

На пути к созданию электронной семантической библиотеки

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

Свердловская областная универсальная научная библиотека им. В.Г. Белинского,
г. Екатеринбург
semantic.adm@gmail.com

Аннотация

В статье излагаются некоторые практические результаты, достигнутые в существующих электронных библиотеках на пути их развития к так называемым «электронным семантическим библиотекам».

Описывается действующая система «Семантик» и ее приложение в фактографическом обслуживании читателей библиотек Свердловской области.

1. Модель объект-свойство в электронной библиотеке (ЭБ)

Наблюдая познавательную деятельность человека, можно сказать, что существенную часть занимает выделение объектов, их связей и свойств в потоке восприятия. Объектное представление может обрабатываться компьютером, и, таким образом, формируется новая информация на основании связей и взаимоотношений между объектами. Описание объектов и их взаимосвязей – часть содержания онтологий. Онтологии определяют концепты и отношения между ними, используемые для описания и представления знаний в какой либо области. Приводимая в разделе 3 система «Семантик» на данном этапе развития (август 2011 г.) не содержит важной части онтологии – ограничений, и в ней не осуществляется вывод одних знаний из других.

Электронную библиотеку, основанную на объектной модели, можно характеризовать тем, что:

- А) Она является хранилищем 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 близки к традиционным интерфейсам информационных систем;

Е) Обеспечено размещение объектов любой сферы деятельности (области знаний).

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. 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 составляют сеть, и ознакомление с материалом может происходить переходами по именованным ссылкам от объекта к объекту. В узлах такой сети находятся объекты действительности – галактики, экскаваторы, 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 представить логические связи объекта с другими объектами, значения его свойств, а также прикрепленные к нему традиционные информационные ресурсы.

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 универсальность [2].

2. Достоинства электронной библиотеки на объектной основе

Универсальность является существенным достоинством. Конкретное наполнение

характеризуется определенным массивом конкретных объектов и их типов. Но то же программное обеспечение, установленное в другом месте, позволит ввести совсем другие объекты другой тематики.

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

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

В системе реализуется **поиск** по объектам и сосредоточение **традиционных материалов** около объекта, которому они посвящены.

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

Известно, что обучение с использованием таких систем 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.

3. «Семантик» - электронная библиотека фактографического обслуживания

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 изображение, 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 изображений 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 C# Microsoft Visual Studio 2010, используемая СУБД – MS SQL Server 2008.

Техническое сопровождение обеспечивается программистами библиотеки им. Белинского.

4. Приложения электронных библиотек

Ниже приведены примеры приложений – две электронные библиотеки, которые, используя свои «инструменты», позволяют работать с объектной моделью. Не обладая приемлемым для публичного использования продуктом, электронные библиотеки моделируют объектную 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].

application oriented on supply of users of traditional libraries with facts about their region according to their queries.

Литература

- [1] Using Dublin Core - Dublin Core Qualifiers. <http://dublincore.org/documents/usageguide/qualifiers.shtml>
- [2] Shahrul Azman Noah, Nor Afni Raziah Alias, Nurul Aida Osman, Zuraidah Abdullah, Nazlia Omar, Yazrina Yahya and Maryati Mohd Yusof. Ontology-Driven Semantic Digital Library, INFORMATION RETRIEVAL TECHNOLOGY, Springer, LNCS, 2010, Volume 6458/2010
- [3] Thanassis Tiropanis, Hugh Davis, Dave Millard, Mark Weal. Semantic Technologies for Learning and Teaching in the Web 2.0 Era: A survey of UK Higher Education. LSL, School of Electronics and Computer Science, University. <http://www.jisc.ac.uk/media/documents/projects/se-mtech-websci-paper.pdf>
- [4] Na Li, Leilei Zhu, Prasenjit Mitra, Karl Mueller, Eric Poweleit, and C. Lee Giles. 2010. oreChem ChemXSeer: a semantic digital library for chemistry. In Proceedings of the 10th annual joint conference on Digital libraries (JCDL '10). ACM, New York, NY, USA, 245-254.
- [5] Patty Kostkova, Gayo Diallo, and Gawesh Jawaheer. 2008. 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 (ESWC'08), Sean Bechhofer, Manfred Hauswirth, Hoffmann, and Manolis Koubarakis (Eds.). Springer-Verlag, Berlin, Heidelberg, 827-831.

On The Way To Semantic Digital Library

© M. Yakovleva, A. Ten, V. Kugler

A few practical results for existing digital libraries being on the way to transform into so called Semantic Digital Libraries are stated out in this article. The digital library “Semantic” has been described with its