

ИНФОРМАЦИОННАЯ СИСТЕМА "ВОСТОКОВЕДЕНИЕ"

Котов А.В.

Центр научных телекоммуникаций
и информационных технологий РАН,
Москва, ул. Вавилова, 42
sanek@ccas.ru

Серебряков В.А.

Центр научных телекоммуникаций
и информационных технологий РАН,
Москва, ул. Вавилова, 42
serebr@ccas.ru

Столяров А.А.

Институт востоковедения РАН,
103031 Москва, Рождественка, 12
astol@orc.ru

Аннотация

В статье рассматривается Информационная система “Востоковедение”, разрабатываемая совместно Центром научных телекоммуникаций РАН и Институтом востоковедения РАН. Целью системы является хранение различной информации о востоковедческих ресурсах (таких как востоковеды, востоковедческие организации, публикации, проекты, сайты, новости, фактографическая информация о восточных артефактах) в электронной форме, обеспечение своих пользователей механизмом загрузки информации в систему с разделением прав доступа и защитой информации, поиск информации, её представление. Описаны метаданные информационной системы “Востоковедение”.

Введение

Во многих российских депозитариях (отделах рукописей и редких книг научных институтов и библиотек, музеях, архивах и т.п.) хранятся бесчисленные источники по истории и культуре народов Востока, являющиеся базой для большинства фундаментальных исследований. Труды, посвященные исследованию этих бесценных ресурсов, составляют основу российской классической востоковедческой школы, чьи научные достижения мирового уровня до сих пор высоко ценятся за рубежом. Ежегодно востоковедческие центры по всей России выпускают огромное количество публикаций, в которых рассматриваются фундаментальные проблемы сегодняшней науки. Все они отличаются высоким уровнем цитируемости.

Между тем, лишь редкие из этих публикаций доходят как до западных, так и до отечественных коллег, становясь библиографической редкостью практически сразу же после выхода в свет. Подавляющее большинство трудов по проблемам востоковедения издается ограниченными тиражами и распространяется среди узкого, заранее известного круга специалистов, не попадая не только в массовые, но и в малотиражные библиографические указатели. Выходом из сложившегося положения может стать перевод большинства

исследований, проводившихся до сих пор традиционными способами, на новый, компьютерно-сетевой уровень.

Современные информационные технологии позволяют создавать автоматизированные информационные системы хранения библиографической, справочной, музейной и прочей информации, которые доступны по сети Интернет и которые позволяют пользователю любого уровня получить оперативный доступ к интересующей его информации. Это, с одной стороны, в значительной мере избавляет исследователей от трудоёмких, ручных и малоинтересных операций, связанных с обработкой информации, а с другой - даёт им в руки мощнейший научно-исследовательский инструмент.

В настоящее время востоковеды предпринимают активные попытки создания специализированных информационных систем, которые, по сути своей, являются узкоспециализированными базами данных, интерфейс и функциональность которых варьируют от системы к системе. Им не присуще свойство совместимости ни на уровне метаинформации, в них содержащейся, ни на уровне функциональности. Подобные специализированные системы ориентированы на решение конкретных востоковедческих проблем. Многие из таких систем описаны в сборниках “Электронные библиотеки и базы данных по истории Евразии в средние века” ([4] – [7]), они также были предметом обсуждения на международных конференциях “Исторические источники Евразийских и Североафриканских цивилизаций: компьютерные подходы” (Софрино – 1995, Звенигород – 1998, Сергиев Посад – 1999, Звенигород – 2001). К специализированным информационным системам описанного типа относятся, например, база данных “Посольские книги России конца XV – начала XVIII вв.”, разработанная Н.М. Рогожиным и А.А. Богуславским ([5]), информационно-поисковая система по библиографии для локальных сетей, представленная А.М. Вайнгаузом и Я.А. Шером ([5]), каталог монгольских фольклорных материалов Архива востоковедов Санкт-Петербургского Филиала Института востоковедения РАН, разработанный И.М. Кульганек и И.М. Голубом ([7]) и др.

Как уже было отмечено, эти системы узко ориентированы на конкретные информационные данные. Такой подход к организации хранения ускоспециализированной информации и специфического доступа к ней обладает своими плюсами и минусами. К плюсам относится то, что такая узкая ориентированность систем на хранение конкретных исторических материалов позволяет наиболее полно и точно описывать эти материалы. К минусам относится то, что теряются возможности по совокупному анализу материалов и сведений из разных отраслей востоковедения, собранных и обработанных востоковедами за многие десятилетия их деятельности. Каждая такая информационная система является замкнутым, никак не связанным с другими ресурсами и системами, хранилищем некоторых специфических данных.

В то же время, в решениях конференций “Исторические источники Евразийских и Североафриканских цивилизаций: компьютерные подходы” выражена “необходимость выработки системы единых подходов и стандартов с целью унификации уже выполненных объемов информации специалистами, работающими в различных востоковедческих центрах России и СНГ”. Одновременно заявлено о развёртывании работ по созданию “серии электронных каталогов востоковедческих центров, а также каталога тюркских, монгольских и тибетских источников, хранящихся в библиотеках, архивах и музеях России и СНГ, для последующего их размещения на Евроазиатском востоковедческом сервере EAOS”. В число объявленных электронных каталогов, в частности, входят “Аннотированный электронный каталог востоковедческих центров России”, “Электронный библиографический указатель российских востоковедов”, “Аннотированный электронный каталог российских коллекций памятников материальной, письменной и духовной культуры народов Востока” и проч. Предполагается, что все эти системы предназначены для

работы с информацией библиографического и справочного типа, и должны удовлетворять принципу универсальности в способах хранения и предоставления данных конечному пользователю.

В 2000 году совместными усилиями Центра научных телекоммуникаций и информационных технологий РАН (ЦНТК РАН) и Институтом востоковедения РАН (ИВ РАН) началась разработка и реализация так называемой информационной системы "Востоковедение", которой посвящен этот доклад.

Общее описание информационной системы "Востоковедение"

Информационная система "Востоковедение" должна решить большинство из проблем, упомянутых выше. Система создается на базе ИСИР РАН ([3]) и Евроазиатского востоковедческого сервера EOAS (<http://www.orient.ru>). Она представляет собой цифровую библиотеку клиент-серверной архитектуры, доступ к которой осуществляется через Интернет при помощи стандартного браузера Интернет. Клиентом для такой системы является стандартный браузер Интернет, версии не ниже 4. Сервер представляет собой Web-сервер (в текущих реализациях – Apache или IIS) и сервер реляционной базы данных (в текущих реализациях – Oracle, MS SQL Server, а также любой, поддерживающий протокол доступа DBI). С точки зрения пользователя информационной системы "Востоковедение", ИС "Востоковедение" является "большой частью" интернет-сайта www.orient.ru, который, помимо, собственно, самой ИС "Востоковедение" содержит еще некоторые вспомогательные вещи, удобные и необходимые для востоковедов, такие как специализированные форумы по различным отраслям и направлениям востоковедения, в том числе, и по метаданным в востоковедении.

Предназначение информационной системы "Востоковедение"

ИС "Востоковедение" предназначена для:

- накопления
- обработки
- хранения
- поиска
- просмотра

информации следующих трех основных типов:

- библиографической;
- справочной;
- фактографической.

Функциональность информационной системы "Востоковедение"

Работа с информационной системой "Востоковедение" состоит из трех основных частей:

- поиск информации, содержащейся в ней;
- просмотр информации, содержащейся в ней;

- внесение и редактирование информации;
- администрирование системы.

Просмотр информации можно производить в результатах поиска по некоторому запросу, при помощи навигации по иерархическим структурам системы (организации, подразделения), или посредством перехода по гипертекстовым ссылкам.

Поиск информации осуществляется по метаполям, описание которых дано в следующем пункте доклада. Для каждого из ресурсов свой вариант поиска по некоторому набору метаполей, относящихся к этому ресурсу. Возможно, также, сужение поиска по одному ресурсу с помощью другого ресурса.

Редактирование информации осуществляется с помощью браузера Интернет посредством web-форм. Внесение информации возможно как при помощи web-форм, так и посредством пакетной загрузки файлов с данными.

Администрирование системы построено по иерархическому принципу согласно иерархии востоковедческих коллективов и организаций. В востоковедческом коллективе или организации может существовать ответственная персона, которая обладает правами на редактирование информации в подразделении, к которому она относится, а также во всех подразделениях, ему рекурсивно подчиненных. Кроме того, эта персона обладает правами на передачу прав другой персоне, ответственной за любое поддерево в дереве, за которое отвечает первая персона.

Описание метаданных информационной системы "Востоковедение"

Информация библиографического типа включает в себя

- метаинформацию о научных публикациях по востоковедческой тематике, о научных периодических изданиях, касающихся востоковедения, книг по востоковедческой тематике. Для каждого ресурса такого типа в системе хранится следующая информация: название, авторы, тип (из специализированного классификатора), год издания, язык, аннотация, ключевые слова, по мнению автора характерные для публикации. В информационной системе "Востоковедение" за основу для хранения такой метаинформации взят распространенный формат Dublin Core ([1]), служащий для описания документов, публикаций и т.п.
- текст самой публикации (текст в общепотребимой форме – TXT, TEX, DOC, PDF, PS и т.д.)

Информация справочного типа включает в себя метаинформацию о:

- востоковедах;
- востоковедческих организациях и их подразделениях;
- проектах, в которых востоковеды принимают участие;
- новостях востоковедческой тематики;
- адресов сайтов в Интернет по востоковедческой тематике.

За основу способа хранения метаинформации о востоковедах и востоковедческих организациях и подразделениях в информационной системе "Востоковедение" принят достаточно распространенный в мире формат vCard, подходящий для представления метаданных о, так называемых, персоналиях, разработанный компаниями Apple, AT&T, IBM и Siemens. vCard - формат для хранения, так называемых, электронных бизнес-карт. vCard удобен для обмена метаинформацией о персоналиях между приложениями и не зависит от конкретного метода передачи информации между этими приложениями. Говоря более конкретно, для хранения метаинформации о востоковедах в информационной системе "Востоковедение" используется

следующее подмножество полей vCard, такое, как фамилия, имя, отчество, звание (из специализированного классификатора), ученые степени (из специализированного классификатора), телефоны, адреса электронной почты, специальности ВАК (из специализированного классификатора), направления деятельности, отрасль востоковедения, где он трудится, ключевые слова, характерные для его деятельности, ее результатов, фотография, организация и подразделение, где работает, его публикации, проекты, в которых он принимал участие. Для хранения информации о востоковедческих организациях и их подразделениях в информационной системе "Востоковедение" используется следующее подмножество полей vCard, такое, как название, сокращенное название, тип организации (из специализированного классификатора), город, в котором расположена, телефоны/факсы, почтовый индекс и адрес, схема проезда, адреса электронной почты, адреса ее www-страниц, направления ее деятельности, ключевые слова, характерные для ее деятельности, ее результатов, историческая справка, ее сотрудники, публикации ее сотрудников, проекты, в которых принимала участие организация (подразделение) или ее (его) сотрудники.

За основу способа хранения метаинформации о проектах в информационной системе "Востоковедение" принят распространенный в мире формат CERIF (The Common European Research Information Format), содержащий достаточно широкий набор атрибутов для описания метаинформации о персонах, организациях, проектах, научных исследованиях и научных публикациях. Для хранения метаинформации о проектах в информационной системе "Востоковедение" используется подмножество полей CERIF, такое, как название проекта, его участники, код в нумерации спонсора проекта, даты начала и, возможно, окончания проекта, востоковедческие организации и востоковеды, принимавшие участие в проекте, публикации, связанные с проектом.

Об адресах Интернет-сайтов по востоковедческой тематике в настоящее время в информационной системе "Востоковедение" хранится следующая метаинформация: название (заголовок) сайта, описание сайта, URI (текст) - [URL-] ссылки на документ, классифицируемые типом формата (HTML, RTF, DOC, PDF, PS, и т.д.), вид (из классификатора, обеспечивающего классификацию, семантическое деление ресурсов, причисленных к этому типу), язык данных, вид представимой информации, описываемой ею деятельности, ключевые слова, интервал дат, в течение которого указатель актуален (затем виден только в "архиве" - попадает "архив"). О новостях востоковедческой тематики в системе хранится следующая метаинформация: название, интервал дат, в течение которого новость актуальна (затем видна только в "архиве" - попадает архив), описание новости, вид новости (из специализированного классификатора, обеспечивающего классификацию, семантическое деление ресурсов, причисленных к этому типу), ключевые слова.

Для представления метаинформации о новостях и сайтах в мире уже, разумеется, существуют общепринятые стандарты. Например, проект Arts Digital, поддерживаемый компанией AOL, предложил свой способ описания новостей и сайтов, на котором пока, на момент написания этого доклада, остановились разработчики информационной системы "Востоковедение". Перечисленные выше атрибуты (поля метаданных) для новостей и сайтов являются, в некоторой степени, отображением моделей из Art Digital на текущие реальные потребности ИС "Востоковедение", которая пока находится (в части описания метаинформации новостей и сайтов) в стадии разработки.

Информация фактографического типа представляет собой метаинформацию об электронных копиях восточных исторических артефактов духовной и материальной культуры. В понятие электронной копии входят сведения обо всех объективных [изображение, размеры, материал, место обнаружения и т.д.] и субъективных [в основном, добытых исследователями на основе трактовки содержания текста,] свойствах исследуемого объекта).

Как уже говорилось выше, востоковедами предпринимаются активные попытки создания узкоориентированных на конкретные отрасли и подотрасли востоковедческих систем. Были перечислены плюсы и минусы таких систем. Одним из недостатков узкой ориентируемости таких систем, как уже отмечалось, - это отсутствие возможности по совокупному анализу материалов и сведений из разных отраслей востоковедения. В частности, у таких систем нет общего поиска информации внутри них самих.

Одно из предназначений информационной системы "Востоковедение" - это предоставить своим пользователям унифицированный способ доступа к узкоспециализированным данным об электронных копиях восточных исторических артефактов духовной и материальной культуры. Для решения этой задачи был проведен анализ существующих в мире информационных систем, претендующих на свою универсальность в области предоставления доступа к различным видам исторической информации. На основе анализа существующих систем подобного рода был выделен проект ECAI (Electronic Cultural Atlas Initiative ([2])). Целью этого проекта служит создание, так называемого, "Электронного атласа мира", содержащего в себе разнообразную информацию о культурном наследии Человечества. Для решения проблемы унификации доступа к, на первый взгляд, очень разнородной информации ECAI предложила выделить у всех исторических ресурсов общий набор метаполей и использовать для хранения и представления этих полей стандарт представления метаданных Dublin Core ([1]). Разработчики информационной системы "Востоковедение" приняли решение действовать таким же образом, как и ECAI, и использовать стандарт Dublin Core для описания некоторого подмножества фактографической метаинформации об электронных копиях восточных исторических артефактов духовной и материальной культуры, которое является инвариантным по отношению к различным специфическим узкоориентированным востоковедческим информационным системам. Ниже приведена таблица с общим описанием метаполей Dublin Core, используемых для представления фактографической метаинформации. В ней не приведены квалификаторы для каждого поля, уточняющие и раскрывающие каждое метаполе.

Название поля	Краткое описание поля
DC.Title	Имя, данное электронной копии.
DC.Creator	Персона или организация, ответственная за создание интеллектуального содержимого электронной копии.
DC.Subject	Ключевые слова или фразы, которые описывают электронную копию.
DC.Description	Текстовое описание содержимого электронной копии.
DC.Publisher	Ответственный за публикацию электронной копии в системе.
DC.Contributor	Персона или организация, внесшие вклад в изучение прообраза электронной копии ресурса.
DC.Date	Дата рождения прообраза электронной копии, и дата создания электронной копии.
DC.Type	Тип электронной копии (из классификатора).
DC.Format	Формат, в котором представлен электронный ресурс (электронная копия артефакта).
DC.Identifier	Уникальный идентификатор электронной копии.
DC.Source	Печатная или электронная публикация, в которой прообраз электронной копии получен или описан.
DC.Language	Язык электронной копии.
DC.Relation	Типизированные взаимосвязи с другими электронными копиями.
DC.Coverage	Место обнаружения прообраза электронной копии.
DC.Rights	Поле, описывающее авторские права на электронную копию.

Каждая запись об отдельной электронной копии восточных исторических артефактов описывается такими атрибутами. Все такие записи связаны между собой и связаны с другими ресурсами типа персона, организация, публикация. Связь с публикацией осуществляется через поле DC.Source, которое представляет

собой печатную или электронную публикацию, в которой ресурс был получен или описан. Связь с персонами и организациями осуществляется через поля DC.Creator и DC.Contributor.

На рисунке 1 изображена схема взаимосвязи отдельных узкоориентированных востоковедческих информационных систем и информационной системы "Востоковедение". В докладе не рассматривается вопрос об экспорте метаданных отдельных специализированных систем в ИС "Востоковедение", так как пока только идет разработка общей концепции этого механизма.

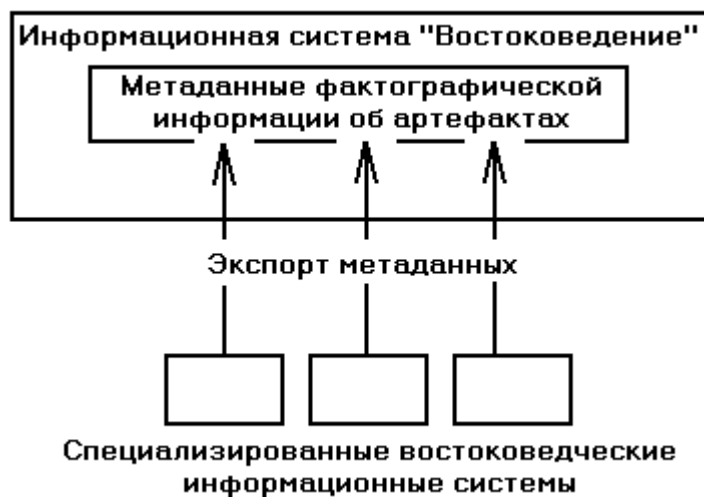


Рисунок 1

По большинству вышеперечисленных атрибутов (метаполей) ресурсов, относящихся к библиографической, справочной и фактографической информации, возможен поиск в системе. Все ресурсы взаимосвязаны между собой. Их связь в достаточно общем виде изображена на рисунке 2. Для хранения такой структуры данных в ИС "Востоковедение" используется, как уже отмечалось, реляционная модель данных.

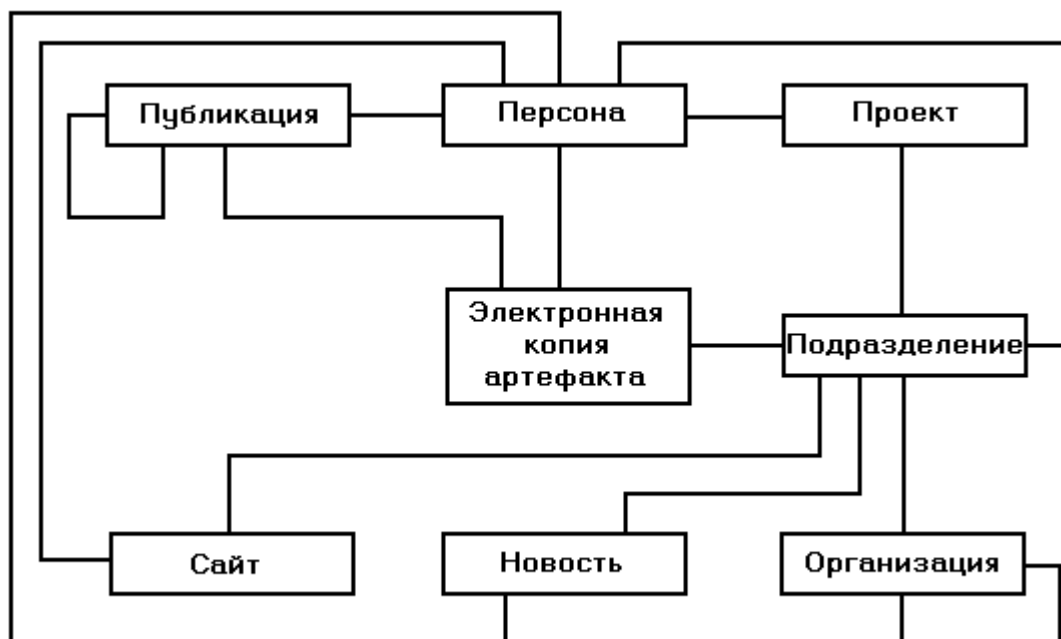


Рисунок 2

Заключение

В статье была описана информационная система “Востоковедение”, рассмотрены её общая архитектура, предназначение, функциональность, метаданные и взаимосвязи между ними. К сожалению, в докладе почти не рассмотрена важная проблема, связанная с хранением информации о местонахождении прообразов электронных копий фактографических ресурсов и датировок, связанных с ними. В настоящее время эти вопросы всё еще находятся в стадии обсуждения, так как информационная система “Востоковедения” находится сейчас в стадии развития.

Список литературы

- [1] Dublin Core Metadata Initiative
http://purl.org/metadata/dublin_core/
- [2] The Electronic Cultural Atlas Initiative Metadata Manual, Version 1.1, March 1999
<http://www.ecai.org/>
- [3] М. В. Кулагин , В. А. Серебряков , А. Н. Бездушный , М. П. Галочкин , А. М. Меденников , С. В. Агошков
Интегрированная Система Информационных Ресурсов (ИСИР) РАН - подход к созданию интегрированных цифровых библиотек. международная научная конференция "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург.
- [4] Базы данных по истории Евразии в средние века. - Вып. 1. - Отв. ред. Столяров А.А. - Институт востоковедения РАН, Ассоциация востоковедов РАН. - М. - 1992.
- [5] Базы данных по истории Евразии в средние века. - Вып. 4-5. - Отв. ред. Столяров А.А. - Институт востоковедения РАН, Ассоциация востоковедов РАН. - М. - 1997.
- [6] Базы данных по истории Евразии в средние века. - Вып. 6. - Отв. ред. Столяров А.А. - Институт

востоковедения РАН, Ассоциация востоковедов РАН. - М. - 1998.

[7] Базы данных по истории Евразии в средние века. - Вып. 7. - Отв. ред. Столяров А.А. - Институт востоковедения РАН, Ассоциация востоковедов РАН. - М. - 1999.