

Министерство науки и высшего образования Российской Федерации
Сибирский федеральный университет

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ

Материалы Международной научно-практической конференции

Красноярск, 25–28 сентября 2023 г.

Электронное научное издание

Красноярск
СФУ
2023

УДК 168.522:004.9(08)
ББК 71я4+32.97я4
И741

Ответственный редактор – Лаптева Марина Анатольевна

И741 Информационные технологии в гуманитарных исследованиях : материалы Междунар. науч.-практ. конф. Красноярск, 25–28 сентября 2023 г. / отв. ред. М. А. Лаптева. – Электрон. дан. (5,6 Мб). – Красноярск : Сиб. федер. ун-т, 2023. – Электрон. опт. диск (CD-ROM). – Систем. требования : PC не ниже класса Pentium I ; 128 Mb Ram ; Windows 98/XP/7 ; Adobe Reader v 8.0 и выше. – Загл. с экрана.
ISBN 978-5-7638-4900-4

В настоящем издании представлены материалы докладов Международной научно-практической конференции «Информационные технологии в гуманитарных исследованиях», которая состоялась в сентябре 2023 г. в Красноярске. Статьи посвящены актуальным вопросам развития цифровых гуманитарных наук в России и мире и размещены в шести разделах, соответствующих основным тематическим направлениям конференции: «Цифровое наследие»; «Цифровая история»; «Digital Humanities в образовании»; «Цифровые проекты»; «Компьютерный анализ текста»; «Пленарные доклады».

Предназначены специалистам, работающим в области цифровых гуманитарных исследований, а также всем, кому интересно современное развитие гуманитарного знания.

Ответственность за аутентичность и точность цитат, имён и иных сведений, а также за соблюдение законов об интеллектуальной собственности несут авторы публикуемых материалов.

УДК 168.522:004.9(08)
ББК 71я4+32.97я4

ISBN 978-5-7638-4900-4

© Сибирский федеральный
университет, 2023

Электронное научное издание

Корректор Л. В. Боос
Компьютерная вёрстка Е. А. Сафиной

Подписано в свет 18.12.2023. Заказ № 20342
Тиражируется на машиночитаемых носителях
Библиотечно-издательский комплекс
Сибирского федерального университета
660041, г. Красноярск, пр. Свободный, 82а
Тел.: (391)206-26-16; <http://rio.sfu-kras.ru>
E-mail: publishing_house@sfu-kras.ru

СОДЕРЖАНИЕ

РАЗДЕЛ 1. ЦИФРОВОЕ НАСЛЕДИЕ.....	7
<i>Гайдук В. Л.</i> Можно ли сделать датасет из архивной базы: опыт Медиатеки ГМИИ им. А. С. Пушкина.....	7
<i>Кожин К. Д., Кижнер И. А.</i> Оценка вариативности описания пропусков в значениях полей метаданных учётных записей объектов коллекции Госкаталога.....	12
<i>Мазин Е. И., Полоян А. В.</i> Digital Humanities в сохранении культуры и традиций: говорящий город.....	20
<i>Гук Д. Ю., Пермяков В. А.</i> Электронная энциклопедия Эрмитажа. Итоги первой пятилетки.....	24
<i>Пиков Н. О., Дворецкая А. П., Меховский В. А., Дяченко А. С.</i> Виртуальная репрезентация сакрального пространства Енисейской Сибири.....	27
<i>Шемелина Д. С., Радченко Л. К., Декунцов Н. С.</i> Архитектурно-градостроительное наследие сибирских оборонительных линий XVIII века: поиск перспектив ревалоризации с помощью геоинформационной модели.....	32
РАЗДЕЛ 2. ЦИФРОВАЯ ИСТОРИЯ.....	40
<i>Бабайцев М. Н.</i> Обработка исторических источников инструментарием OCR, табличных процессоров и ГИС на примере списка населённых мест Тверской губернии.....	40
<i>Владимиров В. Н., Брюханова Е. А., Неженцева Н. В.</i> Информационная система по исторической статистике Алтая: особенности формирования комплекса региональных данных.....	46
<i>Дамдынчап В. М.</i> Пространственная репрезентация и анализ социодемографического пространства территории Тувы советского периода на основе ГИС-технологий.....	51
<i>Данилов Е. В.</i> Опыт применения статистических методов для анализа процессов обращаемости за медицинской помощью в России в конце XIX века.....	60
<i>Саломатина С. А.</i> Расстояния между региональными рынками Российской империи во второй половине XIX века: сетевой и статистический анализ.....	64

<i>Солощенко Н. В.</i> Образ «нового рабочего» пищевой промышленности СССР в годы первой и второй пятилеток в отражении публицистических материалов (опыт сетевого анализа и контент-анализа).....	70
<i>Стальмакова А. В.</i> Проблемы советской истории на страницах калининградской партийной прессы в период перестройки.....	74
<i>Стороженко А. А.</i> Старообрядческие миграции в 30–50-е годы XX века в Сибирь: опыт применения цифровых технологий.....	79
<i>Воронина П. С.</i> База данных «Карбеженцы: люди и судьбы (1920-е годы)»: концепция и структура.....	82
<i>Акашева А. А.</i> Наборы данных, локальные проекты, онлайн-ресурсы об изданиях Первой всеобщей переписи населения 1897 года Российской империи.....	87
РАЗДЕЛ 3. DIGITAL HUMANITIES В ОБРАЗОВАНИИ.....	89
<i>Водолажская Л. Н.</i> Трансдисциплинарный курс «Цифровые и естественно-научные методы в археологии».....	89
<i>Жила В. В., Михайлова М. А., Ермолаев Е. А., Кочнева Я. В.</i> Цифровое гуманитарное обучение как основа разработки учебных программ.....	93
<i>Ишанходжаева З. Р.</i> Цифровые технологии в высшем историческом образовании в Узбекистане.....	98
РАЗДЕЛ 4. ЦИФРОВЫЕ ПРОЕКТЫ.....	102
<i>Брюханова Е. А., Рыгалова М. В.</i> Справочные издания как источники по городскому населению Сибири конца XIX – начала XX века.....	102
<i>Matkin N. A., Klyuev N. A.</i> A city in a mirror of social media: automated analysis approach to urban representations...	107
<i>Мадрахимова Д. С.</i> Применение искусственного интеллекта для экспертной оценки успешности карьеры российских пореволюционных эмигрантов с учётом их дореволюционного положения в академическом сообществе России.....	114
<i>Косовская К. К., Тихонов Г. А., Долгозвягов Д. Д.</i> Цифровые инструменты и ресурсы, применяемые в гуманитарных науках.....	119

<i>Дяченко А. С., Кижнер И. А.</i> Подход к мультимодальному анализу периодического издания «Курьер ЮНЕСКО»: сегментация и извлечение изображений.....	124
<i>Гурьянов И. С.</i> Тенденции развития исторических ГИС.....	134
<i>Акашева А. А.</i> Информационные ресурсы Ассоциации «История и компьютер».....	137
<i>Брылевская Е. В.</i> Некоторый алгоритм для построения семантических сетей.....	139
РАЗДЕЛ 5. КОМПЬЮТЕРНЫЙ АНАЛИЗ ТЕКСТА.....	142
<i>Бутусова М. А.</i> Сетевой анализ речи героев пьес А. П. Чехова «Три сестры» и «Вишнёвый сад»....	142
<i>Ведерникова В. Д., Северина Е. М.</i> Цифровой анализ идиостиля автора и переводчика (на материале текстов произведений Дж. К. Роулинг).....	147
<i>Максименко П. И.</i> Русскоязычная электронная база фанфикшн-текстов: принципы создания и анализ метаданных.....	151
<i>Панкова М. С.</i> Корпус донских говоров как opportunistic corpus.....	160
<i>Белковский В. А., Максаков И. П.</i> Перспективы использования компьютеризированного контент-анализа как метода гуманитарного исследования в современной практике.....	164
<i>Артемьева А. И.</i> Технологии оценки доступности текстовой информации в практике создания безбарьерной коммуникационной среды.....	169
<i>Сальников Е. А., Бонч-Осмоловская А. А.</i> Применение стилометрии для определения сгенерированных текстов.....	176
<i>Долуденко Е. А., Горохова Л. А., Горохова А. В.</i> Опыт корпусного анализа дневниковых записей студентов-переводчиков.....	183
<i>Emmanuel Ngue Um</i> The erasure of Africa in the era of ChatGPT.....	189
<i>Полоян А. В., Приходько К. Д.</i> Составление промпта в нейросети Midjourney для генерации иллюстраций к творчеству А. П. Чехова.....	194

<i>Шерстинова Т. Ю., Кирина М. А., Хлусова Я. К.</i> Корпус русского рассказа как база для проведения социолингвистических исследований русской литературы.....	200
<i>Шилова М. В.</i> Визуализация исторических текстов: выявление географических связей цифровыми средствами.....	212
<i>Синицын В. Ю., Назаровская В. С.</i> Тестирование эффективности больших языковых моделей при выполнении заданий единого государственного экзамена по русскому языку.....	221
<i>Северина Е. М.</i> Проект Chekhov Digital: цифровой индекс имён и названий.....	228
<i>Баранов В. А.</i> Дистрибуция и семантика славянского слова в историческом корпусе.....	232
<i>Михайлова А. Н.</i> Атрибуция мемуаров Г. К. Жукова и К. К. Рокоссовского с помощью метода Delta...	238
РАЗДЕЛ 6. ПЛЕНАРНЫЕ ДОКЛАДЫ.....	246
<i>Бородкин Л. И.</i> Нечёткая кластеризация с элементами обучения в историко-типологических исследованиях.....	246
<i>Норманская Ю. В.</i> Возможности «ЛингвоДока» для выявления степени близости языков и диалектов и описания графико-фонетических особенностей.....	251
<i>Orekhov B. V.</i> Hard cases in cross-corpora drama annotation.....	258

B. V. Orekhov,

Candidate of Philological Sciences, Associate Professor, Senior Researcher
National Research University Higher School of Economics, Moscow, Russia
Institute of Russian Literature (Pushkin House) of the Russian Academy of Sciences,
St. Petersburg, Russia
e-mail: nevmenandr@gmail.com
ORCID: 0000-0002-9099-0436

HARD CASES IN CROSS-CORPORA DRAMA ANNOTATION

The paper deals with the observations that are made during the work on Dracor.org [1]. Originally, Dracor.org is a set of isolated corpora in different languages with a unified TEI markup that allows you to do quantitative drama studies [2]. But since the platform now includes a noticeable set of corpora in German, English, Russian and other languages, it is natural to try to study cross-corpora drama features. These features include characters that appear in different plays and even in different corpora. Usually these are historical personalities (e.g. kings), mythological figures (Achilles, Heracles), popular literary characters (Faust). The idea is to mark all the characters that are mentioned in the plays more than once. We must not forget that Euripides' Achilles (*Iphigenia in Aulis*) is a different character than Kleist's Achilles (*Penthesilea*). Technically, these characters are part of different fictional systems. But in the future we would like to explore the differences of interaction between these characters and the other components of the play. This will help to establish the very difference between fictional systems and the author's style. So we have to mark the heroes with the same name to be able to extract them automatically from the corpus.

All characters in Dracor have their ids, which depend on their names:

```
<person xml:id="Othello_Oth">
  <persName>Othello</persName>
</person>
```

But id in the play is not suitable for cross-corpora annotation tasks. For example, the character "messenger" is found in many plays, and often has the same id, but this is not the case when we are going to mark these characters as the same cross-corpora character. In addition, characters in different plays may occasionally have the same names. That means we need to have a separate attribute for this kind of case.

We have chosen the attribute "ana" for the tag "person". Due to the TEI guidelines this attribute "indicates one or more elements containing interpretations of the element on which the ana attribute appears". This means that we interpret the character of the play as some kind of entity that may appear

in other plays. The most convenient option would be to associate the value of this attribute with the Linked Open Data. We use Wikidata, which provides the URI for the entity we find in different plays.

```
<person xml:id="phaedra" sex="FEMALE" ana="http://www.wikidata.org/entity/Q212818">
```

This approach works very well and allows us to identify the characters in different plays. But there are cases where we can see the inconvenience of Wikidata URI for the purposes of our cross-corpora markup. There are three types of such cases.

1. **Umbra case.** In ancient plays, characters and their shadows are often found separately. Sometimes they are found within the same play. Then they have different dracor ids, but the URI on Wikidata is the same. See Tantalus and Tantali Umbra in *Thyestes* by Seneca.

```
<person xml:id="tantalus" sex="MALE" ana="http://www.wikidata.org/entity/Q515531">
  <persName>Tantalus</persName>
</person>
<person xml:id="tantaliumbra" sex="MALE" ana="http://www.wikidata.org/entity/Q515531">
  <persName>Tantali Umbra</persName>
</person>
```

and Δαρειός and Εἰδωλον Δαρείου in *Persians* by Aeschylus:

```
<person xml:id="eidolon_dareiou" sex="MALE" ana="http://www.wikidata.org/entity/Q44387">
  <persName>Εἰδωλον Δαρείου</persName>
</person>
<person xml:id="dareios" sex="MALE" ana="http://www.wikidata.org/entity/Q44387">
  <persName>Δαρειός</persName>
</person>
```

At the moment we have to mark these characters with a single id. But it's not the best solution.

2. **Ivppiter case.** The purpose of cross-corpora markup is to tie together characters from different plays and different corpora at the same data object. From this point of view, it is obvious that Zeus and Jupiter are the same character, they perform the same function in the story about Hercules. But in Wikidata we find different URIs for Zeus (URL: wikidata.org/wiki/Q34201) and Jupiter (URL: wikidata.org/wiki/Q4649). Thus, our data is separated. The plan is to get all the necessary information about the characters using SPARQL queries and characters ids. Perhaps, this case could be solved by the additional parameter of SPARQL query (P460). But every time this parameter must be different (depends on gods, kings or someone else we deal with).

3. **Talthybius case.** Talthybius has two relevant objects in Wikidata. First of all, it's a character from a myth (URL: wikidata.org/wiki/Q1367860). This URI allows us to connect the character of the tragedy *Troades* by Seneca and *The Trojan Women* by Euripides. But the Euripides tragedy character has his own URI in Wikidata (URL: wikidata.org/wiki/Q60607554). Technically,

its attribute should have this value, but it contradicts the idea of tying the characters of different plays together. Additional SPARQL parameter P1074 ("fictional analog of") could solve this problem. Another solution is to extend Wikidata by ourselves and add Talthybius to the parameter "present in work" (P1441) from Dracor to the certain play objects.

References

1. Skorinkin D. Building a Corpus for the Quantitative Research of Russian Drama: Composition, Structure, Case Studies / D. Skorinkin, F. Fischer, G. Palchikov // Компьютерная лингвистика и интеллектуальные технологии. 2018. Pp. 662–682.
2. Skorinkin D. Measuring the "Epification" of Drama / D. Skorinkin, F. Fischer // Applications in Cultural Evolution: Arts, Languages, Technologies. Tartu: University of Tartu, 2018. Pp. 46–48.