

# Факты — События — Коммуникации

УДК 021:004.9

ББК 78.002(3)

DOI 10.25281/0869-608X-2019-68-1-103-111

М.Е. ШВАРЦМАН, О.Н. КРЫЖАНОВСКАЯ

## Цифровые методы исследования: новый вектор информатизации библиотек

**Реферат.** За рубежом проходит большое количество конференций на тему автоматизации библиотечных процессов. На этих мероприятиях обсуждаются самые актуальные и важные вопросы, встающие перед библиотеками при внедрении современных информационных технологий. В статье приведен обзор четырех основных международных конференций, на которых собираются программисты, лингвисты и библиотекари и делятся опытом и планами на будущее. Приведен анализ изменения тематики конференций за последние годы, сделан вывод о значительном увеличении количества исследований, посвященных семантическому анализу хранящихся в библиотеках данных, отмечены наиболее интересные доклады. В последние годы увеличивается количество докладов, посвященных Linked Open Data (LOD). Однако большинство из этих докладов посвящено публикации данных в LOD, очень мало исследований о масштабах использования уже опубликованных данных. В статье приведены примеры подобных исследований. Постоянный интерес вызывает тема Digital Humanities («Цифровые гуманитарные исследования») и то, что проводимые в ее рамках исследования часто сосредоточены на проблемах современного общества. Отмечено новое направление деятельности электронных библиотек — хранение исследовательских данных и организация научных коммуникаций. Приведены примеры реализации этих задач.

**Ключевые слова:** цифровые гуманитарные исследования, конференции, связанные открытые данные, LOD, программирование, автоматизация библиотеки, автоматизация, обмен опытом, семантический анализ, программное обеспечение, информационное обеспечение, научные коммуникации, TPD, SWIB, ELAG, Code4Lib, профессиональные мероприятия.

**Для цитирования:** Шварцман М.Е., Крыжановская О.Н. Цифровые методы исследования: новый вектор информатизации библиотек // Библиотековедение. 2019. Т. 68, № 1. С. 103—111. DOI: 10.25281/0869-608X-2019-68-1-103-111.



**Михаил Ефремович Шварцман,**  
Российская государственная библиотека,  
отдел исследования компьютерных систем,  
начальник отдела  
Воздвиженка ул., д. 3/5,  
Москва, 119019, Россия  
ORCID 0000-0003-2524-6819;  
SPIN 4277-9746  
E-mail: ShvartsmanME@rsl.ru



**Ольга Николаевна Крыжановская,**  
Российская государственная библиотека,  
отдел исследования компьютерных систем,  
главный специалист  
Воздвиженка ул., д. 3/5,  
Москва, 119019, Россия  
E-mail:  
KryzhanovskayaON@rsl.ru

**Б**ольшинство документов, имеющих в библиотеках, в настоящее время предоставляются пользователям посредством программного обеспечения (ПО). Решения, которые принимаются при выборе ПО для библиотек, во многом определяют успешность их развития, поэтому важно знать состояние рынка ПО и последние его тенденции. Как писал директор по развитию web-технологий Библиотеки Университета Миннесоты Коди Хэнсон (Cody Hanson): «Мы настолько хороши, насколько хорошо наше программное обеспечение» [1].

Один из лучших способов выяснить современные тенденции в любой области — это послушать, о чем говорят на конференциях, посвященных этой теме. Несмотря на то, что физически сложно посетить все достойные мероприятия даже по одной теме, опубликованные в Интернете их программы, тезисы выступлений и презентации вполне могут помочь определить, какие задачи стояли перед специалистами по автоматизации библиотек в 2018 году. Главное — правильно определить список рассматриваемых событий. В данном исследовании выбраны для анализа следующие конференции:

- TPDL 2018 (International Conference on Theory and Practice of Digital Libraries) существует с 2011 г. и является продолжением конференции ECDL (European Conference on Digital Libraries) (<http://www.tpd1.eu/tpdl2018>);

- Code4Lib 2018 — конференция сообществ разработчиков ПО для библиотек, музеев и архивов, которые имеют сильную приверженность открытым технологиям; проект начинался как простой список рассылки в 2003 г., в настоящее время организаторы издают журнал и проводят ежегодные конференции (<https://2018.code4lib.org>);

- ELAG 2018 (European Library Automation Group) — европейская конференция по применению информационных технологий в библиотеках и центрах документации, проводимая ежегодно с 1979 года (<https://www.elag2018.org>);

- SWIB 2018 (Semantic Web in Libraries) — ежегодная конференция, проводимая с 2009 г., посвященная связанным открытым данным в библиотеках (<http://swib.org/swib18>).

Конечно, этот список не является исчерпывающим перечнем конференций по выбранной теме. На многих мероприятиях (IFLA, LIBER, OCLC и т. п.) рассматриваются наряду с прочими и вопросы автоматизации библиотечных процессов. Существуют также многочисленные региональные конференции.

В изменениях тематики докладов, представленных на конференциях, от года к году прослеживается история развития автоматизированных библиотечных систем и проблематика, которая волновала их разработчиков. Некоторые вопросы долгие годы не теряют актуальности. Например, еще на первой конференции ECDL в 1997 г. был представлен доклад про семантические сети [2], а в наше время на эту тему проводится отдельная конференция. Некоторые темы, которые были «модными» еще несколько лет назад, в последние годы уже не обсуждают. Так, до 2015 г. на многих конференциях часто встречались выступления на тему интероперабельности информационных систем, однако после 2015 г. таких докладов практически не встречалось. Последние два года не рассматриваются вопросы организации мультимедийных коллекций, но это не означает, что данные направления больше никого не интересуют. В связи с развитием технологий наблюдается все более узкая специализация разработчиков, и им становятся интересны более специализированные конференции.

Показательными являются изменения в наименовании секций на конференции TPDL. Программный комитет отбирает из большого потока поступающих докладов наиболее актуальные и формирует из них секции с примерно одинаковым количеством выступлений, поэтому изменение состава секций за последние четыре года<sup>1</sup> дает возможность сделать определенные выводы об изменениях актуальных тематик. Самая стабильная проблематика в последние годы — «Цифровые гуманитарные исследования» (Digital Humanities). В рамках этой темы авторы представляют доклады по исследованию проблем общества с использованием современных методов. Отметим выступление Симона Готчалка (Simon Gottschalk) в соавторстве с коллегами из трех европейских стран о стратегиях исследователей, анализирующих событийно-ориентированную многоязычную информацию [3]. В докладе рассматривалось отношение к референдуму по Brexit и выходу США из Парижского соглашения. Междисциплинарные исследовательские группы в течение недели проводили сводный анализ записей об этих событиях в Википедии на разных языках. В ходе этого мониторинга осуществлялось наблюдение за методами, используемыми участниками, и результатами, которые они получили.

Другим постоянным направлением, обсуждаемым на TPDL, является проблема поиска информации. Среди докладов по индексированию

в 2018 г. можно отметить доклад Сусмиты Садху (Susmita Sadhu) и ее коллег из Индийского института технологий, посвященный семантическому анализу запросов. На основе анализа запросов к Национальной электронной библиотеке Индии был разработан алгоритм автоматической сегментации и семантической аннотации длинных запросов [4]. Под семантической аннотацией в данном исследовании понимается отнесение отдельных сегментов к определенным полям метаданных, т. е. выделение в неструктурированном запросе потенциального автора, издателя и т. п.

Поиск информации в коллекциях важен и в контексте взаимодействия электронной библиотеки и пользователя, поэтому на традиционной секции «Взаимодействие с пользователем» (User Interaction) обсуждалось, как, изучая поведение читателя, можно улучшить эффективность поиска. Так, Перти Ваккари (Pertti Vakkari) из Университета Тампере вместе с немецкими коллегами рассказывал об исследовании, позволяющем повысить эффективность поиска информации для копирайтеров на основе анализа используемого ими текста [5]. Это исследование было проведено на основе Webis Text Reuse Corpus (<https://webis.de/data/webis-trc-12.html>). Отметим, что большое количество доступных для анализа текстов в последние годы предоставляет исследователям новые возможности.

Задачи анализа текстов становятся все более значимыми для разработчиков электронных библиотек. Теперь на конференциях по электронным библиотекам все чаще можно встретить доклады, которые ранее звучали только на специализированных конференциях по информационному поиску. И некоторые традиционные секции, оставляя прежние наименования, меняют суть рассматриваемых вопросов. Так, на секции «Метаданные» (Metadata) все меньше внимания уделяется форматам описания, а больше — методам индексирования и обогащения метаданных различными лингвистическими способами. Например, французские ученые из Университета Лиона рассказали о своем успешном опыте обогащения ключевыми словами документов электронной библиотеки [6]. При поиске научных документов из новой области исследователь не всегда может угадать, какими терминами пользовались авторы нужного ему документа. В некоторой степени в обогащении запроса могут помочь словари синонимов. Однако, как выяснилось в результате описываемого эксперимента, еще на 11% успешнее проходит поиск по заранее обработанному массиву, в

котором были проведены сравнение векторных представлений документов и взаимный обмен ключевыми словами для документов с совпадающими векторами. Похожие доклады встречались также и на конференции SWIB. Сотрудники Национальной библиотеки Финляндии делились опытом автоматического определения предметных рубрик из контролируемого словаря на основе ассоциативного подхода к анализу текстов [7] и даже предложили всем желающим проверить их подход на сайте (<http://dev.annif.org/>).

Новые технологии позволяют ставить новые задачи перед электронными библиотеками и как следствие появляются новые секции на конференциях. Хотя проблема устранения неоднозначности в электронных каталогах существовала и ранее, сейчас появились инструменты для ее решения. В 2018 г. на конференции TPDFL начала работать секция «Устранение неоднозначности» (Entity Disambiguation). Из представленных на ней докладов необходимо обратить внимание на исследование Марселя Аккермана (Marcel R. Ackermann) [8], в рамках которого был проведен эксперимент по использованию методов машинного обучения для распознавания разных профилей авторов, имеющих одинаковые имена (омонимы), в DBLP (DataBase and Logic Programming bibliography site) — одной из крупнейших библиографических баз данных по компьютерным наукам. Для этого были сформированы векторные представления профилей 2 млн авторов и проведено обучение на ручной выборке из 24 тыс. профилей.

Проблемы распознавания личности встают не только при анализе текста, но и при работе с большим количеством мультимедийных данных. В электронных библиотеках наряду с текстовыми материалами уже скопилось огромное количество изображений и видеоматериалов об известных людях. Поэтому неудивительно, что исследователи из Университета Ганновера, известного достижениями в области работы с видеоматериалами, провели исследование возможности автоматического определения известных личностей и связей между ними на основе сохраненного в Internet Archive массива новостей, опубликованных на различных немецких сайтах. Доклад о результатах исследования [9] был представлен на секции «Извлечение информации» (Information Extraction).

Интересная ситуация сложилась с секциями, в названиях которых использовался термин «семантика». Последний раз это понятие встречалось в 2016 году. Однако это не означает, что данная тематика потеряла свою

актуальность, напротив — за эти годы она только приобрела большую популярность: во многих исследованиях использовались семантические методы — направление оказалось поделено между многими вышеописанными секциями. При наличии специализированных конференций и очень разноплановых докладов, вероятно, сложно было выделить именно семантическую секцию, и, вероятно, именно поэтому периодически возникали довольно непривычные для TPD L направления. Так, в 2017 г. были организованы секции «Анализ тональности текста» (Sentiment Analysis) и «Корпусы текстов» (Corpora). В целом внимание разработчиков электронных библиотек к чисто техническим аспектам (инфраструктуре, ПО и т. п.) заметно сократилось.

Похожая ситуация сложилась и с термином Linked Open Data (LOD). Связанные данные используются во многих исследованиях, но доклады, посвященные развитию системы LOD, встречаются сейчас только на профильных конференциях. Библиотекари обсуждают применение LOD в рамках своей конкретной задачи. Несмотря на большое количество докладов про LOD, среди них очень мало работ посвящено использованию данных. В основном докладчики рассказывают, как удачно им удалось что-то с чем-то связать и как опубликовать результаты своей деятельности для всеобщего блага. Но после такой публикации никто почему-то не исследует статистику использования LOD. Конечно, ее непросто собрать, но, судя по докладам, представленным на разных конференциях, количество использующих существенно меньше, чем публикующих. В связи с этим отметим ряд докладов на SWIB 2018, в которых описывался реальный опыт использования LOD. Прежде всего это голландский проект AdamLink (<http://data.adamlink.nl>) [10], в котором удалось объединить онтологии и провести трансформации данных, чтобы связать массивы информации о местах, персонах, предметах, организациях в данных архивов, музеев и библиотек в Амстердаме (всего проект охватил 33 участника). Доклад про AdamLink был также представлен и на конференции ELAG 2018. Отметим еще несколько докладов об использовании LOD: Матиаса Вандермессена (Matthias Vandermaesen) из Музея искусств Антверпена об организации «датахаба» для сбора информации по искусству из различных источников [11] и Джейн Стивенсон (Jane Stevenson) об организации хаба архивных данных в Великобритании [12].

Важно отметить, что эта тема настолько актуальна, что на SWIB 2018 был проведен

специальный семинар «От LOD к LOUD: делая данные используемыми» (From LOD to LOUD: making data usable). Дополнительное U расшифровывается как «используемые» (usable). В качестве основных шагов, необходимых для превращения обычных LOD в «используемые», на семинаре предлагались следующие:

- конвертировать RDF-данные в более понятный JSON-LD-формат;
- индексировать данные с помощью Elasticsearch;
- создавать веб-приложения для доступа к данным;
- использовать для работы с данными уже имеющиеся инструменты (Kibana, OpenRefine) [13].

Трудно сказать, приведут ли эти шаги к эффективным результатам, но попробовать стоит.

Тема LOD анализируется и российскими специалистами, ее отражение можно найти в различных публикациях [14; 15].

В последние годы в научных репозиториях и библиотеках появилась новая задача сохранения данных, полученных в результате исследований. Это могут быть приложения к статьям, содержащие экспериментальные данные, результаты опросов для статей по социологии и т. п. Для решения этой задачи нужно использовать новые для библиотечной практики метаданные. Возможно, именно поэтому наполнение секции «Управление данными» (Data Management) на конференции TPD L меняется с годами, и теперь в большей степени уделяется внимание хранению не полных текстов, а исследовательских данных и методам их описания. Так, например, в докладе Клауса-Питера Кляса (Claus-Peter Klas) и Оливера Хопта (Oliver Hopt) рассматривались вопросы применения метаданных для описания результатов эмпирических исследований на основе стандарта Data Documentation Initiative (DDI) [16]. Данный стандарт существует с 2000 г., он не очень широко используется, скорее всего потому, что до сих пор не было создано программной инфраструктуры для его применения. Можно надеяться, что разработанная в Институте социальных наук им. Лейбница (Leibniz Institute for the Social Sciences) архитектура DDI-FlatDB позволит расширить область применения DDI.

Анализируя данные, хранящиеся в библиотеках, следует отметить, что в последнее время большое внимание уделяется вовлечению их в инфраструктуру, обеспечивающую проведение научных исследований, и, в частности,

в виртуальную исследовательскую среду. На конференции SWIB в 2018 г. отдельная секция была посвящена развитию инфраструктуры для научной работы. На заседании этой секции обсуждались вопросы применения LOD для научных порталов, использования технологий Wikidata для сохранения информации о ПО, использовавшемся в проведенных исследованиях, и многие другие вопросы. Особо обратим внимание на сообщение Мартина Шульца (Martin Scholz) из Библиотеки Университета Эрлангена — Нюрнберга о портале WissKI (<http://wiss-ki.eu/>), посредством которого организована виртуальная инфраструктура для коммуникации немецких ученых [17]. На этом портале собираются данные из многих немецких коллекций (в основном гуманитарных) в виде RDF-триплетов и к ним предоставляется доступ по разным каналам (OAI-PMH, SRU в MARCxml, REST-API и т. п.). На основе собранных данных происходит визуализация установленных связей в виде таблиц и графов, т. е. получается национальная электронная библиотека для ученых. Подобную работу проводят и в Италии. В докладе Мишеля Касалини (Michele Casalini) было рассказано о проекте SHARE-VDE, в рамках которого создается инфраструктура для использования библиотечных данных, которые обогащаются нормативной информацией, связями с другими ресурсами, конвертируются в модель BIBFRAME и публикуются в виде LOD [18].

Секция «Ресурсы» (Resources) тоже при измененном названии получила новое содержание. Большинство докладов было посвящено не описанию ресурсов, как было ранее, а оценке их использования. Общее увлечение наукометрией, особенно в последние годы, не обошло стороной и программистское сообщество в библиотеках. Свидетельством является исследование ученых, посвященное установлению корреляции между экспертной оценкой научных работ и их оценкой по библиометрическим показателям [19]. Вопрос действительно актуальный для всего научного сообщества. В последние годы во многих средствах массовой информации появились критика всеобщего увлечения библиометрическими показателями. В представленном докладе делается вывод, что при оценке отдельных статей может наблюдаться низкая корреляция между экспертной оценкой научных работ и их оценкой по библиометрическим показателям, но при оценке публикационной деятельности ученых на институциональном уровне наблюдается высокая корреляция, и она тем выше, чем выше уровень цитирования.

Большое количество доступных, хорошо структурированных метаданных научных статей и докладов на конференциях способствовало исследованиям различных аспектов научных коммуникаций. Вот уже два года на конференции TPDL организуется секция «Научные коммуникации» (Scholarly Communication). В 2018 г. на этой секции немецкими учеными было представлено исследование метаданных научных мероприятий (конференций, симпозиумов, семинаров и т. п.), событий в области информатики, физики, техники и математики [20]. В исследовании рассматривались различные методы оценки успешности этих мероприятий на основе открытых метаданных о конференциях. Были предложены методы как похожие на библиометрическую оценку журналов, так и специфические (непрерывность, географическое и временное распределение и т. п.). Кроме организации коммуникаций ученых, перед библиотеками стоит другая важная задача — организовать коммуникацию библиотеки с сообществом волонтеров, готовых помочь в создании электронных ресурсов. Не случайно на конференции SWIB 2018 одним из ключевых докладчиков была Миа Райдж (Mia Ridge) — куратор цифровых коллекций Британской библиотеки, которая максимально подробно рассказала про основные принципы организации краудсорсинга и тех задачах (добавление метаданных, исправление автоматически распознанного текста и др.), которые могут быть решены этим методом [21].

Несмотря на то, что в данном обзоре рассматриваются выступления на четырех основных конференциях, «программистские» темы звучат и на других мероприятиях. Например, на Code4Lib в 2018 г., как и в прошлые годы, участники обсуждали вполне конкретные технические вопросы, актуальные для библиотек. Перед конференцией были проведены семинары на разные темы от интеграции открытого ПО для дискавери-систем с Summon and Alma до внедрения OpenRefine. На самой конференции тоже можно было встретить весьма разноплановые сообщения. Рассматривались вопросы применения JIRA для администрирования рабочего процесса управления архивом, автоматизации составления примечаний в процессе обслуживания в ПО Voyager и много других важных для практических работников вопросов. Единственный доклад, посвященный LOD [22], также отличался своей практичностью и обилием примеров запросов на Sparql. Среди прочих докладов можно выделить сообщение, посвященное стратегии восстановления дан-

ных с 5-дюймовых дискет. Многие современные пользователи уже не помнят про такой носитель информации, а ведь в фондах хранятся книги с приложениями на этих дискетах, и мало кто задумывается о том, что большинство из них, возможно, и не прочитается сейчас, даже если найдется соответствующий дискет. Одним из самых многообещающих докладов была очередная презентация успехов в разработке FOLIO (<https://www.folio.org/>) [23]. Про эту инициативу представители компании EBSCO, поддерживающей данную разработку, рассказывают на разных профессиональных мероприятиях. Проект открытого ПО, создаваемый совместно библиотекарями, разработчиками и вендорами, уже выпустил первый релиз 25 января 2019 г. и по заявленным целям и предусмотренным функциям, действительно, может быть удачным решением. Но пока по представленной демонстрационной версии трудно сказать что-то определенное. Закончим описание Code4Lib самым важным, на наш взгляд, призывом, прозвучавшим в докладе Джоша Вейцмана (Josh Weisman): «Соблюдать стандарты и не изобретать колесо заново» [24]. В нем он вспомнил многие библиотечные стандарты и вновь вернулся к теме интероперабельности, давно, как мы ранее заметили, не обсуждавшейся на конференциях по электронным библиотекам. И он прав: при всем обилии разработок и разработчиков только соблюдение стандартов поможет построить общее цифровое библиотечное пространство.

### Примечание

- <sup>1</sup> Рассматривается четыре года, потому что в 2014 г. конференция TPDL проводилась совместно с Joint Conference on Digital Libraries. Программный комитет, методы отбора докладов и соответственно их тематика были несколько отличны от обычных для TPDL.

### Список источников

1. *Hanson C.* Opinion [Электронный ресурс] // Cody Hanson : персон. сайт. URL: <https://www.codyh.com/writing/software.html> (дата обращения: 07.02.2019).
2. *Dörr M., Fundulaki I.* A semantic network approach to semi-structured documents repositories // Research and Advanced Technology for Digital Libraries. Berlin, Heidelberg : Springer, 1997. С. 305—324.
3. *Gottschalk S., Bernacchi V., Rogers R., Demidova E.* Towards Better Understanding Researcher Strategies in Cross-Lingual Event Analytics // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 139—151. DOI: 10.1007/978-3-030-00066-0\_12.
4. *Sadhu S., Bhowmick P.K.* Automatic Segmentation and Semantic Annotation of Verbose Queries in Digital Library // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 270—276. DOI: 10.1007/978-3-030-00066-0\_23.
5. *Vakkari P., Völske M., Potthast M., Hagen M., Stein B.* Predicting Retrieval Success Based on Information Use for Writing Tasks // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds) Springer, Cham, 2018. P. 161—173. DOI: 10.1007/978-3-030-00066-0\_14.
6. *Al-Natsheh H.T., Martinet L., Muhlenbach F., Rico F., Zighed D.A.* Metadata Enrichment of Multi-disciplinary Digital Library: A Semantic-Based Approach // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 32—43. DOI: 10.1007/978-3-030-00066-0\_3.
7. *Suominen O.* Annif: leveraging bibliographic metadata for automated subject indexing and classification [Электронный ресурс] // SWIB18: Semantic Web in Libraries : материалы конференции, 26—28 ноября 2018, Бонн, Германия. URL: [http://swib.org/swib18/slides/2\\_suominen\\_annif.pdf](http://swib.org/swib18/slides/2_suominen_annif.pdf) (дата обращения: 07.02.2019).
8. *Ackermann M.R., Reitz F.* Homonym Detection in Curated Bibliographies: Learning from dblp's Experience // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 59—65. DOI: 10.1007/978-3-030-00066-0\_5.
9. *Müller-Budack E., Pustu-Iren K., Diering S., Ewerth R.* Finding Person Relations in Image Data of News Collections in the Internet Archive // Digital Libraries for Open Knowledge. TPDL 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 229—240. DOI: 10.1007/978-3-030-00066-0\_20.

10. *Koster L., Zandhuis I.* Transformations for aggregating Linked Open Data [Электронный ресурс] // SWIB 18 : Semantic Web in Libraries : материалы конференции, 26—28 ноября 2018, Бонн, Германия. URL: [http://swib.org/swib18/slides/2\\_koster\\_transformations.pdf](http://swib.org/swib18/slides/2_koster_transformations.pdf) (дата обращения: 07.02.2019).
11. *Vandermaesen M.* The Datahub Project: Deblending Museum Data [Электронный ресурс] // ELAG 2018, 4—6 июня 2018 г., Прага, Чехия. URL: <http://repozitar.techlib.cz/record/1263> (дата обращения: 07.02.2019).
12. *Stevenson J.* In Out, In Out, and Shake It All About: a Moving Story of Data [Электронный ресурс] // ELAG 2018, 4—6 июня 2018 г., Прага, Чехия. URL: <http://repozitar.techlib.cz/record/1247> (дата обращения: 07.02.2019).
13. *Stegg F., Pohl A., Christoph P.* From LOD to LOUD: making data usable [Электронный ресурс] // SWIB 18 : Semantic Web in Libraries : материалы конференции, 26—28 ноября 2018, Бонн, Германия. URL: <https://hbz.github.io/swib18-workshop/> (дата обращения: 07.02.2019).
14. *Шварцман М.Е., Найдин О.П.* Linked Open Data как средство обогащения поисковых запросов [Электронный ресурс] // Университетская книга. 2015. № 12. С. 66—71. URL: <http://www.unkniga.ru/innovation/tehnology/5467-linked-open-data-kak-sredstvo-obogascheniya-poiskovyh-zaprosov.html> (дата обращения: 07.02.2019).
15. *Лаврёнова О.А., Павлов В.В.* Библиотечно-библиографическая классификация как традиционная система организации знаний в среде открытых связанных данных [Электронный ресурс] // Научно-технические библиотеки. 2017. № 4. С. 44—60. URL: <https://ntb.gpntb.ru/jour/article/view/182> (дата обращения: 07.02.2019). DOI: 10.33186/1027-3689-2017-4-44-60.
16. *Klas C.P., Hopt O.* An Operationalized DDI Infrastructure to Document, Publish, Preserve and Search Social Science Research Data // Digital Libraries for Open Knowledge. TPD 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 94—99. DOI: 10.1007/978-3-030-00066-0\_8.
17. *Scholz M.* Integrating library metadata in a semantic web research environment for university collections [Видеозапись] // YouTube. URL: <https://youtu.be/oO3uj5DTw1s> (дата обращения: 07.02.2019).
18. *Casalini M.* Share virtual discovery environment in Linked Data (SHARE-VDE) [Электронный ресурс] // SWIB 18 : Semantic Web in Libraries : материалы конференции, 26—28 ноября 2018, Бонн, Германия. URL: [http://swib.org/swib18/slides/1\\_casalini\\_share-vde.pdf](http://swib.org/swib18/slides/1_casalini_share-vde.pdf) (дата обращения: 07.02.2019).
19. *Pride D., Knoth P.* Peer Review and Citation Data in Predicting University Rankings, a Large-Scale Analysis // Digital Libraries for Open Knowledge. TPD 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 195—207. DOI: 10.1007/978-3-030-00066-0\_17.
20. *Fathalla S., Vahdati S., Auer S., Lange C.* Metadata Analysis of Scholarly Events of Computer Science, Physics, Engineering, and Mathematics // Digital Libraries for Open Knowledge. TPD 2018. Lecture Notes in Computer Science, vol. 11057 / E. Méndez, F. Crestani, C. Ribeiro, G. David, J. Lopes (eds). Springer, Cham, 2018. P. 116—122. DOI: 10.1007/978-3-030-00066-0\_10.
21. *Ridge M.* Libraries and their communities: from town halls to mobile phones [Электронный ресурс] // SWIB 18 : Semantic Web in Libraries : материалы конференции, 26—28 ноября 2018, Бонн, Германия. URL: [http://swib.org/swib18/slides/2\\_ridge\\_libraries-and-their-communities.pdf](http://swib.org/swib18/slides/2_ridge_libraries-and-their-communities.pdf) (дата обращения: 07.02.2019).
22. *Meyer S.* Enhancing Library Discovery with Linked Open Data [Электронный ресурс]. URL: <https://tinyurl.com/c4l18-smeyer> (дата обращения: 07.02.2019).
23. *Nagy A.* FOLIO Innovation Challenge [Электронный ресурс] // EBSCO : сайт. URL: <https://www.ebsco.com/folio-innovation-challenge> (дата обращения: 07.02.2019).
24. *Weisman J.* For the Love of Standards [Электронный ресурс]. URL: <https://docs.google.com/presentation/d/1vrwtQYTxxMxRtYh4RjtulKaLM-LVUOwYj4riUV8zEIQ/edit> (дата обращения: 07.02.2019).

# Digital Research Methods: the New Vector in Library Informatization

Michael E. Shvartsman <sup>a \*</sup>, Olga N. Kryzhanovskaya<sup>\*\*</sup>

Russian State Library, 3/5 Vozdvizhenka Str., Moscow, 119019, Russia

<sup>a</sup> ORCID 0000-0003-2524-6819; SPIN 4277-9746

E-mail: \* ShvartsmanME@rsl.ru, \*\* KryzhanovskayaON@rsl.ru

**Abstract.** A large number of conferences on automation of library processes are held abroad. These events discuss the most topical and important issues facing libraries in the implementation of modern information technologies. The article provides an overview of four major international conferences, which brought together programmers, linguists and librarians, who shared their experiences and planned for future. The authors present the analysis of changes in the subject of conferences for the recent years, conclude on the significant increase in the number of studies on semantic analysis of data stored in libraries and mark the most interesting reports. In recent years, the number of reports on Linked Open Data (LOD) has been increasing. However, most of these reports focus on the publication of data in LOD, with very little research on the extent of use of already published data. The article provides examples of such studies. The authors highlight the constant interest to the topic of Digital Humanities and to the fact that the research within this area is often focused on the problems of modern society. The article notes the new direction in the activity of electronic libraries — storage of the research data, and gives the examples of implementation of these tasks.

**Key words:** digital humanities, conferences, linked open data, programming, library automation, automation, sharing of experience, semantics, software, information support, TPDF, SWIB, ELAG, Code4Lib, professional events.

**Citation:** Shvartsman M.E., Kryzhanovskaya O.N. Digital Research Methods: the New Vector in Library Informatization, *Bibliotekovedenie* [Library and Information Science (Russia)], 2019, vol. 68, no. 1, pp. 103—111. DOI: 10.25281/0869-608X-2019-68-1-103-111.

## References

1. Hanson C. Opinion, *Cody Hanson: website*. Available at: <https://www.codyh.com/writing/software.html> (accessed: 07.02.2019).
2. Dörr M., Fundulaki I. A semantic network approach to semi-structured documents repositories. *Research and Advanced Technology for Digital Libraries*. Berlin, Heidelberg : Springer, 1997, pp. 305—324.
3. Gottschalk S., Bernacchi V., Rogers R., Demidova E. Towards Better Understanding Researcher Strategies in Cross-Lingual Event Analytics, *Digital Libraries for Open Knowledge. TPDF 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 139—151. DOI: 10.1007/978-3-030-00066-0\_12.
4. Sadhu S., Bhowmick P.K. Automatic Segmentation and Semantic Annotation of Verbose Queries in Digital Library, *Digital Libraries for Open Knowledge. TPDF 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 270—276. DOI: 10.1007/978-3-030-00066-0\_23.
5. Vakkari P., Völske M., Potthast M., Hagen M., Stein B. Predicting Retrieval Success Based on Information Use for Writing Tasks, *Digital Libraries for Open Knowledge. TPDF 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 161—173. DOI: 10.1007/978-3-030-00066-0\_14.
6. Al-Natsheh H.T., Martinet L., Muhlenbach F., Rico F., Zighed D.A. Metadata Enrichment of Multi-disciplinary Digital Library: A Semantic-Based Approach. *Digital Libraries for Open Knowledge. TPDF 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 32—43. DOI: 10.1007/978-3-030-00066-0\_3.
7. Suominen O. Annif: leveraging bibliographic metadata for automated subject indexing and

- classification, *SWIB 18 : Semantic Web in Libraries*. Proc. of the conf., 26—28 November 2018, Bonn, Germany. Available at: [http://swib.org/swib18/slides/2\\_suominen\\_annif.pdf](http://swib.org/swib18/slides/2_suominen_annif.pdf) (accessed 07.02.2019).
8. Ackermann M.R., Reitz F. Homonym Detection in Curated Bibliographies: Learning from dblp's Experience, *Digital Libraries for Open Knowledge. TPD L 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 59—65. DOI: 10.1007/978-3-030-00066-0\_5.
  9. Müller-Budack E., Pustu-Iren K., Diering S., Ewerth R. Finding Person Relations in Image Data of News Collections in the Internet Archive, *Digital Libraries for Open Knowledge. TPD L 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 229—240. DOI: 10.1007/978-3-030-00066-0\_20.
  10. Koster L., Zandhuis I. Transformations for aggregating Linked Open Data, *SWIB 18 : Semantic Web in Libraries*. Proc. of the conf., 26—28 November 2018, Bonn, Germany. Available at: [http://swib.org/swib18/slides/2\\_koster\\_transformations.pdf](http://swib.org/swib18/slides/2_koster_transformations.pdf) (accessed: 07.02.2019).
  11. Vandermaesen M. The Datahub Project: Deblending Museum Data, *ELAG 2018*, Proc. of the conf., 4—6 July 2018, Prague, CH. Available at: <http://repositor.techlib.cz/record/1263> (accessed 07.02.2019).
  12. Stevenson J. In Out, In Out, and Shake It All About: a Moving Story of Data, *ELAG 2018*, Proc. of the conf., 4—6 July 2018, Prague, CH. Available at: <http://repositor.techlib.cz/record/1247> (accessed 07.02.2019).
  13. Steeg F., Pohl A., Christoph P. From LOD to LOUD: making data usable, *SWIB 18: Semantic Web in Libraries*, Proc. of the conf., 26—28 November 2018, Bonn, Germany. Available at: <https://hbz.github.io/swib18-workshop/> (accessed 07.02.2019).
  14. Schvartsman M., Naidin O. Linked Open Data as means of enriching retrieval requests, *Universitetskaya kniga* [University Book], 2015, no. 12, pp. 66—71. Available at: <http://www.unkniga.ru/innovation/tehnology/5467-linked-open-data-kak-sredstvo-obogascheniya-poiskovyh-zaprosov.html> (accessed 07.02.2019) (in Russ.).
  15. Lavrenova O., Pavlov V. Library Bibliographic Classification as a traditional knowledge organization system in the linked open data environment, *Nauchnye i tekhnicheskie biblioteki* [Scientific and Technical Libraries], 2017, no. 4, pp. 44—60. Available at: <https://ntb.gpntb.ru/jour/article/view/182> (accessed 07.02.2019) (in Russ.). DOI: 10.33186/1027-3689-2017-4-44-60.
  16. Klas CP., Hopt O. An Operationalized DDI Infrastructure to Document, Publish, Preserve and Search Social Science Research Data, *Digital Libraries for Open Knowledge. TPD L 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 94—99. DOI: 10.1007/978-3-030-00066-0\_8.
  17. Scholz M. Integrating library metadata in a semantic web research environment for university collections, *YouTube*. Available at: <https://youtu.be/oO3yj5DTw1s> (accessed 07.02.2019).
  18. Casalini M. Share virtual discovery environment in Linked Data (SHARE-VDE), *SWIB 18 : Semantic Web in Libraries*, Proc. of the conf., 26—28 November 2018, Bonn, Germany. Available at: [http://swib.org/swib18/slides/1\\_casalini\\_share-vde.pdf](http://swib.org/swib18/slides/1_casalini_share-vde.pdf) (accessed 07.02.2019).
  19. Pride D., Knoth P. Peer Review and Citation Data in Predicting University Rankings, a Large-Scale Analysis, *Digital Libraries for Open Knowledge. TPD L 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 195—207. DOI: 10.1007/978-3-030-00066-0\_17.
  20. Fathalla S., Vahdati S., Auer S., Lange C. Metadata Analysis of Scholarly Events of Computer Science, Physics, Engineering, and Mathematics, *Digital Libraries for Open Knowledge. TPD L 2018. Lecture Notes in Computer Science*, vol. 11057. Springer, Cham, 2018, pp. 116—122. DOI: 10.1007/978-3-030-00066-0\_10.
  21. Ridge M. Libraries and their communities: from town halls to mobile phones, *SWIB 18 : Semantic Web in Libraries*, Proc. of the conf., 26—28 November 2018, Bonn, Germany. Available at: [http://swib.org/swib18/slides/2\\_ridge-libraries-and-their-communities.pdf](http://swib.org/swib18/slides/2_ridge-libraries-and-their-communities.pdf) (accessed 07.02.2019).
  22. Meyer S. *Enhancing Library Discovery with Linked Open Data*. Available at: <https://tinyurl.com/c4118-smeyer> (accessed 07.02.2019).
  23. Nagy A. FOLIO Innovation Challenge, *EBSCO web-site*. Available at: <https://www.ebsco.com/folio-innovation-challenge> (accessed 07.02.2019).
  24. Weisman J. *For the Love of Standards*. Available at: <https://docs.google.com/presentation/d/1vrtwQYTxxkMxRtYh4RjtulKaLM-LVUOwYj4ri-UV8zEIQ/edit> (accessed 07.02.2019).