

Информационная система трехслойной архитектуры по молекулярной спектроскопии

А.Ю. Ахлестин, Н.А. Лаврентьев, А.И. Привезенцев, А.З. Фазлиев

Институт оптики атмосферы имени В.Е. Зуева СО РАН

lnick@iao.ru, remake@iao.ru, faz@iao.ru

Аннотация

Создана информационная система по количественной спектроскопии, имеющая трехслойную архитектуру. В докладе обсуждаются назначение и реализация двух слоев: информационного слоя и слоя знаний. Информационный слой содержит наборы высказываний, характеризующих свойства информационных ресурсов, которые в системе представляют решения задач спектроскопии. В информационной системе эти высказывания создаются автоматически при решении задач или импорте решений задач. В слое знаний ИС, содержащем онтологию решений задач спектроскопии, высказывания лежат в основе построения индивидов прикладной онтологии.

Важной задачей слоя знаний является автоматическое создание таксономий классов. Эти таксономии могут быть связаны с информационными задачами, которые можно решать, используя факты из слоя данных и приложений. Подобные задачи могут рассматриваться как намерения исследователя по изучению постановок задач для расчетов и измерений. Рассмотрены два типа намерений, для которых возможно автоматическое построение таксономии классов.

1. Введение

Информационная система (ИС) «Молекулярная спектроскопия» предназначена для предоставления доступа к данным, информации и логическим теориям необходимым для решения прикладных и фундаментальных задач ряда предметных областей (оптики атмосферы, спектроскопии, астрономии и т.д.).

Теоретическим основанием для создания информационной системы является допущение о том, что факты, относящиеся к описанию спектральных характеристик молекул можно представлять разными способами. В трехслойной архитектуре ИС пользователю предоставлена возможность работать со слоем данных и приложений, информационным слоем и слоем знаний.

Для работы в слое данных от пользователя требуется знание естественного языка и понимание основ молекулярной спектроскопии. Фактами, относящимися к слою данных, являются факты, связанные со значениями физических величин (уровни энергии, вакуумные волновые числа, интенсивности и т.д.), которые при математическом моделировании представляют решения прямых и обратных задач молекулярной спектроскопии. Факты появляются в ИС при загрузке опубликованных решений задач молекулярной спектроскопии или при решении подобных задач в ИС. Такой механизм извлечения фактов связан с выбранной моделью предметной области в виде двух цепей задач. В этом слое пользователь может выбирать молекулы и просматривать и сравнивать значения физических величин, относящиеся к ним.

Информационный слой ИС содержит описание свойств решений задач аккумулярованных в слое данных. Использование этого слоя полезно для пользователей, понимающих разнообразные ограничения, накладываемые на эти свойства, в том числе те, которые напрямую не относятся к спектроскопии. Например, необходимо понимать суть ограничений на физические величины, вытекающие из правил отбора, ограничений на существование информационных ресурсов и ограничений, накладываемых отношением рефлексивности при построении математических моделей в спектроскопии. В частности, в этом слое анализируются парные отношения между решениями задач. Простым методом анализа парных отношений является расчет стандартных отклонений. Пользователь может визуализировать в ИС свойства решений задач и значения этих свойств. Анализ свойств позволяет разделять ресурсы на достоверные и недостоверные в рамках выбранных пользователем критериев. Целью формирования такого слоя является создание наборов свойств решений задач спектроскопии и для дальнейшей их систематизации.

Работа со слоем знаний требует от пользователя понимания языка спецификаций онтологий OWL DL. Он является интерпретируемой логикой с помощью которой строится логическая теория источников информации о решении задач молекулярной спектроскопии (прикладная онтология). Основным назначением слоя знаний в ИС является систематизация свойств решений задач спектроскопии, в частности, предоставление пользователю

возможность семантического поиска источников информации о спектральных свойствах молекул, и предоставление информационных ресурсов, для которых возможна машинная логическая обработка. В настоящее время основной акцент в формировании таких ресурсов сделан на свойствах, относящихся к проблеме достоверности информационных ресурсов.

2. Постановка задачи

Для исследователей, использующих в своих задачах в качестве входных данных не собственные данные, важным является их достоверность. Однако критерии достоверности для одних и тех же данных в разных прикладных предметных областях могут быть разными. Например, для обратной задачи определения уровня молекулы в спектроскопии важными является наличие достоверных квантовых чисел, а для вычислений спектров поглощения квантовые числа не имеют значения (за исключением случая интерференции спектральных линий).

Экспертные массивы спектральных данных создавались и создаются разными исследователями для решения определенного круга задач. По этой причине большая часть из них содержит значения, которые противоречат значениям, содержащимся в других экспертных массивах (иногда значения не противоречат, но каждое из них является недостоверным). В ряде случаев измеренные с недостаточной точностью данные, входящие в такие массивы, исправляются с целью достижения приемлемого результата в решаемых задачах, например, задачах вычисления потоков уходящего и приходящего излучения в атмосферах планет. Исследователи, сделавшие такие поправки, манипулировали с массивами, содержащими сотни тысяч фактов, извлеченными из нескольких тысяч публикаций. Причины, побудившие их к подобным манипуляциям, не документированы. В прединтернетовский период практически отсутствовала инфраструктура, позволяющая в автоматическом режиме контролировать такого рода изменения.

Появление суперкомпьютеров в 80-х привело к тому, что в конце 90-х массивы данных только для одного изотомера молекулы воды стали содержать более 109 фактов. В ручной обработке данных наступил коллапс. Временным решением проблемы стало многократное увеличение числа экспертов при подготовке массива. Эта мера решила проблему количества данных в экспертных массивах, но не их качества.

В методологии эмпирических наук существует принцип соответствия, согласно которому факты новой логической теории содержат в себе хорошо выверенные факты предшествующей ей теории. Из этого следует вывод, что факты предшествующей теории должны актуализироваться, систематизироваться и быть доступными для исследователей. Практически каждый исследователь систематизирует доступные ему факты при изучении предметной

области. Однако полнота фактов обеспечивается только в случае доступности научному сообществу всех опубликованных фактов. В спектроскопии число фактов столь значительно, что необходима их автоматическая систематизация. Этот вывод инициирует формулировку проблемы построения экспертных массивов данных в молекулярной спектроскопии с помощью автоматизированных распределенных ИС. В таких ИС должны накладываться ограничения, состоящие в том, что все факты массивов должны быть опубликованы и удовлетворяют формальным и неформальным критериям достоверности. Эта задача решается созданием интернет доступной распределенной информационной системы для обеспечения полного набора опубликованных источников данных.

На наш взгляд такие ИС могут заметить устаревающие технологии представления бумажных, и, построенных на соответствующих информационных технологиях цифровых информационных ресурсов, представленных в Интернете. Задачи актуализации и доступа к данным, информации и знаниям предметных областей, содержащимся в таких системах, не являются задачами этих предметных областей. Такого вида задачи относятся к информатике. Появление новых фактов в предметной области, вступающих в противоречие с данными существующими в общем доступе, должно приводить, после их экспертной оценки, к соответствующим поправкам и комментариям. Эти поправки должны касаться всех фактов, независимо от того в каких журналах они опубликованы. Современная структура информационных ресурсов, дифференцирующая их по издательствам не позволяет реализовать эту возможность и представляет собой препятствие, требующее преодоления. Ключевой задачей, способствующей решению проблемы формирования экспертных массивов в предметных областях с большим количеством фактов, является задача автоматической систематизации данных, информации и знаний.

Приведем примеры, показывающие на старение фактов и содержание в статьях недостоверных фактов. Для рассматриваемой ниже предметной области «Количественная спектроскопия» получена статистика о достоверности опубликованных решений задач по вычислению спектральных характеристик молекулы воды [1]. Она дана в Табл.1, в которой для каждого изотомера воды представлено число решений задач, извлеченных из публикаций, а в скобках дано число решений задач, содержащих в настоящее время только достоверные данные. Из более чем 950 решений задач 20% решений содержат недостоверные данные. Во многом такой большой процент решений с недостоверными данными связан с тем, что, по мере совершенствования методов измерений и вычислений, количественные компоненты критериев достоверности становятся более жесткими.

Таблица 1. Статистика публикаций по спектроскопии воды [1].

Вещ-во	Задача Т1.	Задача Т2.	Задача Т3.	Задача Т5.	Задача Т6.	Задача Т7.
H ₂ O	7 (3)	5 (3)	23 (7)	295 (259)	89 (65)	31 (27)
H ¹⁸ OH	4 (4)	5 (4)	16 (14)	50 (44)	54 (24)	17 (17)
H ¹⁷ OH	3 (3)	5 (4)	6 (4)	21 (20)	38 (20)	18 (15)
HOD	2 (1)	4 (3)	5 (3)	9 (9)	89 (40)	31 (27)
H ¹⁸ OD	1 (1)	2 (2)	2 (2)	10 (10)	12 (12)	9 (8)
D ₂ O	2 (2)	3 (2)	3 (2)	8 (7)	38 (31)	19 (10)
D ₂ ¹⁸ O	1 (1)	2 (2)	2 (2)	2 (2)	4 (4)	5 (5)
H ¹⁷ OD	-	2 (2)	2 (2)	6 (6)	3 (1)	3 (3)
D ₂ ¹⁷ O	-	1 (1)	2 (2)	2 (2)	1 (1)	1 (1)
НТО	-	-	-	-	1 (0)	-
<i>Итого</i>	20 (15)	29 (23)	61 (38)	403 (359)	329 (198)	134 (113)

Для решения задачи систематизации, поставленной выше, необходимо построить онтологическую модель предметной области, в которой явно выделены свойства характеризующие качество решений, и, на основании этой модели, построить ИС с трехслойной архитектурой. В нашей трактовке задача построения уровня знаний такой системы и задача систематизации данных, информации и знаний во многом являются эквивалентными.

3. Данные и приложения

В задачах, решаемых в молекулярной спектроскопии, используются структурированные данные, характеризующиеся интенционалом и экстенционалом этих данных. При реализации предметной задачи в виде приложений часть этих данных является для приложений входными и выходными данными. При этом интенционал данных явно представляется в виде схем баз данных, содержащихся в ИС, а экстенционал данных является определенным зафиксированным решением задачи, размещенным в базе данных в соответствии с этой схемой.

Существующие наборы данных параметров спектральных линий HITRAN и GEISA [2,3] содержат в основном данные о решениях задач молекулярной спектроскопии. Интенционал этих наборов данных представляет собой структурные метаданные, содержащие физические величины, характеризующие спектр молекул, погрешности их определения и ссылки на публикации, в которых они опубликованы. Недостатками существующих структурных метаданных является то, что некоторые физические характеристики молекул, используемые при решении прямых и обратных задач молекулярной спектроскопии, не включены в них. Форма представления библиографических ссылок не позволяет определять с помощью SQL-запросов к БД, в которой размещены эти наборы, принадлежность данных к публикации. Это означает, что невозможно при компьютерной обработке данных установить авторов данных и провести корректное сравнение экспериментальных и расчетных данных.

Наполнение информационной системы значениями данных возможно двумя путями: во-первых, через проведение расчётов средствами системы, во-вторых, через загрузку пользовательских данных опубликованных решений предметных задач в ИС. Загружаемые пользователем данные образуют элементарные ресурсы, которые характеризуются первичным источником данных, содержащим библиографическую ссылку. Как правило, данные загружаются в систему, в виде файлов. Структура данных, используемая в файлах, может не соответствовать структуре данных, используемой для их хранения.

Ввод данных в ИВС сформирован в соответствии с иерархией задач. Для каждого класса решения задач созданы отдельные приложения. Общей для всех процедур ввода данных является процедура создания источника данных, с которым связываются загруженные ресурсы. Источник данных характеризуется уникальным идентификатором, названием, связью с публикацией, хранящейся в БД и классом задач.

В случае отсутствия в библиографической БД необходимой публикации пользователь может создать необходимую ему библиографическую ссылку. В системе используются четыре типа публикаций: статья, тезисы докладов, монография и Интернет ссылка. При вводе данных, подготовленных в виде файла, содержащего колонки символов, пользователь с помощью интерфейса описывает интенционал данных. 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 [4, 5] форматирования спектральных данных в молекулярной спектроскопии. Загруженный на сервер файл преобразуется в XML-документ и проверяется на соответствие экстенционала типам данных, описанных в XML-схеме. После разбора XML-документа данные заносятся в БД и формируются связанные с ним метаданные.

Например, уровни энергии молекул определяют-ся в двух задачах молекулярной спектроскопии: при решении прямой задачи на основе методов квантовой механики и при решении обратной задачи на основе подготовленного списка, полученного из обработки результатов измерений. Решение прямой задачи предоставляет информацию об уровнях энергии молекулы, включающую в себя данные о значениях уровней энергии и квантовых числах. Решение обратной задачи включает в себя данные о значениях уровней энергии, квантовых числах, величинах неопределенности значений уровней энергии и числе переходов, использованных для расчета уровня энергии. Перечисленные данные хранятся в базе данных информационной системы. Заметим, что рассчитанные в прямой и обратной задаче уровни энергии имеют разное происхождение,

т.е. являются разными физическими сущностями. Структура массива данных, относящегося к прямой и обратной задаче и загружаемого пользователем в ИВС, обязательно должна содержать уровни энергии и квантовые числа.

XSD-схемы для каждой задачи используются для структуризации данных задач и проверки их значений. Организация XML-схем основана на предположении, что любая физическая задача из информационной модели спектроскопии состоит в изучении определенной молекулы. Поэтому, корневым элементом в документе будет название изотомера молекулы, ему соответствуют название файла содержащего XML-схему.

Далее цепочка переопределений в XML-схеме приводит к выбору элемента из списка возможных задач, что представлено на рис. 1.

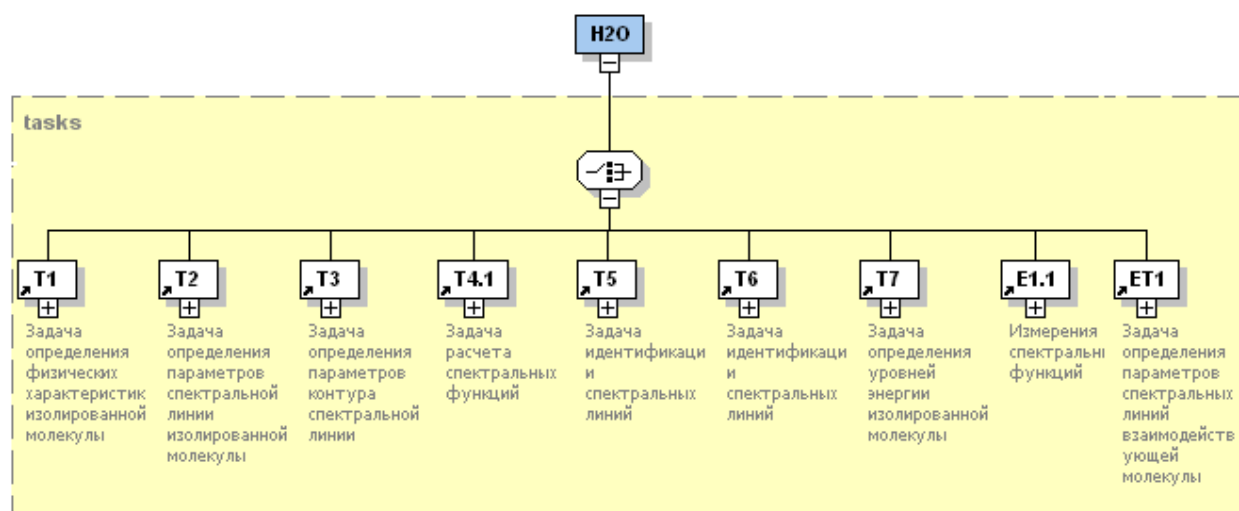


Рис. 1. Визуальное представление XML-схемы для файла H2O.xsd

4. Информационный слой (метаданные)

Физические величины, являющиеся концептами предметной области, могут обладать свойствами, принимающими определенные значения. Эти свойства и их значения могут не относиться к той предметной области, концепты и свойства которой сосредоточены в слое данных и приложений. В молекулярной спектроскопии примером таких свойств являются интервалы изменений физических величин, характеризующие решения задач, в которых они получены, а не сами физические величины. Другим примером может служить перечень физических величин, полученных при решении задачи и называемый структурными метаданными. Здесь под метаданными понимаются свойства и их значения, относящиеся к данным, описанным в слое данных.

В информационном слое описывается расширение предметной области, включающее в себя свойства и концепты, не включенные в рассмотрение при исследовании решений задач спектроскопии в слое данных и приложений. Такое расширение

предметной области зависит от набора информационных задач, которые ставятся при создании ИС и способов описания информационных ресурсов.

В ИС «Молекулярная спектроскопия» предметная область, относящаяся к информационному слою, представляет собой набор свойств и их значений, относящихся к решениям шести задач спектроскопии. К числу этих свойств относятся интервалы изменения физических величин, определенных в соответствующих решениях, число вакуумных волновых чисел, удовлетворяющих правилам отбора, стандартные отклонения решений однотипных задач и т.д.

Для описания предметной области, связанной с этим слоем, в нашей работе используется язык разметки RDF [6]. С формальной точки зрения описание построено в рамках логики предикатов со структурой высказывания вида (subject, predicate, object). Замечательной особенностью такого описания является тот факт, что его большая часть может создаваться автоматически. Заметим, что полная автоматизация построения информационного слоя

достигается в том случае, когда решения задач вычисляются в слое данных и вычислений.

На рис. 2 и 3 приведены субъектно-предикатные структуры, характеризующие два типа наборов свойств решения задач молекулярной спектроскопии. На рис. 2 показан набор свойств, описывающий свойства решения задачи T6. Эти свойства относятся только к решению задачи T6. На рис.4 выделены свойства, связанные одновременно с двумя решениями задачи T6.

Структуры обоих типов представляют собой сущности, которые в информационном слое играют самостоятельную роль, несмотря на то, что отдельные их части могут относиться к данным взятыми из слоя данных.

Каждая из сущностей первого типа (например, рис.2) связан только с той публикацией, в которой оно было впервые опубликовано. Число свойств такой сущности фиксировано моделью и не меняется во времени.

Сущности второго типа состоят из свойств пары решений однотипных задач. К числу этих свойств относятся совпадающие полосы, содержащиеся в этих решениях, максимальное значение стандартного отклонения, число совпадающих по квантовым числам переходов и т.д.

Классификация этих сущностей осуществляется в слое знаний в рамках языка OWL DL [7]. Этому посвящен следующий раздел работы.

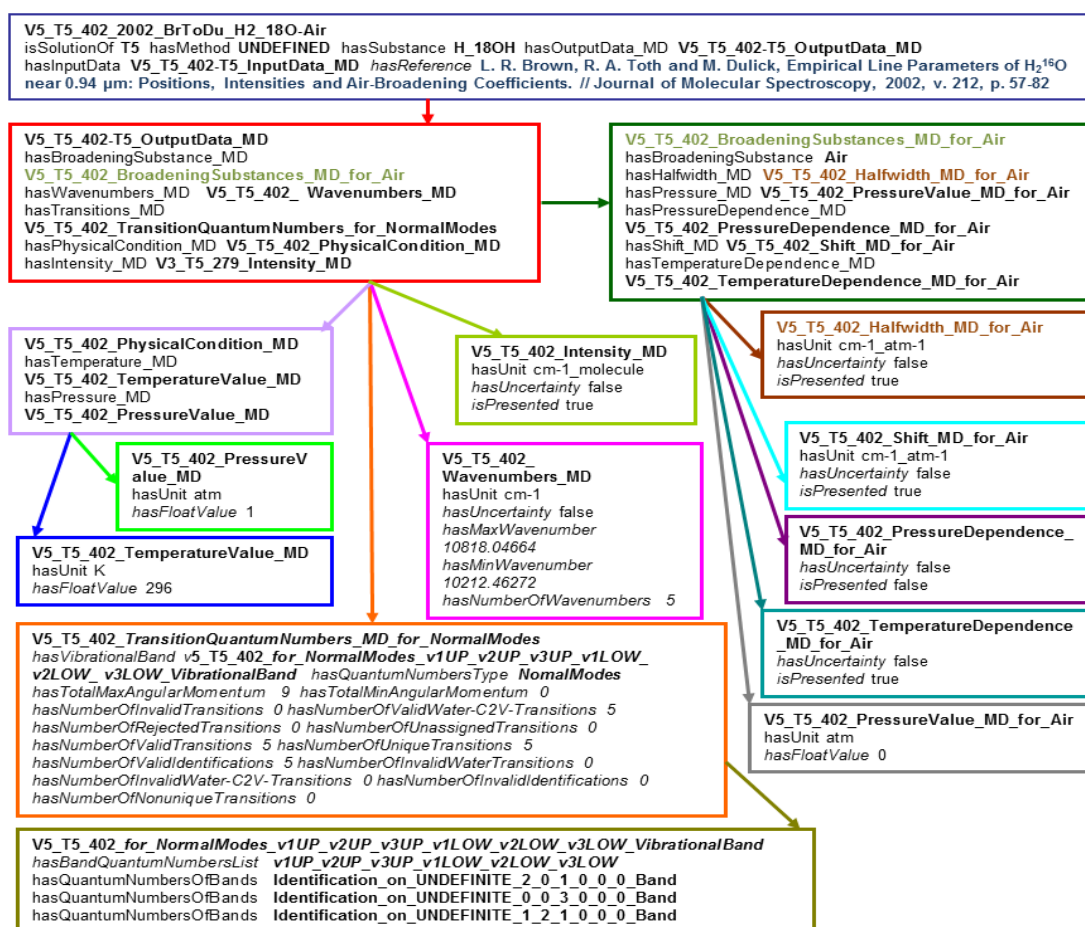


Рис.2. Структура индивида, описывающего свойства решения задачи T5, опубликованного в работе L. R. Brown, R. A. Toth and M. Dulick об индуцированных столкновениях полуширинах и сдвигах спектральных линий изотопомера воды H₂¹⁸O в диапазоне 10200-10800 см⁻¹.

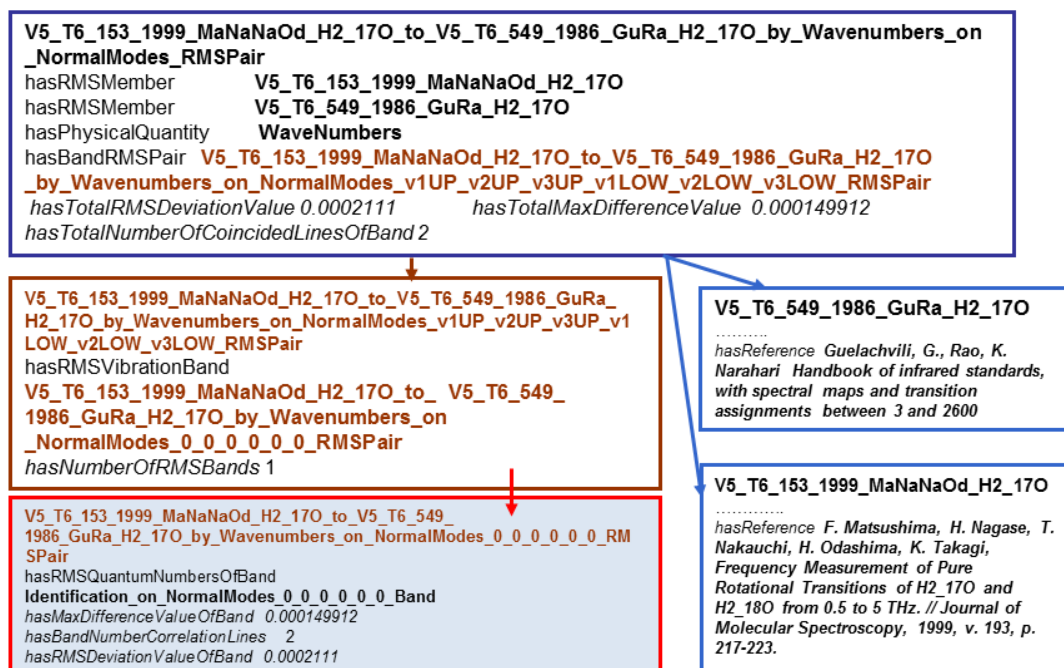


Рис.3. Индивид V5_T6_153_1999_MaNaNaOd_H2_17O_to_V5_T6_549_1986_GuRa_H2_17O_by_Wavenumbers_on_NormalModes_RMSPair», описывающий стандартное отклонение между экспериментальными данными о вакуумных волновых числах молекулы воды во вращательной полосе

Общее число фактов, связанных с сущностями обоих типов, представлено в Таб.2. Как следует из

Табл. 2 число фактов о свойствах решений задач в спектроскопии воды немного превышает 106.

Таблица 2. Статистика высказываний (фактов) в фактологической части базы данных для источников данных по молекуле воды.

Триады \Задачи	Задача T1.	Задача T7.	Задача T2.	Задача T6.	Задача T3.	Задача T5.
A-Box (Independent) $S_{ind} = 451257$	8849	8919	168344	18853	203569	42723
A-Box (Relative) $S_{rel} = 550129$	82746		192114		275269	

5. Слой знаний (прикладные онтологии)

Ключевым моментом в формировании фактологической части онтологии является создание индивидов. Индивиды представляют собой наборы фактов (высказываний), предназначенные (в данной ИС) для систематизации информационных ресурсов.

Приведем для примера число независимых и относительных индивидов для прикладной онтологии молекулы воды. Число независимых индивидов $N_{independent}=1063$, и оно совпадает с общим числом решений задач в предметной области (см. Таб.1). Число индивидов, характеризующих стандартные отклонения, равно $N_{relative}= 6789$.

Проведенная экспертами работа [8, 9] обеспечила полноту набора фактов в фактологической части

базы знаний о свойствах молекулы воды на текущий момент времени для обратных задач спектроскопии воды и их непротиворечивость. Обеспечить полноту фактологической части базы знаний для решений прямых задач не представляется возможным, т.к. большая часть решений не публиковалась и утеряна. Последнее обстоятельство подтверждает необходимость введения ограничения на опубликование решений задач. Без такого ограничения постановка некоторых задач, например, задачи о декомпозиции составных источников данных, теряет смысл.

Построенная на основе модели количественной спектроскопии фактологическая часть базы знаний (A-box) должна быть дополнена таксономической (понятийной) частью базы знаний. Построение таксономии связано с намерением ее создателей. Примерами намерений могут служить задачи каталогизации опубликованных решений задач, задачи

каталогизации всех теоретически возможных колебательных полос молекул и т.д. В спектроскопии, как правило, такие таксономии не содержат бесконечного числа каталогов. Это обусловлено тем фактом, что применение математической модели всегда ограничено и, с определенного момента, каталогизация, основанная на ограничениях, следующих из математической модели, не имеет смысла, т.к. она выходит за рамки применения физической модели, положенной в основу математического описания.

В нашей работе при создании таксономии в слое знаний РИС было проанализировано три группы намерений (см. Рис.4). Две группы связаны с подходом, который принято называть «замкнутый мир», и одна группа с подходом «открытый мир». В работе использовалась следующая трактовка этих терминов. К «замкнутому миру» относятся таксономии,

основанные на полном наборе опубликованных решений задач спектроскопии. Первая группа связана с источниками данных, в которых опубликованы результаты экспериментальных работ (обратные задачи), а ко второй — результаты решения прямых задач. К третьей группе относятся все прочие таксономии, не включающие в себя таксономии первых двух групп. Пересечение классов первых двух групп указывает на часть публикаций, содержащих результаты измерений и расчетов. Дополнение указывает на предсказанные данные, которые не получены в измерениях. Наконец третья группа указывает на возможные ограничения, по которым нет ни теоретических, ни экспериментальных результатов. В количественной спектроскопии наиболее востребованными являются намерения первой группы.

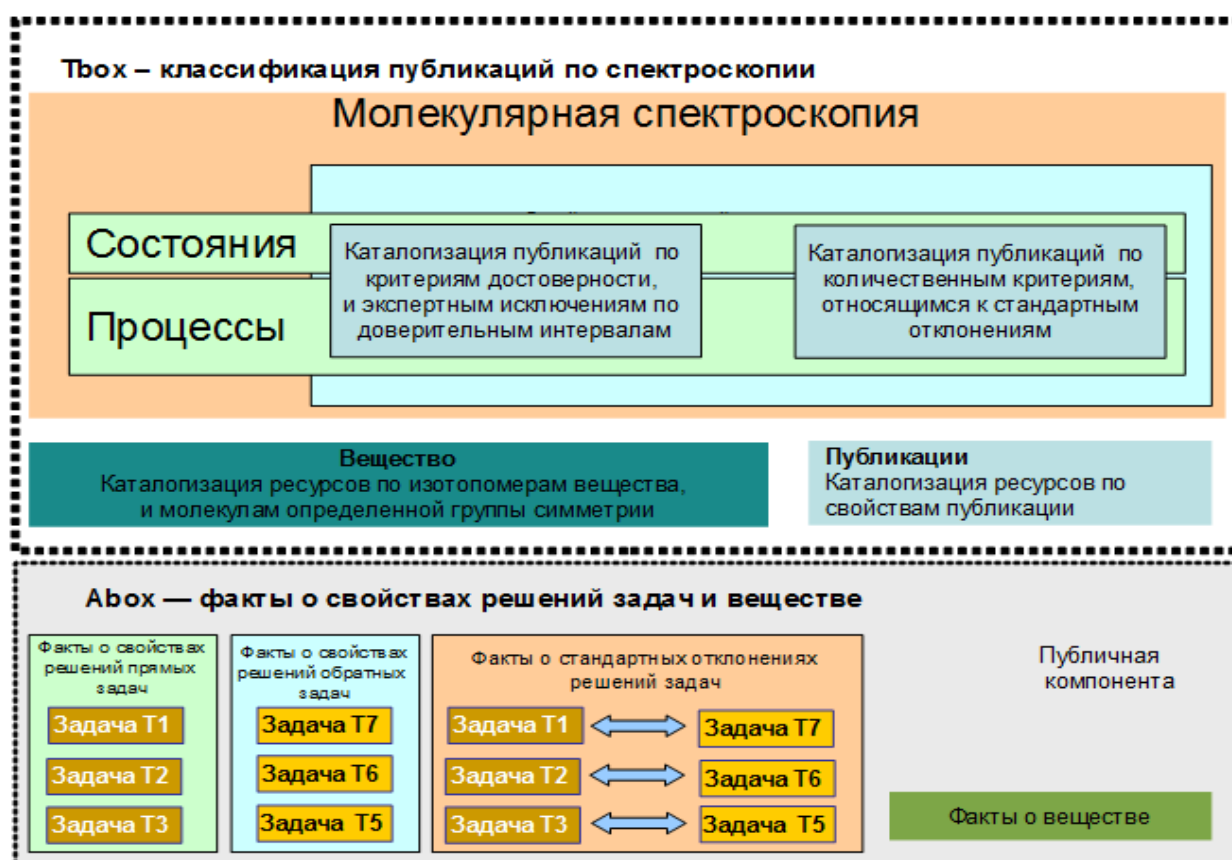


Рис.4. Структура онтологии, ориентированной на систематизацию информационных ресурсов молекулярной спектроскопии по критериям достоверности.

Обратимся к технической стороне формирования таксономии для первых двух групп намерений. В ИС «Молекулярная спектроскопия» автоматическая сгенерированная фактологическая часть предметной области позволяет решить четыре базовых задачи построения таксономий в ИС:

1. Систематизация решений задач по веществам;
2. Систематизация решений задач по физическим величинам и их значениям;

3. Систематизация решений задач по формальным критериям достоверности;
4. Систематизация решений задач по экспертным оценкам.

В работе систематизация проведена с помощью языка разметки OWL. По этой причине понятийная часть базы знаний, связанная с задачами систематизации, определялась в виде прикладных онтологий. Структура прикладной онтологии, формируемой в

РИС по количественной молекулярной спектроскопии, показана на рис. 5.

Приведем пример автоматического формирования таксономии прикладной онтологии. В молекулярной спектроскопии квантовые числа, характеризующие состояния молекулы, разделяют на две группы: колебательные и вращательные. Переход между состояниями молекулы характеризуется двумя наборами квантовых чисел, каждый из которых описывает состояние. Если из набора выделить только колебательные квантовые числа и зафиксировать их значения, то получится набор переходов с одинаковыми значениями колебательных чисел, но разными значениями вращательных. Такой набор переходов называют колебательной полосой. Полосы локализованы по спектру и, вместо указания интервала, можно пользоваться указанием полосы. Колебательная полоса имеет разную структуру по квантовым числам в зависимости от группы симметрии молекулы и типа квантовых чисел. Более

десять групп симметрии и около сотни правил отбора для разных типов квантовых чисел делают задачу ручного построения T-box достаточно утомительной. Число задач и молекул в модели предметной области невелико и фиксировано, тогда как число полос, для которых выполнены измерения, постоянно растет. Если A-box содержит все опубликованные решения задач, то сгенерированные классы будут представлять все непустые классы таксономии. В этом случае не важно, сколько частот в полосе было измерено.

На рис. 5 и 6 показаны интерфейсы для формирования шаблонов классов, на основе которых автоматически формируются части прикладных онтологий. Созданное программное обеспечение, слой данных и информационный слой позволяют формировать прикладные онтологии для решения задач систематизации решений задач спектроскопии, перечисленных выше.

Создание новой онтологии

Имя файла создаваемой онтологии:

Название онтологии по-русски:

Название онтологии по-английски:

Шаблон индивида в формате RDF/XML:

Сохранить

Предпросмотр шаблона

```

1. <?xml version="1.0"?>
2. <!DOCTYPE rdf:RDF [
3.   <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
4.   <!ENTITY owl "http://www.w3.org/2002/07/owl#">
5.   <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
6.   <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
20. <rdf:RDF xml:base="$_NAMESPACE_;" xmlns="$_NAMESPACE_;"
21.   xmlns:owl="owl;" xmlns:rdf="rdf;" xmlns:rdfs="rdfs;"
22.   xmlns:dc="dc;" xmlns:meta="meta;" xmlns:task="task;"
23.   xmlns:t1="t1;" xmlns:t2="t2;" xmlns:t3="t3;"
24.   xmlns:t5="t5;" xmlns:t6="t6;" xmlns:t7="t7;"
25.   xmlns:spectra="spectra;" xmlns:ds="ds;" xmlns:substance="substance;" xmlns:xsd="xsd;">
26.   <owl:Ontology rdf:about="">
27.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T1.owl" />
28.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T2.owl" />
29.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T3.owl" />
30.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T5.owl" />
31.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T6.owl" />
32.     <owl:imports rdf:resource="$_ONTOLOGY_Task_T7.owl" />
33.   </owl:Ontology>
34. </rdf:RDF>

```

Рис.5. Интерфейс для создания кода прикладной онтологии

Список импорта

Показать 20 строк от 0		Всего строк 115	> >>		Настройки
N	Вещество	Задача	Выбрать	Владелец	Триад
1	_13C_17O2	T2	☐	global	64
2	C_17O2	T2	☐	global	145
3	O_13C_18O	T2	☐	global	4649
4	O_13C_17O	T2	☐	global	1552
5	C_18O2	T2	☐	global	2739
6	_17OC_18O	T2	☐	global	213
7	_14CO2	T2	☐	global	132
8	_18O_13C_18O	T2	☐	global	397
9	_13C_17O_18O	T2	☐	global	60
10	OC_18O	T2	☐	global	32566
11	CO2	T2	☐	global	208715
12	O_13CO	T2	☐	global	43985
13	OC_17O	T2	☐	global	12953
14	C_18O2	T6	☐	global	1352
15	_13C_17O_18O	T6	☐	global	185
16	OC_18O	T6	☐	global	5818
17	OC_17O	T6	☐	global	1928
18	O_13C_17O	T6	☐	global	791
19	_17OC_18O	T6	☐	global	1172
20	_18O_13C_18O	T6	☐	global	627
Показать 20 строк от 0		Всего строк 115	> >>		Настройки

Список онтологий

Показать 20		строк от	0	Всего строк 15	Настройки
N	Название онтологии по-русски	Редактировать	Скачать		
1	Источники информации, содержащие только одну конкретную полосу	Редактировать	Скачать		
2	Источники информации, входящие в пары стандартных отклонений для заданных полос	Редактировать	Скачать		
3	Источники информации о переходах в заданных диапазонах	Редактировать	Скачать		
4	Источники информации, содержащие колебательную полосу и решения, полученные определенным методом	Редактировать	Скачать		
5	Источники информации, удовлетворяющие свойствам, связанным с каноничностью	Редактировать	Скачать		
6	Источники информации в которых нет переходов отклоненных экспертами	Редактировать	Скачать		
7	Источники информации, содержащие определенную физическую количественную величину	Редактировать	Скачать		
8	Источники информации, имеющие одинаковые переходы с заданным источником	Редактировать	Скачать		
9	Источники информации в которых все переходы уникальны	Редактировать	Скачать		
10	Источники информации в которых все переходы удовлетворяют правилам отбора (канонические источники информации)	Редактировать	Скачать		
11	Пары коррелирующих источников информации в которые входят только экспериментальные результаты	Редактировать	Скачать		
12	Пары стандартного отклонения, имеющие только одну конкретную полосу	Редактировать	Скачать		
13	Вещества в информационной системе	Редактировать	Скачать		
14	Тест	Редактировать	Скачать		
15	Полосы по которым есть данные только прямых задач	Редактировать	Скачать		

Рис.6. Интерфейсы для выбора молекул и формирования и редактировании прикладных онтологий.

6. Заключение

В работе описаны результаты работ по построению ИС по количественной спектроскопии с трехслойной архитектурой. Уровень знаний этой системы содержит онтологию первичных источников данных, представляющих решения задач спектроскопии. Понятийная часть базы знаний состоит из двух частей. Базовая часть содержит классы, определенные вручную, и представляющие понятия, относящиеся к фундаментальным. Основная задача систематизации, создаваемой в автоматическом режиме, состоит в классификации первичных источников данных по критериям достоверности, основанных на ограничениях относящихся к значениям физических величин. В работе дано краткое описание постановки задачи систематизации и приведены интерфейсы для построения шаблонов классов для автоматического построения прикладных онтологий, имеющих максимальную степень детализации.

Дальнейшее развитие ИС, в части систематизации источников данных, связано с решением задачи систематизации составных источников данных и

информации. Созданное программное обеспечение для проведения декомпозиции составных источников позволяет вычислять значения большей части свойств составных источников данных [10].

Литература

- [1] Привезенцев, А. И. Организация онтологических баз знаний и программное обеспечение для описания информационных ресурсов в молекулярной спектроскопии. Дисс. на соиск. уч. ст. канд. тех. наук. Томск, 2009.
- [2] Rothman, L.S. The HITRAN 2008 molecular spectroscopic database / L.S. Rothman, I.E. Gordon, A. Barbe, et al. // J. Quant. Spectrosc. Radiat. Transfer. 2009. Vol. 110. №. 9. P. 533-572.
- [3] Jacquinet-Husson, N. The GEISA spectroscopic database: Current and future archive for Earth and planetary atmosphere studies / N. Jacquinet-Husson, N.A. Scott, A. Chédin, et al. // J. Quant. Spectrosc. Radiat. Transfer. 2008. Vol. 109. Issue 6. P. 1043-1059.

- [4] Lancashire, T. Davies. Spectroscopic Data: The Quest for a Universal Format // Chemistry International. Boston: International Union of Pure and Applied Chemistry. 2006. Vol. 28. № 1. [Электронный ресурс]. — Режим доступа: http://www.iupac.org/publications/ci/2006/2801/3_lancashire.html
- [5] Dubernet, M.L. Virtual atomic and molecular data centre / M.L. Dubernet, V.Boudon, J.L.Culhane, et al. // J. Quant. Spectros. Rad. Transfer 2010 Vol. 111. P. 2151–2159.
- [6] Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation 10 February 2004. [Электронный ресурс]. — Режим доступа: <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210>.
- [7] OWL Web Ontology Language Semantics and Abstract Syntax / Ed. by P. F. Patel-Schneider, I. Horrocks. [Электронный ресурс]. — Режим доступа: <http://www.w3.org/TR/owl-semantics/>.
- [8] Tennyson, J. IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part I. Energy Levels and Transition Wavenumbers for H₂¹⁷O and H₂¹⁸O / J.Tennyson, P.F.Bernath, L.R.Brown, et al. // J. Quant. Spectrosc. Radiat. Transfer. 2009. Vol. 110. №. 9. P. 573-596.
- [9] Tennyson, J. IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part II: Energy levels and transition wavenumbers for HD¹⁶O, HD¹⁷O, and HD¹⁸O / J. Tennyson, P. F. Bernath, L.R. Brown, et al. // J. Quant. Spectrosc. Radiat. Transfer. 2010. Vol. 111. №. 15. P. 2160-2184.
- [10] Лаврентьев, Н.А. Сравнение спектральных массивов данных HITRAN и GEISA с учетом ограничения на опубликование спектральных данных / Н.А.Лаврентьев, М.М.Макогон, А.З.Фазлиев // Оптика атмосферы и океана. 2011. Т. 24. № 4. С. 279-292.

Three layers architecture of information system on quantitative molecular spectroscopy

A.Yu. Akhlyostin, N.A. Lavrentiev, A.I. Privezentsev,
A.Z. Fazliev

The information system on molecular spectroscopy of three layers architecture is constructed. The purpose and details of realization of two layers (information and knowledge) are discussed in the report. Information layer contains statements set described the properties of information resources which characterize the solutions of spectroscopic tasks. These statements are generated automatically in the system under the computation of the tasks or upload of the published solutions. In the knowledge base the statements are the base for constructing individuals of applied ontology.

The most important aim of the knowledge layer is automatic construction of taxonomy of classes. These taxonomy can be related to information tasks, which can be solved on the base of facts of data layer. Such kind of tasks can be interpreted as the intensions of the investigator to study the tasks which have to be solved (or measured in experimental works). Two types of intensions for which the automatic construction can be realized were considered.