

СТРУКТУРА РЕСУРСОВ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ ПО МОЛЕКУЛЯРНОЙ СПЕКТРОСКОПИИ

Козодоев А.В.¹, Привезенцев А.И.^{1,2}, Фазлиев А.З.¹

1. *Институт оптики атмосферы СО РАН,
634055, г. Томск, пр. Академический, 1, kav@iao.ru, faz@iao.ru*

2. *Томский государственный университет систем управления и радиоэлектроники,
634050, г. Томск, пр. Ленина, 40, remake@iao.ru*

В работе представлен подход к формированию структуры ресурсов информационно вычислительной системы (ИВС) по молекулярной спектроскопии, основанный на анализе иерархии задач молекулярной спектроскопии, задач подготовки данных и задач прикладных предметных областей. Задачи, на решение которых ориентирована ИВС, определяют структуру ресурсов в соответствии с тремя функциональными уровнями: загрузки ресурсов в информационную систему, их хранения и представления ресурсов пользователям

1. Введение

Молекулярная спектроскопия является одним из ключевых разделов оптики, используемых во многих прикладных исследованиях. Информационные системы, ориентированные на работу со спектральными ресурсами, в настоящее время содержат данные о миллионах спектральных линий. Среди данных выделяются экспертные данные о параметрах спектральных линий в банках данных Nitran [1] и Geisa [2], включающие в себя около двух миллионов спектральных линий.

Данные, используемые для решения прикладных задач, в том числе и задач молекулярной спектроскопии, имеют существенные отличия, как по структуре, так и по объёму. Структура данных, используемая в банках данных [1,2], во многом определяет стандартный формат файлов, используемый для обмена между специалистами, использующими спектральные ресурсы. Однако, на наш взгляд, сложившаяся структура данных и ее форма уже не удовлетворяют потребностям пользователей. Данные в этой структуре смешаны с метаданными, что выражается в наличие ссылок на литературу и классов ошибок для физических величин в строке данных. Однако хранение данных в виде файла приводит к резкому сокращению объема спектральных линий в банке данных. В прошлом такой подход был оправдан небольшим объемом данных о параметрах спектральных линий (ПСЛ). В настоящее время, когда рассчитываются ПСЛ для сотни миллионов линий, такая форма хранения данных и принципы работы с ними должны быть изменены. Подходу к решению этой проблемы посвящена данная работа.

Коренное отличие предлагаемого подхода состоит в замене концепции банка данных, используемой в работах [1,2], на концепцию информационно-вычислительной системы (ИВС). Во-первых, в новом подходе понятие данные заменяется понятием ресурсы, которое трактуется как структура, связывающая данные и метаданные, и позволяющая проводить машинную обработку своих составных частей. Во-вторых, приложения, соответствующие задачам молекулярной спектроскопии, и существовавшие ранее отдельно от данных, в ИВС интегрируются в единую систему, в рамках которых происходит динамическое формирование метаданных при выполнении приложений. Задачи построения структуры данных и метаданных, а также описание типов и способов связей между ними, выходят на первый план при концептуальном моделировании ИВС по молекулярной спектроскопии и ее практической реализации.

Построение структуры ресурсов ИВС опирается в нашем подходе на анализ иерархий задач. Эти задачи относятся к самой молекулярной спектроскопии, прикладным предметным

областям (в первую очередь атмосферным наукам и астрономии) и задачам подготовки данных для выполнения приложений. С функциональной точки зрения задачи молекулярной спектроскопии целиком определяют структуры ресурсов, используемые для загрузки в ИВС, и влияют на структуры ресурсов, используемые для хранения в хранилище данных. Задачи подготовки данных обусловлены приложениями, включенными в ИВС, и во многом определяют структуры ресурсов, используемые для хранения и выполнения приложений предметной области. Задачи прикладных областей, в том числе и молекулярной спектроскопии, определяют структуру ресурсов, предоставляемых пользователям.

Во втором параграфе работе дан краткий перечень задач молекулярной спектроскопии и задач подготовки данных. Задачи прикладных предметных областей не обсуждаются. В третьем параграфе представлено описание используемых при загрузке и хранении ресурсов.

2. Иерархии задач

2.1. Задачи молекулярной спектроскопии

Реализация целей, сформулированных во введении, определяется тем, какие ресурсы загружаются пользователями, как они хранятся и используются в ИВС и в каком виде они предоставляются пользователям. Здесь под ресурсами понимаются данные, связанные с аннотациями являющимися наборами тематических метаданных.

Неструктурированный список физических величин, характерный для перечисленных ниже задач молекулярной спектроскопии, представлен ниже.

Вещество, (уровни энергии, частоты перехода, коэффициенты Эйнштейна, параметры контура линий, интенсивности, параметры смешивания линий) (причина, масштаб, неопределенности), статистические веса уровней, законы сохранения (идентификация), ссылки на литературу, мультипольные моменты, и т.д.

Общий подход к классификации задач позволяет разделить все задачи на две группы: прямые и обратные задачи [3-5]. Обратные задачи решаются на основе имеющихся результатов экспериментальных измерений.

К классам атомарных прямых задач относятся следующие классы.

- 1. Задача определения физических характеристик изолированной молекулы (Т1).** Результатом решения этих задач являются вычисленные уровни энергии молекулы, волновые функции, которым соответствуют стационарные состояния молекулы, коэффициенты Эйнштейна и интегралы движения, задающие идентификацию уровней энергии.
- 2. Задача определения параметров спектральной линии изолированной молекулы (Т2).** Результатом решения являются частоты переходов (центры линий). Входными данными для задачи являются уровни энергии, волновые функции и интегралы движения (идентификация).
- 3. Задача определения параметров контура спектральной линии (Т3).** Результатом являются вычисленные полуширины, сдвиги, интенсивности спектральных линий, параметры, характеризующие интерференцию, статистические веса. Входными данными являются частоты переходов, волновые функции, коэффициенты Эйнштейна и др.
- 4. Задача расчета спектральных функций (Т4).** Рассчитываются коэффициенты поглощения, функция пропускания и т.д. при заданных термодинамических и электромагнитных условиях. Входными спектральными данными являются параметры спектральной линии взаимодействующей молекулы.
- 5. Измерения спектральных функций (Е1).** Измеряются спектральные функции. Результатом являются метаданные об условиях проведения эксперимента и значения спектральных функций [4, 5].

Важным фактором является то, что эти задачи образуют иерархию. Например, в простейшем случае, для решения задачи Т3 необходимо иметь решение задачи Т2, или,

иными словами, входные данные задачи Т3 должны включать в себя выходные данные задачи Т2.

К классам атомарных обратных задач относятся следующие классы.

1. **Задача определения параметров спектральной линии взаимодействующей молекулы (ET1).** Входными данными являются измеренные спектральные функции и условия измерения. Результатом решения задачи являются параметры спектральных линий взаимодействующих молекул.
 - 1.1. **Подзадача определения центров спектральных линий (ET1.1).** Результатом решения являются центры спектральных линий (два типа: центры линий, отнесенные к условиям их существования в вакууме, центры линий, отнесенные к конкретным термодинамическим и электромагнитным условиям).
 - 1.2. **Подзадача определения интенсивностей спектральных линий (ET1.2).** Результатом решения являются интенсивности, отнесенные к частотам перехода при заданных термодинамических и электромагнитных условиях.
 - 1.3. **Подзадача определения полуширин, сдвигов и температурных зависимостей полуширин и сдвигов (ET1.3).**
 - 1.4. **Подзадача определения параметров смещения линий (ET1.4).**
 - 1.5. **Подзадача определения коэффициентов Эйнштейна (ET1.5).** Результатом являются коэффициенты Эйнштейна, отнесенные к центрам линий
2. **Задача идентификации спектральных линий (T5).** Результатом является установление связи между центрами спектральных линий и квантовыми числами.
3. **Задача определения уровней энергии изолированной молекулы (T6).** Результатом является список уровней энергии с приписанными к ним квантовыми числами.

В обратных задачах также существует иерархия, приводить которую здесь не будем.

Для описания предметных ресурсов такие иерархии задач позволяют построить онтологию задач предметной области, которая используется в ИВС для формирования базы знаний по молекулярной спектроскопии.

Вторая группа задач обусловлена действиями, связанными с подготовкой ресурсов для решения задач (выполнения приложений).

К классам таких задач, связанным со структурами ресурсов молекулярной спектроскопии, относятся следующие задачи, полный список которых, обусловлен конкретным назначением системы и вряд ли представляется необходимым описывать его здесь в полной мере.

1. **Задача формирования составных источников данных.** Входными данными являются ресурсы хранилища данных, содержащего оригинальные ресурсы. Выходные данные – составные источники информационных ресурсов, относящимся к параметрам спектральных линий, спектральным функциям и уровням энергии.
2. **Задача визуализации метаданных при организации сравнения данных пользователя между собой и данными из хранилища данных.**
3. **Задачи визуализации данных** (параметры спектральных линий, уровни энергии, спектральные функции).
4. **Задача организации связи между данными и метаданными в молекулярной спектроскопии.**

Рамки статьи не позволяют дать детальное описание классов задач относящихся к предметным областям для которых спектральные ресурсы являются входными данными.

3. Структура ресурсов в ИВС по молекулярной спектроскопии

В наших работах [2-5] приведено описание структуры ресурсов в молекулярной спектроскопии для ряда задач и приведены некоторые примеры формирования аннотаций. Используемые при описании схемы данных и метаданных можно найти в Интернете по адресам

- <http://saga.atmos.iao.ru/xml/xsd/2004-absorption-use-ver2.xsd>

- <http://saga.atmos.iao.ru/rdfs/absorption-coefficient.rdfs>

3.1. Загрузка ресурсов

В ИВС коллективного пользования загрузка пользовательских данных является основным способом наполнения системы новыми элементарными данными. Как экспериментальные, так и расчетные данные относящиеся к любому из перечисленных выше классов задач, могут попасть в ИВС только в результате загрузки пользователем. При такой процедуре загрузки отсутствует связь этих данных и другими данными, уже загруженными в ИВС. Загружаемые пользователем данные связываются только со своими аннотациями и образуют элементарные ресурсы. Эти ресурсы характеризуются библиографической ссылкой.

Как правило, данные загружаются в виде. Структуры данных, используемые в файлах, могут не соответствовать структурам данных, используемым для хранения ресурсов. Отметим, что наиболее распространенной структурой данных используемой в загрузочном файле являются колонки, строки с фиксированными позициями данных и деревья, размеченные с помощью языка разметки XML. Существуют и иные способы форматирования спектральных данных [7].

Загружаемые ресурсы содержат данные о фундаментальных характеристиках молекул (задачи T1 и T5), параметрах спектральных линий (задачи T2, T3, T6, ET1.1-1.5) и спектральных функций (T4, E1). Интенсионалы этих данных и структура аннотаций частично описана в [3-5].

После своей загрузки ресурсы должны иметь статус персональных, т.е. быть доступными только пользователю, выполнившему загрузку.

3.2. Хранение ресурсов

На этом уровне формирование структуры ресурсов в ИВС зависит от перечня задач, которые пользователь может решать в ней. Рассмотрим на примере задачи формирования составных ресурсов механизмы изменения структуры ресурсов.

На рис. 1 показана последовательность формирования экспертного информационного ресурса. Каждый пользователь по определенной процедуре может рекомендовать свои элементарные ресурсы, загруженные в ИВС, для публикации. Статус рекомендованного для публикации ресурса означает невозможность его изменения. Рекомендованные к публикации ресурсы становятся доступны экспертам. Отобранные экспертами ресурсы приобретают статус опубликованных, помещаются в хранилище данных и становятся общедоступными.

На основе опубликованных ресурсов все пользователи, в том числе и эксперты, могут формировать составные ресурсы в рамках правил, поддерживаемых в ИВС [8]. Правила должны обеспечивать механизм создания составного ресурса. Формирование хранилища данных решает проблему непрозрачности процедуры формирования ресурсов, имеющих в банках данных Nitran и Geisa. Созданные экспертами составные ресурсы могут также использоваться пользователями для их задач.

Прозрачность формирования ресурсов экспертами обеспечивается с помощью создания соответствующего набора метаданных, характеризующего этот процесс [8].

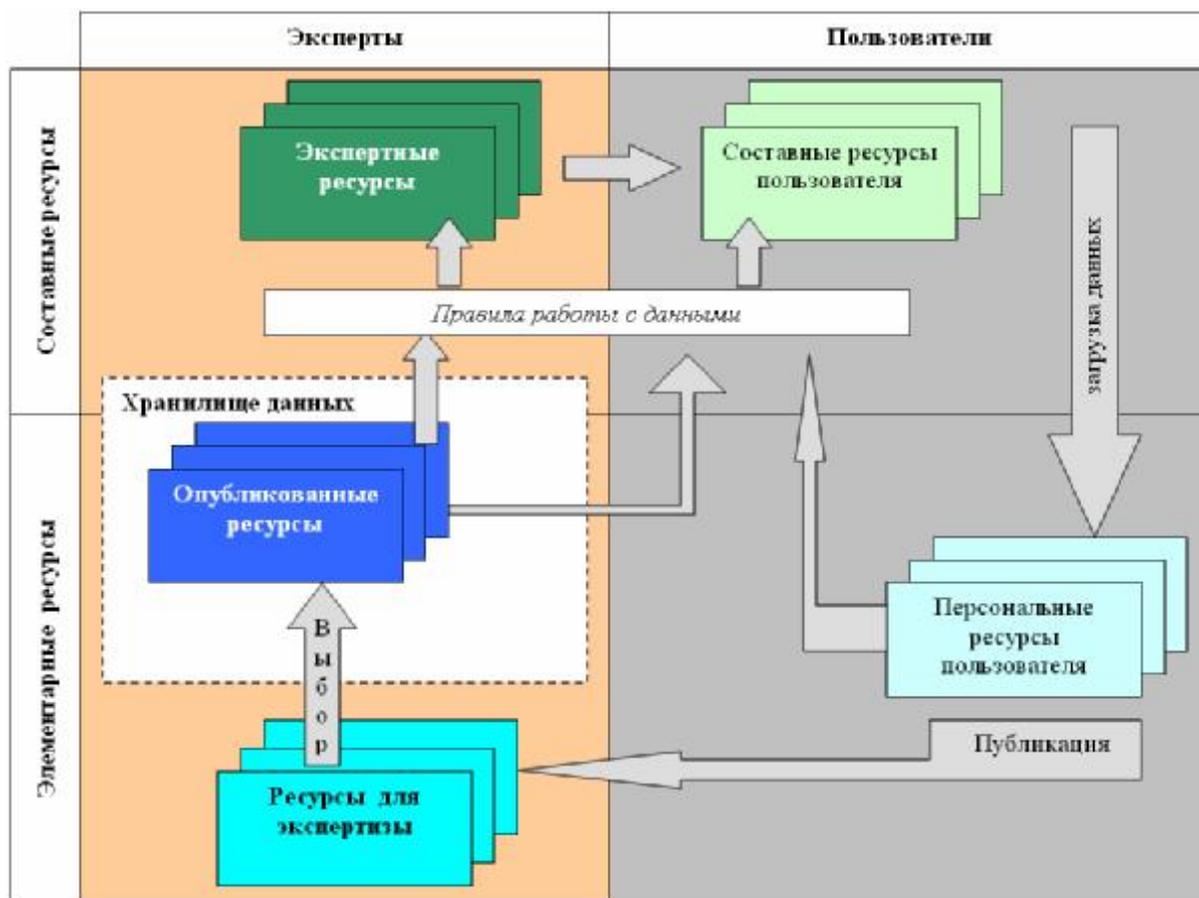


Рис. 1. Схема процессов для работы с данными

4. Заключение

Структура информационных ресурсов в ИВС по молекулярной спектроскопии определяется тремя группами задач: задачами молекулярной спектроскопии, задачами подготовки данных и задачами предметных областей в которых эти ресурсы могут быть использованы [3-6]. Примером ИВС в которой использован предлагаемый подход к формированию ресурсов является созданная в ИОА СО РАН распределенная ИВС (<http://saga.atmos.iao.ru>).

Авторы благодарны РФФИ (грант 05-07-90186), а также Творогову С.Д., Быкову А.Д. и Родимовой О.Б. за помощь в работе и консультации.

Литература

1. L.S. Rothman, D. Jacquemart, A. Barbe, D.Chris Benner, M. Birk, L.R. Brown, M.R. Carleer, C. Chackerian, Jr, K. Chancea, V. Dana, V.M. Devi, J.-M. Flaud, R.R. Gamache, A. Goldman, J.-M. Hartmann, K.W. Jucks, A.G. Maki, J.-Y. Mandin, S.T. Massie, J. Orphal, A. Perrin, C.P. Rinsland, M.A.H. Smith, J. Tennyson, R.N. Tolchenov, R.A. Toth, J. Vander Auwera, P. Varanasi, G. Wagner, The HITRAN 2004 Molecular Spectroscopic Database, J. Quant. Spectrosc. Radiat. Transf., v.96, p.139-204, 2005, (HITRAN, <http://www.hitran.com>).
2. Jacquinet-Husson N., Arie E., et al, The 1997 spectroscopic GEISA databank, JQSRT, v.62, pp. 205-254, 1999. (<http://www.ara.polytechnique.fr>).
3. Быков А.Д., Воронин Б.А., Козодоев А.В., Лаврентьев Н.А., Родимова О.Б., Фазлиев А.З. Информационная система для решения задач молекулярной спектроскопии. 1. Структура информационных ресурсов // Оптика атмосферы и океана, т. 17, № 11, с.921-926, 2004
4. Козодоев А.В. Привезенцев А.И., Фазлиев А.З., Аннотирование информационных ресурсов в распределенной информационной системе "Молекулярная спектроскопия"

- Труды 7 Всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции ”, Ярославль, с.80-86, 2005.
5. Козодоев А.В. Привезенцев А.И., Фазлиев А.З., Организация информационных ресурсов в распределенной информационно-вычислительной системе, ориентированной на решение задач молекулярной спектроскопии Вычислит. технологии, Спец. выпуск, т.10, с.82-91, 2005.
 6. Фазлиев А.З., Описание информационных ресурсов по молекулярной спектроскопии средствами платформы XML, Вычислит. технологии, Спец. выпуск, т.10, ч.1, с.39-46, 2005
 7. R. Lancashire, T. Davies Spectroscopic Data: The Quest for a Universal Format, Chemistry International, Vol. 28, No. 1, 2006,
http://www.iupac.org/publications/ci/2006/2801/3_ref5.html
 8. Козодоев А.В., Фазлиев А.З. Информационная система для решения задач молекулярной спектроскопии. 2. Операции преобразования наборов параметров спектральных линий.// Оптика атмосферы и океана, т. 18, № 09, с.760-764, 2005.

THE STRUCTURE OF THE RESOURCES FOR INFORMATION-COMPUTATIONAL SYSTEM ON MOLECULAR SPECTROSCOPY

A.V. Kozodoev¹, A.I. Privezentsev^{1,2}, A.Z. Fazliev¹

1. Institute of Atmospheric Optics, 1, Akademichesky Avenue, Tomsk 634055, Russia,

2. Tomsk State University of Control Systems and Radio Electronics, Tomsk,

kav@iao.ru, remake@iao.ru, faz@iao.ru.

The work presents an approach to structure the resources on the information-computational system (ICS) on molecular spectroscopy based on the hierarchy analysis of the problems of molecular spectroscopy, the problems of data preparation, and those of applied subject domains. The problems the ICS is supposed to solve determine the structure of the resources in connection with three functional levels: download of the resources to the information system, their storage, and provision to users.