

# ОРГАНИЗАЦИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ В РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЕ, ОРИЕНТИРОВАННОЙ НА РЕШЕНИЕ ЗАДАЧ МОЛЕКУЛЯРНОЙ СПЕКТРОСКОПИИ

А. В. Козодоев

*Институт мониторинга климатических  
и экологических систем СО РАН, Томск, Россия*

ПРИВЕЗЕНЦЕВ А.И.

*Томский государственный университет  
систем управления и радиоэлектроники, Томск, Россия*

А.З. ФАЗЛИЕВ

*Институт оптики атмосферы СО РАН, Томск, Россия*

e-mail: kav@iao.ru, remake@iao.ru, faz@iao.ru

The work describes data and metadata used in the ICS “Atmospheric Spectroscopy”. Organization of metadata exchange in the distributed information system for molecular spectroscopy is discussed.

## Введение

Молекулярная спектроскопия является одним из широко используемых во многих прикладных исследованиях разделов физики. Предметом изучения молекулярной спектроскопии являются спектральные свойства молекул. Детальное изучение спектральных свойств молекул не закончено до сих пор. Связано это с тем обстоятельством, что в расчетах физических характеристик атмосферы используются сотни тысяч линий, каждая из которых описывается десятком параметров. Работа с такими массивами данных требует, с одной стороны, предметной систематизации данных, а с другой — программных средств для их обработки. Систематизация спектральных данных проведена теоретиками несколько десятков лет назад. Массивы данных о спектрах молекул концентрируются в банках данных (БД), среди которых наиболее известны HITRAN [1] и GEISA [2]. Процесс наполнения БД далек от завершения. Современная техника измерения спектров позволяет получать данные с большей точностью и в тех диапазонах длин волн, в которых ранее измерения не проводились. Количество расчетных спектров растет еще более стремительно, например, за последние десять лет количество данных по молекуле воды увеличилось более чем в сто

раз. Растет также число исследовательских групп. Все это указывает на необходимость использования современных информационных технологий для коллективной работы с информацией.

В России работы по созданию информационных ресурсов в области молекулярной спектроскопии ведутся в Институте оптики атмосферы СО РАН с начала 80-х годов [3]. Качественный скачок в создании информационных систем (ИС) произошел с появлением персональных компьютеров в начале 90-х годов, когда появились системы (например, [4]), работающие на клиентском месте. На следующем этапе, когда был создан доступный в Интернете информационный ресурс (<http://spectra.iao.ru>) [5], опирающийся на БД HITRAN и GEISA, Интернет-технологии позволили сделать значительный шаг в развитии информационно-вычислительных систем (ИВС) коллективного использования по молекулярной спектроскопии.

В последние три года произошло переосмысление принципов создания ИВС как в целом в информатике [6], так и в частности в молекулярной спектроскопии [7]. Стала явной необходимость создания информационных ресурсов, содержащих не только параметры спектральных линий, но и ряд других физических величин, связанных с молекулярной спектроскопией. Структура этих данных описана ниже. Более того, потребовалось расширение набора метаданных, используемого в БД HITRAN и GEISA. Стало очевидным, что для коллективной работы в ИВС необходимо предоставить пользователю возможность самостоятельного формирования структуры массивов спектральных данных и их наполнения конкретными значениями, проведения на их основе расчетов и сравнения результатов с наборами экспериментальных данных. Наша работа в этом направлении состоит, в первую очередь, в создании системы ввода данных произвольной структуры.

Стоит отметить, что современный подход к анализу структуры ИВС с доступом в сети Интернет выделяет три ее уровня: уровень данных и вычислений, информационный уровень и уровень знаний. Такой подход тесно связан с трактовкой информационного ресурса как связанных между собой данных и метаданных. Работа с метаданными относится к информационному уровню [6].

Данные — это критический ресурс, который необходимо сохранять, защищать и организовывать доступ к ним для всех пользователей и в любое время суток. Приемлемая и эффективная стратегия хранения данных основана на использовании распределенных архивов, управляемых теми, кто обладает предметным знанием о данных [8]. Следует отметить также, что в силу доступности сети Интернет значительный импульс получила коллективная работа с данными и знаниями. Одной из форм такой работы являются распределенные информационно-вычислительные системы (РИВС). В статье представлена организация обмена метаданными в РИВС “Молекулярная спектроскопия”.

Построение современных ИВС предусматривает создание ресурса как самоописывающегося документа, т.е. документа, содержащего данные наряду с метаданными, формируемыми в процессе создания документа пользователем. В таких системах значительно возрастает роль метаданных, которые, с одной стороны, упрощают работу по внутреннему обмену данными в ИВС, с другой стороны, обмен метаданными в распределенных системах обладает большей текучестью, а значит, облегчает работу пользователя по поиску ресурсов в специализированной сети. Кроме того, метаданные, становясь доступными поисковым машинам, позволяют проводить семантический разбор информации в глобальной сети. Последнее обстоятельство важно в связи с тем, что большинство существующих в сети Интернет информационных систем являются документно-ориентированными и они слабо обеспечены метаданными. Так, в Web сейчас находятся петабайты информации, а

поиск информации крайне неэффективен или просто невозможен (как правило, доступен контекстный поиск). Результаты обработки пользовательских запросов поисковыми машинами характеризуются высоким уровнем информационного шума [9]. Более того, большинство ресурсов, а это информационные ресурсы “скрытого” Web (около 70 % всех ресурсов), остается с точки зрения организации их структуры и семантики “черным ящиком”. Пользователи сети Интернет обладают весьма ограниченными средствами доступа к таким ресурсам. Их структура и содержание им неизвестны, отсутствуют возможности навигационного доступа, доступа с помощью поисковых машин. Наиболее “закрыты” среди информационных ресурсов базы данных, интегрированные в сайты [9]. Появление в течение последних пяти лет ряда открытых стандартов W3C [10, 11] по описанию структур данных, метаданных и онтологий позволяет существенно изменить ситуацию в этом направлении. Выполненное в данной работе описание метаданных полностью опирается на эти стандарты.

## 1. Структура информационных ресурсов в молекулярной спектроскопии

Данные, относящиеся к молекулярной спектроскопии, можно разделить на три группы: фундаментальные характеристики молекул, параметры спектральных линий и значения спектральных функций.

Теоретическое изучение спектров основано на знании фундаментальных характеристик молекул. С их помощью рассчитываются параметры спектральных линий, применяемые для нахождения спектральных функций, например коэффициента поглощения.

Экспериментальное изучение параметров спектральных линий и спектральных функций всегда было задачей спектроскопии. Ясно, что даже одна и та же физическая измеренная и рассчитанная величина будет обладать разным набором метаданных. Например, для экспериментальных данных ключевыми являются разрешение и точность измерения величины, тогда как для расчетных важны происхождение исходных параметров, на основе которых проводится расчет, методы и физические модели, используемые в расчете.

Среди выделенных трех групп данных основной объем данных связан с параметрами спектральных линий молекул. Для атмосферных наук, в частности, наиболее важны спектральные данные о почти полусотне молекул. Отметим, что структура данных, присущая БД HITRAN и GEISA, ориентирована на решение прикладных задач, и набор физических величин в них соответствует области их применения — спектроскопии атмосферных газов.

### 1.1. Структуры данных

Анализ предметной области показывает, что в молекулярной спектроскопии существуют следующие смежные структуры данных.

1. Фундаментальные характеристики молекулы, определяющие энергию молекулы. В зависимости от способа описания ими могут быть либо параметры полного молекулярного гамильтониана (потенциальная энергия, дипольный момент), либо параметры эффективного гамильтониана (вращательные, центробежные и резонансные постоянные, параметры эффективного дипольного момента). К ним необходимо добавить квадрупольные, октупольные моменты молекул и другие параметры, характеризующие межмолекулярное взаимодействие в газах.

2. Параметры спектральных линий, распределенные по группам:

— параметры изолированной спектральной линии (интенсивность, центр линии, энергия нижнего уровня, статистический вес верхнего и нижнего состояний, момент перехода и т. д.);

— параметры идентификации (колебательная и колебательно-вращательная идентификация);

— параметры, обусловленные столкновениями (полуширина, сдвиг давлением, температурная зависимость полуширины и т. д.).

3. Спектральные функции, к которым относятся коэффициент поглощения, функция пропускания, сечение поглощения и т. д.

Существующая иерархия этих структур данных является структурой типа матрешки. Фундаментальные характеристики молекул необходимы для расчета данных более высокого порядка, а именно параметров спектральных линий, а они, в свою очередь, для расчета спектральных функций. Очевидно, что при формировании самоописывающегося документа структура метаданных также обладает структурой типа матрешки. Ниже детально описаны только параметры спектральных линий.

## 1.2. Параметры спектральных линий

### 1.2.1. Физические величины и их атрибуты

За основу набора физических величин, используемого в качестве параметров спектральной линии, нами взят набор из банка данных HITRAN. Мы расширили его, дополнив полуширинами, обусловленными столкновениями с рядом молекул (Ar, Ne, H<sub>2</sub>O и т. д.), а также ввели ряд атрибутов у таких величин, как центр линии, интенсивность и полуширина. К числу этих атрибутов относятся способ получения физической величины (экспериментальный, расчетный или синтетический), ее шкала значений (абсолютная или относительная) и величина ошибки (класс ошибки или абсолютная величина). Подход, используемый в БД HITRAN, не предусматривает явного задания этих атрибутов, но, например, неявно информацию о способе получения физической величины можно извлечь из библиографии, размещенной в нем. Полный список параметров спектральных линий и их атрибутов представлен ниже.

Параметры спектральной линии изолированной молекулы:

— центр линии (экспериментальный, расчетный или синтетический);

— интенсивность (экспериментальная, расчетная или синтетическая, абсолютная или относительная);

— энергия нижнего уровня;

— статистический вес нижнего уровня;

— идентификация.

Параметры спектральной линии неизолированной молекулы (газ):

— сдвиг давлением;

— температурная зависимость полуширины;

— полуширина, обусловленная самоуширением, уширением другими веществами или их смесями (экспериментальная, расчетная или синтетическая).

Сопутствующие данные:

— библиографическая ссылка для каждой линии;

— точность (класс ошибки или абсолютная ее величина) для центра линии, интенсивности и уширения.

### 1.2.2. Идентификация

Отдельно следует выделить идентификацию спектральной линии. Этот параметр представляет собой набор символов, уникально идентифицирующих каждую спектральную линию данной молекулы или изотопомера. По своему происхождению спектральная линия связана с переходом молекулы из одного состояния в другое. Состояние молекулы характеризуется уровнем энергии, который, в свою очередь, можно связать с колебательными и колебательно-вращательными движениями атомов в молекуле. Каждый уровень энергии идентифицируется двумя наборами символов, один из которых определяет тип колебаний атомов в молекуле, а другой — тип вращательного движения атомов в молекуле. В силу того, что спектральная линия связана с двумя состояниями молекулы, для ее идентификации используются четыре группы квантовых чисел: квантовые числа верхнего и нижнего колебательных состояний и квантовые числа верхнего и нижнего вращательных состояний. Число квантовых чисел, необходимое для идентификации колебательного и вращательного уровней, зависит от группы симметрии молекулы. Отметим, что в зависимости от физических условий молекула может менять симметрию. В нашей работе для идентификации колебательного или вращательного уровней молекулы могут быть использованы до двенадцати параметров, каждый из которых содержит до двух символов (колебания) или до трех символов (вращения). В таблице приведены квантовые числа

The screenshot shows the 'Atmospheric Spectroscopy' software interface. At the top, there are navigation tabs: 'Введение', 'База данных', 'Молекула', 'Спектры газов', 'Измерения', and 'Информация'. The 'Молекула' tab is active. Below the tabs, there is a search bar with 'spectr » molecule' and a language selector 'Rus | Eng'. A sidebar on the left contains a tree view for 'Описание вещества' with options like 'Спектроскопическая модель', 'Группа симметрии', 'Класс молекулы', 'Описание вещества', and 'Добавление описания'. The main area displays a table with 20 rows of molecular data. The table has columns for 'N', 'Вещество', 'Спектроскопическая модель', 'Группа симметрии', 'Класс для молекулы', 'Поля колебательной идентификации', and 'Поля вращательной идентификации'. Below the table, there are controls for 'Показать 10 рядов от 10' and 'Всего 79' items, along with navigation arrows and a 'Настройки' button.

N	Вещество	Спектроскопическая модель	Группа симметрии	Класс для молекулы	Поля колебательной идентификации	Поля вращательной идентификации
11	CO	Двухатомные и линейные молекулы с целым J	$C_{infv}$	Двухатомные молекулы	$\nu_1$	*, Br, J
12	CH <sub>4</sub>	Сферические волчки	$T_d$	Пятиатомные и многоатомные молекулы	$\nu_1, \nu_2, \nu_3, \nu$	J, K <sub>a</sub> , K <sub>c</sub>
13	O <sub>2</sub>	Триплетные сигма основные электронные состояния	$D_{infh}$	Двухатомные молекулы с разными электронными уровнями	$X\nu_1$	*, Br, F, Br, N, Br, J, Sym
14	NO	Дублетные П основные состояния (полуцелое J и полуцелое F)	$C_{infv}$	Двухатомные молекулы	$Xi\nu_1$	, F', F
15	NO <sub>2</sub>	Асимметричные волчки	$C_{2v}$	Нелинейные трехатомные молекулы		
16	NH <sub>3</sub>	Симметричные волчки	$C_{3v}$	Пирамидальные четырехатомные молекулы	$\nu_1, \nu_2, \nu_3$	J, K <sub>a</sub> , K <sub>c</sub>
17	HNO <sub>3</sub>	Асимметричные волчки	$C_s$	Пятиатомные и многоатомные молекулы		
18	OH	Дублетные П основные состояния (полуцелое J и целое F)	$C_{infv}$	Двухатомные молекулы с П-дублетом электронного состояния	$X_i, \nu_1$	Br, J, Sym
19	HF	Двухатомные и линейные молекулы с целым J	$C_{infv}$	Двухатомные молекулы		
20	HCl	Двухатомные и линейные молекулы с целым J	$C_{infv}$	Двухатомные молекулы	$\nu_1$	Br, F

Рис. 1. Интерфейс для работы с описанием молекул, необходимым при обработке ввода идентифицированных линий молекулы.

Пример параметров идентификации для молекул различных групп симметрии

Группа симметрии	Идентификация колебательных состояний	Идентификация вращательных состояний
$C_{2v}$	$V'_1 V'_2 V'_3$	$J' K_a K_c$
$D_{\infty h}$	$V'_1 V'_2 l'_2 V'_3 r'$	$B_r J'' Sym$
$C_{\infty v}$	$V'_1$	$B_r J''$
$C_{3v}$	$V'_1 V'_2 l'_2 V'_3 l'_3 V'_4 l'_4 l'$	$J' K_a K_c$

для нескольких групп симметрии, используемые для идентификации уровней молекул. В настоящее время в ИВС включены молекулы тринадцати групп симметрии.

На рис. 1 показана таблица, в которой для каждой молекулы, спектр которой содержится в ИВС, определены спектроскопическая модель, класс, группа симметрии и квантовые числа для ее колебательного и вращательного состояний. На основании этой таблицы при вводе пользователем идентифицированных уровней молекулы осуществляется проверка корректности ввода данных. Подробнее процедура ввода данных пользователем описана ниже.

### 1.2.3. Подсистема ввода данных. Параметры спектральных линий

Ввод данных пользователем осуществляется в несколько этапов (рис. 2). На первом этапе создается описание данных, включающее ряд стандартных метаданных для описания ресурсов и термодинамические условия, при которых получен массив чисел, вводимый пользователем. На втором этапе пользователь формирует структуру данных для ввода. Обязательными элементами этой структуры являются центры линий и их интенсивности. На третьем этапе пользователь указывает значения атрибутов, выбранных им для ввода физических величин. Входные данные пользователь организует в виде набора колонок в файле. Описание последовательности колонок в файле он определяет на втором этапе ввода данных. После загрузки файла на сервер он преобразуется в XML-документ, данные которого проверяются на соответствие значений ограничениям XML-схемы. Корректный документ подлежит дальнейшей обработке и размещению в БД. Ресурс, введенный поль-

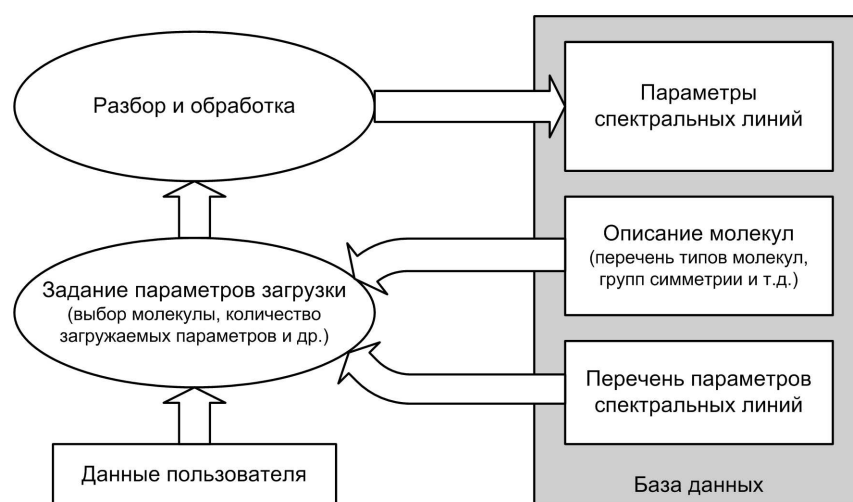


Рис. 2. Общая схема загрузки данных.

зователем, доступен только пользователю.

В настоящее время операции с метаданными для БД и со XML-схемой не связаны между собой. Схема создается вручную, а для создания перечня параметров спектральных линий и их атрибутов создан web-интерфейс.

Подсистема ввода данных имеет ряд ограничений. Во-первых, за один раз можно ввести данные только по одной молекуле. Во-вторых, поддерживаются только два типа структуры данных в файле: “колоночный” файл, в котором строка разбита на колонки строго фиксированной ширины и нет разделителей между ними, и CSV, в котором ширина колонки не фиксирована и требуются разделители между колонками. В-третьих, порядок квантовых чисел при вводе данных об идентификации молекул жестко фиксирован. После загрузки данные становятся доступными для выполнения с ними любых операций в ИВС.

## 2. Метаданные в молекулярной спектроскопии

Как отмечено выше, большая часть имеющихся в Интернете информационных систем является документно-ориентированной. В научной вычислительной системе результаты вычислений можно также представлять в виде документов, содержащих метаданные. Выбор набора метаданных определяется задачами, требующими их формирования. Например, для описания документа как абстрактного информационного ресурса ограничиваются форматом Dublin Core (DC) [12], для описания структуры документа используют формат PRISM [13] и т. д. Эти уровни абстракции достаточно далеко от предметной области.

В настоящее время описание метаданных для молекулярной спектроскопии отсутствует. Ниже мы кратко остановимся на описании подхода, использованного нами при формировании таких метаданных на примере метаданных для спектральной функции.

Основой ИВС являются данные и вычисления. Операции с данными проводятся в рамках физической и математической модели предметной области. Ключевым моментом для молекулярной спектроскопии является существование двух способов получения данных: экспериментальный и расчетный. Для каждого из этих способов сформированы метаданные. Метаданные для экспериментальных данных включают описание устройств, условий эксперимента и т. д. Метаданные для расчетных данных тесно связаны с метаданными, используемыми при описании физической и математической моделей молекулярной спектроскопии. Опуская обсуждение вопроса о полноте метаданных предметной области, заметим, что в нашей работе каждый документ обеспечивается двумя типами метаданных: форматным (DC) и предметным. Далее, для краткости, используется термин “аннотация” для обозначения набора нескольких типов метаданных, связанных с документом. Основное назначение аннотации — это создание возможностей для машинного семантического разбора информационных ресурсов научной ИВС.

### 2.1. Формирование и отображение метаданных

Рассмотрим поэтапную организацию работы с метаданными в распределенной системе “Молекулярная спектроскопия” на примере расчета коэффициента поглощения: от создания документа и его метаданных до механизма обмена метаданными.

Физическая и математическая модели, используемые для расчета коэффициента поглощения, определяют следующие функции для программной реализации:

- ввод и регистрацию данных по коэффициенту поглощения;
- обеспечение средств для расчета коэффициента поглощения с возможностью сравнения полученных результатов с результатами других источников;
- хранение информационных ресурсов в формате XML, а их аннотаций — в виде RDF-описаний в общем реестре;
- обмен аннотациями по коэффициенту поглощения в РИВС “Молекулярная спектроскопия”.

В РИВС “Молекулярная спектроскопия” созданы следующие модули работы с информационными ресурсами по коэффициенту поглощения:

- модуль формирования данных и метаданных по коэффициенту поглощения;
- модуль расчета коэффициента поглощения;
- модуль сравнения рассчитанных данных с результатами других источников;
- модуль отображения данных;
- модуль отображения метаданных;
- модуль поддержки обновлений аннотаций в РИВС;
- сборщик “мусора” в реестре аннотаций и БД;
- модуль обмена аннотациями в РИВС.

Структура информационного ресурса, содержащего экспериментальные данные по коэффициенту поглощения, представлена на рис. 3.

Работа с метаданными по коэффициенту поглощения для экспериментальных данных состоит из двух этапов. Первый этап включает в себя формирование метаданных на основе опубликованного описания эксперимента. Второй этап состоит в формировании метаданных для результатов научного эксперимента. Запись аннотации в виде RDF-документа проводится на втором этапе. Проверка корректности вводимых данных выполняется по типовой схеме: данные пользователя преобразуются в XML-документ, и при разборе используется соответствующая XML-схема. После разбора данные заносятся в БД.

При формировании RDF-документа по схеме Dublin Core используются данные о пользователе и ряд технических данных, имеющих в РИВС. Метаданные для ресурса (DC) и метаданные для коэффициента поглощения формируют аннотацию ресурса. Аннотации ресурсов собраны в РИВС в реестре аннотаций.

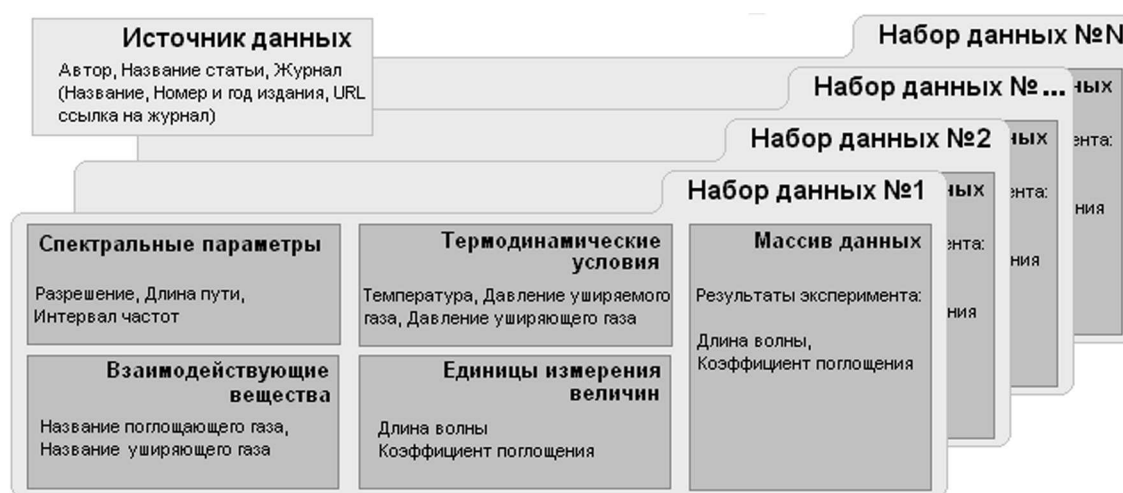


Рис. 3. Метаданные для описания экспериментальных данных по коэффициенту поглощения.



В РИВС метаданные для коэффициента поглощения, измеренного в эксперименте, формируются вручную, т. е. через формы вносятся пользователем или администратором. Метаданные для коэффициента поглощения, рассчитанного в РИВС, формируются автоматически при работе пользователя в диалоговой системе. На рис. 4 показан ряд форм, используемых в диалоговой системе. К числу метаданных для рассчитанного коэффици-

**а Выберите вещество** ( Выбрано: CO<sub>2</sub> )  
 Выбрать изотоп  
 Показать [20] рядов от [0] Всего 37 > >>

N	Выбор	Вещество	Название
1	<input type="radio"/>	H <sub>2</sub> O	Вода
2	<input checked="" type="radio"/>	CO <sub>2</sub>	Углекислый газ (диоксид углерода)
3	<input type="radio"/>	SO <sub>2</sub>	Двуокись серы
4	<input type="radio"/>	O <sub>3</sub>	Озон
5	<input type="radio"/>	N <sub>2</sub> O	Закись азота
6	<input type="radio"/>	CO	Угарный газ (оксид углерода)
7	<input type="radio"/>	CH <sub>4</sub>	Метан
8	<input type="radio"/>	O <sub>2</sub>	Молекулярный кислород
9	<input type="radio"/>	NO	Оксид азота
10	<input type="radio"/>	NO <sub>2</sub>	Двуокись азота
11	<input type="radio"/>	NH <sub>3</sub>	Аммиак
12	<input type="radio"/>	HNO <sub>3</sub>	Азотная кислота
13	<input type="radio"/>	OH	Гидроксил
14	<input type="radio"/>	HF	Фтористоводородная кислота
15	<input type="radio"/>	HCl	Хлористый водород
16	<input type="radio"/>	HBr	Бромистоводородная кислота
17	<input type="radio"/>	HI	Йодная кислота

**б Задайте параметры**

Спектральный диапазон (см <sup>-1</sup> )	[3400] - [3500]
Температура (°K)	[300]
Давление	[1]
Интенсивность	Значения <input type="button" value="больше"/> 1e-34

**в Выберите источники**  
 Показать [20] рядов от [0] Всего 1

N	Выбор	Название	Число линий в заданном диапазоне	Число линий в полосах
1	<input type="checkbox"/>	NITRAN (test)	425	2042

**г**

Величина обрывания контура (см <sup>-1</sup> ):	[100]
Тип контура:	<input type="button" value="Контур Лорентца"/>
<input type="radio"/> Фактор разбиения:	[ ]
<input checked="" type="radio"/> Число разбиений:	[500]
<input type="radio"/> Загрузить файл:	[ ] <input type="button" value="Browse..."/>

Сохранить результат

Рис. 4. Расчет коэффициента поглощения: *а* — выбор вещества; *б* — задание спектрального интервала и термодинамических параметров; *в* — выбор источника данных для параметров спектральных линий; *г* — задание количества точек расчета, типа контура спектральной линии и величины обрезания контура.

**Аннотация** DublinCore XML

Вещество		Термодинамические условия	
Поглощающий газ	CO <sub>2</sub>	Температура (°K)	613
Уширяющий газ	Ar	Давление (atm)	0.0000011-59
<b>Массив данных</b>		Давление уширяющего газа (atm)	0.0000011-48
Частоты перехода (число значений) (см <sup>-1</sup> )	7	<b>Спектральные параметры</b>	
Коэффициент поглощения (число значений) (Δm <sup>-2</sup> см <sup>-1</sup> , exp)	7	Разрешение (см <sup>-1</sup> )	0
Ошибки	No	Длина пути (m)	4.4
<b>Литература</b>		Интервал частот (см <sup>-1</sup> )	2400-2450
Авторы	Boissoles J., Menoux V., LeDoucen R., Boulet C., Robert D		
Название статьи	Collisionally induced population transfer effect in infrared absorption spectra. III. Temperature dependence of absorption in the Ar-broadened wing of CO <sub>2</sub> nu <sub>3</sub> band		
Журнал, год издания, том, страницы	J.Chem.Phys. 93, No.4, 2217-2221 (1990)		
Комментарий	Полное давление, up to 60 bar Давление поглощающего газа 20% CO <sub>2</sub> + 80% Ar		

Рис. 5. Представление аннотации для коэффициента поглощения.

ента поглощения относятся спектральный диапазон, температура, давление, ограничение на интенсивность линии, тип контура, величина обрывания контура, способ разбиения спектрального интервала при расчете и т. д.

Результат расчета коэффициента поглощения может быть представлен в виде html-страницы, содержащей его табличное или графическое представление, а также ссылки на аннотацию для этого информационного ресурса. На рис. 5 показано табличное представление аннотации для коэффициента поглощения, измеренного в эксперименте. При отображении аннотации формируются ссылки на метаданные по схеме DC и XML-документ, содержащий данные эксперимента.

## 2.2. Работа с метаданными в распределенной ИВС

В ИВС в зависимости от способа обмена данными выделяют два типа узлов: базовые узлы и узлы-поставщики. Между базовыми узлами происходит обмен ресурсами, а между узлами-поставщиками и базовыми узлами — обмен аннотациями ресурсов. Базовые узлы этой системы расположены в Томске, Санкт-Петербурге, Нижнем Новгороде и Москве. В настоящее время функционируют два базовых узла с адресами <http://saga.atmos.iao.ru/> и <http://saga.molsp.phys.spbu.ru/>. Создание узлов-поставщиков ориентировано на организации, являющиеся собственниками информационных ресурсов в области молекулярной спектроскопии, например кафедры вузов. Они обеспечиваются типовым программным обеспечением для поддержки узла-поставщика.

Для работы с аннотациями в ИВС используются реестры аннотаций, которые по сути являются хранилищем web-контента. На практике для структурирования такого хранилища используются XML-стандарты агрегирования данных RSS и Atom [14, 15]. В нашей работе используется типовой для RDF контейнер Bag.

На рис. 6, а схематически представлена структура реестра аннотаций в ИВС “Молекулярная спектроскопия”. Сводный реестр аннотаций узла в распределенной системе

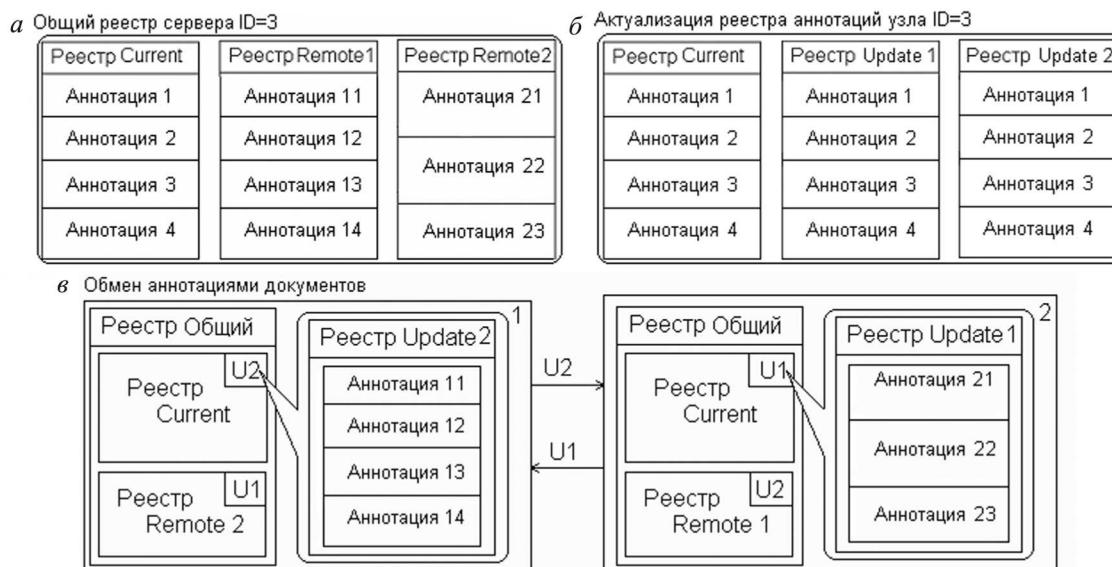


Рис. 6. Реестр аннотаций (Current, Remote) в ИВС (а); реестры аннотаций, актуализирующиеся при работе с документом (Update) (б); обмен аннотациями документов между двумя узлами ИВС (в).

состоит из реестра аннотаций данного узла (Current) и реестров аннотаций удаленных узлов (Remote) распределенной системы. Для администратора узла они различаются правами доступа. Для реестра аннотаций данного узла возможна полноценная работа с XML-документами и их аннотациями (загрузка, редактирование). Для реестров удаленных узлов возможен только просмотр XML-документов и их аннотаций.

Модуль поддержки обновлений аннотаций документов в РИВС используется для формирования на текущем узле реестров обновлений (Update N) для всех удаленных базовых узлов РИВС. Эти реестры обновлений содержат актуализированные аннотации данного узла (рис. 6, б).

Модуль обмена аннотациями необходим для синхронизации содержания реестров базовых узлов в РИВС. Синхронизация происходит путем обмена содержимым реестров обновлений. Обмен содержимым реестров обновлений происходит по pull технологии, т. е. по запросу с удаленного базового узла ему передается предназначенный для него реестр обновлений (рис. 6, в).

Программное обеспечение для формирования данных и аннотаций написано на PHP с использованием типовых модулей для работы с MySQL и XML, а также RAP (RDF API for PHP). При создании модулей работы с реестром аннотаций применялась технология сервлетов, использующая API Jena и JDOM [16–18]. Генерация HTML-страниц из XML-документов осуществляется с помощью XSLT-преобразований. Для обмена аннотациями используется технология Web-services, а именно подсистема SOAP на Java-Axis [19].

## Заключение

Описана структура информационных ресурсов создаваемой распределенной информационно-вычислительной системы “Молекулярная спектроскопия”. Предложена структура данных для молекулярной спектроскопии, на примере параметров спектральных линий показан механизм ввода данных пользователя в РИВС, описаны метаданные для коэффициента поглощения и процедура обмена метаданными.

Авторы благодарят чл.-корр. РАН С.Д. Творогова, А.Д. Быкова и Б.А. Воронина за консультации и помощь при определении структуры данных в молекулярной спектроскопии и Н.А. Лаврентьева за реализацию программ для расчета коэффициентов поглощения газов.

## Список литературы

- [1] HITRAN, <http://www.hitran.com>.
- [2] JACQUINET-HUSSON N., ARIE E. ET AL. The 1997 spectroscopic GEISA databank // JQSRT. 1999. Vol. 62. P. 205–254. (<http://www.ara.polytechnique.fr>).
- [3] ВОЙЦЕХОВСКАЯ О.К., МАКУШКИН Ю.С., ПОПКОВ А.И. и др. Структура и принципы реализации подсистемы формирования банка параметров спектральных данных // Тез. докл. 6-го Всесоюз. симп. по молекулярной спектроскопии высокого и сверхвысокого разрешения. Томск, 1982. Ч. 2. С. 42–44.
- [4] GOLOVKO V.F., NIKITIN A.V., CHURSIN A.A., TYUTEREV VL.G. Information system AIRSENTRY for modeling atmospheric IR-spectra and radiation transmission in the atmosphere // Proc. 2nd Intern. Workshop. ADBIS'95. M., 1995. Vol. 2. P. 12–14.

- [5] БАБИКОВ Ю.Л., БАРЬ А., ГОЛОВКО В.Ф., ТЮТЕРЕВ В.Г. Интернет-коллекции по молекулярной спектроскопии // Тр. 3-й Всерос. конф. по электронным библиотекам. Петрозаводск, 2001. С. 183–187.
- [6] DE ROURE D., JENNINGS N., SHADVOLT N. A Future e-Science Infrastructure, Report commissioned for EPSRC/DTI Core e-Science Programme, 2001. 78 p.
- [7] РОДИМОВА О.Б., ТВОРОГОВ С.Д., ФАЗЛИЕВ А.З. Онтология по молекулярной спектроскопии атмосферных газов // Тр 5-й Всерос. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. СПб., 29–31 октября 2003. С. 211–215.
- [8] PRESERVING Scientific Data On Our Physical Universe, National Academy Press, Washington, D.C., 1995.
- [9] КОГАЛОВСКИЙ М.Р. Перспективные технологии информационных систем. М.: ДМК Пресс; Компания АйТи, 2003. 288 с.
- [10] RDF/XML Syntax Specification (Revised), W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210>
- [11] OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation 10 February 2004. <http://www.w3.org/TR/2004/REC-owl-semantics-20040210/>
- [12] THE Dublin Core Metadata Initiative. <http://dublincore.org/>
- [13] PUBLISHING Requirements for Industry Standard Metadata. <http://www.prismstandard.org/>
- [14] RSS Specification. <http://blogs.law.harvard.edu/tech/rss>
- [15] THE Atom Project. <http://www.intertwingly.net/wiki/pie/FrontPage>
- [16] МАК-ЛАХЛИН Б. Java и XML. СПб.: Символ-Плюс, 2002. 544 с.
- [17] ХАБИБУЛЛИН И.Ш. Разработка Web-служб средствами Java. СПб.: БХВ-Петербург, 2003. 400 с.
- [18] ХОЛЛ М. Сервлеты и Java Server Pages. СПб.: Питер, 2001. 496 с.
- [19] THE Apache Software Foundation. <http://ws.apache.org/axis/>

*Поступила в редакцию 18 марта 2005 г.*