

ОПТИЧЕСКИЕ МОДЕЛИ И БАЗЫ ДАННЫХ ОПТИЧЕСКОЙ ИНФОРМАЦИИ
ОБ ОКРУЖАЮЩЕЙ СРЕДЕ

УДК 535.343.4+004.78

**Информационная система
для решения задач молекулярной спектроскопии.
4. Переходы в молекулах симметрии C_{2v} и C_s**

Н.А. Лаврентьев, А.И. Привезенцев, А.З. Фазлиев*

*Институт оптики атмосферы им. В.Е. Зуева СО РАН
634021, г. Томск, пл. Академика Зуева, 1*

Поступила в редакцию 5.08.2008 г.

Описана часть информационной системы по молекулярной спектроскопии, построенная на основе прикладной онтологии задач. Приведено описание информационных моделей двух задач (прямой и обратной) по нахождению переходов, используемых для построения информационной системы. Наполнение базы данных, связанных с этими задачами, проведено для молекулы воды и ее изотопомеров.

Ключевые слова: молекула воды, переходы, информационная система.

Введение

Информация о переходах между уровнями энергии в изолированной молекуле является ключевой в цепочке решения прямых и обратных задач в молекулярной спектроскопии. Частота, характеризующая переход, с одной стороны, является наиболее востребованной спектральной характеристикой в смежных предметных областях, изучающих потоки радиации, с другой — основой для получения фундаментальных характеристик молекулы: уровней энергии. Идентифицированные уровни энергии, полученные из обработки экспериментальных спектров, лежат в основе проверки корректности приписывания квантовых чисел и построения поверхности потенциальной энергии. Последнее, на наш взгляд, является одной из важных технических задач современной молекулярной спектроскопии. Полученные как в прямых, так и обратных задачах решения требуют проверки на достоверность, в первую очередь на формальные ограничения [1], накладываемые математическими моделями.

Важным является факт открытости источников информации о переходах и возможности проверки достоверности данных на любом этапе работы с ними. Для этой цели нами создано программное обеспечение, позволяющее собирать данные о переходах, проверять их на ограничения, следующие из математических моделей, и на соответствие типам данных, хранить их в реляционной базе данных, доставлять по требованию пользователя в виде файлов разного формата и представлять в табличном и графическом виде.

Главная особенность созданной информационной системы (ИС) состоит в предоставлении пользователю возможности загрузки массива данных в информационную систему и его отнесения к одной из молекул, публикации в печати или URI информационного ресурса, размещенного в сети Интернет, обеспечении этих данных набором семантических метаданных. В описываемой ИС семантические метаданные формируются автоматически при загрузке данных. Они позволяют решить две задачи: задачу семантического поиска информационных источников, имеющихся в ИС, и задачу систематизации информационных ресурсов, загруженных пользователем в систему с использованием общедоступной базы знаний.

В разд. 1 приведены физические характеристики прямой и обратной задач по нахождению частот перехода и обратной задачи по получению идентифицированного набора частот из экспериментальных данных, структура метаданных для молекул симметрии C_{2v} и C_s . В разд. 2 обсуждается процедура загрузки данных пользователями. Более детально она представлена в [2]. В разд. 3 описаны приложения для проведения поиска по ограниченному набору физических характеристик, доставки данных пользователю, сравнению данных из разных источников информации и их представлению в табличном и графическом виде. Это описание посвящено ИС W@DIS (Water internet-accessible Distributed Information System) [3], ориентированной на работу с данными и метаданными, относящимися к молекуле воды и ее изотопомерам.

1. Данные и метаданные о переходах

В рамках выбранной информационной модели молекулярной спектроскопии [4] переходы между уровнями молекул определяются в двух задачах

* Николай Александрович Лаврентьев; Алексей Иванович Привезенцев; Александр Зарипович Фазлиев (faz@iao.ru).

молекулярной спектроскопии: при решении прямой задачи методом эффективного гамильтониана или вариационным методом и при решении обратной задачи подгонкой подготовленного списка переходов (пиклиста), полученного из обработки результатов измерений к некоторому реперному для данной молекулы расчету.

Решение прямой задачи определяет вакуумные частоты переходов, квантовые числа, связанные с ними, и коэффициенты Эйнштейна. Решение обратной задачи дает вакуумные частоты переходов, поправки к ним, обусловленные экспериментальной точностью измерений, и квантовые числа, характеризующие переход. Поскольку приписывание квантовых чисел к вычисленным из результатов измерений частотам перехода является не формализованным процессом, а оценочным, то результаты такой оценки часто бывают противоречивыми. Пример такого расхождения результатов приписывания квантовых чисел приведен ниже. Факт расхождения обнаруживается при вычислении среднеквадратических отклонений данных для пар источников информации. Нахождение коэффициентов Эйнштейна при решении обратной задачи, как правило, не проводят. Это обусловлено значительными ошибками в экспериментальном определении интенсивностей, достигающих сотен процентов.

В традиционной молекулярной спектроскопии значение полного углового момента может меняться только на единицу, что связано с рассмотрением в ней только однофотонных процессов при поглощении и испускании света молекулами:

$$J'' \rightarrow (J' \text{ или } J''). \quad (1)$$

Для молекул симметрии C_{2v} и C_s для каждого уровня энергии должно выполняться ограничение

$$k_a + k_c = J \text{ или } J + 1. \quad (2)$$

При переходе с нижнего уровня на верхний должно выполняться ограничение на вращательные квантовые числа k_c :

$$|k'_c - k''_c| = 2n - 1, \quad n = 1, 2, 3, \dots \quad (3)$$

Для молекул симметрии C_{2v} (H_2O , D_2O , $H_2^{17}O$, $H_2^{18}O$, $D_2^{17}O$, $D_2^{18}O$) должно выполняться ограничение на колебательные и вращательные квантовые числа

$$v''_3 + v'_3 + k''_a + k'_a = 2n - 1, \quad (4)$$

связанное с запретом переходов между уровнями энергии пара- и орто-воды.

В настоящее время заполнение базы данных ИС W@DIS данными о переходах в молекуле воды является наиболее полным среди всех известных нам информационных систем. Работа по сбору данных выполнена группой IUPAC [5], предоставившей нам данные для занесения в ИС. В ИС W@DIS собраны данные из более чем 250 источников по переходам в молекуле воды и ее изотопомерах.

Расчетные данные, занесенные в ИС W@DIS, взяты из работ [6–9]. Интервалы частот, покрываемые ими, показаны в таблице.

Молекула	Интервал частот, cm^{-1}	Ссылка
H_2O	0–30000	[6, 7]
$H_2^{17}O$	0,452854–25400,16924	[7]
$H_2^{18}O$	0,189153–25399,985226	[7]
HDO	0,02752–19966,75785	[8]
D_2O	0,306822–15980,26341	[9]

Решения обратных задач содержат фрагменты спектра, полный обзор этих частей дан в работе [5].

В ИС W@DIS наряду со значениями физических величин содержатся семантические метаданные, образующие базу знаний, в которую включены наборы утверждений об индивидах, описывающих источники информации, и таксономии классов.

Общий подход к построению метаданных для описания решений вышеуказанных задач частично описан в [10]. При выборе метаданных для описания задач нахождения вакуумных частот перехода в молекуле в качестве количественных метаданных использовались экстремальные значения некоторых физических величин и число ошибок, совершенных авторами публикации по разным ограничениям, наложенным использованной ими математической моделью молекулы.

Для источников информации, используемых при загрузке данных о переходах, применялись пять уровней описания:

1. *Описание атрибутов источника информации*: название источника данных, ссылка на публикацию, комментарий, тип задачи и название вещества, уровни энергии которого относятся к источнику данных.

2. *Описание интенционала данных, представленных в источнике информации*: вакуумные частоты переходов, коэффициенты Эйнштейна, неопределенности в нахождении вакуумной частоты и коэффициентов Эйнштейна, квантовые числа переходов (в нотации нормальных мод и нотация Barber и Tennyson BT2).

3. *Метаданные о задаче, связанной с источником информации*, определяются в онтологиях задач спектроскопии и содержат информацию о входных данных и методе решения задачи.

4. *Количественные метаданные о характеристиках, описывающих переходы*: минимальное и максимальное значение вакуумной частоты в массиве данных, входящих в информационный источник, число переходов, наличие погрешностей, наличие коэффициента Эйнштейна в источнике информации, число переходов с уникальным набором квантовых чисел, число переходов с неполными квантовыми числами.

5. *Количественные метаданные*, характеризующие ошибки в квантовых числах, используемых при идентификации переходов (см. ограничения, определенные формулами (1)–(4)).

На рис. 1 показан скриншот, содержащий все типы метаданных, предоставляемые пользователю при работе с источниками информации о переходах в ИС W@DIS. Представленные метаданные содержат гиперссылки на табличное отображение соответствующих метаданных. Например, набор символов [T] указывает на переход к табличному представлению.

Вещество	Выходные данные	
D ₂ O	Частотные характеристики	
Использованный метод	Единица измерения	cm ⁻¹
UNDEFINED	Минимальная частота	110.586802
Публикация	Максимальная частота	413.794939
R. Paso and V.-M. Horneman, High-resolution rotational absorption spectra of H ₂ ¹⁶ O, HD ¹⁶ O, and D ₂ ¹⁶ O between 110 and 500 cm ⁻¹ . // Journal of Optical Society of America, B, 1995, v. 12, no. 10, p. 1813-1838.	Число длин волн	215 [T]
Incorrect assignment 0 0 0 13 3 12 0 0 0 13 2 11 161.037344	Вакуумная частота	true
Pure rotational spectra of ¹⁶ O isotopic species of H ₂ O, HDO, and D ₂ O were measured with a high-resolution Fourier-transform spectrometer in order to obtain better calibration line positions for high-resolution spectroscopy. The number of assigned lines for HDO was 380 and for H ₂ O it was ~310 between 170 and 500 cm ⁻¹ . For D ₂ O, ~220 lines in the 110-420-cm ⁻¹ region were assigned. The calibration was performed with the highly accurate n ₂ band of carbonyl sulfide.	Точность	false
	Коэффициент Эйнштейна	
	Единица измерения	s ⁻¹
	Наличие	false
	Точность	false
	Переходы	
	Тип квантовых чисел	NormalModes
	Число полос	1 [T]
	Число переходов с уникальной идентификацией	215 [T]
	Число переходов с не уникальной идентификацией	0 [T]
	Число переходов без квантовых чисел	0 [T]
	Число уровней с правильными вращательными квантовыми числами ($k_a + k_c = J$ or $J+1$)	214 [T]
	Число уровней с неправильными вращательными квантовыми числами ($k_a + k_c \neq J$ or $J+1$)	1 [T]
	Число разрешенных переходов ($J' \rightarrow J''$ or $J'' \pm 1$)	215 [T]
	Число запрещенных переходов ($J' \rightarrow J''$ or $J'' \pm 1$)	0 [T]
	Число разрешенных переходов для воды ($ k_c' - k_c'' = 2n+1, n=1,2,3,..$)	215 [T]
	Число запрещенных переходов для воды ($ k_c' - k_c'' = 2n$)	0 [T]
	Число разрешенных переходов для воды(C _{2v}) ($v_3' + k_a' + v_3'' + k_a'' = 2n+1$)	215 [T]
	Число запрещенных переходов для воды(C _{2v}) ($v_3' + k_a' + v_3'' + k_a'' = 2n$)	0 [T]

Рис. 1. Метаданные источника информации, содержащего переходы D₂O. Среди метаданных стоит отметить переход с некорректными квантовыми числами

На рис. 1 переход по такой ссылке в строке «Число переходов с неправильными квантовыми числами ($k_a + k_c \neq J$ или $J + 1$)» приведет к переходу на страницу, содержащую в таблице строку с неправильным набором квантовых чисел. Если число некорректных отнесений не велико, то они перечисляются в комментарии к источнику информации.

2. Загрузка данных о переходах

Для сбора данных создана система загрузки с веб-интерфейсом. Загрузка переходов состоит из четырех этапов:

1. Выбор источника информации.
2. Задание молекулы и числа сущностей, значения которых находятся в загружаемом файле.
3. Описания позиций значений выбранного списка загружаемых сущностей (рис. 2).
4. Задание адреса нахождения файла с данными на стороне пользователя.

Для загрузки данных в ИС пользователь должен создать источник информации, содержащий библиографическую ссылку на публикацию. На следующем шаге необходимо указать молекулу, к которой относится массив данных. Система

позволяет определять следующие ошибки формирования файла пользователя:

- несоответствие типов данных, использованных пользователем в файле;
- некорректность структуры физических сущностей, описанной пользователем при подготовке загрузки;

– несоответствие допустимым интервалам значений для физических величин и т.д.

На рис. 3 показана схема базы данных, представляющая связи таблиц, в которой хранятся решения задач T2 и T5. Упрощение касается перечня используемых нотаций квантовых чисел и связи с библиографической базой данных.

№	Физическая величина	Начальная позиция колонки	Конечная позиция колонки
1	Частота перехода	1	14
2	Квантовые числа перехода (Нормальные моды)	14	91
3	Коэффициент Эйнштейна	92	100
4	Ошибка для частоты перехода	101	110
5	Ошибка коэффициента Эйнштейна	111	121
6	Квантовые числа перехода (BT2)	122	140

Ok

Источник информации	2008_Marvel_H2O
Вещество	H ₂ O
Схема данных	Частота перехода (1,14) Квантовые числа перехода (Нормальные моды) (14,91) Коэффициент Эйнштейна (92,100) Ошибка для частоты перехода (101,110) Ошибка коэффициента Эйнштейна (111,121) Квантовые числа перехода (BT2) (122,140)

Рис. 2. Описание колонок файла, содержащего выходные данные задачи T5

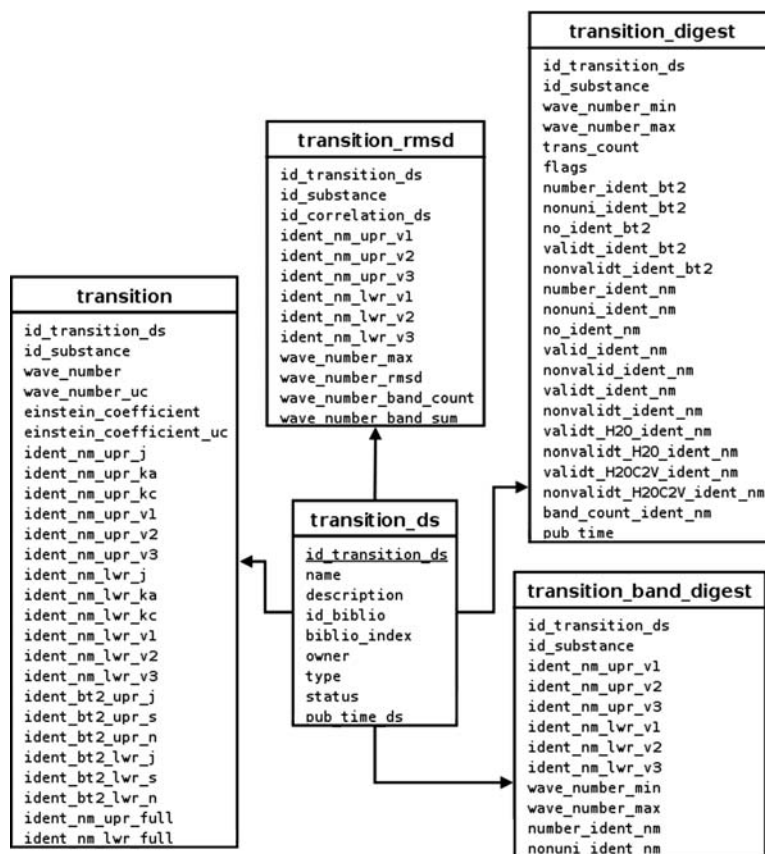


Рис. 3. Структура базы данных для хранения данных о переходах молекул симметрии C_{2v} и C_s

3. Поиск, представление, доставка и сравнение вакуумных частот в ИС W@DIS

Вакуумные частоты перехода, связанные с источниками информации, относящиеся к прямой или обратной задаче, можно сравнивать между собой как по значению частоты, так и по квантовым числам в нотациях Normal Modes или VT2. Такое сравнение важно при анализе экспериментальных данных, число которых на порядок меньше расчетных данных.

Пользователь может экспортировать данные об вакуумных частотах переходов молекул при просмотре или сравнении источников информации, используя для этой цели их табличное представление. Экспортированным может быть как весь массив данных, так и его часть, предварительно выбранная в таблице. Доставка данных осуществляется по сети Интернет в текстовом формате или в формате данных, используемых программой обработки таблиц Excel.

Прежде чем проводить сравнение, пользователь должен найти требуемые источники информации.

В ИС W@DIS существуют два механизма поиска:

– атрибутный поиск по значению таких атрибутов, как название молекулы, интервал значений полного углового момента, нотации квантовых чисел и т.д.;

– семантический поиск информационных ресурсов в базе знаний, сформулированной в виде явной спецификации молекулярной спектроскопии с помощью языка OWL [11].

При атрибутном поиске в базе данных пользователь перечисляет значения атрибутов физических сущностей или интервалов их изменений с помощью статического интерфейса, фиксированного разработчиками ИС. После выбора необходимого числа источников информации в ИС можно проводить их попарное табличное или графическое сравнение частот переходов по их значениям или просматривать данные отдельного источника информации. Для пары источников информации можно вычислять среднеквадратические отклонения как по полосам, так и по всему массиву данных. На рис. 4 приведен пример такого сравнения. Видно, что расхождение результатов работ может быть значительным.

Источник информации слева							Источник информации справа						
2005_MiMeStTa_D2O							2006_ZoOvShPo_D2O						
S.N. Mikhailenko, G.Ch. Mellau, E.N. Starikova, S.A. Tashkun, V.I.G. Tyuterev, Analysis of the first triad of interacting states (020), (100), and (001) of D ₂ ¹⁶ O from hot emission spectra. // Journal of Molecular Spectroscopy, 2005, v. 233, p. 32-59.							N.F.Zobov, R.I.Ovsannikov, S.V.Shirin, O.L.Polyansky, J.Tennyson, A.Jankas and P.F. Bernath, Infrared emission spectrum of hot D ₂ O. // Journal of Molecular Spectroscopy, 2006, v. 240, no. 1, p. 112-119.						

Число совпавших по квантовым числам вакуумных частот	(Вакуумная частота) _{max}
1298	72.14407

Показать	5	строк от	0	Всего строк 11669	>	>>	Настройки
----------	---	----------	---	-------------------	---	----	-----------

Вакуумная частота перехода w ₁ (см ⁻¹)	v ₁ ^{up}	v ₂ ^{up}	v ₃ ^{up}	j ^{up}	k _a ^{up}	k _c ^{up}	v ₁ ^{low}	v ₂ ^{low}	v ₃ ^{low}	j ^{low}	k _a ^{low}	k _c ^{low}	Вакуумная частота перехода w ₂ (см ⁻¹)	w ₁ -w ₂ [z-a]
2251.00707	0	2	0	17	15	3	0	0	0	18	16	2	2178.863	72.1441
2858.73997	0	2	0	16	14	2	0	0	0	16	13	3	2810.9334	47.8066
2259.62068	0	2	0	15	14	2	0	0	0	16	15	1	2214.0946	45.5261
2856.8959	0	2	0	15	14	2	0	0	0	15	13	3	2811.384	45.5119
2240.66575	0	2	0	16	13	3	0	0	0	17	14	4	2200.3937	40.272

Показать	5	строк от	0	Всего строк 11669	>	>>	Настройки
----------	---	----------	---	-------------------	---	----	-----------

Спектральные полосы (v ₁ ,v ₂ ,v ₃) ^{верх} - (v ₁ ,v ₂ ,v ₃) _{ниж}	(Вакуумная частота) _{ско}
Все полосы	3.30242343874
0 0 1 - 0 0 0	0.78388113939
0 2 0 - 0 0 0	11.09430052799
1 0 0 - 0 0 0	0.94947867438

Рис. 4. Сравнение значений вакуумных частот и вычисление среднеквадратических отклонений. Приведенный пример указывает на несоответствие частот перехода. Последнее может быть связано с некорректным отнесением частот к квантовым числам

Для организации поиска, позволяющего однозначно находить источники информации, соответствующие запросам пользователя, необходимо, с одной стороны, создать основанный на логике язык спецификации молекулярной спектроскопии, а с другой – предоставить пользователю ИС средства для построения запроса на поиск источников информации. Некоторые детали реализации поиска приведены в работе [12].

Заключение

Источники информации, описанные в статье, ассоциированы с задачами определения вакуумных частот перехода, отнесения их к квантовым числам и нахождения коэффициентов Эйнштейна. Наполнение базы данных ИС W@DIS [3, 13] расчетными и экспериментальными данными, относящимися к вышеупомянутым задачам, позволяет рассматривать ее в настоящее время как прототип информационной системы, ориентированной на представление данных о спектральных параметрах изолированной молекулы произвольной симметрии и знаний об источниках информации, в которых эти данные опубликованы.

Авторы благодарны РФФИ за финансирование работы, а также профессорам J. Tennyson и A. Császár за обсуждение структуры данных переходов и типов метаданных, Н.Ф. Зобову за предоставленные результаты вычислений параметров переходов изотопомеров молекулы воды, А.Ю. Ахлестину за реализацию приложения для представления среднеквадратических отклонений.

1. Furtenbacher T., Császár A.G., Tennyson J. MARVEL: measured active rotational-vibrational energy levels // *J. Mol. Spectrosc.* 2007. V. 245. N 2. P. 115–125.
2. Козодоев А.В., Козодоева Е.М., Привезенцев А.И., Фазлиев А.З. Система ввода и обмена данными в распределенной информационной системе «Молекулярная спектроскопия» // Труды XIII Байкальской Всероссийской конф. «Информационные и математические технологии в науке и управлении». Часть 2. Иркутск: ИСЭМ СО РАН, 2008. С. 156–162.
3. Fazliev A.Z., Császár A.G., Tennyson J. W@DIS: Water spectroscopy with a Distributed Information System // *Proc. of the 10 HITRAN Database Conference.* 2008. P. 38–39.

4. Bykov A.D., Fazliev A.Z., Filippov N.N., Kozodoev A.V., Privezentsev A.I., Sinita L.N., Tonkov M.V., Tretyakov M.Yu. Distributed information system on atmospheric spectroscopy // *Geophys. Res.* 2007. Abstracts 9. 01906.
5. Tennyson J., Bernath P.F., Brown L.R., Campargue A., Carleer M.R., Császár A.G., Gamache R.R., Hodges J.T., Jenouvrier A., Naumenko O.V., Polyansky O.L., Rothman L.S., Toth R.A., Vandaele A.C., Zobov N., Daumont L., Fazliev A.Z., Furtenbacher T., Mikhailenko S.N., Shirin S.V., Voronin B.A. IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part I. Energy Levels and Transition Wavenumbers for H₂¹⁷O and H₂¹⁸O // *J. Quant. Spectrosc. and Radiat. Transfer.* (In press).
6. Barber R.J., Tennyson J., Harris G.J., Tolchenov R.N. A high accuracy computed water line list – BT2 // *Mon. Notic. Roy. Astron. Soc.* 2006. V. 368. N 3. P. 1087–1094.
7. Zobov N.F., Ovsyannikov R.I., Shirin S.V., Polyansky O.L. The Assignment of Quantum Numbers in the Theoretical Spectra of the H₂¹⁶O, H₂¹⁷O and H₂¹⁸O Molecules Calculated by Variational Methods in the Region 0–26 000 cm⁻¹ // *Opt. and Spectrosc.* 2007. V. 102. N 3. P. 348–353.
8. Schwenke D.W. and Partridge H. Convergence testing of the analytic representation of an ab initio dipole moment function for water: Improved fitting yields improved intensities // *J. Chem. Phys.* 2000. V. 113. N 16. P. 6592–6597.
9. Shirin S.V., Zobov N.F., Polyansky O.L. Theoretical line list of D₂¹⁶O up to 16,000 cm⁻¹ with an accuracy close to experimental // *J. Quant. Spectrosc. and Radiat. Transfer.* 2008. V. 109. N 4. P. 549–558.
10. Козодоев А.В., Привезенцев А.И., Фазлиев А.З. Информационная система для решения задач молекулярной спектроскопии. 3. Уровни энергии молекул // *Оптика атмосф. и океана.* 2007. Т. 20. № 9. С. 805–809.
11. OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation 10 Feb 2004 / Patel-Schneider, Hayes, Horrocks, eds. <http://www.w3.org/TR/owl-semantic/>
12. Привезенцев А.И., Фазлиев А.З. Применение семантических метаданных для систематизации информационных ресурсов в молекулярной спектроскопии // Труды XIII Байкальской Всероссийской конф. «Информационные и математические технологии в науке и управлении». Часть 2. Иркутск: ИСЭМ СО РАН, 2008. С. 171–176.
13. Информационная система W@DIS, <http://wadis.saga.iao.ru>

N.A. Lavrentyev, A.I. Privezentsev, A.Z. Fazliev. Informational system for the solution of molecular spectroscopy problems. 4. Transitions in molecules of C_{2v} and C_s symmetry.

A part of a molecular spectroscopy informational system has been designed and implemented on the basis applied ontology of tasks. A description of informational models of two tasks (direct and inverted) of finding the transitions, used for the creation of the informational system, has been presented. The filling of the data bases related to these tasks was done for water molecules and water isotopomers.