

А. В. Козодоев, Е. М. Козодоева

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

kav@iao.ru, klen@iao.ru

УНИВЕРСАЛЬНЫЙ МОДУЛЬ «УНАРНЫЕ ОПЕРАЦИИ» В ИС «МОЛЕКУЛЯРНАЯ СПЕКТРОСКОПИЯ»

Описан подход, использованный при разработке и реализации модуля, выполняющего унарные операции над данными в ИС «Молекулярная спектроскопия». Приводится краткое описание унарных операций с учетом особенностей предметной области. Описан подход, использованный в программной реализации, позволяющий применять один набор скриптов для работы с имеющимися базами данных по различным веществам и несколькими типами спектроскопических данных.

Ключевые слова: структуры данных, количественная спектроскопия, унарные операции, базы данных.

Введение

В естественных науках, таких как спектроскопия, химия, генетика и др., которые опираются на значительные объемы данных, важную роль играют структуры и типы данных, используемые предметными приложениями, вычисляющими количественные характеристики. Универсальная детальная классификация структур данных предметной области, применимая ко всем наукам, вряд ли возможна. Однако некоторые ее элементы, относящиеся к систематизации информационных ресурсов, например, первичные и составные источники данных, применяются в значительном числе предметных областей. Среди составных источников данных выделяют экспертные данные, основная цель которых – предоставить исследователю прикладных наук достоверные и согласованные данные, которым он может доверять [1]. Как правило, объем экспертных данных на порядок ниже объема подобных по структуре данных предметной области.

Полностью автоматизировать построение экспертных данных на практике невозможно в силу того, что некоторые принимаемые экспертами решения не поддаются алгоритмизации. По этой причине эксперты применяют средства манипуляции данными, которые можно частично автоматизировать, опираясь на анализ алгебры операций, характерной для манипуляций данными в конкретной предметной области [2].

В статье рассмотрены манипуляции с данными, основанные на использовании пары унарных операций, а также подход к созданию настраиваемого под структуры данных количественной спектроскопии программного модуля «унарные операции».

Постановка задачи

В количественной спектроскопии значительный объем исследований связан с измерениями и вычислениями. Предсказанные переходы являются результатами теоретических расчетов (глобальная подгонка или *ab initio* расчеты) и характеризуются набором значений физи-

ческих величин. Измеренные переходы состоят из характеристик переходов, определенных при обработке экспериментальных данных и эталонных переходов, вычисленных из уровней энергии молекул по правилу Ридберга – Ритца. Необходимым условием для использования в приложениях первичных предсказанных и измеренных данных является их согласование. Последнее не всегда достижимо в полном объеме и создает трудности для алгоритмизации процесса построения экспертных данных.

В зависимости от решаемых задач в спектроскопии используются разные структуры данных. Структура экспертных спектральных данных, относящихся к процессам поглощения, формировалась последние сорок лет (см. последние версии [3; 4]) и ориентирована на их использование в задачах переноса излучения. В работах [5]¹ предложен формат передачи спектральных данных, а в [6] – структуры спектральных данных для изучения свойств решений шести задач спектроскопии, которые применимы для построения экспертных наборов данных разного назначения.

Использование упомянутых в [6] структур данных означает, что полученные в результате экспериментов уровни энергии, вакуумные волновые числа и параметры профилей линий представляют собой набор решений трех обратных задач спектроскопии, а предсказанные (рассчитанные) уровни энергии, вакуумные волновые числа и параметры профилей линий – набор решений трех прямых задач спектроскопии. Поскольку структуры решений прямых и обратных задач подобны, то необходимо рассмотреть операции манипуляции с данными только трех структур.

Решение каждой задачи в ИС «Молекулярная спектроскопия» представляет собой независимый информационный объект, называемый набором данных, содержащий данные из первичных или составных источников данных. Манипуляции над наборами данных являются сложными операциями, которые строятся путем комбинирования базовых операций. Базовые операции делятся на унарные и бинарные по количеству массивов данных, участвующих в операции. С учетом особенностей предметной области к базовым унарным операциям относятся операции:

- 1) выбор / удаление данных, удовлетворяющих условию;
- 2) выбор / удаление атрибутов из массива данных.

Необходимо рассмотреть применимость этих операций к данным молекулярной спектроскопии с учетом особенностей предметной области.

После того как рассмотрены особенности структур данных и особенности выполнения унарных операций, требуется создать программный модуль, который позволит автоматизировать часть неформализуемых операций при составлении экспертных массивов данных в молекулярной спектроскопии.

Для достижения полного доверия по критерию опубликования [1] к созданным экспертным данным необходимо, чтобы создание таких экспертных наборов данных основывалось на унарных операциях только с наборами данных первичных источников, их частями или наборами, производными от первичных данных. В остальном пользователь не ограничен в применении унарных операций.

Ввиду наличия данных, содержащих решения прямых и обратных спектроскопических задач по уровням энергии, спектральным переходам и профилям спектральных линий для достаточно большого количества веществ и их изотопологов, требуется создать программный модуль для выполнения унарных операций, имеющий возможность настройки под соответствующие структуры данных, так как создание большого количества отдельных модулей со схожей функциональностью требует больших трудозатрат. Настройка модуля должна позволять распространять его функциональность на данные, которыми будет пополняться ИС по веществам, ранее в ней не представленным.

Структура спектральных данных в количественной спектроскопии

Как упоминалось ранее, структура спектральных данных зависит от того, для решения каких задач они предназначены. В экспертных банках данных [3; 4] структура данных опреде-

¹ The Virtual Atomic and Molecular Data Centre (VAMDC). URL: <http://www.vamdc.eu>

ляется структурой входных данных для программ расчета потоков излучения в атмосферах планет и представляется набором строк. В задачах расчета эталонных уровней энергии и вакуумных волновых чисел используется структура данных, характеризующая решения этих двух задач спектроскопии [7; 8].

Решение обратной задачи описания состояний молекулы включает значения уровней энергии и квантовые числа состояний. Обратная задача описания переходов изолированной молекулы в качестве решения имеет вакуумные волновые числа и квантовые числа переходов. Решения обратных задач описания переходов молекулы в газовой среде характеризуются квантовыми числами и набором значений параметров контура спектральной линии. Физические величины, входящие в решение первых двух задач, обязательно должны присутствовать в них. Решения третьей задачи должны содержать квантовые числа переходов и хотя бы один из параметров контура спектральной линии.

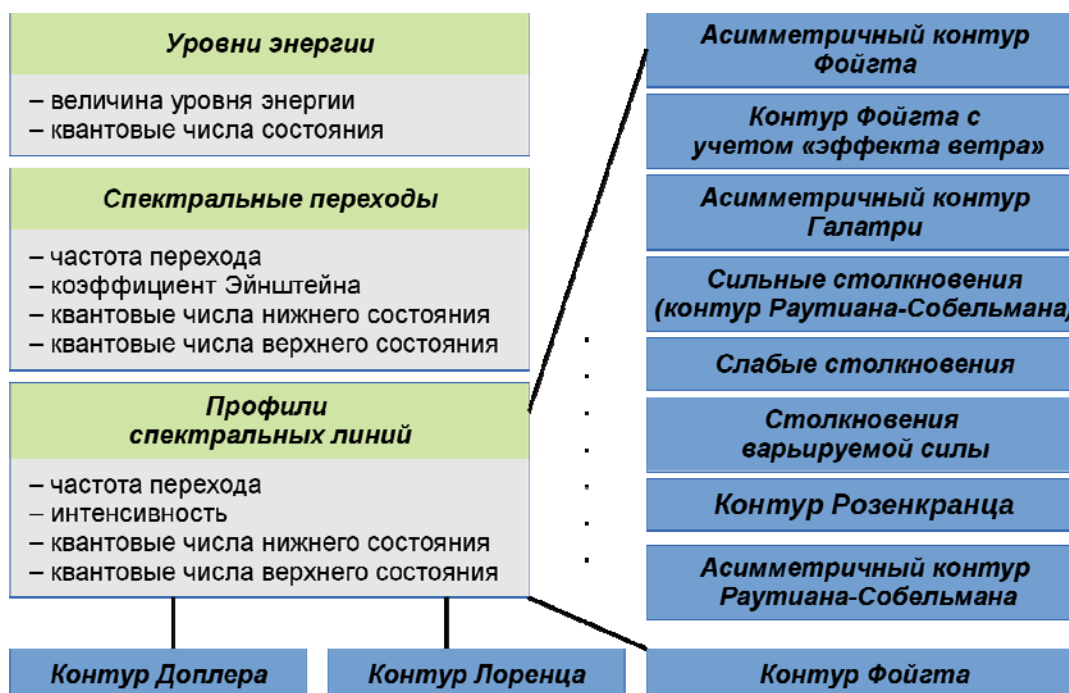


Рис. 1. Структуры данных решений спектроскопических задач

Рассматриваемые нами прямые и обратные задачи спектроскопии имеют сходные структуры данных. Следовательно, достаточно рассмотреть операции манипуляции с данными только трех структур (рис. 1).

Большая часть физических величин, входящих в структуру спектральных данных, не зависит от выбора конкретной молекулы. Квантовые числа состояния молекулы являются единственным набором физических величин, которые меняются при переходе от молекулы одного спектрального класса молекул к другому, а также от n -атомной к m -атомной молекуле. Среди наборов квантовых чисел выделяют подмножества, определяющие так называемые спектральные полосы. Эти подмножества также меняются в зависимости от конкретной молекулы.

Квантовые числа зависят от физического процесса, в котором участвует вещество, и его электронного состояния. Мы рассматриваем квантовые числа в нормальных модах, как наиболее часто используемые для веществ в условиях, близких к атмосферным и находящимся в основном электронном состоянии. В таких квантовых числах выделяют две части: колебательные и колебательно-вращательные квантовые числа. Колебательные квантовые числа являются подмножеством квантовых чисел, определяющих спектральные полосы переходов.

Мы используем квантовые числа, предложенные в проекте VAMDC для процессов поглощения молекул в основном электронном состоянии[5]².

В качестве примера различий в квантовых числах у разных веществ возьмем состояния молекул основных изотопологов CO₂ и H₂O. Квантовые числа, описывающие состояние CO₂, обозначаются как $\nu_1, \nu_2, l_2, \nu_3, r, \text{Sym}, J$. Первые пять являются колебательными квантовыми числами. Для H₂O квантовые числа следующие: $\nu_1, \nu_2, \nu_3, J, K_a, K_c$, а колебательными являются первые три.

Определение унарных операций в количественной спектроскопии

Для формального определения операций воспользуемся аппаратом алгебры кортежей (АК) [9]. Приведем определения используемых понятий.

Алгебра кортежей – алгебраическая система, носитель которой – совокупность многоместных отношений, выраженных в специфических структурах (элементарный кортеж, C-кортеж, C-система, D-кортеж, D-система), называемых АК-объектами.

Под *атрибутом* понимается имя некоторого свойства системы или ее части, представленное множеством непосредственно заданных или вычисляемых значений (*доменом*). Возможны случаи, когда одному и тому же домену соответствует несколько атрибутов. Тогда такие атрибуты относятся к одному *сорту*. Если задано некое отношение, то его определяющий признак – совокупность присущих ему атрибутов. Последовательность этих атрибутов в АК называется *схемой отношения*.

Элементарный кортеж – это последовательность элементов, каждый из которых есть элемент *домена* соответствующего *атрибута* из схемы отношения. Например, если задан элементарный кортеж $T[XYZ] = (a, b, c)$, то подразумевается, что $a \in X$, $b \in Y$ и $c \in Z$. В АК элементарные кортежи принадлежат множествам (отношениям), выраженным другими типами АК-объектов.

Компонентами называются множества, представляющие собой подмножества доменов соответствующих атрибутов. Среди компонент имеются две *фиктивные* компоненты: полная и пустая компоненты. В этой работе будет использоваться только одна – *полная* компонента. Полная компонента «*» равна домену соответствующего (по месту ее расположения в кортеже) атрибута.

C-кортежем называется заданный в определенной схеме отношения кортеж компонент. C-кортеж эквивалентен некоторому множеству элементарных кортежей – это множество можно перечислить, если вычислить декартово произведение компонент C-кортежа. Для изображения C-кортежей используются прямые скобки. Например, $R[XYZ] = [ABC]$ означает, что $A \subseteq X$, $B \subseteq Y$, $C \subseteq Z$ и $R[XYZ] = [A \times B \times C]$.

C-системой называется множество однотипных C-кортежей, которые записываются в виде матрицы, ограниченной прямыми скобками. Строки этой матрицы содержат C-кортежи:

$$Q[XYZ] = \begin{bmatrix} A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{bmatrix}.$$

C-система эквивалентна множеству элементарных кортежей. Это множество равно объединению множеств элементарных кортежей, принадлежащих соответствующим C-кортежам. Например, C-системе можно представить как множество элементарных кортежей, вычисляемое по формуле $Q[XYZ] = (A_1 \times B_1 \times C_1) \cup (A_2 \times B_2 \times C_2)$.

Проекция создает из отношения $Q[X']$ отношение $Q_p[X' \setminus Y']$, где для множеств атрибутов X' и Y' справедливо $Y' \subset X'$.

Для операции *селекции* необходимо задать ограничения на значения одного или нескольких произвольных атрибутов схемы отношения, чтобы затем выбрать кортежи, удовлетворяющие этим ограничениям. Пусть ограничение на значения любого атрибута задается в виде условий $X\Delta c$, связанных операторами конъюнкции и дизъюнкции, в комплексе отра-

² The Virtual Atomic and Molecular Data Centre (VAMDC). URL: <http://www.vamdc.eu>

жающих требуемое ограничение на значение атрибута, где c – элемент или множество элементов сорта, соответствующего данному атрибуту, Δ – один элемент из множества отношений сравнения $\{<, \leq, =, \neq, \geq, >\}$. Применимость отношений сравнения к конкретному атрибуту определяется типом домена этого атрибута. Обозначим ограничение на значения атрибута с доменом X как X_c . Тогда ограничение можно выразить как C -кортеж, у которого выбранные атрибуты X_i представлены элементами или подмножествами соответствующих сортов (соответствующими условиям X_{i_c}), а остальные атрибуты данной схемы отношения – фиктивными компонентами. Например, для отношения $P[XYZV]$ C -кортеж $Q = [* \{a\} * \{d, f\}]$ является селектором с ограничениями $(Y = a)$ и $(V = d \text{ или } V = f)$. Селекция данных по таким ограничениям формируется при пересечении этого C -кортежа с отношением P . Операция удаления сводится к пересечению $P \cap \bar{Q}$. Если результат селекции обозначить как $R_s = P \cap Q$, а результат удаления как $R_d = P \cap \bar{Q}$, то $R_d \cup R_s = P$. Таким образом, из операций селекции и удаления достаточно одной, так как результат второй можно получить, инвертировав условие.

С точки зрения АК таблицы в СУБД могут быть представлены как множество элементарных кортежей или как C -система, где все компоненты есть одноэлементные множества. Для простоты рассмотрим случай, когда все данные представлены в одной таблице.

Теоретические понятия АК согласуются с практикой использования СУБД. Строка в таблице базы данных соответствует элементарному кортежу. C -система, как множество элементарных кортежей, соответствует содержимому таблицы базы данных или результату SQL-запроса. Условия на значения атрибутов кортежа в селекции соответствуют условиям на значения элемента строки таблицы в SQL-запросе. Проекция в СУБД определяется перечнем колонок таблицы в SQL-запросе.

Как упоминалось ранее, нам необходимы две унарные операции – проекция и селекция. Применение операции селекции, описанной выше, к данным количественной спектроскопии не имеет ограничений, накладываемых предметной областью. В то время как применение операции проекции имеет ограничения, так как некоторые атрибуты являются обязательными, например, квантовые числа, а другие не имеют смысла без базового атрибута. Так, величина ошибки полуширины не имеет смысла без значения самой полуширины. Также ограничением на применение операций селекции и проекции можно назвать необходимость сохранения неизменным исходного набора данных, т. е. в результате применения любой унарной операции должен получаться новый набор данных.

Реализация модуля «Унарные операции»

Описанные выше унарные операции над наборами данных реализованы в виде отдельного модуля ИС «Молекулярная спектроскопия» (<http://www.saga.iao.ru>). Прототип модуля унарных операций, на примере данных CO_2 и задачи переходов в изолированной молекуле, был описан в [10]. В качестве дальнейшего развития ИС мы предполагали расширение возможностей этого модуля в сторону работы с данными по различным веществам и их изотопологам. Ввиду необходимости выполнения манипуляций со спектральными данными, относящимися к решениям разных задач количественной спектроскопии, требования к функционалу модуля были расширены работой с данными различных структур, которые определяются соответствующими задачами, рассмотренными ранее.

Таким образом, универсальность модуля охватывает три различные структуры данных с изменчивой частью в виде квантовых чисел для разных веществ. Далее описывается подход, позволивший реализовать такую универсальность.

Как было сказано, данные различных задач имеют свою структуру (см. рис. 1). В дополнение к задаче структура данных определяется квантовыми числами, присущими конкретному веществу. Таким образом, комбинация задачи и вещества определяет множество элементов структуры данных, соответствующих квантовым числам, физическим характеристикам и другим данным. Обычно знания о них используются в качестве априорных при написании программы, что приводит к жесткой привязке элементов структуры данных и связанных с ними обработчиков в программе и не позволяет использовать программу для других струк-

тур данных. Этого можно избежать, если связь элементов структуры данных и соответствующих им обработчиков указать в качестве свойств этих элементов, создав отдельное описание структуры данных вне программы.

Для конкретной задачи и вещества нам известны как элементы структуры, так и действия, которые выполняются над ними. Количество видов действий при выполнении унарных операций ограничено (конечно и невелико). Соответственно, можно вынести признаки таких действий в свойства элементов структуры. Таким образом, подразумеваемые знания об элементах указываются явно и являются настройкой (конфигурационной информацией) для работы программы. Степень детализации и уровень абстракции при выделении таких свойств определяются функциональностью разрабатываемого модуля, и для получения искомой универсальности далеко не все действия требуется выносить в настраиваемые свойства элементов структуры. Для реализации унарных операций над наборами спектроскопических данных были выделены свойства, необходимые для сопоставления элементов структуры данных и полей в таблицах базы данных, преобразования квантовых чисел в человекочитаемый вид, группировки квантовых чисел при выборке данных по спектральной полосе, задания типа условия, накладываемого на данные при селекции данных из набора данных (задание диапазона значения величины или выбор значения из списка), и некоторые другие.

В программах модуля, использующего такое описание структуры данных в качестве настроек, не заложено информации о том, что какой-то конкретный элемент множества является, например, квантовым числом. Программы модуля перебирают элементы конфигурационного массива и выполняют действия, соответствующие значениям свойств этих элементов. Таким образом, мы получили требуемую степень универсальности и, что более важно для нас, возможность расширения области применимости разработанного модуля дополнением элементов конфигурационного массива данными, описывающими новые структуры данных при наполнении ИС данными по новым веществам.

Интерфейс созданного модуля представлен на рис. 2–4. На первом этапе (см. рис. 2) пользователь системы выбирает тот источник данных, над набором данных которого будут производиться унарные операции. Источник выбирается из списка источников данных, сформированного по заданным пользователем типу спектроскопической задачи, веществу или изотопологу и ключевым словам для контекстного поиска по аннотациям публикаций.

На втором этапе (см. рис. 3) пользователь формирует набор условий для значений тех физических величин, которые входят в решение выбранной им на первом этапе спектроскопической задачи, и составляет логическое выражение из набора условий с помощью логических операторов «И», «ИЛИ» и «НЕ», которое отражает общее ограничение, накладываемое на набор данных.

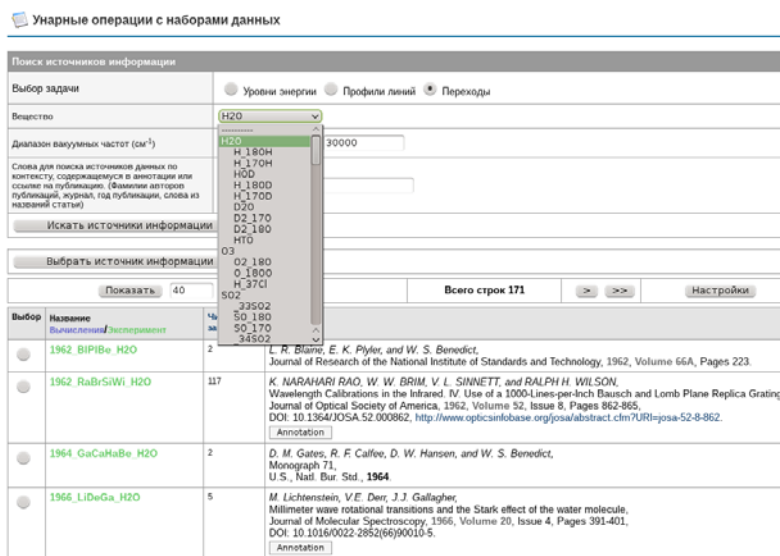


Рис. 2. Интерфейс выбора набора данных

Условия для унарных операций

Shanshan Yu, John C. Pearson, Brian J. Drouin, Marie-Aline Martin-Drumel, Olivier Pirali, Michel Vervloet, Laurent H. Coudert, Holger S.P. Muller, Sandra Brunken.
 2012_YuPeDrMa_H2O
 Measurement and analysis of new terahertz and far-infrared spectra of high temperature water,
 Journal of Molecular Spectroscopy, 2012, Volume 279, Pages 16–25,
 DOI: 10.1016/j.jms.2012.07.011, http://dx.doi.org/10.1016/j.jms.2012.07.011.

Annotation

Выбор ограничений	
Физические величины	Ограничения
Колебательная полоса	0 2 0 - 0 2 0
Полный угловой момент нижнего состояния $J_{\min} < J < J_{\max}$	1 20 (1 - 26)
Вакуумные волновые числа $\omega_{\min} < \omega < \omega_{\max}$	400 (9.85717 - 687.61435)
Неопределенность вакуумных волновых чисел $\Delta\omega_{\min} < \Delta\omega < \Delta\omega_{\max}$	(0.0008 - 0.02)

Сохранить ограничение

Таблица ограничений				
Имя ограничения	Колебательная полоса	Полный угловой момент нижнего состояния (J)	Вакуумные волновые числа (ω)	Неопределенность вакуумных волновых чисел ($\Delta\omega$)
C0	0 0 0 1 0			
C1	0 2 0 0 2 0	1 - 20	400 -	

Очистить таблицу ограничений

Формирование условий и выполнение унарных операций

Условие: C0 OR C1

--- NOT OR AND BACK

Данные удовлетворяющие сформированному условию:

Выбрать Исключить

C0
C1

Рис. 3. Интерфейс задания условий для унарной операции

Показать 10 строк от 0		Всего строк 629		Настройки	
исключить/оставить	Квантовые числа	Вакуумные волновые числа (ω)	Неопределенность вакуумных волновых чисел ($\Delta\omega$)	Выбранные колонки	
<input checked="" type="checkbox"/>	0 2 0 7 2 5 0 2 0 7 5 2	401.14868	0.0008	<input checked="" type="checkbox"/>	исключить/оставить
<input checked="" type="checkbox"/>	0 2 0 17 3 14 0 2 0 18 4 15	401.8306	0.0008	<input checked="" type="checkbox"/>	Квантовые числа
<input type="checkbox"/>	0 2 0 8 4 5 0 2 0 9 5 4	402.92637	0.0008	<input checked="" type="checkbox"/>	Вакуумные волновые числа (ω)
<input type="checkbox"/>	0 2 0 6 6 0 0 2 0 7 7 1	405.03481	0.0008	<input checked="" type="checkbox"/>	Неопределенность вакуумных волновых чисел ($\Delta\omega$)
<input type="checkbox"/>	0 2 0 7 5 3 0 2 0 8 6 2	406.54199	0.0008	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	0 2 0 18 5 14 0 2 0 19 4 15	407.79716	0.0008	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	0 2 0 19 3 17 0 2 0 20 2 18	409.62144	0.0008	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	0 2 0 20 1 19 0 2 0 21 2 20	409.69432	0.0008	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	0 2 0 20 2 19 0 2 0 21 1 20	409.80903	0.0008	<input type="checkbox"/>	
<input checked="" type="checkbox"/>	0 2 0 19 2 17 0 2 0 20 3 18	410.08299	0.0008	<input type="checkbox"/>	

Показать 10 строк от 0

Всего строк 629

Настройки

Выбор

Сохранение источника данных в ИС с учётом выбранных колонок и строк.

Название сохраняемого источника данных:
 NewSet from 2012_YuPeDrMa_H2O

Сохранить созданный источник данных в ИС

Рис. 4. Интерфейс просмотра результатов унарной операции и сохранения нового источника данных

Далее производится операция выбора или удаления, в зависимости от решения пользователя, тех данных из набора данных, которые удовлетворяют сформированному ограничению. Результат операции выводится пользователю в табличном виде (см. рис. 4), где в первой колонке и первой строке имеются элементы, позволяющие производить выбор колонок (проекция) и дополнительный выбор строк для сохранения результата в новый набор данных.

Разработанное программное обеспечение выполнено с применением языка программирования PHP. В качестве веб-сервера используется Apache. Хранение и обработка данных в ИВС осуществляется с помощью реляционной СУБД MySQL.

Заключение

В работе представлен подход к созданию универсального модуля, выполняющего унарные операции над наборами данных, содержащими решения таких прямых и обратных спектроскопических задач, как уровни энергии, спектральные переходы и профили спектральных линий, для различных веществ. Дано краткое описание созданных интерфейсов ИС для выполнения унарных операций, позволяющих исключать или выбирать данные из исходного набора данных, удовлетворяющие условиям, наложенным пользователем. Результат выполнения операций сохраняется в новый набор данных, обеспечивая сохранность исходного набора данных.

Следующими направлениями развития ИС «Молекулярная спектроскопия» в области манипуляций со спектроскопическими данными предполагается создание модуля бинарных операций, основанного на аналогичных принципах.

Список литературы

1. *Лаврентьев Н. А., Макогон М. М., Фазлиев А. З.* Сравнение спектральных массивов данных HITRAN и GEISA с учетом ограничения на опубликование спектральных данных // Оптика атмосферы и океана. 2011. Т. 24, № 4. С. 279–292.
2. *Козодоев А. В., Фазлиев А. З.* Информационная система для решения задач молекулярной спектроскопии. 2. Операции преобразования наборов параметров спектральных линий // Оптика атмосферы и океана. 2005. Т. 18, № 9. С. 760–764.
3. *Jacquinot-Husson N., Crepeau L., Armante R. et al.* The 2009 edition of the GEISA spectroscopic database // Journal of Quantitative Spectroscopy and Radiative Transfer. 2011. Vol. 112, iss. 15. P. 2395–2445.
4. *Rothman L. S., Gordon I. E., Barbe A. et al.* The HITRAN 2008 molecular spectroscopic database // Journal of Quantitative Spectroscopy and Radiative Transfer. 2009. Vol. 110, iss. 9. P. 533–572.
5. *Dubernet M.-L., Boudon V., Culhane J. L. et al.* Virtual Atomic and Molecular Data Centre // Journal of Quantitative Spectroscopy and Radiative Transfer. 2010. Vol. 111. P. 2151–2159. doi:10.1016/j.jqsrt.2010.05.004
6. *Bykov A. D., Fazliev A. Z., Filippov N. N., Kozodoev A. V., Privezentsev A. I., Sinitsa L. N., Tonkov M. V., Tretyakov M. Yu.* Distributed information system on atmospheric spectroscopy // Geophysical Research Abstracts, SRef-ID: 1607-7962/gra/EGU2007-A-01906. 2007. Vol. 9. P. 01906.
7. *Tashkun S. A., Perevalov V. I., Teffo J.-L.* Global Fittings of the Vibrational – Rotational Line Positions of the $^{16}\text{O}^{12}\text{C}^{17}\text{O}$ and $^{16}\text{O}^{12}\text{C}^{18}\text{O}$ Isotopic Species of Carbon Dioxide // J. Mol. Spectrosc. 2001. Vol. 210. P. 137–145.
8. *Tibor Furtenbacher, Attila G. Császár.* MARVEL: Measured active rotational-vibrational energy levels. II. Algorithmic improvements // Journal of Quantitative Spectroscopy and Radiative Transfer. 2012. Vol. 113, iss. 11. P. 929–935.
9. *Кулик Б. А., Зуенко А. А., Фридман А. Я.* Алгебраический подход к интеллектуальной обработке данных и знаний. СПб.: Изд-во Политехн. ун-та, 2010. 235 с.
10. *Козодоев А. В.* Унарные операции над источниками данных в информационной системе CaD@DIS // XIV Рос. конфю с участием иностранных ученых «Распределенные информационные и вычислительные ресурсы» (DICR-2012): Материалы конф. (Новосибирск, Россия, 26–30 ноября 2012). Новосибирск: ИВТ СО РАН, 2012. URL: <http://conf.nsc.ru/dicr2012/ru/reportview/141076>

A. V. Kozodoev, E. M. Kozodoeva

*V. E. Zuev Institute of Atmospheric Optics of Siberian Branch of the Russian Academy of Science
1 Academician Zuev Square, Tomsk, 634021, Russian Federation*

kav@iao.ru, klen@iao.ru

**EXTENSIBLE MODULE «UNARY OPERATIONS» IN THE INFORMATION SYSTEM
«MOLECULAR SPECTROSCOPY»**

The implementation of the module which performs unary operations on the data in the IS «Molecular spectroscopy» is described in the paper. Given a brief description of unary operations, taking into account characteristics of the subject area. Describes the approach used in the software implementation that allows you to use one set of scripts to work with existing databases on various substances, and several types of spectroscopic data.

Keywords: data structure, quantitative spectroscopy, unary operations, database.

References

1. Lavrent'ev N. A., Makogon M. M., Fazliev A. Z. Comparison of HITRAN and GEISA spectral data, based on taking into account the existent constraints. *Atmospheric and oceanic optics*, 2011, vol. 24, no. 4, p. 279–292 (in Russ.).
2. Kozodoev A. V., Fazliev A. Z. Information system for molecular spectroscopy. 2. Array operations for transformation of data on spectral line parameters. *Atmospheric and oceanic optics*, 2005, vol. 18, no. 09, p. 680–684.
3. Jacquinet-Husson N., Crepeau L., Armante R., et al. The 2009 edition of the GEISA spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2011, vol. 112, iss. 15, p. 2395–2445.
4. Rothman L. S., Gordon I. E., Barbe A. et al. The HITRAN 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2009, vol. 110, iss. 9, p. 533–572.
5. Dubernet M.-L., Boudon V., Culhane J. L., et al. Virtual Atomic and Molecular Data Centre. *Journal of Quantitative Spectroscopy and Radiative Transfer*. 2010, vol 111, p. 2151–2159.
6. 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. *Geophysical Research Abstracts*, 2007, vol. 9, p. 01906.
7. Tashkun S. A., Perevalov V. I., Teffo J.-L. Global Fittings of the Vibrational – Rotational Line Positions of the $^{16}\text{O}^{12}\text{C}^{17}\text{O}$ and $^{16}\text{O}^{12}\text{C}^{18}\text{O}$ Isotopic Species of Carbon Dioxide. *J. Mol. Spectrosc.*, 2001, vol. 210, p. 137–145.
8. Furtenbacher T., Császár A. G., MARVEL: Measured active rotational–vibrational energy levels. II. Algorithmic improvements. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 2012, vol. 113, iss. 11, p. 929–935.
9. Kulik B. A., Zuenko A. A., Fridman A. Ya. Algebraic approach to intelligent processing of data and knowledge, St. Petersburg Polytechnic University Publishing House, 2010.
10. Kozodoev A. V. Unary operations on data sources in the in the information system CaD@DIS. DICR-2012 conference proceedings, Novosibirsk, Russia, 2012, ICT SB RAS. URL: <http://conf.nsc.ru/dicr2012/ru/reportview/141076>