

ВИЗУАЛИЗАЦИЯ В НАУЧНОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКЕ**Филозова Ирина Анатольевна¹, Заикина Татьяна Николаевна²**

¹Старший преподаватель, научный сотрудник;
ГОУ ВПО «Международный Университет природы, общества и человека «Дубна»,
Институт системного анализа и управления;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Объединенный институт ядерных исследований,
Лаборатория информационных технологий;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: fia@jinr.ru.

²Инженер-программист;
Объединенный институт ядерных исследований,
Лаборатория информационных технологий;
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, 6;
e-mail: ztanya@jinr.ru.

Современная ситуация развития технологий электронных библиотек (ЭБ) позволяет говорить о накоплении огромного объема контента, для исследования необходим переход на новое качество пользовательского интерфейса. Использование методов и средств визуализации в электронных библиотеках открывает пользователям новые возможности “увидеть” их содержимое с разных точек зрения и существенно повышает информативность и эффективность ЭБ за счет привлечения эмоциональных образов и ассоциаций положительного пользовательского опыта. Обсуждается развитие сервисов визуализации в научных электронных библиотеках.

Ключевые слова: научная публикация, документ, электронные библиотеки, визуализация, информационный поиск, навигация, семантические связи.

VISUALIZATION FOR THE SCIENTIFIC DIGITAL LIBRARY**Filozova Irina¹, Zaikina Tatiana²**

¹Senior teacher, scientist;
Dubna International University of Nature, Society, and Man,
Institute of system analysis and management;
141980, Dubna, Moscow reg., Universitetskaya str., 19;
Join Institute for nuclear Research,
Laboratory of Information Technologies;
141980, Dubna, Moscow reg., Joliot-Curie str., 6;
e-mail: fia@jinr.ru.

²Engineer-programmer;
Join Institute for nuclear Research,
Laboratory of Information Technologies;
141980, Dubna, Moscow reg., Joliot-Curie str., 6;
e-mail: ztanya@jinr.ru.

Current situation of development of digital libraries (DLs) technology suggests the accumulation of the large content amount. There is necessary to shift at a new quality of the user interface for the effective DL research. Using the methods and tools of the visualization in digital libraries offers new capabilities to «see» their content with different points of view that increases the informativeness and efficiency of digital libraries by bringing emotional images and associations of a positive user experience. We discuss the development of visualization services for scientific digital libraries.

Keywords: scientific publication, document, digital libraries, visualization, information search, navigation, semantic relations.

Введение

Электронные библиотеки представляют собой сложные распределенные информационные системы, предназначенные для хранения, обработки, многоаспектного поиска, распространения и анализа информационных ресурсов. Т.е. ЭБ обладают качествами, присущими всем информационным системам (хранилища данных, поиск и навигация, развитый набор пользовательских сервисов), а также имеют свои особенности. Специфика электронной библиотеки определяется ее наполнением (контентом), формируемым исходя из информационных потребностей ее пользователей.

1. Информационные объекты научной электронной библиотеки и связи между ними

В научном мире главным критерием профессиональной состоятельности ученого и исследователя служат его публикации. Соответственно, информационными ресурсами научной электронной библиотеки (НЭБ) являются научные и научно-технические тексты таких литературных жанров, как статья, препринт, отчет, монография, диссертация, автореферат диссертации, труды конференций и пр. Продукцией научно-исследовательской деятельности, научного и инженерного творчества являются издания с установленными семантическими связями на другие научные тексты через механизм цитирования. В процессе научной деятельности формируются, например, связи научного вывода (отражают заимствование, развитие, подтверждение, уточнение, коррекцию объекта, на который направлена связь), связи использования (отражают характер использования объекта, на который направлена связь, для получения нового результата), связи мнений и оценок (характеризуют профессиональные мнения или оценки о научных текстах) [5]. Описание этих связей и их свойств открывает новые возможности для исследования корпуса документов ЭБ. Публикация, а также коллекции публикаций содержат и другие семантические связи, например, связи, отражающие логику изложения авторской мысли в рамках данной публикации, темы, предметной области. Эти связи также могут быть описаны на основе качественного анализа научных текстов и исследованы [1]. Можно выделить также связи, возникающие между персоналиями — авторами и соавторами публикаций, их принадлежность к организации (месту работы автора), участие в коллаборациях, экспериментах, проектах и пр. Одним словом, в рамках данной ЭБ формируются персональные профили, которые служат основой научной социальной сети. Структурированное описание подобного рода связей между объектами электронных библиотек и их поддержка открывают новые возможности для исследования и повышают качество ЭБ. Такое семантическое структурирование фонда ЭБ ведет к появлению многоуровневой сетевой структуры, которая может быть положена в основу навигационного механизма ЭБ. Эта структура также дает возможность для новых наукометрических измерений и исследования корпуса знаний электронных библиотек.

2. Визуализация и электронные библиотеки

Визуализация, являясь особой формой пользовательского интерфейса, обеспечивает обнаружение различного рода связей между документами электронной библиотеки, облегчая их исследование. Традиционно визуализация играла второстепенную роль при анализе данных. Современные же исследования признают визуализацию эффективным средством анализа данных различной природы. Цели анализа и используемые графические представления этих данных могут быть различны и зависят от информационных потребностей и задач пользователей. Визуализация в ЭБ имеет целью сделать использование ресурсов электронных библиотек более эффективным [2,3]. Визуализация направлена на отображение результатов взаимодействия пользователя с доступными сервисами электронной библиотеки. Визуализация позволяет пользователю находить альтернативные интерпретации данных, получать интерпретацию отобранных ресурсов с различных точек зрения, не прибегая к другим средствам анализа. Этот процесс предполагает максимально дружественный и минимально сложный пользовательский интерфейс, в котором пользователь управляет представлением данных щелчком мыши на элементах визуализации, перемещая и перетаскивая их и/или выбирая пункты ме-

ню. Такое взаимодействие представляет собой этап итерационного анализа данных и является частью навигационного механизма, помогающего пользователю исследовать объекты ЭБ.

2.1. Визуализация результатов поиска

В настоящий момент электронные библиотеки в основном используют стандартный способ вывода результатов поиска, однако, такие поисковые интерфейсы испытывают недостаток в возможности поддерживать исследование информации. Имея стандартный поисковый интерфейс, ученым и исследователям трудно получить цельное представление о содержании и структуре электронной библиотеки. Актуальность проблемы подтверждается тем, что поисковые сервисы продолжают динамично развиваться, предлагая разные подходы усовершенствования стандартного способа вывода результатов. Из предлагаемых решений можно выделить три основных подхода. Первый направлен на улучшение интерфейса для выдачи поискового результата. Второй подход делает упор на усовершенствовании и развитии механизма поиска, не ограничиваясь вычислением релевантности документа поисковому запросу. Третий подход основан на структуризации процесса поиска, т.е. категоризации содержимого – музыка, видео, блоги, изображения, справочники и т. д. Известен также вариант визуализации в виде вывода в результатах поиска скриншотов найденных web-страниц. Среди проектов, внедряющих визуализацию результатов поиска, относительно широкую известность имеют такие информационные системы, как распределенная сеть медиаресурсов Snap (<http://www.snap.com/>), поисковая система search engine KartOO (<http://www.kartoo.com/>), интерактивная поисковая система Quintura see&find (<http://www.quintura.ru/>). Известен также метод WebSOM, разработка группы ученых Хельсинкского Технологического Университета (<http://www.cis.hut.fi/research/som-research/som.shtml>), предназначенный для публикации документоориентированных баз данных в виде карты плотностей на плоскости [4]. Компания Tianamo предлагает поисковый интерфейс (<http://www.tianamo.com/>), использующий возможности анимационной визуализации — интерактивный 3D-интерфейс для отображения данных результатов web-поиска. Поисковые результаты в виде точек накладываются на плоскую поверхность на расстоянии, соответствующему тому, насколько близки между собой результаты поиска.

2.2. Визуализация навигации

Графический навигационный механизм обеспечивает удобный обзор ресурсов электронной библиотеки и простоту просмотра и тесно связан с визуализацией результатов поиска, если рассматривать поиск как часть навигационного механизма (поиск в режиме просмотра). Навигационный граф должен быть настраиваемым, отображать существующие связи объекта как исходящие, так и входящие, иметь систему фильтров по заданным свойствам, например, включение/отключение связей заданной категории.

2.3. Визуализация данных

При исследовании фондов научной информации часто требуется интерпретировать информационные ресурсы с различных точек зрения. Например, получить распределение публикаций по предметным рубрикам/тематикам. На рис.1. представлена визуализация, разработанная для репозитория Открытого доступа JINR Document Server, выполненная с помощью метода древовидной карты для отображения публикаций по тематике [6].

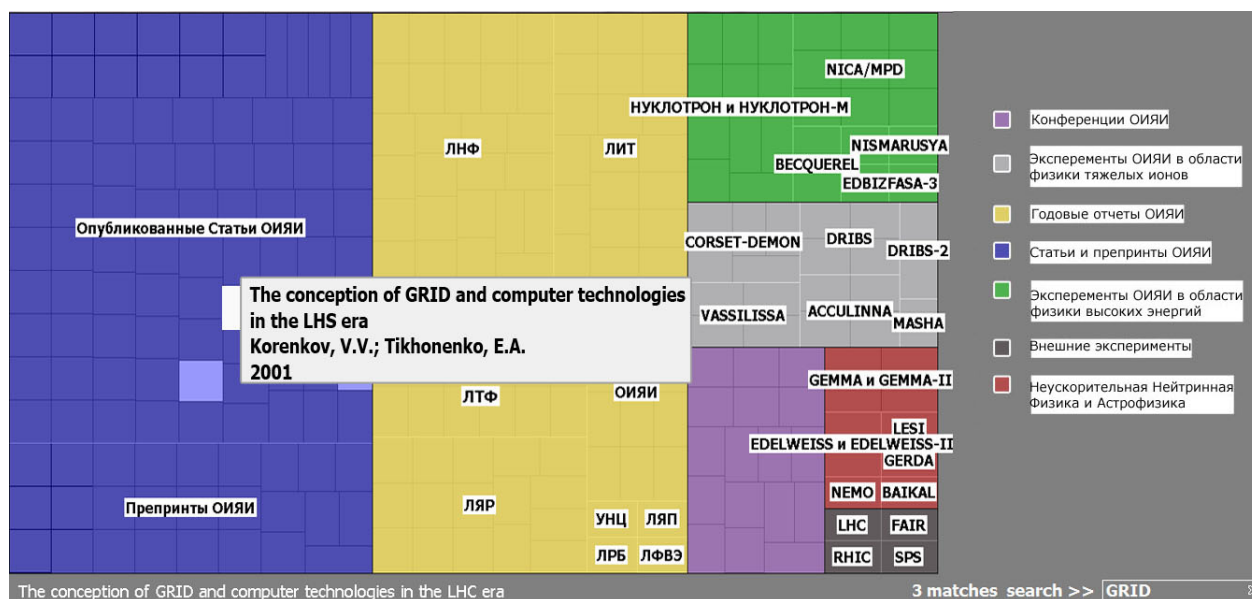


Рис. 1. Визуальное представление информационных ресурсов JDS с помощью метода древовидной карты

Каждая тематика отображается соответствующим цветом. Чем больше пространства окрашено определенным цветом, тем больше публикаций в данной тематике. Данная визуализация наглядно отображает распределение публикаций по тематикам.

Следующий уровень иерархии отображается в виде записи, и конечные публикации отображаются в виде интерактивных прямоугольников. Представление также содержит строку поиска, найденная публикация выделяется соответствующим цветом. Публикация является ссылкой на саму себя. При наведении курсора на публикацию появляется информация об её авторе, соавторе, дате опубликования. При нажатии на правую клавишу мыши можно масштабировать изображение.

Разновидностью визуализации использования является визуализация статистики и результатов мониторинга. Большинство электронных библиотек имеют настраиваемый модуль статистики, обеспечивающий сбор статистики по таким показателям, как количество обращений, количество и интенсивность загрузок, наиболее популярные документы (хиты), прирост записей за определенный временной период и пр.

2.4. Визуализация научной социальной сети

Во многих ЭБ предусмотрен также инструментарий для оценки и обсуждений документов читателями. Модули, обеспечивающие взаимодействие читателей (обмен сообщениями, создание групп по интересам, по принадлежности к какому-либо научному проекту или эксперименту), пользователей и модераторов (отправка уведомлений, служебные рассылки) позволяют учитывать социальные особенности сети, возникающей в рамках данной ЭБ. Пользователи имеют возможность взаимодействовать между собой посредством сервиса для обсуждения публикации, образуя сеть коммуникаций в научном сообществе в рамках информационной системы. Анализируя эти связи, можно получать дополнительную важную информацию о характере этих взаимодействий, например, по интенсивности связей между авторами, работающими в одном научном направлении (теме, проекте), можно судить о степени активности развития данного направления и пр. Визуализация социальной научной сети позволяет делать выводы о характере взаимодействия акторов в рамках данной ЭБ, не прибегая к другим методам анализа графа.

На рисунке 2 представлена визуализация на основе метода радиального графа, отражающая связи между авторами и соавторами публикаций в репозитории Открытого доступа JINR Document Server [6].

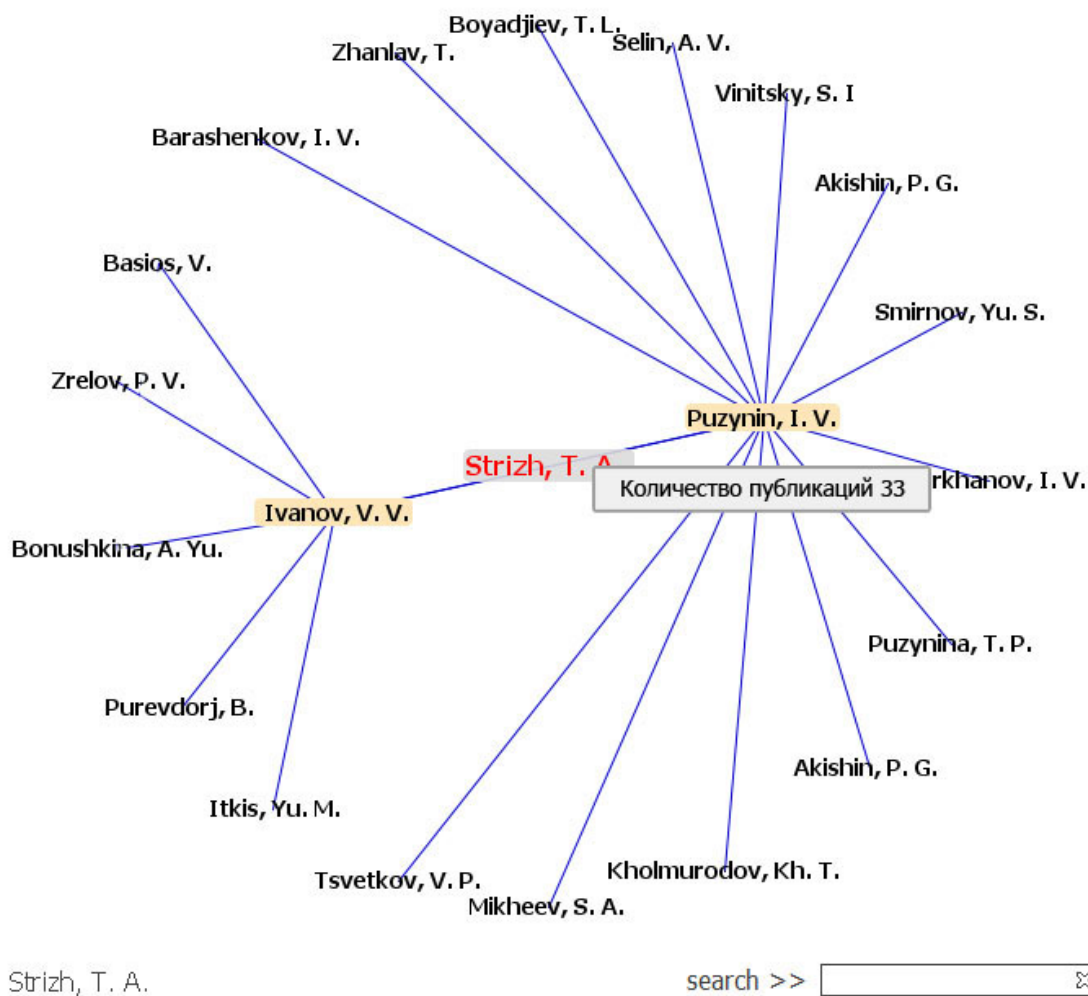


Рис. 2. Визуальное представление информационных ресурсов JDS с помощью метода радиального графа

Автор располагается в центре, на второй окружности – соавторы, на третьей – соавторы соавторов. Автор и его соавторы подсвечиваются определенным цветом. При наведении указателя мыши на автора, высвечивается информация о количестве его публикаций в репозитории. Поддерживается масштабирование и панорамирование, простой поиск.

Заключение

Визуализация в электронных библиотеках служит источником информации для пользователя, не требуя от него теоретических знаний и специальных навыков работы, предоставляя визуальные образы и представления, благодаря которым исследователи могут расширять, углублять, уточнять или получать новые знания в данной предметной области. Доступ к информационным хранилищам через визуальные интерфейсы позволит пользователям более быстро и естественно достигнуть желаемого результата в рамках данной электронной библиотеки: осуществить быстрый отбор релевантных ресурсов и исследовать их характеристики, отслеживать свой личный фактор влияния (вклад автора в развитие научных знаний) и/или фактор влияния группы, а также развитие различных областей знаний и разработок.

Список литературы

1. Добрынин В. Н., Филозова И. А. Поиск в научной электронной библиотеке на основе логико-семантической сети Вопрос-Ответ-Реакция // Труды XII Всероссийской научной конференции RCDL'2010 «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Казань: Казанский университет, 2010. – С. 301-308.
2. Judith Gelernter. Visual Classification with Information Visualization (Infoviz) for Digital Library Collections. – New Jersey: Knowledge Organization 34(3), 2007.
3. Panayiotis Zaphiris, Kulvinder Gill, Terry Hoi-Yan Ma, Stephanie Wilson, Helen Petrie. Exploring the use of Information Visualization for Digital Libraries. – London, United Kingdom: Center for Human Computer Interaction Design, 2005.
4. В. Плешко. Визуальные средства поиска информации в документоориентированных базах данных. [Электронный ресурс]. URL: <http://citforum.ru/ofis/ofis96/125.shtml>.
5. Паринов С.И., Когаловский М.Р. Технология семантического структурирования контента научных электронных библиотек // Труды XIII Всероссийской научной конференции RCDL'2011 «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – Воронеж: Воронежский университет, 2011. – [Электронный ресурс]. URL: <http://rcdl.ru/doc/2011/paper13.pdf>.
6. Filozova I.A., Korenkov V.V., Musulmanbekov G. Towards Open Access Publishing at JINR. – Proceedings of XXII International Symposium on Nuclear Electronics and Computing (NEC'2009). – Varna, Bulgaria, Sept. 7-14, 2009. – Dubna: JINR, 2010. – Pp.124-128. – (JINR ; E10,11-2010-22).