

УДК 004

Информационные ресурсы по спектроскопии в ИОА СО РАН

А.Ю. Ахлестин, С.С. Воронина, Н.А. Лаврентьев, А.З. Фазлиев*

*Институт оптики атмосферы им. В.Е. Зуева СО РАН
634055, г. Томск, пл. Академика Зуева, 1*

Поступила в редакцию 13.03.2015 г.

Информационные ресурсы по спектроскопии составляют наиболее значительную часть ресурсов ИОА СО РАН. Более того, информационные системы Spectra, S&MPO, W@DIS и Hitran on the Web, являющиеся основной частью спектроскопических ресурсов, созданные с помощью Web-технологий, предоставляют широкому кругу исследователей верифицированные данные, информацию, характеризующую свойства этих данных, и знания, позволяющие обрабатывать эту информацию на понятийном уровне. Дано краткое описание этих систем. Наиболее подробно представлена информационная система W@DIS, основным назначением которой является анализ качества спектральных данных с разной степенью детализации.

Ключевые слова: спектроскопия, информационные ресурсы, Spectra, S&MPO, W@DIS, Hitran on the Web; spectroscopy, information resources, Spectra, S&MPO, W@DIS, Hitran on the Web.

Введение

Результаты научных исследований, как правило, представляют в публикациях. Результатами могут быть факты, относящиеся к предметной области, связанные с ними термины и понятия, методы решения задач и алгоритмы, реализующие их. Одним из приемов научных исследований является агрегация данных, информации, знаний и алгоритмов их обработки. Подобные агрегации приводят к созданию информационных систем (ИС). При организации доступа к ним по сети Интернет эти системы становятся частью цифровой науки (e-Science). Инфраструктура цифровой науки детально исследовалась в начале 2000-х гг. [1]. Основой для описания ресурсов в e-Science стали языки разметки, разработанные W3C [2].

Для описания информационных ресурсов в спектроскопии языки разметки начали использовать в середине 2000-х гг. [3, 4]. Например, XML-схемы данных применяли для стандартизации представления данных [5] и для элементарного контроля качества импортируемых в информационные системы спектральных данных [6]. В конце 2000-х гг. европейское сообщество спектроскопистов начало выполнять проект VAMDC [7], в рамках которого для организации доступа к ресурсам распределенной системы разнородных баз данных применялась схема XSAMS [8].

Настоящая статья посвящена описанию развития информационных ресурсов (ИР) по спектроскопии в Институте оптики атмосферы СО РАН (далее Институт) за последние 5 лет и дополняет работу [9]. Не проводя детальный обзор ресурсов, дадим характеристику информационных систем по спектроскопии.

В первом разделе статьи определен термин «информационные ресурсы» [9], описана инфраструктура информационных ресурсов, характерная для цифровой спектроскопии [1].

Во втором разделе кратко представлены иерархия задач [10], используемая нами при систематизации ИР в ИС W@DIS, и схема потоков работ в спектроскопии [11]. Приведено также количество публикаций спектроскопистов Института, распределенное в соответствии с задачами спектроскопии.

В третьем разделе рассмотрены назначение четырех ИС по спектроскопии, а также основные спектроскопические и информационные задачи, решаемые ими. К числу этих задач относятся представление и анализ качества спектральных данных, вычисление спектральных функций.

В заключении обсуждены основные результаты работы и перспективы развития ИС W@DIS.

1. Информационные ресурсы

В научных работах исследователи разных предметных областей используют термин «информационные ресурсы», часто понимая под ним объекты, семантика которых противоречива. Мы используем термин «информационные ресурсы», оп-

* Алексей Юрьевич Ахлестин (lexa@iao.ru); Светлана Станиславовна Воронина (vss@iao.ru); Николай Александрович Лаврентьев (lnick@iao.ru); Александр Зарипович Фазлиев (faz@iao.ru).

ределенный стандартом RFC 2396 [12] – «Ресурсом может быть любая вещь, обладающая индивидуальностью» (A resource can be anything that has identity). Из такого определения, например, следует, что неопубликованные результаты работ, доклады, обсуждения на семинарах и диспуты не являются информационными ресурсами.

Ресурсы могут включать в себя данные, информацию (свойства данных), знания (системы классификации), относящиеся к разным предметным областям, а также программное обеспечение (ПО) для вычислений и манипуляций с данными. Заметим, что одни и те же данные, их свойства и системы их классификации могут иметь разные представления, выбор их определяется прикладными задачами, в которых они используются. Для представления информационных ресурсов используются как общепринятые математические модели, так и оригинальные модели предметных областей.

Стандарт RFC 2396 использован при создании ряда языков спецификации информационных ресурсов, к числу которых относятся XML [13], RDF [14] и OWL 2 DL [15]. Для разметки спектральных информационных ресурсов (СИР) язык XML использовался в проекте VAMDC [8] для формализации структур данных, передаваемых по запросам пользователей, а язык OWL 2 DL применялся в ИС W@DIS для описания свойств данных и создания классификаторов спектральных ресурсов [16]. В настоящей статье мы используем определение ИР в узком смысле, ограничивая его ресурсами сети Интернет. Именно такие web-ресурсы являются основными в информационных системах Института.

Для цифровой науки и, в частности, цифровой спектроскопии принципиальным является представление статической части ресурса в виде структуры: данные, свойства данных (метаданные) и классификаторы (онтологии). Другой значимый компонент информационных ресурсов – это ПО, используемое для функционирования информационных систем. Основными действиями, реализуемыми с помощью ПО, являются вычисления (спектральные функции, качество данных и т.д.) и манипуляции данными (импорт и экспорт данных, унарные операции с данными и т.д.). Программное обеспечение можно разделять по назначению, т.е. для работы исследователя или программного агента, который обрабатывает данные автоматически. Поскольку для работы программного агента требуется более сложное программное обеспечение, то его создание требует значительных усилий и финансовых ресурсов. Положительными чертами использования агента являются существенно более высокий уровень автоматизации работы информационной системы и значительное уменьшение рутинной части работы исследователя.

Базовая проблема информационных ресурсов – это контроль их качества. Для контроля качества используются критерии достоверности данных, оценки доверия к данным и метрики, характеризующие свойства данных и классификаторов.

2. Решения задач количественной спектроскопии

Методы спектроскопических исследований взаимодействия молекул с излучением зависят от того, какие физические процессы являются предметами исследования. Подавляющее число работ, проведенных в Институте, связано с абсорбционной спектроскопией. Эти работы сфокусированы на построении математических моделей молекул, вычислении значений физических характеристик процессов испускания и поглощения, деталях формирования контура спектральных линий и решении проблемы непрерывного поглощения. Результаты исследований в этих направлениях составляют основные СИР, производимые в Институте.

В Институте созданы четыре информационные системы по спектроскопии (Spectra [17], Nitran on the Web [18], S&MPO [19] и W@DIS [20]), доступные в сети Интернет. В основном эти ИС используют статические информационные ресурсы, представляющие собой решения спектроскопических задач. На упрощенной схеме связей между прямыми и обратными задачами абсорбционной спектроскопии [10] (рис. 1) указаны обозначения задач (T1–T7, ET), которые используются при классификации спектроскопических ИР. В созданных ИС наряду со спектроскопическими задачами решаются информационные задачи: контроль качества и свойств данных, классификаторов результатов исследований, согласование данных, построение классификаторов ресурсов и т.д.

Среди перечисленных информационных задач выделим задачу автоматического построения классификатора при импорте данных в информационную систему. Созданный классификатор позволяет ответить на две группы вопросов. К первой группе относятся вопросы качества публикаций, из которых извлечены данные, ко второй – вопросы качества описания опубликованных характеристик состояний и переходов атмосферных молекул. Например, классификатор позволяет быстро найти ответы на вопросы, относящиеся к конкретному состоянию или массиву данных в определенной публикации. В ИС [17–19] используются экспертные или первичные данные и имеются приложения для решения только одной из восьми задач спектроскопии – задачи T4 (см. рис. 1). Информационная система W@DIS [20] при анализе качества данных использует решения задач T1–T3, T5–T7.

На рис. 2 показаны потоки работ, характерные для абсорбционной спектроскопии. В слое данных представлены только те, которые являются решением задач. Направление стрелок указывает на происхождение или использование данных. Стрелка, выходящая из прямоугольника, соответствующего задаче (например, «Глобальная подгонка»), указывает на решение этой задачи, а стрелки, выходящие из прямоугольника «Эталонные (эмпирические) уровни энергии», – на задачи, в которых эти данные используются в качестве входных. Заметим,



Рис. 1. Упрощенная иерархия задач спектроскопии, используемая при систематизации информационных ресурсов



Рис. 2. Упрощенная схема потоков работ с информационными ресурсами в спектроскопии

что все задачи, кроме «Манипуляции данными», алгоритмированы. Отсутствие алгоритма составления экспертных данных вынуждает использовать оценки доверия таким данным, например по критерию опубликования [22].

Основными форматами, в которых создаются СИР в Институте, являются публикации и размещенные на ftp [23] Института файлы, содержащие результаты вычислений параметров спектральных линий. Молекулы, свойства которых изучаются

в Институте, перечислены в табл. 1, которая также содержит распределение публикаций по информационным и спектроскопическим задачам за последние 5 лет.

Таблица 1
Статистика публикаций по спектроскопии
в ИОА СО РАН

Молекула	T1, T7	T2, T6	T3, T5	ИЗ
H ₂ O	5	35	19	3
H ₂ S	2			2
SO ₂	1	1		1
CO ₂	1	25	10	1
CH ₄	2	15	9	
C ₂ H ₂		3		
CO		2	2	
OCS				1
(O ₂) ₂	—	1		
NO ₂		1		1
N ₂ O		9		
PH ₃				1
O ₃		9	3	
O ₂		3		
CH ₃ F	1			
CH ₃ Cl			2	
NH ₃				1
CH ₃ CN			1	

За период 2010–2014 гг. было опубликовано более двухсот работ сотрудников Института. Большая часть этих работ посвящена решению прямых и обратных задач спектроскопии для 18 молекул и их изотопологов. Институт является правообладателем более 10 баз данных по спектральным свойствам атмосферных молекул.

Свыше половины публикаций, вошедших в табл. 1, связаны с обработкой результатов измерений, причем 15% из них представляют собой результаты измерений, выполненных в Институте. К информационным задачам (ИЗ), решенным за этот период, относятся задачи систематизации и анализа качества данных, построения метаданных и онтологий СИР, состояний и переходов молекул.

3. Информационные системы по спектроскопии

Информационные системы по спектроскопии Института предназначены для обработки данных, генерации информации и представления знаний, полученных при исследовании процессов взаимодействия вещества с излучением. Результаты этой обработки применяются в предметных областях, в которых объектом изучения являются как фундаментальные характеристики вещества (уровни энергии, структура вещества и т. д.), так и более сложные физические и химические процессы. При их исследовании спектральные информационные ресурсы используются в качестве входных данных для решения задач.

Основными задачами ИС Spectra и Nitran on the Web являются: представление для исследователей экспертных данных для более чем 50 молекул и некоторых их изотопологов, оригинальных расчетных данных, полученных в ИОА СО РАН, а так-

же вычисление спектральных функций как для однородного по составу газа, так и для смеси газов.

Информационная система S&MPO предлагает инструментальные средства для решения тех же информационных и спектроскопических задач, что и ИС Spectra, но только для молекулы озона и ее изотопологов. Особенностью этой системы является предоставление более детальной информации о решениях спектроскопических задач T1–T2 и математических моделях, используемых в расчетах.

Информационная система W@DIS предназначена для решения задач агрегации спектральных ресурсов, согласования значений идентичных физических величин, анализа качества ресурсов, классификации ресурсов, генерации машинно-обрабатываемой информации с целью построения экспертных массивов спектральной информации для разных прикладных областей.

Все рассмотренные ИС предлагают интерфейсы для работы с ними исследователей, но только ИС W@DIS формирует онтологии для работы программных агентов.

3.1. Представления данных и вычисление спектральных функций

Первые три информационные системы (Spectra, Nitran on the Web, S&MPO) предназначены для решения исследователями двух информационных задач: представления данных и вычисления спектральных функций для процессов поглощения и испускания. Приложения ИС Spectra ориентированы на графическое и табличное представление параметров спектральных линий, аккумулированных в экспертных массивах данных (NITRAN [18], GEISA [24]), результатов расчетов Патриджа–Швенке [22], сечений поглощения (NITRAN [18]), а также вычисление и представление спектральных функций, рассчитанных с помощью экспертных данных.

Спектральные функции можно рассчитывать как для однородного по составу газа, так и для смеси газов. Информационная система Nitran on the Web является клоном ИС Spectra, в котором для тех же целей берутся только экспертные данные NITRAN [18]. В обеих системах при расчетах спектральных функций могут применяться контуры Фойгта, Лоренца и Гаусса и аппаратные функции семи типов. По своему происхождению используемые в этих системах данные являются решениями задач T1–T3, T5–T7 и ET, а также задач вычисления эмпирических уровней энергии и вакуумных волновых чисел (см. рис. 2). Выбор того или иного набора экспертных данных лежит на исследователе. Сложность выбора обусловлена тем, что используемые экспертные и расчетные данные противоречат друг другу. Критерием выбора данных являются доверие к авторам экспертных или расчетных данных или совпадение вычисленных спектральных функций с результатами измерений. Качественные свойства спектральных данных в этих системах приведены в публикациях, а описание количественной стороны анализа качества экспертных данных, как правило, не полное.

Достоинством ИС S&MPO по сравнению с двумя предыдущими ИС является детальное описание молекулы озона в основном электронном состоянии (структура, спектроскопические константы, потенциальные функции, дипольные моменты). Это описание при определенной степени профессиональной подготовки поможет сделать выбор между экспертными данными и данными по озону из этой системы в пользу последних при вычислении спектральных функций.

Использование этих систем эффективно, когда исследователь уверен в достоверности используемых данных и доверяет им в рамках критериев к точности спектральных функций, характерных для прикладной предметной области, в которой они используются. Дополнительным ограничением является то, что при определенных термодинамических условиях для некоторых атмосферных молекул (например, диоксид углерода, метан и т.д.) используемые в ИС Spectra контуры спектральных линий неадекватно описывают эти спектральные линии.

3.2. Анализ качества информационных ресурсов

Решение вычислительных задач существенно зависит от качества входных данных, математической модели молекул и моделей физических про-

цессов. Физические и математические модели накладывают ограничения на решение спектроскопических задач. Примером таких ограничений являются ограничения на квантовые числа состояний и правила отбора для переходов в молекулах.

Назначением ИС W@DIS является предоставление исследователям возможности контроля качества собственных ИР по абсорбционной спектроскопии. Для анализа они могут использовать полные наборы данных для определенных молекул, имеющиеся в ИС. В табл. 2 представлены молекулы и число публикаций, из которых извлечены данные о спектральных характеристиках. В колонках «Число источников данных или публикаций» указано число источников или публикаций, а в круглых скобках – интервал времени опубликования статей, из которых они извлечены. Кроме перечисленных молекул в цифровой библиотеке содержатся данные о сечениях поглощения еще более 200 молекул. В ИС содержится 11 тыс. библиографических ссылок, большая часть из которых связана с источниками данных.

В табл. 2 стехиометрические формулы для молекулярных комплексов представлены в условном виде. Например, формула $(\text{H}_2\text{O})_n$ относится к множеству комплексов, содержащих атомы или молекулы, а также в обязательном порядке молекулу воды n раз.

Таблица 2

Перечень молекул и их изотопологов и число связанных с ними источников данных, извлеченных из публикаций для двух групп задач спектроскопии в ИС W@DIS

Молекула	Число источников данных или публикаций		Молекула	Число источников данных или публикаций	
	ET	T1–T3, T5–T7		ET	T1–T3, T5–T7
H ₂ O	52 (1950–2013)	~750 (1924–2015)	PH ₃		~110 (1933–2014)
CO ₂	38 (1940–2013)	~600 (1931–2015)	SF ₆	15 (1966–2011)	74 (1933–2014)
O ₃	93 (1932–2011)	309 (1926–2015)	H ₂ S	21 (1964–2011)	~110 (1929–2014)
N ₂ O	24 (1949–2011)	331 (1931–2015)	HCOOH	31 (1929–2012)	75 (1938–2012)
CO	15 (1955–1993)	368 (1930–2015)	HOBr	9 (1994–2011)	10 (1989–2010)
CH ₄	31(1950–2007)	~600 (1925–2015)	C ₂ H ₄	21 (1949–2011)	192 (1934–2015)
O ₂	182 (1940–2011)	303 (1924–2015)	CH ₃ OH	10 (1958–2012)	~400 (1928–2015)
NO	20 (1928–1986)	310 (1927–2015)	CH ₃ Br	23 (1951–2011)	89 (1928–2012)
SO ₂	41 (1962–2011)	~220 (1934–2014)	CH ₃ I	29 (1937–2011)	98 (1931–2011)
NO ₂	57 (1934–2011)	220 (1928–2014)	CH ₃ F		149 (1928–2014)
NH ₃	25 (1955–2007)	~250 (1925–2015)	CH ₃ CN	6 (1985–2012)	75 (1935–2015)
HNO ₃	6 (1993–2006)	73 (1952–2012)	HNCO		59 (1950–2007)
OH	3 (1984–1995)	147 (1926–2011)	C ₃ H ₈	8 (1962–2012)	21 (1941–2010)
HF	2 (1981–1985)	123 (1919–2013)	(CH ₃) ₂ CO		36 (1934–2013)
HBr	14 (1935–2011)	212 (1919–2013)	SiH ₄		55 (1934–2012)
HI	18 (1935–2011)	48 (1935–2013)	GeH ₄		37 (1935–2014)
ClO	27 (1950–2011)	37 (1948–2001)	HCl	20 (1948–2011)	198 (1919–2014)
OCS	22 (1966–2011)	140 (1932–2014)	CF ₄	15 (1965–2011)	41 (1952–2012)
H ₂ CO	28 (1959–2012)	236 (1932–2013)	CF ₃ H	5 (1973–2013)	69 (1941–2012)
HOCl	17 (1936–2011)	43 (1951–2009)	(H ₂ O) _n		260 (1968–2014)
N ₂	24 (1952–1999)	240 (1927–2014)	(CO ₂) _n		103(1968–2015)
HCN		178 (1924–2013)	(CH ₄) _n		19 (1976–2013)
CH ₃ Cl	29 (1930–2011)	128 (1928–2014)	(O ₂) _n		44 (1951–2012)
C ₂ H ₂	14 (1953–2011)	~300 (1928–2015)	(N ₂) _n		36 (1965–2012)
C ₂ H ₆	16 (1953–2010)	150 (1928–2015)	Всего 49	1012	~9034

Для анализа качества данных в ИС W@DIS используется набор интерфейсов, предоставляющих пользователю доступ к вычисленным значениям свойств источников данных по разным спектроскопическим задачам и к системе классификации информационных ресурсов по спектроскопии, основанной на результатах анализа качества данных. Для исследователей эти результаты представляются в табличном виде, а для программных агентов в форме OWL-онтологий [25].

Предметами анализа качества являются извлеченные из них источники данных, а также состояния и переходы в молекулах. На рис. 3 показан интерфейс для анализа бинарных отношений между источниками данных, имеющих идентичные переходы, три условных спектра.

Верхний спектр имитирует ситуацию, для которой корректным считается источник данных 1988_MaChFlCa_H₂O_ucl. Порядок следования цветов нарушен, и это означает, что потенциально для некоторых переходов квантовые числа в источнике данных 2008_ToTe_H₂O_ucl приписаны некорректно. Нижний спектр имитирует ситуацию, при которой корректным считается источник данных 2008_ToTe_H₂O_ucl. Средний спектр соответствует естественному порядку следования цветов. Детальное описание интерфейса можно найти в работе [25].

На рис. 3 (цв. вклейка) представлена таблица, содержащая значения среднеквадратического отклонения (СКО) для пары источников данных, значения максимальной разности ($\Delta\omega_{\max}$) среди всех пар идентичных переходов и факторы упорядочения (A_{00}, A_{10}, A_{01}).

При решении задачи классификации ИР используется визуализация парных отношений в форме графа, представляющего, например, значения среднеквадратических отклонений, удовлетворяющих критерию, характеризующему допустимую величину отклонения одним цветом, а недопустимую – другим [26]. Регламентация свойств источников данных, извлеченных из публикаций, описана в работе [27].

Результаты анализа качества данных состояний и переходов молекул актуализируются автоматически при импорте в ИС каждого нового источника данных. Они кодируются на языке спецификации онтологий OWL 2 DL. Исследователь может просмотреть онтологию с помощью редактора Protege. Результаты этого анализа, в свою очередь, также являются информационным ресурсом, качество которого анализируется с помощью метрик онтологии [28]. Некоторые из этих метрик (число индивидов и число высказываний) приведены в табл. 3.

Поясним термины, используемые в табл. 3. Источник данных представляет собой решение спектроскопической задачи, извлеченное из одной статьи. Оно связано с одной молекулой и одним методом решения задачи. Примерами высказываний являются: «квантовое число J равно 4» или в более сложных случаях – «Источник данных 1988_MaChFlCa_H₂O_ucl извлечен из публикации 1988_MaChFlCa». В последнем высказывании значением свойства «извлечен из публикации» является индивид 1988_MaChFlCa_H₂O_ucl, представляющий собой набор высказываний. В данном частном случае для описания индивида 1988_MaChFlCa_H₂O_ucl используются высказывания: «статья 1988_MaChFlCa опубликована в журнале JOSRT», «статья 1988_MaChFlCa имеет авторов J.Y. Mandin, J.P. Chevillard, J.M. Flaud and C. Samy-Peyret» и т.д. Подавляющее число высказываний связано с анализом относительных свойств источников данных (среднеквадратических отклонений, разупорядочения и максимальных разностей между значениями физических величин идентичных состояний и переходов, содержащихся в источниках данных).

В табл. 3 представлено число источников данных, индивидов и высказываний, относящихся к задаче классификации ИР по абсорбционной спектроскопии молекул для некоторых молекул. Детали решения этой информационной задачи опубликованы в [16].

Таблица 3

Количество источников данных, индивидов и высказываний, относящихся к решению задачи классификации публикаций по молекулярной спектроскопии

Молекула	Число источников данных			Число индивидов			Число высказываний		
	T1, T7	T2, T6	T3, T5	T1, T7	T2, T6	T3, T5	T1, T7	T2, T6	T3, T5
H ₂ O	158	520	556	20401	207157	439693	103818	849518	2916767
CO ₂	0	382	457	0	374848	359901	0	1360670	998623
N ₂ O	0	197	68	0	39403	83725	0	190876	491384
SO ₂	14	116	77	174	42026	3935	1000	96341	25542
NH ₃	0	99	0	0	4491	0	0	32323	0
OCS	0	103	47	0	8244	16739	0	43142	92331
C ₂ H ₂	0	40	0	0	2023	0	0	9891	0
PH ₃	0	43	21	0	2023	1592	0	6221	9442
H ₂ S	71	82	10	5231	2176	885	29022	12278	3881
Всего	243	1582	1236	25806	672391	902535	133840	1377260	4537970

В полный набор данных по девяти молекулам входит 3061 источник данных. Эти источники содержат как первичные расчетные и экспериментальные данные, так и экспертные, эталонные и другие типы данных. Формальное описание качества данных для этих молекул содержит 1,6 млн уникальных терминов (индивидов). Эти термины входят в 6 млн высказываний, содержащихся в онтологии ИР по спектроскопии для девяти молекул. Формально каждое высказывание состоит из трех слов. Из сравнения числа слов в сгенерированной онтологии с числом слов 4-томного романа «Война и мир» следует, что онтология ИР по числу слов в 90 раз больше.

В процессе создания ИС W@DIS были систематизированы опубликованные спектральные данные для молекул H₂O [29], H₂S [30], SO₂ [31], CO₂ [32], NH₃ [33], PH₃ [34] и проведена оценка доверия историческому ряду экспертных данных [11, 24].

Заключение

Основной частью спектроскопических информационных ресурсов Института являются четыре информационные системы: Spectra, S&MPO, W@DIS и Nitran on the Web. Выбор данных, использованных в этих системах, и их согласование осуществлены сотрудниками Института совместно с рядом коллег из европейских университетов и организаций США. Программное обеспечение для решения спектроскопических и информационных задач разработано и реализовано сотрудниками Института. Часть СИР, используемых в этих системах, доступна через интерфейсы виртуального центра атомных и молекулярных данных [7]. За последние 5 лет в Институте опубликовано более 200 статей по спектроскопии.

Информационные системы Spectra и Nitran on the Web используют последние версии экспертных данных или расчетов параметров спектральных линий, полученные в Институте. Наряду с экспертными данными HITRAN и GEISA для вычисления спектральных функций можно использовать набор данных UC&MPO. В этой системе размещены данные о структуре молекулы озона и о спектроскопических константах для основного электронного состояния, потенциальные функции, дипольные моменты, моменты переходов, колебательные и вращательно-колебательные энергии и волновые функции и т.д.

Информационная система W@DIS используется для интеграции разнородных спектральных данных в единую систему, позволяющую импортировать, экспортировать и сравнивать данные, вычислять свойства импортированных данных, создавать системы классификации всех импортированных данных, предоставлять исследователям интерфейсы для контроля качества данных и представлять агентом описание свойств данных для их автоматической обработки.

В настоящее время слабым местом всех описанных систем является невозможность автоматического

интегрирования вычислительных и информационных приложений, входящих в эти ИС, в сложные потоки работ, в том числе в потоки работ по спектроскопии.

Авторы благодарны РФФИ за финансирование работы, выполненной в рамках гранта № 13-07-00411.

1. De Roure D., Jennings N., Shadbolt N. A Future e-Science Infrastructure, in book: Grid Computing: Making the Global Infrastructure a Reality. 2003. P. 437–470.
2. World Wide Web Corporation, <http://www.w3.org/standards/semanticweb/>
3. Родимова О.Б., Творогов С.Д., Фазлиев А.З. Онтология по молекулярной спектроскопии атмосферных газов: Труды конф. // 5-я Всерос. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». СПб., октябрь, 2003. С. 211–215.
4. Moreau N., Dubernet M.L., Muller H. VO access to CDMS spectroscopic database / M. Guainazzi, P. Osuna, editors // Astronomical spectroscopy and virtual observatory: Proc. EURO-VO workshop, held at the European Space Astronomy Centre of ESA. Villafraanca del Castillo, Spain, March, 2007. European Space Agency, 2008. P. 195.
5. Osuna P., Guainazzi M., Salgado J., Dubernet M.-L., Roueff E. Simple Spectral Lines Data Model, Version 1.0, IVOA Recommendation 09 December 2010 / Editors: P. Osuna, J. Salgado.
6. 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 атмосф. и океана. 2004. Т. 17, № 11. С. 921–926.
7. Dubernet M.L., Boudon V., Culhane J.L., Dimitrijevic M.S., Fazliev A.Z., Joblin C., Kupka F., Leto G., Le Sidaner P., Loboda P.A., Mason H.E., Mason N.J., Mendoza C., Mulas G., Millar T.J., Nucez L.A., Perevalov V.I., Piskunov N., Ralchenko Y., Rixon G., Rothman L.S., Roueff E., Ryabchikova T.A., Ryabtsev A., Sahal-Brechot S., Schmitt B., Schlemmer S., Tennyson J., Tyuterev V.G., Walton N.A., Wakelam V., Zeippen V. Virtual atomic and molecular data centre // J. Quant. Spectrosc. Radiat. Transfer. 2010. V. 111, N 15. P. 2151–2159.
8. Rixon G., Dubernet M.L., Piskunov N., Walton N., Mason N., Le Sidaner P., Schlemmer S., Tennyson J., Akram A., Benson K., Bureau J., Doronin M., Endres C., Heiter U., Hill C., Kupka F., Nenadovic L., Marquart T., Mulas G., Ralchenko Y., Shih A., Smith K., Schmitt B., Witherick D., Boudon V., Culhane J.L., Dimitrijevic M.S., Fazliev A.Z., Joblin C., Leto G., Loboda P.A., Mason H.E., Mendoza C., Millar T.J., Nucez L.A., Perevalov V.I., Rothman L.S., Roueff E., Ryabchikova T.A., Ryabtsev A., Sahal-Brechot S., Tyuterev V.G., Wakelam V., Zeippen C.J. VAMDC – The Virtual Atomic and Molecular Data Centre – A new way to disseminate atomic and molecular data – VAMDC level 1 release // AIP Conf. Proc. 2011. V. 1344. P. 107–115.
9. Фазлиев А.З. Развитие информационных систем в ИОА СО РАН // Оптика атмосф. и океана. 2009. Т. 22, № 10. С. 988–992.

10. *Bykov A.D., Fazliev A.Z., Filippov N.N., Kozodoev A.V., Privezentsev A.I., Sinitza L.N., Tonkov M.V., Tret'yakov M.Yu.* Distributed information system on atmospheric spectroscopy: Proc. SPIE // Int. Sympos. High Resolution Mol. Spectrosc. 2006. V. 6580. P. 65800W.
11. *Fazliev A., Privezentsev A., Tsarkov D., Tennyson J.* Ontology-Based Content Trust Support of Expert Information Resources in Quantitative Spectroscopy // Knowledge Engineering and the Semantic Web: Communications in Computer and Information Science. 2013. V. 394. P. 15–28.
12. *Berners-Lee T., Fielding R., Irvine U.C., Masinter L.* RFC 2396. Uniform Resource Identifiers (URI): Generic Syntax, <https://www.ietf.org/rfc/rfc2396.txt>
13. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures*, W3C Recommendation 5 April 2012, <http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/> Editors (Version 1.1): Shudi (Sandy) Gao, C. M. Sperberg-McQueen, Henry S. Thompson.
14. *RDF 1.1 Concepts and Abstract Syntax*, W3C Recommendation 25 February 2014, <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>, Editors: Richard Cyganiak, David Wood, Markus Lanthaler.
15. *OWL 2 Web Ontology Language Direct Semantics*, W3C Recommendation 27 October 2009, <http://www.w3.org/TR/2009/REC-owl2-direct-semantics-20091027/>, Editors: Boris Motik, Peter F. Patel-Schneider, Bernardo Cuenca Grau, Contributors: Ian Horrocks, Bijan Parsia, Uli Sattler.
16. *Привезенцев А.И., Царьков Д.В., Фазлиев А.З.* Базы знаний для описания информационных ресурсов в молекулярной спектроскопии. 3. Базовая и прикладная онтологии // Электронные библиотеки. 2012. Т. 15, вып. 2.
17. *Михайленко С.Н., Бабиков Ю.Л., Головкин В.Ф.* Информационно-вычислительная система «Спектроскопия атмосферных газов». Структура и основные функции // Оптика атмосфер. и океана. 2005. Т. 18, № 9. С. 765–776.
18. *Rothman L.S., Gordon I.E., Babikov Y., Barbe A., Benner D.C., Bernath P.F., Birk M., Bizzocchi L., Boudon V., Brown L.R., Campargue A., Chance K., Coudert L., Devi V.M., Drouin B.J., Fayt A., Flaud J.-M., Gamache R.R., Harrison J., Hartmann J.-M., Hill C., Hodges J.T., Jacquemart D., Jolly A., Lamouroux J., LeRoy R.J., Li G., Longo D., Mackie C.J., Massie S.T., Mikhailenko S., Müller H.S.P., Naumenko O.V., Nikitin A.V., Orphal J., Perevalov V., Perrin A., Polovtseva E.R., Richard C., Smith M.A.H., Starikova E., Sung K., Tashkun S., Tennyson J., Toon G.C., Tyuterev V.G., Vander Auwera J., Wagner G.* The HITRAN 2012 Molecular Spectroscopic Database // J. Quant. Spectrosc. Radiat. Transfer. 2013. V. 130. P. 4–50.
19. *Babikov Y.L., Mikhailenko S.N., Barbe A., Tyuterev V.G.* S&MPO – An information system for ozone spectroscopy on the WEB // J. Quant. Spectrosc. Radiat. Transfer. 2014. V. 145. P. 169–196.
20. *Привезенцев А.И., Козодоев А.В., Лаврентьев Н.А., Ахлестин А.Ю., Фазлиев А.З., Császár A.G., Tennyson J.* Информационная система «W@DIS», ориентированная на представление семантических метаданных в спектроскопии воды // Оптическая спектроскопия и стандарты частоты. Атомная и молекулярная спектроскопия. Томск: Изд-во ИОА СО РАН, 2009. С. 432–443.
21. *Лаврентьев Н., Макогон М., Фазлиев А.* Сравнение спектральных массивов данных HITRAN и GEISA с учетом ограничения на опубликование спектральных данных // Оптика атмосфер. и океана. 2011. Т. 24, № 4. С. 279–292.
22. *FTP ИОА СО РАН*, <ftp://ftp.iao.ru>
23. *Spectra*, <http://spectra.iao.ru>, Hitran on the Web, <http://hitran.iao.ru>, S&MPO, <http://smo.iao.ru>, W@DIS, <http://wadis.saga.iao.ru>
24. *Jacquinet-Husson N., Crepeau L., Armante R., Boutamine C., Chédin A., Scott N.A., Crevoisier C., Capelle V., Boone C., Poulet-Crovisier N., Barbe A., Campargue A., Benner Chris D., Benilan Y., Bézard B., Boudon V., Brown L.R., Coudert L.H., Coustenis A., Dana V., Devi V.M., Fally S., Fayt A., Flaud J.-M., Goldman A., Herman M., Harris G.J., Jacquemart D., Jolly A., Kleiner I., Kleinböhl A., Kwabia-Tchana F., Lavrentieva N., Lacome N., Xu L.-H., Lyulin O.M., Mandin J.-Y., Maki A., Mikhailenko S., Miller C.E., Mishina T., Moazzen-Ahmadi N., Müller H.S.P., Nikitin A., Orphal J., Perevalov V., Perrin A., Petkie D.T., Predoi-Cross A., Rinsland C.P., Remedios J.J., Rotger M., Smith M.A.H., Sung K., Tennyson J., Toth R.A., Vandaele A.-C., Auwera J.V.* The 2009 edition of the GEISA spectroscopic database // J. Quant. Spectrosc. Radiat. Transfer. 2011. V. 112, N 15. P. 2395–2445.
25. *Ахлестин А.Ю., Лаврентьев Н.А., Привезенцев А.И., Фазлиев А.З.* Базы знаний для описания информационных ресурсов в молекулярной спектроскопии. 5. Качество экспертных данных // Электронные библиотеки. 2013. Т. 16, № 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part4/AKLPF>.
26. *Апанович З.В., Винокуров П.С., Ахлестин А.Ю., Привезенцев А.И., Фазлиев А.З.* Визуализация парных отношений источников данных в количественной спектроскопии // Мат-лы 15-й Всерос. конф. «Интернет и современное общество». СПб., 2012. С. 7–15
27. *Апанович З.В., Винокуров П.С., Ахлестин А.Ю., Привезенцев А.И., Фазлиев А.З.* Цифровая библиотека научных статей по количественной спектроскопии // Труды 14-й Всерос. научн. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». RCDL'2012. Переяславль, 2012. С. 53–59.
28. *Воронина С.С., Привезенцев А.И., Царьков Д.В., Фазлиев А.З.* Различные онтологические представления предметной области // Мат-лы 4 Всерос. симпозиума «Инфраструктура научных информационных ресурсов и систем». СПб., 2014. Т. 2. С. 5–17.
29. *Tennyson J., Bernath P.F., Brown L.R., Campargue A., Császár A.G., Daumont L., Gamache R.R., Hodges J.T., Naumenko O.V., Polyansky O.L., Rothman L.S., Vandaele A.C., Zobov N.F., Dénes N., Fazliev A.Z., Furtbacher T., Gordon I.E., Hu S.-M., Szidarovszky T., Vasilenko I.* A IUPAC critical evaluation of the rotational-vibrational spectra of water vapor. Part IV. Energy levels and transition wavenumbers for D₂¹⁶O, D₂¹⁷O, and D₂¹⁸O // J. Quant. Spectrosc. Radiat. Transfer. 2014. V. 142. P. 93–108.
30. *Половцева Е.Р., Лаврентьев Н.А., Воронина С.С., Науменко О.В., Фазлиев А.З.* Информационная система для решения задач молекулярной спектроскопии. 5. Колебательно-вращательные переходы и уровни энергии молекулы H₂S // Оптика атмосфер. и океана. 2011. Т. 24, № 10. С. 898–905.
31. *Voronina S.S., Akhlestin A.Yu., Kozodoev A.V., Lavrentev N.A., Privezentsev A.I., Fazliev A.Z., Naumenko O.V.* Systematization of Published Spectral Data on Sulfur Dioxide Molecule and its Isotopologues // Proc. SPIE XX-th Int. Sympos. Atmos. Ocean Opt. Atmos. Phys. 2014. V. 9292.

32. *Laurentiev N.A., Privesentsev A.I., Filippov N.N., Fazliev A.Z.* Complete set of published spectral data on CO₂ molecule // Abstracts of the 22-nd Colloquium on High Resolution Molecular Spectroscopy. Dijon: Dicolor Groupe, 2011. P. 353.
33. *Voronina S.S., Yurchenko S.N., Fazliev A.Z.* Systematization of the published spectroscopic parameters of ammonia // Abstracts of the 22-nd Colloquium on High Resolution Molecular Spectroscopy. Dijon: Dicolor Groupe, 2011. P. 163.
34. *Akhlyostin A.Y., Voronina S.S., Privesentsev A.I., Fazliev A.Z.* Systematization of published data on phosphine isotopologues // Proc. SPIE. Irkutsk, Russian Federation, 2–6 July 2012. V. 8696. P. 8696–8638.

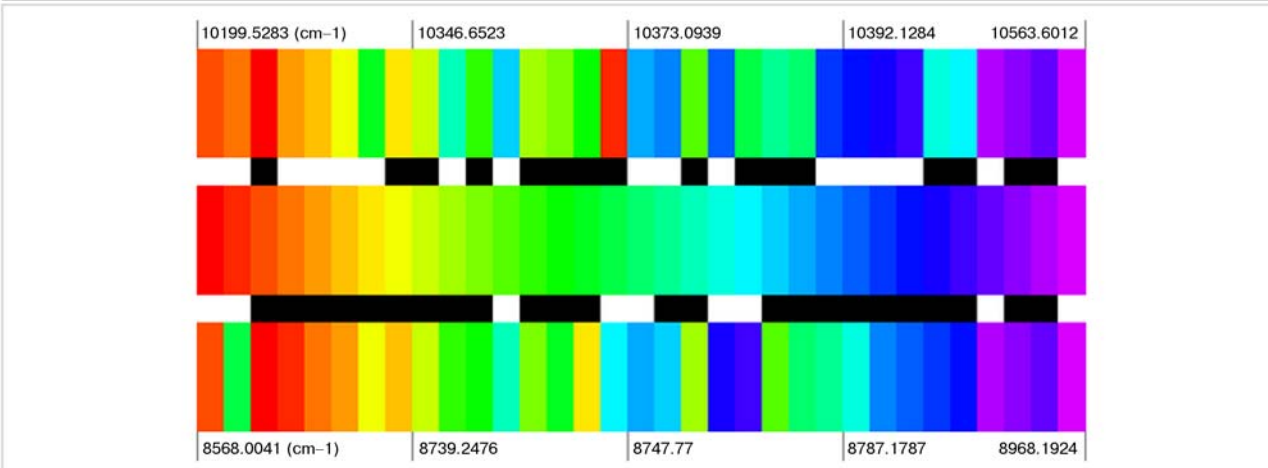
A.Yu. Akhlyostin, S.S. Voronina, N.A. Laurentiev, A.Z. Fazliev. **Spectroscopic information resources at Institute of Atmospheric Optics SB RAS.**

Information resources on spectroscopy are the most significant part of Institute of Atmospheric Optics SB RAS information resources. Furthermore, information systems Spectra, S&MPO, W@DIS, and HITRAN on the Web, developed via web-technologies, provide for a wide spectrum of researchers with access to verified data, information about this data, and knowledge, which allow processing this information on conceptual level. This article presents a brief description of these information systems. The focus of the work is on the information system W@DIS designed for qualitative analysis of spectroscopic data with different degrees of detailing.

Переходы. Представление парных отношений источников данных

[2008_Tote_H2O_ucl](#)

R. Tolchenov and J. Tennyson,
 Water line parameters from refitted spectra constrained by empirical upper state levels: Study of the 9500–14500 cm⁻¹ region,
 Journal of Quantitative Spectroscopy and Radiation Transfer, 2008, Volume 109, Issue 4, Pages 559-568,
 DOI: 10.1016/j.jqsrt.2007.08.001.



[1988_MaChFICa_H2O_ucl](#)

Mandin J.Y., J.P.Chevillard, J.M.Flaud and C.Camy-Peyret,
 H₂¹⁶O : Line positions and intensities between 8000 and 9500 cm⁻¹: the second hexad of interacting vibrationsl states: {(050), (130), (031), (210), (111), (012)},
 Canadian Journal of Physics, 1988, Volume 66, no. 11, Pages 997-1011,
 DOI: 10.1139/p88-162.

Таблица сравнения источников данных

Число идентичных переходов в полосах	$\Delta\omega_{\max}$	Вакуумные волновые числа СКО	A ₀₀	A ₀₁	A ₁₀
33	1.79e+3	1.62e+3	74	16	24

Таблица сравнения источников данных по идентичным колебательно-вращательным полосам

ν_1^i	ν_2^i	ν_3^i	ν_1^r	ν_2^r	ν_3^r	Число идентичных переходов в полосах	$\Delta\omega_{\max}$	Вакуумные волновые числа СКО	A ₀₀	A ₀₁	A ₁₀
1	2	1	0	0	0	33	1.79e+3	1.62e+3	0	16	24

Рис. 3. Интерфейс для анализа парных отношений между источниками данных