

13. *Официальный сайт проекта Primos.* [Электронный ресурс]. Режим доступа: <http://www.primos.nl/>. — Загл. с экрана.
14. *Официальный сайт проекта Ignite.* [Электронный ресурс]. Режим доступа: <http://www.confio.com/>. — Загл. с экрана.
15. *Официальный сайт проектов Zero Impact Enterprise и Performance Monitor.* [Электронный ресурс] Режим доступа: <http://www.sqlpower.com/>. — Загл. с экрана.

## ТЕКСТ И МУЛЬТИМЕДИА В ЭЛЕКТРОННОЙ СЕМАНТИЧЕСКОЙ БИБЛИОТЕКЕ

*М.В. Яковлева, А.К. Тен, В.М. Куглер*

*Свердловская областная универсальная научная библиотека им. В.Г. Белинского  
Екатеринбург*

### ВВЕДЕНИЕ

Современные научные исследования и разработки предполагают наличие развитой системы справочников. Объектный подход и движение в направлении электронных семантических библиотек позволяют улучшить индексацию и поиск материалов.

### ЧАСТЬ I. НА ПУТИ К ЭЛЕКТРОННОЙ СЕМАНТИЧЕСКОЙ БИБЛИОТЕКЕ

#### **Признаки электронной семантической библиотеки (ЭСБ)**

По мнению авторов, существенные признаки ЭСБ лежат в области модели мира, способов работы с ней и удобных интерфейсах пользователей:

- А) Это хранилище объектов с их свойствами;
- Б) Между объектами установлены связи и отношения;
- В) Объекты можно искать и просматривать;
- Г) К объектам присоединены традиционные формы: тексты, картинки, видео, аудио;
- Д) Экранные интерфейсы пользователя понятны, не содержат специальных терминов и удобны;
- Е) Универсальность: обеспечено размещение объектов любой сферы деятельности (области знаний).
- Ж) ЭСБ может содержать ограничения на объекты, связи и прочее, дополняющие модель объектов;
- З) Отдельные свойства, связи объектов могут вычисляться (выводиться) из других свойств и связей.

Представление объекта в ЭСБ определяется пунктами (А)-(З). Объекты характеризуются свойствами. Например, у пещер есть свойства «Доступность» и «Высотная отметка входа», а у людей – «Дата рождения». Возможности формировать у объекта требуемый набор свойств – одна из характеристик универсальности ЭСБ. Свойства (атрибуты) используются, чтобы изложить информацию, легко структурируемую. Обычно свойство, например, «Протяженность», используется как характеристика для многих объектов.

Между объектами могут существовать связи. Например, объект Иванов Иван Иванович может быть связан с объектом Фабрика «Одежда» свойством-ссылкой Директор, а объект Статья «История Москвы» с объектом Красная площадь свойством «О чем». В целом, объекты, представленные в ЭСБ, составляют сеть, и ознакомление с материалом может происходить переходами по именованным ссылкам от объекта к объекту. В узлах сети находятся не только документы, но большей частью объекты действительности – галактики, события, изобретения, люди и т.д.

Поиск объекта реализуется на основе его наименования, класса, к которому он относится, условий, накладываемых на значения свойств.

Экранные интерфейсы, то есть то, как пользователь видит представление объекта, позволяют передать логические связи объекта с другими объектами, значения его свойств, а также прикрепленные к нему традиционные информационные ресурсы. Пользователи с большим энтузиазмом встретят оформление объектов, похожее на представление страниц в традиционном вебе.

Универсальность определяется свойствами программного обеспечения. Например, в системах, основанных на Dublin Core [1], все сосредоточено на объектах типа документ: книгах, статьях и т.д. Чтобы индексировать содержание документов, необходима универсальная онтология [2].

## Достоинства ЭСБ

**Универсальность** является существенным достоинством. Конкретное наполнение ЭСБ характеризуется определенным массивом конкретных объектов и их информационных моделей. Но то же программное обеспечение, установленное в другом месте, позволит ввести совсем другие объекты другой тематики со своими связями и отношениями.

**Подвижность** представления объекта. Развитие информационного материала об объекте повлечет добавление новых свойств, связей, традиционных файлов, но та модель, что была изначально, сохраняется, как часть новой.

**Структурированность.** Связи между реальными объектами действительности в ЭСБ явно выделены. Их можно наблюдать, можно переходить по ним, составляя в уме целостную картину моделируемой действительности.

**Поиск** по объектам и сосредоточение **традиционных материалов** около объекта, которому они посвящены.

Считается, что семантические инструменты полезны в **организации учебного процесса**, формировании учебных программ, подборе групп для выполнения совместной проектной работы [3].

Известно, что обучение с использованием ЭСБ является эффективным за счет связей между объектами, а также между объектами и традиционными файлами.

Универсальность и инструменты ввода, поиска и просмотра создают возможность **создания справочных систем** без привлечения программистов. В то же время наличие логических связей между объектами в системе формирует у участников таких проектов развитое представление о множестве взаимосвязей объектов действительности.

Объекты действительности, их свойства и связи, классификация объектов, аспектов, связей являются материалом для индексации традиционных источников.

Добавленные к объектам традиционные источники (статьи, фотографии, видео) позволяют пользователю получить картину знаний об объекте.

### **«Семантик» - электронная библиотека фактографического обслуживания на объектной основе.**

Работы над системой начаты в 2007 году. Коллектив разработчиков: Свердловская областная универсальная научная библиотека им. В.Г. Белинского в партнерстве с Уралмультимедиацентром УГТУ –УПИ и ГОУ ЦПК Центр "Учебная книга". С 2009 г. работа продолжается коллективом библиотеки им. Белинского.

Цель создания «Семантик»: фактографическое обслуживание реальных и виртуальных читателей библиотеки, информационные потребности которых состоят не только в получении конкретной книги (статьи), но и в решении конкретных проблем, получении ответов на свои вопросы.

В системе хранятся объекты действительности с их названиями и синонимами. Объекты обладают свойствами, значения которых могут быть текстовыми, целочисленными, датой, ссылкой на другой объект. Объекты распределены на типажи, и это позволяет за типом закрепить рекомендуемый набор свойств.

Свойства, значений которых может быть у одного объекта несколько, выделены отдельно. Их рабочее название – группы, групповые свойства. В интерфейсе объекта они выглядят как меню. Как и в свойствах, набор групп объекта формируется на основании его типа. Например, у типажа «город» могут быть группы «Почетные граждане», «Общественные объединения»; у «школы» – «Преподаватели», «Учащиеся» и т.п. Выбор конкретной группы у объекта вызывает страницу, где перечислены все элементы группы, например, все преподаватели данной школы, или все общественные организации в этом городе. Имена объектов, связанные с группой, являются также и кнопкой вызова страницы этих объектов (на экран конкретной организации, конкретного человека).

Для полноты информационного представления, к объектам можно присоединять файлы с традиционным содержанием: тексты (статьи, библиографические списки), картинки, видео и т.д.

Интерфейс конкретного объекта состоит из элементов: название объекта, главное изображение, главная статья, меню групповых свойств, файлы для скачивания, www-ссылки и значения свойств объекта.

В данный момент в опытной эксплуатации находится проект «Информация о городах Свердловской области»: <http://semantic.uraic.ru/sv-obl/>. Основное содержание: города, реки, горы, люди, памятники, предприятия, события и другое – объекты географического, исторического, культурного, социального значения на территории Свердловской области. Максимальное наполнение системы разноплановыми объектами позволит в будущем организовать узко тематические проекты. Например, «Природные территории и Красная книга Свердловской области» (проект должен быть готов осенью 2011 г.), «Памятники истории и культуры Свердловской области», «Краеведы и краеведческие коллекции Урала» и др. Одни и те же объекты могут участвовать в нескольких тематических подпроектах.

Редакторы контента проекта – специалисты библиотек Свердловской области (в основном, библиографы). Их профессиональные знания, навыки работы с информацией, доступность к первоисточникам позволяют создавать фактографически точный контент.

Техническое решение проекта: программные модули выполнены в С# Microsoft Visual Studio 2010, используемая СУБД – MS SQL Server 2008.

Техническое сопровождение обеспечивается программистами библиотеки им. Белинского

### **Приложения электронных семантических библиотек.**

Ниже приведены примеры приложений. Не обладая приемлемым для публичного использования продуктом ЭСБ, электронные библиотеки моделируют семантическую функциональность с помощью вспомогательных подсистем.

oreChem ChemXSeer: семантическая библиотека для химии. Моделируются химические статьи метаданными. Используются тэги, основанные на химической онтологии[4].

National electronic Library of Infection (UK). Для поиска и просмотра статей понятия выбираются из онтологии. Обрабатываются более узкие, широкие и связанные понятия. Пользователи библиотеки проиндексированы элементами онтологии и для них формируются рекомендуемые списки[5].

## **ЧАСТЬ II. ТЕКСТЫ И МУЛЬТИМЕДИА. ПОСТЫ**

Объектная модель в системах семантической направленности создает возможность для индексации традиционных материалов объектами реального мира. Будем рассматривать один из вариантов организации связей текстов и мультимедиа с объектами – через посты. Мы предлагаем использовать посты (термин наш для применения в данном контексте) в качестве оболочек для сборки текста и мультимедиа. Через эту оболочку организуется связь с индексами.

Пост состоит из необязательных элементов: заголовка, текста, картинки, виджета видео (tube), файла для скачивания, www-ссылок, списка библиографии, ссылок на другие посты, являющиеся его частями или продолжающие его тему. Каждый пост сопровождают индексы: «объекты», «характер информации», «даты».

Индекс «объекты» состоит из пар: (роль объекта в данном посте, ссылка на объект). По ссылке на объект можно перейти на экран с полным описанием объекта, содержащим его связи, свойства и связанные с ним документы. Роли объектов – это, например, автор, участник (событий, описанных в посте), исполнитель (поручения, описанного в посте) и другие.

Индекс «характер информации» состоит из списка типа: официальный документ, рассказ, пословица, цитата, закон и т.д. Этим индексом характеризуется форма материала, объем, а также роль в социуме.

Индексом дат является совокупность пар (роль даты, дата). Позволяет сопоставить текст (мультимедиа) с датой с указанием, что именно происходило. Роль даты – это дата создания, подписания, исполнения (поручения в тексте), открытия (выставки) и т.д.

Экран поиска постов имеет окна для слов, которые встречаются в тексте поста, указатель, где их искать (заголовок, текст). Также можно указать объект, с которым связан пост и его роль; выбрать характер информации поста; указать условия на даты и их роли.

Находясь на экране, где представлен конкретный объект, можно выбрать посты, связанные с ним, и их содержимое – текст и мультимедиа. Тем самым осуществляется индексация последних.

Совокупность постов при наличии контекстного поиска составляет портал (ворота, вход) в пространство объектов - более «размытый», чем при поиске от названия объекта и при условиях на свойства объектов.

Наличие постов (мультимедиа, текстов) оптимизирует поиск объектов и позволяет представить модель мира человеку в переплетении полностью структурированного формата (объекты) с традиционно линейным (текстово-изобразительным) представлением. Это улучшает восприятие информационного материала и расширяет круг людей, способных эффективно работать с информацией, представленной в электронной библиотеке.

## **ЛИТЕРАТУРА**

1. *Using Dublin Core – Dublin Core Qualifiers* [Электронный ресурс] // The Dublin Core Metadata Initiative : [сайт]. – 1995–2011. – URL: <http://dublincore.org/documents/usageguide/qualifiers.shtml> (Дата обращения: 22.08.11).
2. *Shahrul Azman, N. Ontology-Driven Semantic Digital Library* [Электронный ресурс] // Lecture Notes in Computer Science. – 2010. – V 6458. – URL: <http://www.springerlink.com/content/978-3-642-17186-4/#section=828477&page=1&locus=-3> (Дата обращения: 22.08.11).

3. *Tiropanis, T.* Semantic Technologies for Learning and Teaching in the Web 2.0 Era: A survey of UK Higher Education [Электронный ресурс] // JISC : [сайт]. – 2011. – URL: <http://www.jisc.ac.uk/media/documents/projects/semtech-websci-paper.pdf> (Дата обращения: 22.08.11).
4. *Li, N.* oreChem ChemXSeer: a semantic digital library for chemistry // In Proceedings of the 10th annual joint conference on Digital libraries (JCDL '10). – 2010. – P. 245–254.
5. *Kostkova, P.* User profiling for semantic browsing in medical digital libraries // In Proceedings of the 5th European semantic web conference on The semantic web: research and applications. – 2008. – P. 827–831.