterns of continuous joint appointments, when the low-rank persons were serving under the command of the same upper-rank service men. Though this picture is far more complicated than binary ‘patron ‘protégé’ model. Low-ranking nobles seem able to choose from several possible connections; the total number of available promoters varies from 1 to 5. The *Oprichina* has severely lessened

---

<sup>1</sup> Though, there are some studies of the social history of Russian nobility in Late Middle Ages and Early Modern Times which are not based on the “State’s view”, but rather on the records of personal relations and corporate mentalities, for example, see Kollmann 1999

<sup>2</sup> The following editions were used: Razryadnaya kniga 1475–1605. Vol. 1–3 Moskva, 1977–1982; Razryadnaya kniga 1475–1598. Moskva, 1966; Razryadnaya kniga 1550–1636 gg. T. 1, Moskva, 1975. For the most recent and complete overview of deployment registers as a historical source see (Anchimuk 2005)

these possibilities and drove ordinary service men to stricter personal relations with their high-ranking connections. Among all the factors suggesting a probability for a joint official military or civil appointment the geographical factor was the most significant. That clearly shows that the geographical clannishness and neighbour connections played the strong role in career strategies in 16th century.

We can also try to compare the two synchronous networks within the studied period: *Oprichnina* and its institutional descendant – the “Separate court” (*Osobiy dvor*) court of Ivan the Terrible and the *Zemshina*, which has preserved the traditions and personal composition of the ‘old’ Court (see Stanislavskiy 2004).

Some brief observations on the character of observed differences can be made. The Social Network of the *Oprichnina* Court seems to be less centralized and more ‘archaic’, which is a pretty much surprising result. The established understanding of *Oprichnina* and the so-called Separate Court of Ivan the Terrible is that they were more ‘modern’, less ‘traditionalized’ and more centralized. The *Oprichnina* is always pictured as a result of the Tsar’s vision of the more efficient instrument: free of traditional limitations, informal ties, conspiracies etc. In order to compose such a Court the Tsar introduced a lot of ‘new men’ into the highest levels of hierarchies. We expect that the new Court would be based on the principles of centralization, personal loyalty to the Sovereign. It seems though that driven away from their neighbourhoods, un-connected courtiers had a tendency to form limited quantity of personal relations, in other words, they tried to stick with one patron instead of forming spread networks of kinship and neighbourhood relations.

## References

1. *Anchimyuk Yu. V.* 2005. *Chastnye razryadnye knigi s zapisyami za poslednyuy chetvert’ XV – nachalo XVII vekov.* Moskva: Drevlehranilishe.
2. *Kollmann N. Shields.* 1999. *By Honor Bound: State and Society in Early Modern Russia.* Cornell University Press.
3. *Krom M.* 2010. “Vdovstvuiushchee tsarstvo”: politicheskiĭ krizis v Rossii 30-40-kh godov XVI veka. Moskva: Novoe literaturnoe obozrenie.
4. *Pavlov A.P.* 1992. *Gosudarev dvor i politicheskaja bor’ba pri Borise Godunove (1584–1605 gg.).* Nauka.
5. *Razryadnaya kniga 1475–1605.* Vol. 1–3. Moskva, 1977–1982;
6. *Razryadnaya kniga 1475–1598.* Moskva, 1966;
7. *Razryadnaya kniga 1550–1636 gg.* T. 1, Moskva, 1975.
8. *Stanislavskiy.* 2004. *Trudy po istorii gosudareva dvora v Rossii XVI–XVII vekov.* Moskva: RGGU.

## **АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ КАК ИНСТРУМЕНТ ДЛЯ ИССЛЕДОВАНИЯ СОЦИАЛЬНОЙ ИСТОРИИ ГОСУДАРЕВА ДВОРА В XVI В.**

*Успенский В.С.*

Национальный исследовательский университет  
«Высшая школа экономики» (Санкт-Петербург), Россия,  
Санкт-Петербург, ул Промышленная, 17, [vuspenskiy@hse.ru](mailto:vuspenskiy@hse.ru)

Структуру т. н. служилого сословия в России эпохи позднего средневековья и ее социальную историю историки обычно реконструируют в институциональном ключе: вслед за изменениями номенклатуры иерархий в официальных документах меняется и его устройство: от первых попыток стратификации (статей т. н. «Тысячной книги», князей и детей боярских дворовых «Дворовой тетради» 1550-х гг. XVI в.) до классической структуры московских и городских чинов допетровского периода. Применение методики анализа социальных сетей дает возможности для альтернативного взгляда на способы организации служилого сословия. Методика позволяет приблизиться к пониманию внутреннего устройства служилой корпорации: подтверждается предположение, что разрядные дьяки совершали такие назначения не случайным образом, опираясь только лишь на местнические счета аристократов, а действовали в соответствии с личными связями, устоявшимися отношениями внутри служилой корпорации.

*Key words:* Иван Грозный, царский двор, Московия, опричнина, земщина.

УДК 930.1; 314.1

## **НЕГЭНТРОПИЙНЫЙ ПОДХОД К ОПИСАНИЮ СОЦИАЛЬНЫХ РЕВОЛЮЦИЙ (НА ПРИМЕРЕ РОССИИ И СССР)**

*Чечулин В.Л.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, [chechulinvl@mail.ru](mailto:chechulinvl@mail.ru)

Рассмотрен негэнтропийный подход к описанию социальных причин некоторых исторических событий, использующий интерпретацию теоремы Алесковского о связи мер информации и энтропии. Указано на объективную закономерность революционной ситуации в России перед 1917 г. ввиду невозможности правящих классов продуцировать социальный порядок при наличии оснований продуцирования социального порядка основной массой населения. Указаны социальные причины распада СССР.

*Ключевые слова:* теорема Алесковского о связи мер информации и энтропии; копирование информации, негэнтропия, социальный порядок, революция в России 1917 г., распад СССР.

В работе [1] описан подход к анализу социального поведения людей, использующий теорему Алесковского о связи мер информации и энтропии [2], из которой следует, что для копирования информации (при поддержании социального порядка и воспроизводства поколений) необходимо потребление определённой доли негэнтропии, которая в рационе питания получается от определённых растительных продуктов [см. 1]. Потребление негэнтропии положительно коррелирует с традиционностью внутрисемейных отношений, низким уровнем преступности. Этот естественно-научный подход применим и к анализу исторических событий, и необходим в современности, как отмечал В.В. Путин в «Послании к Федеральному собранию»: «2017 год – год 100-летия Февральской и Октябрьской революции. <...> российское общество нуждается в объективном, честном, глубоком анализе этих событий» [3].

По данным, приведенным в работе Ю.И. Кирьянова [4; см. также 1], в России в 1913 г. доля потребления негэнтропии крестьянами и рабочими (90 % населения) была позволяло считать 0,85, что делало класс рабочих и крестьян обладающим материальными основаниями для упорядочивающего социального поведения. У правящих же классов доля потребления негэнтропии была меньшей, чем 0,5, что делало их неспособными к поддержанию социального порядка [1, с. 93–94]. В.И. Ульянов обозначил это состояние в 1913 г. так: «Для революции недостаточно того, чтобы низы не хотели жить, как прежде. Для неё требуется ещё, чтобы верхи не могли хозяйничать и управлять как прежде» [5]. Массы обладали основаниями упорядочивающего социального поведения, что и определило построение социалистического государства и его защиту от внешних и внутренних врагов.

Революционный посыл иссяк. К 1960 г. доля потребления негэнтропии в СССР упала ниже 0,7, что влекло невозможность поддержания плановой экономики (для равновесия которой необходима доля больше 0,7 [6]), затем — реформы, введшие разрушительные для системы в целом элементы рынка (т. н. хозрасчёт). К концу 1980-х гг. доля потребления негэнтропии упала ниже 0,5 [1, с. 91], вследствие чего стало невозможным сохранение социального порядка, выросла преступность, распался СССР, произошла реставрация капитализма.

Негэнтропийный подход позволяет понять причины социальных переворотов.

### Библиографический список

1. Чечулин В.Л., Богомякова В.С. Негэнтропия и социальные факторы (модели и анализ). Пермь: изд-во Перм. гос. ун-та, 2016. 130 с.
2. Алесковский В.Б. Путь разработки технологии, не вредящей природе // Журнал прикладной химии. 2002. Т. 75, № 5. С. 706–713.

3. Путин В.В. Послание Федеральному собранию (2016 г.) URL: <http://www.kremlin.ru/events/president/news/53379> (дата обращения: 01.03.2017).
4. Кирьянов Ю.И. Жизненный уровень рабочих России (кон. XIX – нач. XX в.). М., 1979.
5. Ленин В.И. Маёвка революционного пролетариата // ПСС. Т. 23. 300 с.
6. Чечулин В.Л. Об интерпретации основного логистического уравнения в терминах теоремы Алесковского о связи мер информации и энтропии // Университетские исследования: интернет-журн. Перм. гос. ун-т. ПГНИУ. Пермь, 2015. С. 222–223.

## **NEGENTROPY APPROACH TO THE DESCRIPTION OF SOCIAL REVOLUTIONS (ON THE EXAMPLE OF RUSSIA AND THE USSR)**

*Chechulin V.L.*

Perm State University, 15, Bukireva st., Perm, Russia, [chechulinvl@mail.ru](mailto:chechulinvl@mail.ru)

The paper describes the content of the negentropy approach to the description of social reasons of some historical events using interpretation of Aleskovsky's proposition of connection between the measures of information and entropy. The author states the objective inevitability of a revolutionary situation in Russia before 1917 because of ruling classes' impossibility to produce a social order whereas the rest of population was able to produce it. The social causes of the collapse of the USSR are indicated.

*Key words:* Aleskovsky's proposition of connection between the measures of information and entropy, coping of information, negentropy, social order, revolution in Russia of 1917, the collapse of the USSR.

УДК 070:94(47)"17/1917"(520)

## **РЕПРЕЗЕНТАЦИЯ ЯПОНИИ В ПЕРМСКОЙ ГУБЕРНСКОЙ ПЕРИОДИКЕ В ГОДЫ ПЕРВОЙ МИРОВОЙ ВОЙНЫ<sup>1</sup>**

*Ященко Ю.В., Гагарина Д.А.*

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15,  
[yaschenko.j@mail.ru](mailto:yaschenko.j@mail.ru), [dinara@psu.ru](mailto:dinara@psu.ru)

В работе проанализированы содержание и динамика образа Японии в пермской губернской периодике в годы Первой мировой войны. Изучены публикации семи газет, издававшихся в указанный период в Перми. Количество статей, в которых обсуждается или упоминается Япония, превысило 90. Исследование выполнено на основе информационной системы «Пермская губернская периодика: 1914–1922» (<http://permnewspapers.ru/>), позволяющей формировать выборки источников по заданным условиям. Анализ показал, что в пермской периодике единого взгляда на Японию не было, хотя не было и значительных

---

© Ященко Ю.В., Гагарина Д.А., 2017

<sup>1</sup> Публикация подготовлена при финансовой поддержке гранта РГНФ (в настоящее время РФФИ) № 16-11-59009.

различий. И на досоветском, и на советском этапах наблюдается неоднозначная интерпретация Японии. Для советского этапа свойственны более негативные характеристики государственного строя и внешней политики Японии.

*Ключевые слова:* Первая мировая войны, газеты, Япония, Пермская губерния.

Русско-японские отношения вызывают большой интерес у исследователей из различных областей гуманитарной науки. В то же время восприятие Японии в России изучено недостаточно [1], в частности, его особенность в период Первой мировой войны [2, 3]. Несмотря на поражение России в Русско-японской войне, Япония не была исключена из стран, привлекавших внимание российского общества в годы Первой мировой войны, тем более, что в этой войне Япония стала союзницей России. Это проявлялось на уровне информации не только в центральной печати, но и в губернских периодических изданиях, что подтверждается и анализом пермских газет.

Цель данного исследования – выявление содержания, особенностей и динамики образа Японии в пермской губернской периодике в годы Первой мировой войны. Источниковой и инструментальной основой стала информационная система «Пермская губернская периодика: 1914–1922» (<http://permnewspapers.ru/>), разработанная Центром цифровой гуманитаристики Пермского университета. В ходе исследования были проанализированы выпуски семи газет, издававшихся в Пермской губернии с 1 января 1914 г. по 1 января 1919 г.:

1. «Пермские губернские ведомости»,
2. «Пермская жизнь»,
3. еженедельник «Пермская земская неделя»,
4. «Пермский вестник Временного правительства»,
5. «Бюллетень Пермского губернского военно-революционного комитета»,
6. «Известия Пермского губернского исполнительного комитета Советов рабочих и солдатских депутатов»,
7. «Известия исполнительного комитета Совета железнодорожных депутатов Пермской дороги».

Анализ показал, что количество статей, в которых обсуждается или упоминается Япония, превысило 90., а тех статей, в которых использован топоним «Япония», т. е. содержание территориально относится к этой стране, более 30. При этом октябрь 1917 г. делит рассматриваемый период на два этапа – досоветский и советский.

На первом этапе Япония упоминается в рамках военных сводок или телеграмм как одна из союзных стран, отмечается ее участие в различных боях против германских захватчиков. В то же время имеют место и другие темы и сюжеты. Так, не раз были встречены «перепечатки» из японских газет (заметим, что в нашем случае речь идет о

губернской периодике), где Япония говорила о себе как о стране, поддерживающей Россию в борьбе с Германией. Интересным представляется то, что в губернской периодике имеют место не дублированные материалы из столичных изданий, а оригинальные публикации, посвященные более серьезным вопросам взаимоотношений России и Японии. Со стороны России наблюдается более жесткая риторика по отношению к Японии, которую можно рассматривать как отголоски поражения в Русско-японской войне и его последствий. Например, в названиях и текстах статей встречаются формулировки: «японские хищники», «японские захватчики», «японские аппетиты» и т. д. В контексте таких публикаций обсуждаются претензии Японии на полуостров Сахалин, что не могло не вызвать негативной реакции со стороны российского общества. Но осуждением японских требований и территориальных претензий диалог в периодических изданиях не ограничивается. В газетах встречаются материалы, посвященные агрессии Японии против Китая, являвшегося в тот момент, согласно характеристикам пермской периодики, ослабленной страной, неспособной к противостоянию и подвергавшейся нападениям со стороны японских соседей, явно рассчитывающих на территориальные приращения. Особенно усилилась данная тематика в информации о Японии после установления советской власти и превращения большевистской прессы в официальные издания. Большевистская пресса освещала проблему агрессии Японии против Китая уже с позиций критики и осуждения политики японского империализма и милитаризма. Поддержка Китая рассматривалась как важный фронт борьбы с империализмом и милитаризмом.

В целом можно говорить, что и на досоветском, и на советском этапах наблюдается неоднозначная интерпретация Японии, зависящая во многом от содержания конкретного сюжета и его цели, если публикации не касались «проблемных» вопросов, т. е. Китая и территориального спора, то информация губернской периодики носила в основном нейтральный характер.

Анализ также показал, что большая часть публикаций о Японии представляет собой небольшие сообщения, посвященные военной тематике. Однако к концу 1917 – началу 1918 г. больше внимания стало уделяться негативным проявлениям внешней политики Японии, критика ее приобрела более жесткий оттенок. В 1918 г. в пермской периодике стали обсуждаться агрессивность Японии и возможная угроза японского милитаризма для других государств, включая Россию и Китай.

Таким образом, в российской, в том числе в пермской прессе, единого взгляда на Японию не существует, но вместе с тем нет и значительных различий. По мере того, как Первая мировая война для Рос-

сии перестает быть важным направлением во внешней политике, а власть в стране меняется, в периодической печати в отношении Японии развивается более жесткая критика. Прежде всего негативно оцениваются и подвергаются осуждению режим и территориальные амбиции восточного соседа. Изменяются со сменой строя в России и те аспекты, которые рассматривались в контексте русско-японских отношений, акцент переносится с союзных взаимоотношений и прозападной позиции Японии на территориальный вопрос, при этом, ситуация усугубляется идеологическим подтекстом. Это подтверждает наличие двух этапов в репрезентации Японии региональной периодической печатью: досоветский и советский, причем для последнего свойственны более жесткая критика, негативные оценки и характеристики государственного строя и внешней политики Японии.

### **REPRESENTATION OF JAPAN IN THE PERM PROVINCE PERIODICALS IN THE YEARS OF THE FIRST WORLD WAR**

*Yaschenko Ju. V., Gagarina D. A.*

Perm State University, 15, Bukireva st., Perm, Russia,  
yaschenko.j@mail.ru, dinara@psu.ru

The paper analyzes the content and dynamics of the image of Japan in the Perm provincial periodicals during the First World War. The publications of seven newspapers published in this period in Perm were studied. Japan is discussed or mentioned in more than ninety articles. The research was carried out on the basis of the information system Perm Province Periodicals: 1914–1922 (<http://permnewspapers.ru/>), which makes it possible to generate samples of sources according to specified conditions. The analysis showed that there was no unified view of Japan in the Perm periodicals, although there were no significant differences. Both in the pre-Soviet and Soviet stages there is an ambiguous interpretation of Japan. For the Soviet stage, there are more negative characteristics of the state system and Japan's foreign policy.

*Key words:* First World War, newspapers, Japan, Perm province.

УДК 930.272

### **READING BETWEEN THE LINES – MULTISPECTRAL IMAGING OF ANCIENT MANUSCRIPTS**

*Brenner S., Sablatnig R.*

Computer Vision Lab, Institute of Computer Aided Automation, TU Wien, Austria, Vienna, Favoriten st., 9/183-2 A-1040, sbrenner@caa.tuwien.ac.at

Ancient manuscripts are a valuable source of information for historians and philologists. However, many of those texts are in a bad condition and barely readable, either due to palimpsesting or due to natural degradation over the centuries caused by

mold or fading of the ink. We are presenting a set of tools and methods for a Multispectral Imaging pipeline that aims at readability enhancement of degraded handwritings. These methods cover the necessary steps from multispectral image acquisition, through registration and dimensionality reduction, up to readability enhancement. The resulting digital images are a valuable aid for the analysis and deciphering of historic manuscripts.

*Ключевые слова:* multispectral imaging, readability enhancement, historic manuscripts.

## Introduction

Ancient manuscripts are a valuable source of information for historians and philologists. However, many of those texts are in a bad condition and barely readable, either due to palimpsesting or due to natural degradation over the centuries caused by mold or fading of the ink.

Multispectral Imaging (MSI) refers to the combination of a series of images of the same object using different bands of the electromagnetic spectrum, potentially stemming from different imaging devices. For the analysis of handwritings and paintings, these spectral bands typically range from ultraviolet (UV) to infrared (IR) [1,2]. Through MSI the different responses of different materials to specific wavelengths, that are invisible to the human observer, can be detected; for example, the differences between background and faded ink in a manuscript [2].

Within the last ten years we have participated in three inter-university projects dedicated to the digital preservation and restoration of historic handwritings. In collaboration with philologists, material analysts and restorers, we have developed a repertoire of tools and techniques for acquisition, processing and enhancement of multispectral images of such manuscripts. The proposed report gives a comprehensive overview of our current MSI processing pipeline, starting from multispectral image acquisition, through registration and source unmixing, up to readability enhancement.

### Multispectral image acquisition

A portable MSI system tailored for the *in situ* acquisition of manuscripts was developed. In its current form it is capable of producing reflectography images in eleven spectral bands from 365nm (UV) to 970nm (IR), as well as UV fluorescence images, backlight images and conventional color images for *facsimile* editions in an automated capturing process.

The system has been field-tested in libraries and monasteries around Europe and in Egypt.

### Image registration

A pixel-accurate registration of the layers of a multispectral image is required for all further processing steps. This necessity is obvious for images taken with different devices. But also images from the same device might require exact registration, due to different light refractions in optical filters

and chromatic aberrations. A two-step registration process has been successfully applied. In the first optional step images from different devices are coarsely registered with a feature based registration algorithm [3]. In the second step various imperfections are accounted for with a pixel accurate deformable fine registration algorithm that was originally developed for the registration of multi-modal medical images [4].

#### Dimensionality reduction

A raw MSI data set is usually difficult to interpret directly, because the multiple layers contain partly correlating, partly complementary information and are difficult to view simultaneously. Dimensionality reduction techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Linear Discriminant Analysis (LDA) are used to compress the information contained in the dataset to a small number of components, which ideally emphasize the differences between the individual sources in the document. We have shown that LDA is partly superior to PCA and ICA [5], but it requires prior labelling of pixels in background, text or underwriting regions, which in turn requires other document analysis techniques such as ruling detection.

To further improve the readability of the texts, additional post-processing steps are applied before delivering the results to the interested humanist; particularly Contrast Limited Local Histogram Equalization [6] to increase the image contrast and the combination of different resulting channels or components to a single pseudo color image, which enables the display of multiple facets of the data simultaneously.

#### Conclusion

The Multispectral Imaging pipeline described in this report has been in practical use for digital conservation and enhancement of ancient manuscripts in the past years. It has proven a significant aid in humanists' attempts of deciphering degraded handwritings.

#### Библиографический список

1. *Cosentino A.* Identification of pigments by multispectral imaging; a flowchart method? *Heritage Science* 2014 2: 8. doi:10.1007/s11694-014-9282-8.
2. *Easton R.L., Christens-Barry W.A., Knox K.T.* Spectral image processing and analysis of the Archimedes Palimpsest, *2011 19th European Signal Processing Conference*, Barcelona, 2011, pp. 1440–1444.
3. *Lowe D.* Distinctive image features from scale-invariant keypoints, *International journal of computer vision* 60 (2), 91–110.
4. *Heinrich M. et al.* MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, *Medical Image Analysis* 2012 Oct; 16(7):1423-35.
5. *Hollaus F., Gau M., Sablatnig R.* Enhancement of Multispectral Images of Degraded Documents by Employing Spatial Information, *Proceedings of the International Conference on Document Analysis and Recognition 2013*, 145–149.
6. *Zuiderveld K.* Contrast Limited Adaptive Histogram Equalization. *Graphic Gems IV*. San Diego: Academic Press Professional, 1994. 474–485.

## **ЧТЕНИЕ МЕЖДУ СТРОК: МНОГОСПЕКТРАЛЬНЫЕ ИЗОБРАЖЕНИЯ ДРЕВНИХ РУКОПИСЕЙ**

*Бреннер С., Саблатник Р.*

Лаборатория компьютерного зрения, Институт компьютерной автоматизации, Австрия, Вена, ул. Фаворитен, 9/183-2 А-1040, [sbrenner@caa.tuwien.ac.at](mailto:sbrenner@caa.tuwien.ac.at)

Древние рукописи являются ценным источником информации для историков и филологов. Тем не менее, многие из этих текстов находятся в плохом состоянии и плохо читаемы либо из-за естественного угасания источника (плесень или выцветание чернил), либо из-за того, что источник является палимпсестом. В работе представлен набор инструментов и методов для мультиспектральной съемки, предназначенных для повышения читабельности угасших рукописных текстов. Эти методы покрывают все необходимые этапы процесса от спектральной съемки, регистрации и сокращения размерности до повышения читабельности. Полученные цифровые изображения способствуют дальнейшему изучению и анализу исторических рукописей.

*Key words:* мультиспектральная съемка, улучшение читабельности, исторические рукописи.

УДК 930:91.528:004

## **SPACE AND THE REPRESENTATION OF SPACE DIGITIZATION AND ANALYSIS OF THE CENTRAL EUROPEAN CENSUS 1869–1910**

*Goderle W.*

University of Graz, Austria, Graz, Attemsgasse, 8/IV 8010, [wolfgang.goederle@uni-graz.at](mailto:wolfgang.goederle@uni-graz.at)

The paper deals with the representation of historical statistical data in GIS and the problems that arise when linking geographical and numerical data extracted from historical sources in contemporary digital representation systems. Particularly for the 19<sup>th</sup> century the smoothing of geo-data and numerical data equals an intervention into primary sources, as form and content of the data cannot be separated easily. Such an intervention threatens to shift certain meanings of historical representations by prolonging contemporary ways of looking upon such data retrospectively into the past. This applies particularly to seemingly “objective” classes of data, such as maps and statistics. The paper aims to raise awareness for these issues by exploring a case study regarding the late Habsburg Empire between roughly 1870 and 1910.

*Key words:* Historical statistics, historical GIS, Central Europe, Habsburg Empire, Social History, Imperial History, Late Modern History, 19<sup>th</sup> century.

In the course of the 19<sup>th</sup> century, the rise of modern statehood transformed the surface of Europe. By the middle of the century, mapping and counting were among those activities, administrations pursued most perseveringly and insistently.

In the Habsburg Empire, the first census to cover the entire territory of the state in a uniform way, was conducted in 1848, but its results proved most unsatisfactory. Another census in 1857 failed. The tradition of the modern census began in 1869 and delivered reliable data on the total population and its development for almost half a century on the basis of a decennial survey.

The data raised in these census operations formed the basis of the GAFP (Graz Austrian Fertility Project in analogy to the PEFP – Princeton European Fertility Project) research project, which investigated into the core parameters of demographic evolution in this part of Central Europe on the level of the political district.

Therefore, the entire census data was digitized from 2009 onwards and integrated into a database system. In the next step, database contents were analysed in order to reconstruct the demographic behavior of the population of nowadays Austria between roughly 1870 and 1910. The analysis focused on four major dimensions of demographic behavior: Fertility, mortality, nuptiality and migration. Whereas fertility, mortality nuptiality and can be deduced from the data provided by contemporary administrations in a satisfactory way, migration constituted a more difficult field of research. Migration could not be recorded in a direct way, thus we had to rely on indirect calculation (population balance) and on the little information the census contained.

We imagined that mapping the data we had might help us in understanding the major processes behind the phenomenon under scrutiny. Archival research provided us with map material displaying political district borders for 1885, 1900 and 1914. The project team built a database allowing for a historical reconstruction of the mapspace required for the census years and map data and census data were linked.

Although the analytical process turned out very positive and the maps produced contributed essentially to the understanding of the phenomenon of migration as represented in the census sources, we were astonished by the degree of inaccuracy we had to accept in order to be able to produce these maps. The fact, that modern globes provided by software solutions such as ArcGIS differ considerably from the surface projections used in 19<sup>th</sup> century cartography led to distortion in the display of geo-referenced historical map material. As this effect affects only the optical representation of the results, its consequences were negligible.

Less negligible, however, was the fact, that both, the users and the producers of the historical map material disposed of a different set of ideas

and concepts on how space was supposed to be imagined and used. Thus, our reconstruction not only distorted the optics of the historical map material, but it also interfered with a certain concept of space and its appropriation and use.

Finally, it turned out that any exact reconstruction of political districts could be achieved only under extreme efforts. Administrative borders and entities were constantly changing, despite the impression of continuity and stability they might evoke. But the Habsburg Empire featured a total of more than 400 political districts and several thousand communes, and major changes were taking place on a yearly basis. Whether the central administrative institutions of the state themselves disposed of up-to-the-minute information concerning the current status quo of its administrative structure is to a certain extent questionable. Looking upon the map material the latent conflict materializes for the historian dealing with these issues.

In my presentation I will present some of the results of the research project, and I will tackle the issues that may arise when dealing with historical map material particularly from the 19<sup>th</sup> and early 20<sup>th</sup> centuries, when maps looked already quite similar to those in use today, though they differed considerably in detail yet.

### **ПРОСТРАНСТВО И РЕПРЕЗЕНТАЦИЯ ПРОСТРАНСТВА ОЦИФРОВКИ И АНАЛИЗА ПЕРЕПИСЕЙ ЦЕНТРАЛЬНОЙ ЕВРОПЫ 1869–1910**

*Гедерле В.*

Университет Граца, Австрия, Грац, пер. Аттемс, 8/IV 8010,  
wolfgang.goederle@uni-graz.at

В статье рассматриваются вопросы представления исторических статистических данных в ГИС, а также проблемы, которые возникают при связывании географических и количественных данных, извлеченных из исторических источников в современных цифровых системах репрезентации данных. В частности, для данных XIX века сглаживание гео-данных и числовых данных эквивалентно вмешательству в первичные источники, поскольку форма и содержание данных не могут быть легко разделены. Такое вмешательство угрожает сместить определенные значения исторического представления данных посредством увеличения продолжительности современных способов обозрения таких данных в ходе ретроспективного анализа прошлого. Это особенно относится к, казалось бы, «объективным» типам данных, таким как карты и статистические данные. Целью данной работы является повышение осведомленности по этим вопросам на основе изучения Поздней Империи Габсбургов между 1870 и 1910 гг.

*Key words:* историческая статистика, исторические ГИС, Центральная Европа, Империя Габсбургов, социальная история, имперская история, поздняя новейшая история, XIX век.

## ACCESS TO CENSUS MICRODATA AND AGGREGATES<sup>1</sup>

*Thorvaldsen G.*

Norwegian Historical Data Centre, University of Tromsø, 9037 Tromsø,  
Norway, [gunnar.thorvaldsen@uit.no](mailto:gunnar.thorvaldsen@uit.no)

With the digital computer came a revolution in access to census data in many formats. Many statistical agencies have made their statistical publications from censuses available as scanned images on CD-roms or via the Internet. Increasingly, however, researchers prefer to compute their own aggregates from microdata on the individual level, since this provides more fine-grained statistics, can be easier to compare over time and avoids ecological fallacies. Fortunately, most archives have understood that it is sound source protection to spread copies and most statistical agencies have understood that their own economic interests should not block researchers' access to the microdata. All together, these databases contain in excess of two billion individual records, which can be analyzed online or more thoroughly after downloading selected parts. The optimal solution is to link census microdata with vital registers and thus create population registers.

*Key words:* census, microdata, Internet, record linkage, population registers.

Until the 1960s, the only form of access to information from censuses was on paper or microfilm, the latter relevant for copying census manuscripts, books being the normal way for each statistical bureau to publish the aggregates. Main variables were also available in comparative international volumes, made available by Mitchell since 1978[1]. With the digital computer came a revolution in access to census data in many formats. The simplest to use are the search engines popular among genealogists for finding information about individuals; these are also used in academic research to identify groups of persons under study. Such research will usually cater for statistical information, including the scanned versions of the above-mentioned volumes with aggregates, which are now more easily available via the Internet than the book versions in libraries. Digital aggregate data was first collected as variables about administrative units such as municipalities, often combining census data with information from other sources such as vital registers. The many border changes of these units create problems if studying development over time, however.

This is one of the reasons why access to individual level census data has become so widespread for statistical research both historically and for studies of the present-day situation. When such materials are younger than

---

© Thorvaldsen G., 2017

<sup>1</sup>The work was supported by the RSF, grant No. 16-18-10105 "Ethno-religious and demographic dynamics in mountainous Eurasia around 1900. A comparison of the Urals and Scandinavia".

72 to 100 years – depending on country – they must be anonymized by removing names, birthdates and addresses, and de-identified by making it impossible to identify anybody indirectly. Older, historical censuses can be used as complete versions, however, and in this sense the work to spread nominative census taking from the mid-19<sup>th</sup> century still carries fruit. The digital versions have been encoded so that accidental differences in spelling etc are removed, variables are constructed for the study of family structure, the formats are harmonized with respect to the coding of variable values and the record structure is as similar as possible [2]. This work has been carried furthest by the Minnesota Population Center in order to facilitate comparative research across time and space. Here all the surviving nominative US censuses from 1850 to the present have been made available either as full-count or representative samples, with or without identifying information in the IPUMS – International Public Use Microdata Samples [3]. Another major database is the IPUMS-International with census samples of households from countries on all continents after the 1950s. Among the exceptions are countries where the statistical agencies have not yet understood that the significant advantages of the microdata versions far outweigh the unlikely risk that users can identify anybody in the samples and anyway must sign a declaration of non-disclosure. Thus, the problem is not differences created by the lawmakers, but rather that similar laws of statistics are interpreted differently. The third major integrated database is the NAPP – North Atlantic Population Project with full-count censuses covering North America, Scandinavia and the UK from 1801 to 1910 [4]. Such compilation of joint resources enable the transcription or transfer of an increasing number of censuses into full-count digital sources rather than samples, enabling also the detailed study of small population groups. All together, these databases contain in excess of two billion individual records, which can be analyzed online or more thoroughly after downloading selected parts [5].

In addition to earlier in-house transcriptions, the Minnesota Population Center collects its census materials in cooperation with census bureaus, genealogical companies and population researchers in most countries of the world. Among the partners on the historical censuses, mainly from the 19<sup>th</sup> century, this book has focused on North America and Scandinavia. Like in the US, the start was made by researchers who turned source materials into individual level census data with specific research questions in mind. Increasingly, the projects turned into specific infrastructure undertakings focusing on the transcription, encoding and standardizing of census materials. These databases are then built with an eye to making historical census data available both for their own research and for other researchers. In Canada, representative samples or full-count versions of all the censuses from 1851 are now available, either for downloading or analysis via specially protected networks in so-called data enclaves. Norway has put five full-count nomina-

tive censuses online for the period 1801 to 1910 while those from 1960 onwards are only available as deidentified files sold by Statistics Norway to bona fide researchers. In a project to build a historical population register from 1801 onwards, work is now underway on scanning the remaining paper-based censuses in order to have them transcribed for integration through record linkage together with the nominative vital records [6].

### References

1. *Mitchell B. R.* International historical statistics: Europe 1750–1993 / B.R. Mitchell. London: Macmillan, 1998.
2. *Thorvaldsen G.* Номинативные источники в контексте всемирной истории переписей: Россия и Запад // Известия Уральского федерального университета. Серия 2: Гуманитарные науки. 2016. № 18(3). С. 9–29.
3. *McCaa R.* Thanks to 70 years of Inter American Statistical cooperation, the world's largest integrated census microdata dissemination site [www.ipums.org/international](http://www.ipums.org/international) // Estadística, 2013. № 65(184). P. 31–45.
4. *Roberts E.* The North Atlantic Population Project: An Overview // Historical Methods, 2003. № 36(2). P. 80–88.
5. *Ruggles S.* Big Microdata for Population Research // Demography, 2014. № 51(1). P. 287–297.
6. *Thorvaldsen G.* Using NAPP Census Data to Construct the Historical Population Register for Norway. HistoricalMethods. 2011. 44(1). Pp. 37–47.

### ДОСТУП К МИКРОДАНЫМ ПЕРЕПИСЕЙ И АГРЕГИРОВАННЫМ ДАННЫМ

*Торвальдсен Г.*

Норвежский исторический центр, Университет Тромсе,  
9037 Тромсе, Норвегия, [gunnar.thorvaldsen@uit.no](mailto:gunnar.thorvaldsen@uit.no)

В эпоху цифровых технологий произошла и революция в доступе к данным переписей населения в различных форматах. Многие статистические агентства опубликовали данные переписей в виде отсканированных изображений на Сd-дисках или в Интернете. Однако все чаще исследователи предпочитают рассчитывать собственные совокупности микроданных на индивидуальном уровне, так как это обеспечивает более точную статистику и дает возможность сравнить динамику и избежать «экологических заблуждений». Большинство архивов признали, что распространение копий способствует защите первоисточника; а большинство статистических агентств – что их экономические интересы не должны ограничивать доступ исследователей к микроданным. В общей сложности эти базы данных содержат свыше двух млрд отдельных записей, которые могут быть проанализированы как в режиме онлайн, так и более тщательно после скачивания выбранных частей. Оптимальным решением является связывание микроданных переписей с актами гражданского состояния и создание таким образом регистров населения.

*Ключевые слова:* переписи, микроданные, Интернет, связывание данных, регистры населения.

*Научное издание*

**ЦИФРОВАЯ ГУМАНИТАРИСТИКА:  
РЕСУРСЫ, МЕТОДЫ, ИССЛЕДОВАНИЯ**

Материалы международной научной конференции  
(г. Пермь, 16–18 мая 2017 г.)

Часть 1

Редактирование материалов и верстка оригинал-макета издания  
выполнены ООО «Эдумаркет»

---

Подписано в печать 24.04.2017. Формат 60×84/16  
Усл. печ. л. 10,18. Тираж 180 экз. Заказ \_\_\_\_

---

Издательский центр  
Пермского государственного  
национального исследовательского университета.  
614990, г. Пермь, ул. Букирева, 15

Отпечатано на ризографе  
ООО Учебный центр «Информатика»  
614990, г. Пермь, ул. Букирева, 15