

МЕТОДЫ И СИСТЕМЫ АВТОМАТИЗАЦИИ. ОБРАБОТКА ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

УДК 539.534

О.К. Войцеховская, С.В. Кузнецов

ПРОГРАММНАЯ ОБОЛОЧКА ОТКРЫТОЙ МНОГОФУНКЦИОНАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПО СПЕКТРОСКОПИИ ВЫСОКОГО РАЗРЕШЕНИЯ ДЛЯ ИВМ-СОВМЕСТИМЫХ КОМПЬЮТЕРОВ

Описывается информационно-поисковая система (ИПС) по спектральным характеристикам атмосферных и примесных газов в широком диапазоне температур и частот для ИВМ-совместимых компьютеров. Разработаны модель и логическая структура открытой оболочки для многофункциональной ИПС, формирующей проблемно ориентированные банки данных (по параметрам спектральных линий, по коэффициентам поглощения, по молекулярным константам и т.д.).

Организация ИПС осуществляется на основе индексно-последовательных файлов с множественными индексами (ключами), как это предлагается в сетевых системах управления базами данных (СУБД). Основное внимание уделено архитектуре ИПС и организации интерфейса пользователя. Предлагаются способы организации банков данных ИПС и возможные операции с базами данных по экспериментальным и рекомендованным значениям параметров спектральных линий, с архивами молекулярных и спектроскопических констант. Подчеркивается, что реляционные и ассоциативные связи доминируют в предлагаемой ИПС.

Развитие научных исследований в настоящее время невозможно без применения современных информационных технологий, переводящих научный процесс на более высокий уровень. Постоянное совершенствование персональных компьютеров значительно расширяет технологические возможности информационно-поисковых систем (ИПС) и требует соответствующей адаптации ранее созданных. Появление новых подходов в организации структур данных (см., например, [1]) требует их внедрения в организацию ИПС и для задач спектроскопии высокого разрешения.

В данной статье описывается модель и логическая структура открытой оболочки для многофункциональной ИПС, одновременно формирующая проблемно ориентированные банки данных (по параметрам спектральных линий, по коэффициентам поглощения, по молекулярным константам и т.д.), что является дальнейшим развитием ИПС, ранее реализованной на ЭВМ класса <Эльбрус> [2]. Основное внимание обращено на описание модели и логической структуры ИПС.

ИПС состоит из отдельных подсистем, реализующих свои алгоритмы обработки и интерфейсы пользователя, и объединяется одним общим меню выбора режима работы. Для расчетных задач формируются выборки из базы данных и запоминаются в переходном файле. С целью экономии оперативной памяти при запуске расчетной задачи ядро ИПС выгружается из памяти и загружается задача. Все расчетные задачи по окончании своей работы инициируют загрузку ядра программной оболочки ИПС или той подсистемы ИПС, с которой связана данная задача. Непосредственная запись результатов расчетных задач в базу данных не предусмотрена, поскольку операция включения требует довольно большого времени и будет замедлять процесс расчета. Поэтому включение рассчитанных данных в базу производится через промежуточный файл, который сначала анализируется, а затем принимается решение о приеме полученных данных в базу (полностью, частично или отвергнуть).

Обращение к ИС начинается с главного меню, в котором предлагается запустить одну из подсистем СУБД. Каждая подсистема СУБД имеет свой пользовательский интерфейс в виде меню либо в виде таблиц-списков некоторых параметров базы данных, в которых требуется провести установки ключевых параметров, позволяющих сузить рассматриваемое множество записей в базе данных и установить режимы обработки всех используемых полей (размерности, типы ошибок, условия среды и т.п.). Первоначальный запуск и подготовка ИПС к работе с пользователями начинаются с определения основных справочных данных для интерфейса

пользователя и расчетных задач (классы молекул, коды молекул, коды изотопов, физические величины и их размерности и т.п.). Эти операции осуществляются администратором системы. Все остальные банки данных являются пользовательскими и могут изменяться по их усмотрению. Следует иметь в виду, что понятия (спецификации и т.п.), используемые для расчетных задач, могут добавляться, но изменение или удаление этих понятий может привести к неверной интерпретации входных данных в ранее разработанных задачах.

1. Операции с базами данных по экспериментальным и рекомендованным параметрам спектральных линий (ПСЛ)

Сначала заметим, что все поля в банках данных по экспериментальным и рекомендованным ПСЛ имеют фиксированные размерности физических величин. Эти размерности определяются при первоначальном запуске СУБД и в дальнейшем изменениям не подлежат. Таким образом, интерфейс должен предусматривать работу по коррекции и вводу ПСЛ в тех единицах измерения, которые используются авторами в работах либо более удобны для пользователя.

Блок ввода и коррекции экспериментальных ПСЛ позволяет собирать информацию в режиме ручного ввода из любых источников информации. При этом перед началом ввода необходимо определить следующие неизменяемые в процессе ввода характеристики: единицы измерения полей ПСЛ, типы ошибок отдельных полей, температуру и давление среды; добавить или присоединить номер литературного источника, определить изотопическую модификацию ушряемого газа и газовый состав ушряющей среды. В результате пользователь работает только в указанных условиях и не видит остальных ПСЛ из архива, не соответствующих этим условиям. Предусмотрена возможность фиксации колебательного перехода для ускорения процесса ввода ПСЛ в пределах одной колебательно-вращательной полосы. В этом банке данных допускается определение нескольких вариантов ПСЛ при заказанных условиях и фиксированном вращательном переходе.

Аналогично устроен блок ввода и коррекции рекомендованных ПСЛ для ручного режима обработки, но предполагается наличие стандартных условий среды, и поэтому характеристики среды, погрешности определения параметров не используются. Выборки из банка данных по экспериментальным ПСЛ производится по следующим параметрам поиска: интервал частот для центров линий, интервал по интенсивностям, номер литературного источника, класс и код ушряемой молекулы, код изотопической модификации, температурный интервал, интервалы значений давления среды. Все эти условия отбора независимы, за исключением класса, кода молекулы и кода изотопической модификации (код молекулы можно определить только при классе, отличном от нуля, код изотопической модификации можно задать только при коде молекулы, не равном нулю).

В последующем, если это будет необходимо, можно расширить возможности фильтрации данных из банка ПСЛ, но в этом случае построение выборок будет производиться несколько медленнее, поскольку в некоторых ситуациях при построении выборки придется просматривать весь архив.

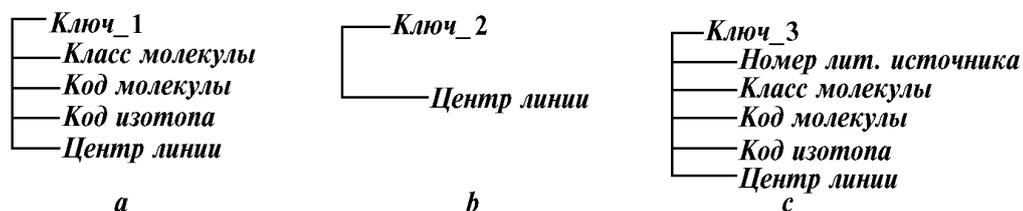


Рис. 1. Ключевые файлы для быстрого поиска в архивах ПСЛ: *a* – ускоренный поиск по классификации соединения; *b* – по спектральному интервалу; *c* – по библиографии

Аналогично строятся выборки и для рекомендованных ПСЛ, но здесь используются стандартные условия среды ($P = 1$ атм, $T = 296$ К), и поэтому установка интервалов температуры и давления не требуется. Для скоростного построения вышеуказанных выборок используются ключи (рис. 1), позволяющие ускорить процесс установки в файле на первую запись с заказанными значениями ключевых параметров и завершить процесс просмотра файла при выходе совокупности ключевых параметров в записи за пределы указанных значений.

На данных диаграммах поле с более низким приоритетом фильтрации находится соответственно ниже. Порядок следования параметров в выборке определяется используемым ключом для чтения записей и порядком следования параметров в ключе. Например, если в ключе 1 задан только класс, то читаются только те записи, у которых класс совпадает с установленным. Коды молекул будут выстраиваться в порядке возрастания, то же самое произойдет с кодами изотопов, при этом записи будут фильтроваться таким образом, чтобы центры линий попадали в указанный интервал частот и порядок следования записей в выборке будет определяться возрастающими значениями центров при фиксированных значениях класса, код молекулы и кода изотопа. При увеличении кода изотопа центры опять выстроятся в порядке возрастания и т.д. На рис. 2 показано формирование архивов для экспериментальных ПСЛ. Организация архивов для рекомендованных и экспериментальных ПСЛ идентична, но в последнем архиве не требуются ссылки на литературный источник, на поля, определяющие условия среды (температура, давление, код уширяющего газа), и на ошибки измерения центров, интенсивностей, полуширин и сдвигов. Единственное дополнительное поле заполняется признаком способа получения ПСЛ для данного кортежа (расчет или эксперимент).



Рис. 2. Структура организации базы экспериментальных ПСЛ

Для архивов рекомендованных ПСЛ предусматриваются режимы включения в банки данных расчетных ПСЛ с режимом замещения записей или отдельных полей записи, статистическая обработка экспериментальных ПСЛ и включение их в архив с полным замещением записи или с замещением некоторых полей. При этом ПСЛ, включенные ранее в архив на основе экспериментальных ПСЛ (или их отдельные поля), не могут замещаться рассчитанными значениями. С этой целью в архиве рекомендованных ПСЛ введен в кортеж признак, характеризующий способ получения каждого информационного поля, что помогает при операциях замещения ПСЛ.

2. Операции с библиографическими данными

Библиографическая справочная система служит для сбора информации по используемым литературным источникам при формировании базы данных. Она может включать в себя любую публикацию или же представлять собой замечания пользователя, сопровождаемые обязательным уникальным номером для вводимой справки. Библиографическая справка представляет собой кортеж, описываемый блоком 8 в диаграмме (рис. 3).

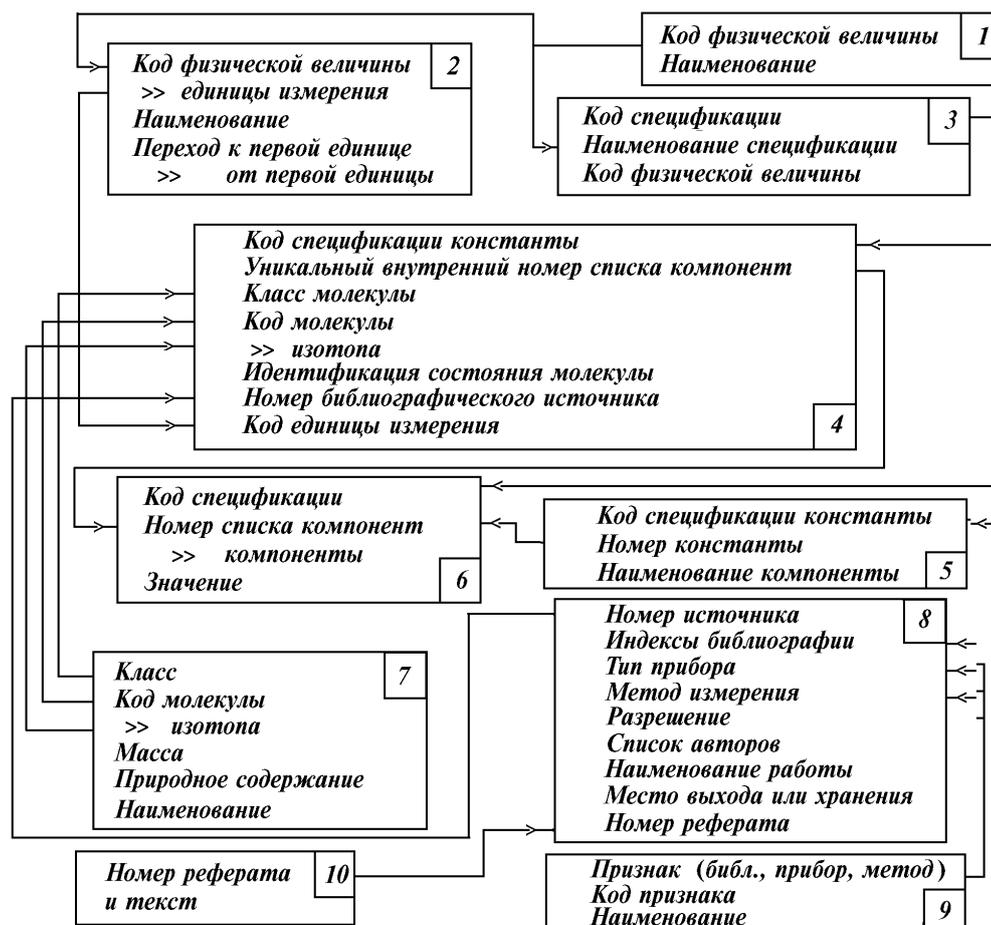


Рис. 3. Структура базы данных по молекулярным и спектроскопическим параметрам

В реальной работе у каждого пользователя будет скапливаться несколько сотен библиографических справок, так что потери дисковой памяти здесь незначительны. Эта подсистема поставляет номера библиографических ссылок для кортежа ПСЛ и, кроме этого, может быть использована для построения списка используемой пользователем литературы с некоторыми классификационными признаками, по которым можно проводить просмотр списка литературы по мере надобности и добавлять новые источники.

3. Архивы универсальных, молекулярных и спектроскопических констант

Блок формирования и поддержки банков физических, молекулярных, спектроскопических констант и констант межмолекулярного взаимодействия необходим для организации потока начальных данных к расчетным задачам по получению ПСЛ и др. Здесь выделяются 4 уровня данных: универсальные физические константы, молекулярные константы, спектроскопические константы, константы межмолекулярного взаимодействия.

Универсальные константы – это те физические константы, которые получаются в теоретических соотношениях и не связаны с условиями проведения эксперимента. Одна и та же константа может приводиться в различных единицах измерения, следовательно, кортеж, соответствующий универсальной константе, должен содержать уникальный внутренний номер константы, наименование, значение и код единицы измерения, который определяется с помощью списков физических величин и единиц измерения. Поскольку каждая универсальная константа является уникальной величиной, то следует хранить только по одному кортежу на каждую константу. Заметим, что каждая константа приводится в каком-либо литературном источнике, следовательно, к кортежу можно присоединить номер библиографической ссылки, где указывается ее происхождение.

Молекулярные константы – это те константы, которые связаны только с молекулой (масса, потенциальные функции, моменты инерции и т.д.) и электронным состоянием, в котором данная молекула находится. Эти константы определяются посредством нескольких кортежей, связи между которыми иллюстрируются диаграммой (см. рис. 3).

В табл. 1 описывается назначение каждого из используемых кортежей. При этом каждый прямоугольник описывает один кортеж, а входящая стрелка характеризует тот факт, что данный параметр передается из другого кортежа и является порожденным.

Для молекулярных констант в кортеже 5 (рис. 3) идентификация состояний всегда имеет нулевое значение, для спектроскопических констант этот параметр определяется колебательным состоянием молекулы, а для констант межмолекулярного (или переходов из одного колебательного состояния в другое для одной молекулы) добавляются еще четыре поля, определяющие состояние второй молекулы или второе состояние исходной молекулы (класс, коды молекулы и изотопа, идентификация колебательного перехода). Архив значений компонент для молекулярных, спектроскопических констант и констант межмолекулярного взаимодействия является общим. Следовательно, кортеж 6 является основным архивным кортежем. Различие между типами констант в этом архиве определяется признаком, который входит в ключ и позволяет в каждом режиме работы использовать только один вид констант. На диаграмме в кортежах этот признак не приводится.

Таблица 1

№ кортежа	Назначение
1	Физические величины
2	Соотношения между единицами измерения
3	Спецификация типа константы (списка констант)
4	Описание варианта списка констант
5	Заголовки компонент списка констант
6	Архивная запись значений компонент
7	Описание классов молекул и изотопов
8	Описание литературного источника
9	Описание библиографических индексов, типов приборов и методов измерений
10	Описание присоединяемых рефератов

4. Организация передачи данных из ИС

Блоки формирования данных для расчетных задач должны развиваться и добавляться по мере расширения круга используемых расчетных задач. Следует иметь в виду, что непосредственное обращение расчетных задач к архивным файлам во многих случаях нежелательно по той причине, что скорость непосредственного доступа к архивным данным намного меньше той, что позволяет организовать алгоритм расчетной задачи при оптимальном использовании оперативной памяти и дискового пространства. С этой целью в ИПС имеются возможности по построению выборок из архивов в максимально приспособленном для расчетных задач виде. При этом выборки можно формировать в текстовых или двоичных файлах.

При проектировании новой задачи следует учитывать наличие в базе данных необходимой информации и при недостатке этой информации в базе необходимо доопределить новые понятия, константы (добавляя, но не удаляя старые) и только затем приступать к реализации алгоритма ввода исходных данных в задачу и алгоритма вывода результатов в файл стандартной структуры, который в последующем будет приниматься в базу данных.

5. Занесение расчетных данных в ИПС

Блоки приема расчетных данных в ИС организуются по мере развития ИПС и связывают между собой банки данных и полученные в результате расчета данные. При построении банков рекомендованных значений отдается предпочтение экспериментальным значениям параметров, следовательно, кортежи для рекомендованных ПСЛ должны содержать внутри себя признак, определяющий способ получения параметра. В данный момент используются 4 па-

раметра, которые могут измеряться в эксперименте: центр, интенсивность, полуширина, сдвиг. Индексация параметра осуществляется в соответствии с табл. 2.

Общее значение параметра получается в результате суммирования приводимых признаков из табл. 2. Например, 11 будет выражать тот факт, что центр, интенсивность и сдвиг получены в результате обработки эксперимента, тогда как полуширина была получена в процессе расчетов, следовательно, при анализе кортежа, в котором имеется такой признак, мы полуширину можем заменять на новое расчетное значение, а остальные параметры оставляем без изменений.

Таблица 2

Параметр	Принадлежность	
	Эксперимент	Расчет
Центр	1	0
Интенсивность	2	0
Полуширина	4	0
Сдвиг	8	0

Аналогичные признаки можно включать и в кортежи молекулярных и спектроскопических констант. Правила формирования признаков у них могут быть другими в зависимости от числа используемых компонент в константе.

Заключение

Описанные принципы организации интерфейса пользователя и архитектуры ИПС могут быть реализованы на любом алгоритмическом языке. Наиболее быстро этот процесс осуществляется на основе СУБД PARADOX и CLARION, ориентированных на создание баз данных. Другие системы программирования баз данных, известные нам, не перспективны из-за более медленной обработки записей в архивах и неудобных интерфейсов программиста.

Авторы данной статьи создали информационную систему по спектроскопии высокого разрешения на ЭВМ IBM PC AT 286/287 с помощью аппарата СУБД CLARION. При этом разработаны архитектура и интерфейс ИПС, ориентированный на фундаментальные и прикладные задачи молекулярной спектроскопии и оптики газовых сред.

1. Tsichritzis D., Lochovsky F. DATA MODELS, New Jersey, 1982, 381 P.

2. Войцеховская О.К., Розина А.В., Трифонова Н.Н. Информационная система по спектроскопии высокого разрешения. Новосибирск: Наука, 1988. 150 с.

Институт оптики атмосферы
СО РАН, Томск

Поступила в редакцию
28 декабря 1993 г.

O.K. Voitsekhovskaya, S.V. Kuznetsov. Software Environment of an Open Multifunctional IBM PC Compatible Information System for High.

In this paper we describe an IBM PC compatible information retrieval system (IRS) for spectral characteristics of the atmospheric and trace gases in wide frequency and temperature ranges. A model and logic structure of an open software environment for a multifunctional IRS developed are described.

The structure of the IRS is arranged according to indexed sequential method as it is recommended for network data-base management systems. Special attention we pay to the IRS architecture and man-machine interface. We also propose the ways of arranging data banks of IRS and recommend some possible operations with data bases on the experimental and recommended values of spectral line parameters and with the archives of molecular and spectroscopic constants. It is characteristic of this IRS that relational and associative relationships dominate in this IRS.