

УНАРНЫЕ ОПЕРАЦИИ НАД ИСТОЧНИКАМИ ДАННЫХ В ИНФОРМАЦИОННОЙ СИСТЕМЕ CaD@DIS

Козодоев А.В.

Институт оптики атмосферы им. В.Е. Зуева СО РАН, Томск, Россия

kav@iao.ru

Аннотация

В работе описывается подход использованный при разработке и реализации модуля выполняющего унарные операции над источниками данных в ИВС «Молекулярная спектроскопия». Приводится описание унарных операций с учётом особенностей предметной области.

Введение

Важную роль в молекулярной спектроскопии играют экспертные массивы данных, которые включают данные как о переходах предсказанных расчётами, так и эталонные экспериментальные данные. Существующие экспертные массивы имеют противоречия из-за недостоверных и неопубликованных данных [1]. Таких противоречий можно избежать если формировать экспертные массивы на основе выверенных наборов предсказанных переходов и эталонных экспериментальных данных. Которые, в свою очередь, формируются из первичных источников данных [2]. Процесс формирования состоит из операций преобразования наборов данных [3], которые делятся на бинарные и унарные по числу наборов данных участвующих в операции.

Для реализации вышеупомянутых операций в рамках информационно-вычислительной системе (ИВС) требуется создание соответствующего программного обеспечения. Первым шагом в этом направлении является реализация унарных операций, позволяющих исключить данные, не удовлетворяющие требованиям пользователя. Далее описывается подход использованный для создания модуля унарных операций в ИВС CaD@DIS, которая является частью ИВС «Молекулярная спектроскопия».

Операции над данными в молекулярной спектроскопии

При работе с массивами спектроскопических данных эксперт часто выполняет операции, которые можно описать как применение набора ограничений к величинам имеющимся в этих массивах. Набор таких операций для данных в молекулярной спектроскопии ограничен и определён спецификой предметной области [3]. В рамках данной предметной области используются три типа ограничений: ограничения на спектральную полосу, ограничения на вращательную часть идентификации спектрального перехода и ограничения на физические величины (частота перехода, коэффициент Эйнштейна и др.).

При описании операций с данными будем придерживаться терминологии принятой для описания ER-модели [4].

К числу унарных операций относятся выборка и проекция. Выборка – выбор из отношения кортежей, удовлетворяющих заданному условию. Новое отношение будет иметь

ту же степень, а кардинальность будет меньше или равной кардинальности исходного отношения. Проекция – выбор из кортежей значений определённых атрибутов. При этом в зависимости от конкретной задачи спектроскопии удаление дубликатов строк может не производиться. Новое отношение будет иметь такую же кардинальность, а степень меньше или равную степени исходного отношения.

Унарные операции применимы к источникам данных, содержащих строки произвольной структуры (в рамках понятий, связанных с параметрами спектральных линий).

Определение унарных операций в ИС

Ограничения накладываемые, в рамках данной предметной области, на физические величины носят интервальный характер, т.е. задаются минимальное и максимальное значения параметра, причем границы интервалов включаются ($0 \leq J \leq 5$). Если при формировании набора ограничений пользователь не указал, для какой спектральной полосы они применимы, то ограничения накладываются для всех полос, входящих в состав источника данных (ИД). В предметной области имеется два типа ограничений – ограничения на ИД и на полосу (часть ИД).

Для составления ограничений накладываемых на данные пользователю доступно два логических операции – конъюнкция (AND или &) и инверсия (NOT или \neg), операция дизъюнкции (OR или +) исключена, т.к. при формировании ограничений на полосу, ИВС предоставляет полную информацию о выбранной полосе, следовательно, наложенные ограничения всегда будут верны, с точки зрения их корректности для рассматриваемого источника информации, и, следовательно, применение данной операции теряет свой смысл. Истинность элементарных высказываний, формируемых пользователем, обусловлена проверками интервальных значений, входящих в состав реализованного программного обеспечения (ПО).

Введём оператор Delete[B], где B логическое высказывание сформированное пользователем. Выполнение данного оператора будет приводить к удалению заданных спектральных полос, с наложенными на них ограничениями, из источника данных, или удалению кортежей не удовлетворяющих ограничениям значений физической величины. Пусть a_i и a_j – спектральные полосы из источника данных A , $a_i, a_j \subseteq A$, $a_i \cap a_j = \emptyset$, n – элементарное ограничение, состоящие из отношений $<$, $>$ либо $=$. Обозначим $\{n_0\}_{a_i}$ – логическое высказывание для i -ой полосы. Тогда $\text{Delete}[\{n_0\}_{a_i}] = A \setminus a_i^{n_0}$, где $a_i^{n_0}$ – подмножество полосы a_i удовлетворяющие ограничениям n_0 , \setminus – операция дополнения. Полученный результат будем трактовать следующим образом – “Удалить из источника данных A полосу a_i с наложенным на нее ограничением n_0 ”. Введем $\{N\}_{a_i} = \{n_0 \& \dots \& n_k\}_{a_i}$ – логическое высказывание для i -ой полосы, содержащее набор элементарных ограничений. На основании выдвинутых суждений сформулируем ряд основных правил, используемых в ИС при выполнении унарных операций:

1. $\text{Delete}[\{N\}_{a_i} \& \{M\}_{a_j}] = (A \setminus a_i^N) \setminus a_j^M$
2. $\text{Delete}(\neg \{N\}_{a_i}) = a_i^N$
3. $\text{Delete}(\neg \{N\}_{a_i} \& \neg \{M\}_{a_j}) = (a_i^N \cup a_j^M)$

$$4. \text{Delete}(\neg\{N\}_{a_i} \& \{M\}_{a_i}) = a_i^N$$

Эти правила позволяют реализовать унарную операцию удаления кортежей.

Реализация модуля унарных операций в ИВС

Описанная выше унарная операция над источниками данных реализована в виде отдельного модуля ИВС коллективного использования «Молекулярная спектроскопия» (<http://www.saga.iao.ru>). При разработке программного обеспечения учитывалось два основных аспекта: создание интуитивно понятного для пользователя интерфейса и реализации программной части таким образом, чтобы выполнение большей части функций модуля производилось на стороне клиента Web браузером, что позволяет снизить нагрузку на сервер.

Необходимо отметить, что при выполнении операций исходные данные и источник информации, не модифицируется, а происходит копирование данных в новый пользовательский источник информации.

Унарные операции над данными

2003_TaPeTeBy_a_CO2	S.A. Tashkun, V.I. Perevalov, J.-L. Teffo, A.D. Bykov, N.N. Lavrentieva, CDSD-296, The carbon dioxide spectroscopic databank: version for atmospheric applications, Abstracts of the XIV Symposium on High Resolution Molecular Spectroscopy, Krasnoyarsk, Russia, July 6-11, 2003, IAO SB RAS, Tomsk, 2003,
---------------------	--

Выбор ограничений и выполнение унарных операций								
Физические величины				Ограничения				
Колебательная полоса				0 0 0 1 1 ----- 0 0 0 0 1				
Ветвь Br	Полный угловой момент $J_{\min} < J < J_{\max}$	Симметрия ϵ						
Вакуумные волновые числа $\omega_{\min} < \omega < \omega_{\max}$								
Неопределенность вакуумных волновых чисел $\Delta\omega_{\min} < \Delta\omega < \Delta\omega_{\max}$								
Коэффициент Эйнштейна $E_{\min} < E < E_{\max}$								
Неопределенность коэффициента Эйнштейна $\Delta E_{\min} < \Delta E < \Delta E_{\max}$								
Ok								

Таблица ограничений								Очистить таблицу ограничений
Ограничение(я)	Колебательная полоса	Ветвь (Br)	Полный угловой момент (J)	Симметрия (ε)	Вакуумные волновые числа (ω)	Неопределенность ω (Δω)	Коэффициент Эйнштейна (E)	Неопределенность E (ΔE)
Band_1	0 0 0 1 1 ----- 0 0 0 0 1	-	0<J<10	-	-	-	-	-
Band_2	0 0 0 1 1 ----- 0 0 0 0 1	-	-	-	2340<W<2350	-	-	-
Band_3	0 0 0 1 1 ----- 0 0 0 0 1	-	0<J<10	-	2340<W<2350	-	-	-

DELETE(Band_3	NOT	And	Применить ограничения
NOT Band_3	Удалить высказывания в DELETE			

Рисунок 1: Интерфейс выбора условий унарной операции

В качестве примера работы со спектральными переходами на рисунке 1 представлен интерфейс позволяющий формировать условия используемые в унарных операциях. В этой части система позволяет задавать ограничения на полосу или на значения физических

величин. Затем строится структура высказываний с помощью операций NOT и OR в операторе DELETE. После выполнения оператора DELETE результат можно просмотреть (рис.2) и, при желании, сохранить в новом источнике информации.

DELETE(NOT Band_3)																								
Показать 60 строк от 0										Всего строк 7			Настройки											
ν_1^{up}	ν_2^{up}	ν_3^{up}	ν_4^{up}	ν_5^{up}	ν_6^{up}	ν_7^{up}	ν_1^{low}	ν_2^{low}	ν_3^{low}	ν_4^{low}	ν_5^{low}	ν_6^{low}	ν_7^{low}	ν_8^{low}	ν_9^{low}	ε	Вакуумные волновые числа (ω)	Неопределенность ω (Δω)	Коэффициент Эйнштейна (E)	Неопределенность E (ΔE)				
<input type="checkbox"/>	0	0	0	1	1	P	0	0	0	0	1	10	e				2341.061784	<input type="checkbox"/>	<input type="checkbox"/>	0.04949	<input type="checkbox"/>			
<input type="checkbox"/>	0	0	0	1	1	P	0	0	0	0	1	8	e				2342.727032			0.04887				
<input type="checkbox"/>	0	0	0	1	1	P	0	0	0	0	1	6	e				2344.367764			0.04789				
<input type="checkbox"/>	0	0	0	1	1	P	0	0	0	0	1	4	e				2345.983954			0.04608				
<input type="checkbox"/>	0	0	0	1	1	P	0	0	0	0	1	2	e				2347.575575			0.04143				
<input type="checkbox"/>	0	0	0	1	1	R	0	0	0	0	1	0	e				2349.916884			0.1035				
Показать 60 строк от 0										Всего строк 7			Настройки											
Выполнение операций над строками и столбцами. Сохранение источника данных в ИС.															Удалить строку(и) или столбец(ы)					Сохранить созданный источник данных в ИС				

Рисунок 2: Интерфейс просмотра результатов унарной операции

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

Заключение

В работе представлен подход к созданию модуля выполняющего унарную операцию удаления. Дано краткое описание созданных приложений ИВС для выполнения унарных операций, позволяющих исключать данные, не удовлетворяющие требованиям пользователя.

Разработанный интерфейс и программное обеспечение позволяет пользователям в рамках ИВС составлять собственные источники данных с использованием операции удаления. Стоит отметить, что введенная процедура составления собственных источников данных, обеспечивает выполнение жесткого условия о сохранности первичных источников информации.

Следующими направлениями развития ИВС «Молекулярная спектроскопия» в области манипуляций со спектроскопическими данными будет дальнейшее развитие модуля унарных операций для работы с другими веществами и создание модуля бинарных операций.

Авторы благодарны РФФИ за финансирование работы (грант 11-07- 00660-а).

ЛИТЕРАТУРА

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, 24, 4, 279-292.

2. Быков А.Д., Науменко О.В., Сеница Л.Н. и др. Информационные аспекты молекулярной спектроскопии, Томск, Из-во ИОА СОРАН, 2008, 356С.
3. Козодоев А.В., Фазлиев А.З. Информационная система для решения задач молекулярной спектроскопии. 2. Операции преобразования наборов параметров спектральных линий. // Оптика атмосферы и океана, том 18, 2005г., № 09, стр.760-764
4. Chen P.P. The Entity-Relationship Model: Toward a Unified View of Data // ACM Transactions on Database Systems, Vol.1, No.1, March 1976, pp. 9-36.