

Информационная система по проблеме Тунгусского явления

© Марчук А.Г.¹, Караваева А.Г.², Привезенцев А.И.³, Родимова О.Б.³, Фазлиев А.З.³

1. Институт систем информатики СО РАН, Новосибирск
2. Государственный архив Томской области, Томск
3. Институт оптики атмосферы СО РАН, Томск
faz@iao.ru

Аннотация

В данной работе описана информационная система, содержащая в себе, в настоящее время, несколько тысяч документов гуманитарного и научно-популярного характера, связанных с историей исследования нерешенной проблемы двадцатого века: Тунгусского феномена. Представлены некоторые детали системы управления контентом, средства для формирования коллекций изображений и системы подготовки метаданных, использованных при ее создании. В заключительной части работы предложен подход к систематизации материалов по данной проблеме, имеющих научно-исследовательский характер.

1 Введение

Вот уже скоро сто лет как наука получила интригующий вызов в лице Тунгусского феномена и не ответила на него до сих пор. В течение 80 лет посещают место происшествия многочисленные экспедиции, имеются свидетельства сотен очевидцев, собраны сотни проб почв и растений, сделаны сотни лабораторных анализов, предложены десятки гипотез, трактующих те или иные стороны явления. Но полученные ответы порождают новые вопросы, а суть случившегося ясным летним днем 1908 года в Тунгусской тайге ускользает, оставляя ощущение тайны.

Ушло довоенное поколение исследователей, связанное, прежде всего, с именем Л.А.Кулика, оставив статьи в журналах, ставших библиографической редкостью. На грани ухода поколение 60-80-х, на долю которого пришлось большая часть полевой и лабораторной работы. Ее результаты, частично опубликованные в основном в сборниках и трудах конференций, остались в личных архивах членов экспедиций. Написанные этими людьми в конце прошлого века обобщающие монографии изданы мизерными тиражами. А тайна

осталась. Теперь появится новое поколение, выдвигающее забытые гипотезы и стремящееся повторить уже сделанную работу.

Целью работы является сохранение пройденного в рамках информационной системы “Тунгусский феномен”, содержащей документы, связанные с тунгусской проблемой. Не исключено, что решение Тунгусской загадки станет доступно лишь будущим поколениям и собранные материалы лягут в основу ее решения. Известно, что упоминание в летописях о небесных гостях помогло определить период обращения комет. В летописной древности люди пользовались пером и зрением, но даже это способствовало решению астрономической проблемы.

2 Структура информационной системы

Известно, что построение системы основано на анализе ее элементов и связей между ними. Документно-ориентированная информационная система “Тунгусский феномен” построена в соответствии со схемой, представленной на рис. 1. Ниже дано краткое описание элементов и их групп, отображенных на этом рисунке.

Источниками документов являются электронная коллекция [1], собранная участниками КСЭ, некоторые фонды Государственного архива Томской области [2, 3], газетные и иные публикации, частные архивы и т.д. В таблице 1 дана интервальная оценка количества документов, включенных в систему (за исключением монографий и научных статей) и их структура.

События, относящиеся к истории изучения феномена, разделены на две группы. К первой из них относятся экспедиционные работы, которые сами являются сложными событиями, включающими в себя маршруты, опросы, взятие проб и т.д., и конференции и семинары. Общим для этого рода деятельности является то, что ее результаты оформлялись в виде научных статей и монографий или первичных материалов, сохранившихся как в личных, так и государственных архивах. Ко второй группе относятся события косвенно связанные с изучением ТФ. Понимание событий этой группы дает

информацию о том, кем были исследователи и их помощники. Документы этой группы связаны, главным образом, с Комплексной Самодеятельной Экспедицией (КСЭ).

Люди, принимавшие участие в изучении Тунгусского феномена, разделены на две группы. В первую из них входят те, кто создавал документы естественно-научного содержания. Они в системе названы исследователями. К числу помощников отнесены участники экспедиций, не вошедшие в число исследователей. Общее число участников

экспедиций, идентифицируемых фамилией, превышает 1000 человек. Экспедиции разбиты по годам их проведения. Было проведено 7 экспедиций АН СССР, 47 – КСЭ, 21 самодеятельная экспедиция из Москвы, а также экспедиции Золотова А. и некоторые другие, в том числе итальянские и немецкая. В системе установлены связи между исследователями и организациями, в которых они работали. Организации также связаны с экспедициями отношением «организатор экспедиции».

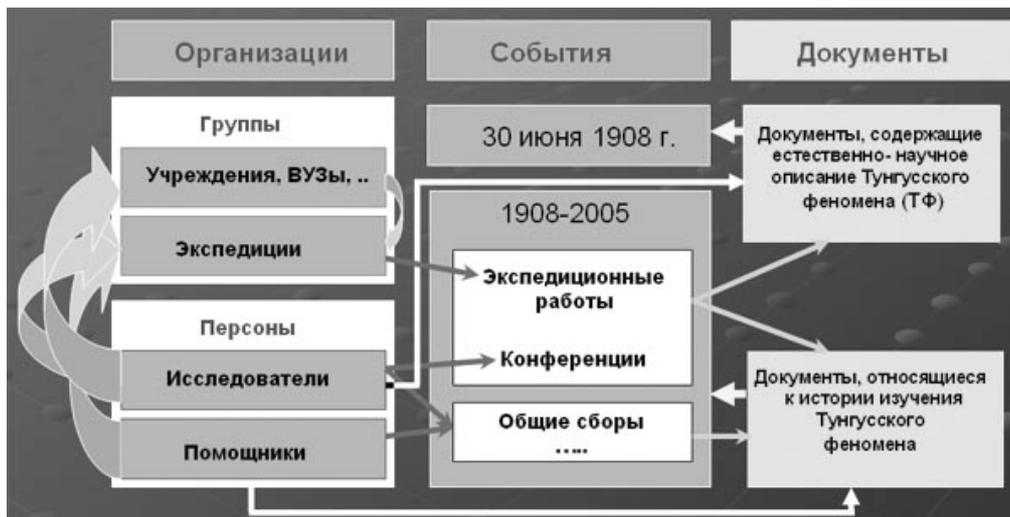


Рис.1. Концептуальная схема информационной системы «Тунгусский Феномен»

Таблица 1. Оценка количества документов

	От 10 до 100	От 100 до 1000	От 1000 до 10000
Документы, содержащие естественно-научное описание феномена	Монографии	Доклады и тезисы Публикации в журнале «ТВ»	Научные статьи Первичные материалы (в листах)
Документы по истории изучения феномена	Проза, Дневники Фильмы Курумники Альбомы Звуковые файлы	Стихи и песни Газетные публикации Архивные материалы (листы)	Фотографии и их описания

Доступные пользователям сети Интернет документы на сайте «Тунгусский феномен» (<http://tunguska.tsc.ru>) расположены в четырех разделах.

Раздел «Архивные документы» представляет материалы, содержащиеся в ГАТО в фондах Н.В.Васильева, Ю.А.Львова и КСЭ.

Раздел «Исследования» содержит монографии, труды конференций, журнал «Тунгусский вестник», выпускаемый коллективом КСЭ и библиографию работ по Тунгусской проблеме.

Раздел «КСЭ» посвящен Комплексной самодеятельной Экспедиции, возникшей в Томске в 1959 году и занимающейся исследованием Тунгусской проблемы. КСЭ - неформальная организация, не имеющая ни устава, ни оформленного членства, - существует уже почти

полстолетия. Раздел содержит материалы об экспедициях, проведенных в рамках КСЭ – их состав и тематическую направленность, сведения о людях, принимавших участие в делах КСЭ, фотоархив, альбомы с фотографиями, принадлежащие членам КСЭ, рукописные журналы, выпускаемые в КСЭ дважды в год, называемые «Курумниками» и представляющие самодеятельное графическое и литературное творчество людей, объединенных общим делом.

Раздел «Лирика» включает стихи, песни и прозу, созданные членами КСЭ, фильмы, посвященные исследованиям Тунгусской проблемы, а также газетные и журнальные публикации, вышедшие с 20-х годов по настоящее время.

В процессе заполнения коллекции документов выделились два основных направления работы: люди и первичные материалы

исследований. Пожалуй, только большой объем просмотренных и включенных в коллекцию документов (см. таблицу) позволил представить, как многих затронула эта проблема, масштабы борьбы идей, самоотверженность и бескорыстие, преданность делу, которое стало для некоторых делом жизни и для большинства – самым значительным, с чем пришлось столкнуться. Эта сторона Тунгусской проблемы – яркая страница истории науки и истории возникновения уникального социального феномена КСЭ. Работа в этом направлении будет продолжена в рамках имеющейся коллекции документов в разделе «Персоналии».

3. Система управления контентом

В информационных системах, использующих архитектуру клиент-сервер, существует три рабочих потока [4]. Здесь рабочий поток понимается как задача, содержащая подзадачи, координируемые пользователем. Первый рабочий поток связан с действиями пользователя на клиентском месте, второй – с действиями, производимыми пользователем на сервере и третий – с действиями при обработке запроса пользователя на сервере. Они относятся к разным абстрактным уровням моделирования предметной области.

В представленной ниже документо-ориентированной информационной системе, для пользователя, имеющего права только на чтение, первый рабочий поток (действия на клиентском месте, не требующие обращения к серверу) отсутствует, тогда как для пользователя с административными правами, обеспечивающего наполнение информационной системы ресурсами, на этом абстрактном уровне доступны веб-приложения, часть из которых будет описана ниже.

Второй абстрактный уровень действий связан с обращением пользователя к серверу. Этот уровень обеспечивает навигацию по информационным ресурсам с учетом прав доступа к ним. Навигация осуществляется по каталогу, называемому далее меню. Инициация пункта меню приводит к генерации информационного ресурса на стороне сервера в виде html-документа. Доступ к ресурсам на стороне сервера зависит от прав, которыми обладает пользователь. Наряду с меню, характеризующим каталогизированные ресурсы, в информационной системе существует ряд типизованных действий, требующих обращения к серверу и тесно связанных с каждым пунктом меню. К числу этих действий, доступных только администраторам, относятся создание пункта меню и описание метаданных для ресурса, связанного с пунктом меню. Просмотр перечня документов всей информационной системы, просмотр каталога метаданных и распечатка текущей страницы доступны всем пользователям. Эти действия

вынесены на отдельную панель управления (см. пиктограммы рис. 5 и 6).

Третий рабочий поток связан с обработкой запроса пользователя к серверу. В системе использованы два модуля (ядра системы) для обработки запроса. Модуль, запрограммированный на языке скриптов PHP, занимается обработкой всех запросов к серверу за исключением запросов к метаданным. Ключевыми действиями, выполняемыми в этом ядре, являются аутентификация пользователей, поддержка сессий, инициализация приложений, формирование документов и коллекций изображений. Обработку запросов к метаданным осуществляет ядро, реализованное средствами Java. Такое смешение подходов обусловлено эволюцией развития создаваемого нами программного обеспечения для построения информационных систем.

В систему управления контентом входит набор приложений, обеспечивающий построение меню, формирование текстовых ресурсов, формирование коллекции изображений и формирование метаданных. Модули для работы с текстами и изображениями реализованы средствами PHP, а модуль для формирования наборов метаданных и поддержки работы с каталогом ресурсов реализован на Java.

Остановимся на кратком описании этих приложений.

3.1. Построение меню

Меню или часть каталога доступных пользователю информационных ресурсов формируется каждый раз при обращении к серверу пользователя с учетом его прав доступа к ИР. Конструирование каталога осуществляется с помощью веб-приложения.

Каталог в системе отождествляется со структурой дерева, каждый узел которого обладает рядом атрибутов перечисленных ниже.

Атрибуты узла дерева (каталог)

1. Директория – идентификатор узла;
2. Заголовок меню
3. Заголовок страницы
4. Комментарий к странице
5. Тип приложения, подключаемого к узлу дерева (Контент, Коллекция фотографий, Новости, Гостевая книга, Обратная связь, Сервлет, Паспорт)
6. Статус узла (доступен или недоступен)
7. Скрыть пункт меню – запрет отображения пункта меню в интерфейсе пользователя вне зависимости от его прав
8. В новом окне – предоставление возможности отображения вызываемого ресурса в новом окне браузера
9. Язык – указание языка на котором описан пункт меню
10. Ссылка – преобразование пункта меню в статическую гиперссылку.

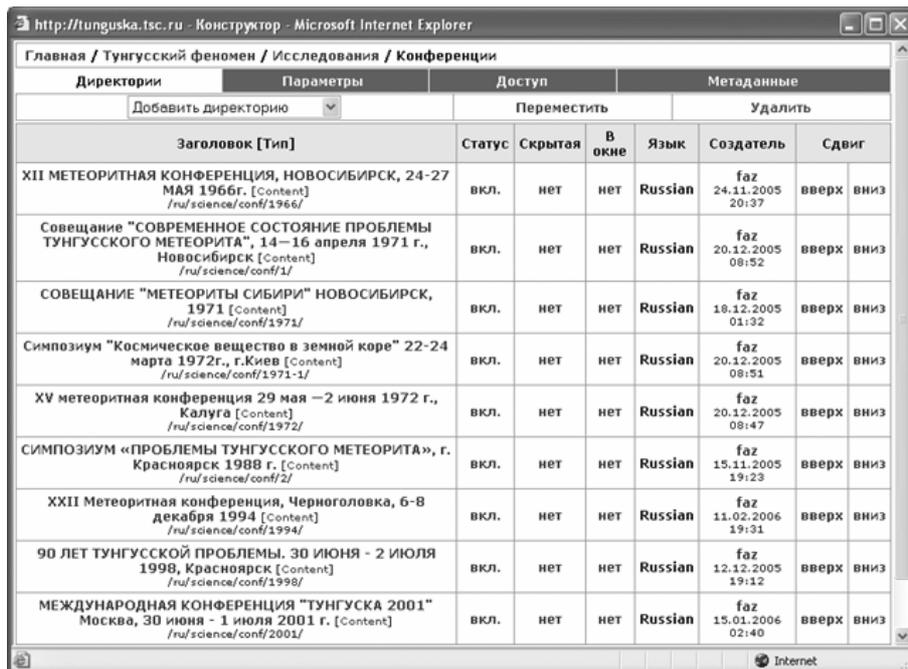


Рис.2. Административный интерфейс для формирования пользовательского меню для навигации в ИС

На рис. 2 показан один из интерфейсов приложения, используемого для создания каталога.

3.2. Формирование текстовых документов

Особенность формирования текстовых документов состоит в возможности установления связей в тексте между символом персоны или

события со структурированными данными о персоне или событии, посредством формирования гиперссылки (в автоматическом или ручном режиме) на соответствующий информационный ресурс, а также между персоной и экспедицией, между исследователем и учреждением.



Рис.3. Веб-интерфейс редактора для занесения и редактирования документов

На рис. 3 показан интерфейс приложения для создания текстовых документов, позволяющего интегрировать в него изображения, формировать гиперссылки и выполнять иные действия ассоциированные с пиктограммами, используемыми во многих типовых редакторах.



3.3. Коллекции изображений

Процесс формирования коллекции изображений включает в себя создание пункта меню, определяющего положение коллекции в каталоге информационных ресурсов, загрузки и краткого описания изображений в соответствии с шаблоном, детального описания каждой фотографии в рамках возможностей, имеющихся в приложении для формирования текстовых документов, геометрического расположения изображений на плоскости отображения.

Созданные этими модулями текстовые и графические ресурсы содержат часть метаданных (создатель ресурса, время создания и т.д.). Формирование метаданных по схеме DC [5] и принятого в ГАТО набора метаданных, ориентированного на описание единицы хранения, осуществляется в третьем модуле. Каталог метаданных об информационных ресурсах создан на основе концептуальной схемы, представленной на рис. 1.

Расположение изображения:	После техник Карадашев
Название документа:	Осушение Суловской воронки
Скрыть изображение:	<input type="checkbox"/>
Автор документа (фотография, изображение) или лицо, предоставившее документ:	Ромейко В.А.
Место хранения оригинала документа:	http://www.tunguska.ru
Время создания документа (начальная дата) (день/месяц/год):	Нет / Нет / 1929
Период создания оригинала документа (конечная дата) (день/месяц/год):	Нет / Нет / Нет
Место создания оригинала документа:	Суловская воронка

Рис.4. Форма для описания фотографии



Осушение Суловской воронки
 Место хранения оригинала документа - МР://www.tunguska.ru
 Время создания документа - 1929
 Преподобная изображения - Суловская воронка

У Л.А.Жульки было твердое убеждение, что метеорит лежит на дне воронки расположенной как на Южном, так и на Северном Бортах. Идея прокопать траншею и таким образом осушить одну из них не давала ему покоя ни днем, ни ночью... Весной 1929 года работы по осушению Суловской воронки шли полным ходом. В центре сидит Л.А.Жульке почти все время участники всегледания.

Помнить

Рис.5. Фотография, представленная на отдельной странице вместе с ее метаданными и описанием

1929. Начало работ в эпицентре катастрофы

Тунгусский феномен » КСЭ » Фотоархив » 1920-1929 » 1929

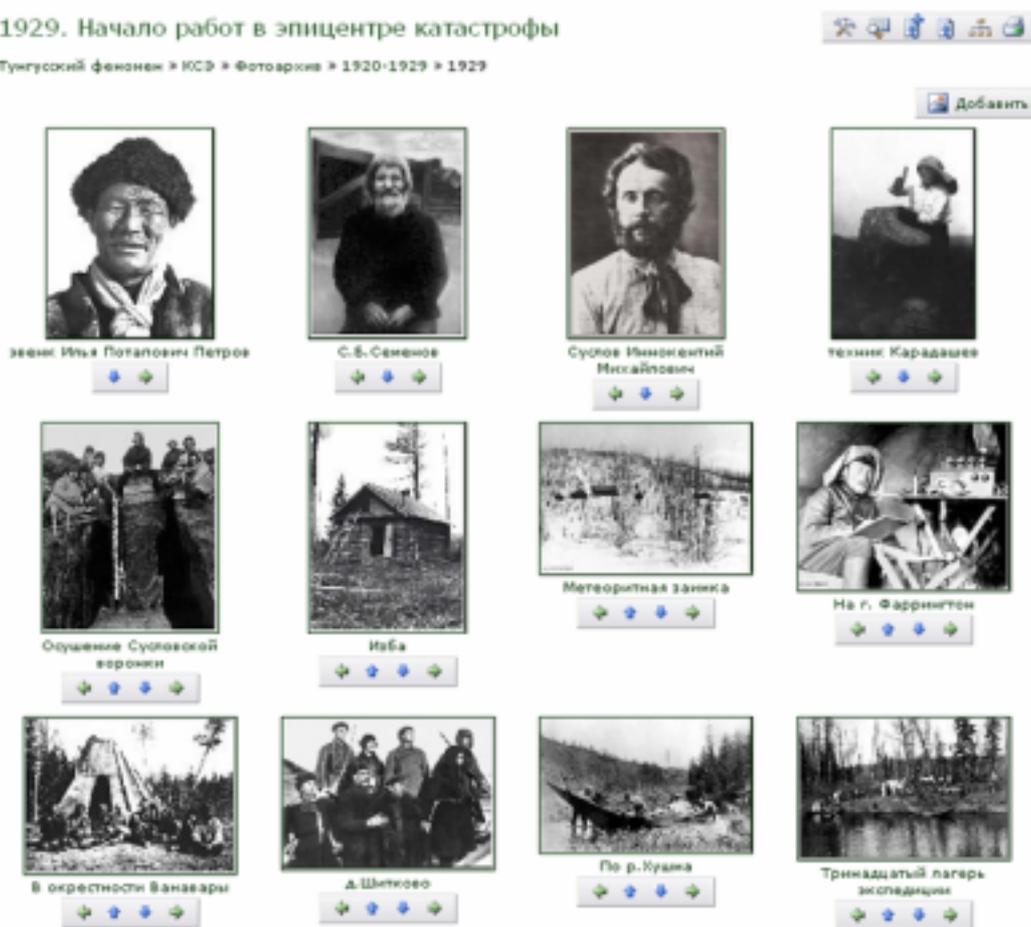


Рис.6. Коллекция фотографий. Представленный на рисунке интерфейс администратора показывает элементы интерфейса, позволяющие перемещать изображения по вертикали и горизонтали.

3.4. Метаданные и аннотации

Документы на сайте «Тунгусский феномен» имеют метаданные. Последние формируются в рамках традиционного подхода с использованием жестких схем, определяемых стандартами и рекомендациями. Формирование метаданных в рамках спецификации W3C RDF/OWL предоставляют возможности для их машинного семантического разбора в рамках технологий Semantic Web. Метаданные информационного ресурса (ИР) представляют собой RDF-документы, описывающие данные, а также связанные с ними RDF-схемы.

Первым шагом при создании метаданных является выбор схемы метаданных. Например, для описания документа, как абстрактного информационного ресурса, можно ограничиться схемой Dublin Core (DC), для описания документа музейного экспоната использовать схему CIMI и т.п. Но стандартный подход с использованием жестких схем метаданных, не всегда применим для работы со сложными документами. В каждой предметной области уже сложилось определенное семантическое описание материализованного аналога информационного ресурса. Например, ниже

представлена предметная схема для архивной описи, используемая специалистами в государственных архивах РФ. Эта предметная схема включает в себя свойства:

- индекс;
- дата события;
- рубрика;
- подрубика;
- место события;
- содержание;
- название архива;
- название фонда;
- название структурной части;
- номер фонда;
- и т. д.

Нами использован следующий подход к организации метаданных. В обязательном порядке формируются метаданные по схеме Dublin Core, которыми можно описать любой абстрактный информационный ресурс, затем, если необходимо, - другие схемы метаданных (CIMI, PRISM). Набор метаданных разных схем, связанный с информационным ресурсом, будем назвать его аннотацией.

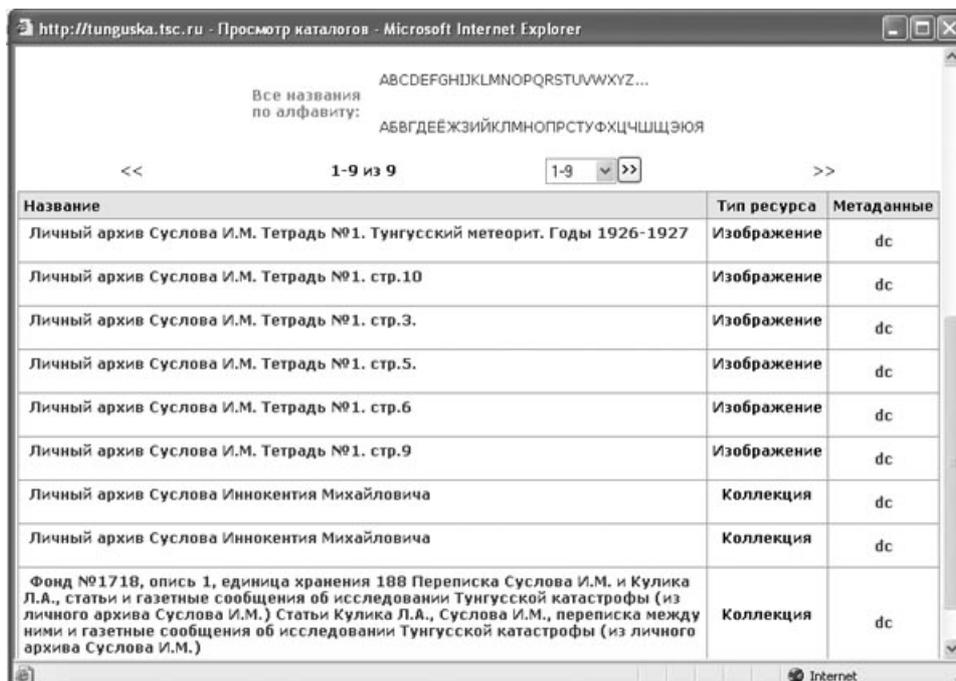


Рис. 7. Интерфейс для работы с каталогом аннотаций

Первичные задачи, которые можно решать с помощью аннотаций – это классификация, поиск и сравнения информационных ресурсов. Задачами второго этапа использования метаданных являются задачи, связанные с машинным семантическим разбором метаданных, т.е. задачи Semantic Web .

Работа с метаданными в рамках сайта «Тунгусский феномен» состоит из двух этапов: подготовки метаданных и их просмотра. Формированием метаданных занимается публикатор, первоначально он должен заполнить поля HTML-формы для схемы Dublin Core (название, автор, тема и др.), но некоторые из полей генерируются автоматически (публикатор, дата, язык). В результате отправки формы на сервер создается XML-документ для проверки достоверности вводимых данных по соответствующей XML-схеме. После этого происходит запись аннотации в виде RDF-документа в файл и параллельно занесение этих данных в базу знаний. База знаний представляет собой объединение всех аннотаций информационных ресурсов и предметных знаний (схем).

На рис. 7 показана часть интерфейса для работы пользователя с каталогом аннотаций.

Реализации модулей работы с метаданными осуществлена в рамках технологии Java - сервлеты, API для работы с RDF/OWL – Jena, API для работы с XML - JDOM. Для создания Web-интерфейса производилась динамическая генерация HTML-страниц из XML-документов посредством XSLT-преобразований. Хранение реестров аннотаций в виде триад осуществляется в СУБД MySQL 4.0.

4. Перспективы

Выполненная работа связана с документами, относящимися к истории изучения Тунгусского Феномена. По количеству документов и деталям описания событий, немногие нерешенные научные проблемы документированы столь подробно.

Первичные материалы, имеющиеся в распоряжении Комплексной Самодеятельной Экспедиции, содержат данные, по меньшей мере, по 15 научным направлениям, среди которых назовем, например, каталог показаний очевидцев, каталоги по ожогу и вывалу леса, данные по пожарам, приросту деревьев, и т.п. Чтобы пользователь мог продуктивно работать с этими данными, необходим выход за рамки коллекции документов и создание информационно-вычислительной системы, позволяющей географическую привязку данных и хотя бы минимальную статистическую обработку. Кроме того, необходимо описание основных теоретических работ с перечислением приближений и допущений, использованных авторами. В этом случае информационная система могла бы служить основой для сравнения данных разных авторов между собой, а также, возможно, для получения новой информации путем наглядного сопоставления данных из разных предметных областей, например, по вывалу и мутационному эффекту и т.п., что затруднительно сделать без информационной системы.

Предполагаемая информационная система необходима как инструмент для решения, по крайней мере, трех естественно-научных проблем. Это, прежде всего, проблема самого Тунгусского феномена. Далее, следует помнить, что Тунгусское явление - это самое крупное документированное

столкновение Земли с космическим объектом. Помимо выяснения свойств такого объекта, что само по себе представляет непосредственный физический интерес, мы оказываемся перед феноменом, важность которого еще в достаточной мере не оценена человечеством. Речь идет об опасности столкновения Земли с космическими объектами в будущем, что, оказывается, не является столь уж невероятным. С конкретными чертами такой опасности поможет ознакомиться информационная система на примере Тунгусской катастрофы. И, наконец, район Тунгусской катастрофы, являющийся сейчас Государственным заповедником [6], в силу исторически сложившихся причин оказался очень подробно исследованным с позиций самых разных дисциплин. Это и генетические исследования коренного населения, и анализ мутаций древесной растительности, и элементный состав почв и торфяников, и данные по термолюминесценции и палеомагнетизму почв и минералов, и геологическое строение Великой котловины. Перечисленное выше, помимо своеобразного растительного и животного мира региона, делает Тунгусский заповедник экологическим объектом, возможно, единственным на Земле, на примере которого могут быть построены экологические модели с уникальной детализацией его свойств и параметров. Данные для решения такой задачи можно будет найти в информационно-вычислительной системе планируемой для создания к столетию падения «Тунгусского метеорита».

5. Заключение

Авторы благодарят Российский Гуманитарный Научный Фонд за финансирование работы в рамках гранта №05-03-12324в и участников комплексных самостоятельных экспедиций Плеханова Г.Ф., Бояркину А.П., Черникова В.М., Бидюкова Б.Ф., Кривякова В.В., Сапожникова В.А. и других за помощь в работе.

Отметим, что на основе описанной в данной работе технологии созданы информационные системы Томского научного центра СО РАН [7,8] и Государственного архива Томской области [9].

Литература

- [1] А.П. Бояркина, В.Кривяков, Г.Ф.Плеханов, О.Б. Родимова, В.А. Сапожникова, А.З. Фазлиев, Электронная коллекция материалов комплексной самостоятельной экспедиции по изучению Тунгусского явления, Труды Всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”, Пушино, 2004, с.
- [2] Государственный архив Томской области. Путеводитель. М., Звенья, 2001, 89бс.

- [3] А.Г. Караваева, А.Г. Марчук, В.И. Привезенцев, О.Б. Родимова, А.З. Фазлиев, Опыт по формированию электронного архива на примере материалов КСЭ, Материалы Второй Всероссийской научно-практической конференции “Документ в парадигме междисциплинарного подхода”, под ред. проф. О.А.Харусь, Томск: Томский гос. ун-т, 2006, с.119-123.
- [4] Yu Lei, Munindar P.Singh, *A Comparison of Workflow Metamodel*, Proceedings of the ER'97 Workshop on Behavioral Models and Design Transformations: Issues and Opportunities in Conceptual Modeling 6 - 7 November 1997, UCLA, Los Angeles, California (<http://osm7.cs.byu.edu/ER97/workshop4/ls.html>)
- [5] Dublin Core, <http://purl.org/dc/>
- [6] Тунгусский заповедник. Биоценозы тайги и влияние на них экстремальных природных факторов, Труды ГПЗ «Тунгусский», часть 1, Томск, Из-во Том. ун-та., 2003, 294 с.
- [7] Баряхнин В.Б., Бычков И.В., Гуськов А.Е., и др., Распределенный виртуальный музей науки и техники СО РАН, Труды 1-ой международной конференции «Системный анализ и информационные технологии», 12-16 сентября 2005 г., Переяславль-Залесский, 2005, с. 41-45
- [8] Томский сегмент Виртуального музея науки и техники СО РАН, <http://www.tsc.ru/ru/tsc/vm/>
- [9] Государственный архив Томской области, <http://www.gato.tomica.ru/>

Information System “Tunguska Phenomenon”

Marchuk A.G.¹, Karavaeva A.G.², Privezetsev A.I.³, Rodimova O.B.³, Fazliev A.Z.³

1. A.P.Ershov Institute of Informatics Systems SB RAS, Novosibirsk
2. State Archive of Tomsk Region, Tomsk
3. Institute of Atmospheric Optics SB RAS, Tomsk faz@iao.ru

Our paper reports development of the information system that currently holds several thousands of documents related to the history of a major unsolved problem of the XX century: the Tunguska meteorite. The report covers details of the content management system, toolkits to form image collections and metadata preparation systems. An approach to systemize the science data relevant to this problem is suggested.