



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

**Систематизация научных  
графических ресурсов по  
молекулярной спектроскопии**

***Рекомендуемая форма библиографической ссылки***

Ахлестин А.Ю., Лаврентьев Н.А., Фазлиев А.З. Систематизация научных графических ресурсов по молекулярной спектроскопии // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18-23 сентября 2017 г., г. Новороссийск). — М.: ИПМ им. М.В.Келдыша, 2017. — С. 34-42. — URL: <http://keldysh.ru/abrau/2017/39.pdf>  
doi:[10.20948/abrau-2017-39](https://doi.org/10.20948/abrau-2017-39)

Размещена также [презентация к докладу](#)

# Систематизация научных графических ресурсов по молекулярной спектроскопии

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

*Институт оптики атмосферы им. В.Е. Зуева, Томск, Россия*

**Аннотация.** Данный доклад рассматривает проблему введения в научный оборот программно обрабатываемых результатов, представленных в статьях в форме научной графики. В работе исследуется представление измеренных и предсказанных спектральных функций, относящихся к молекулярным комплексам.

**Ключевые слова:** научная графика, информационная система, коллекция спектральных функций, слабосвязанный молекулярный комплекс

## Введение

Целью данной работы является проведение систематизации спектральных функций и спектральных характеристик слабосвязанных молекулярных комплексов, представленных в научных публикациях в графическом виде. Ключевой термин, используемый в работе – источник данных о графически представленных спектральных функциях. Такой источник данных в созданной системе содержит в себе рисунок, извлеченный из статьи и цифровой образ рисунка. Цифровой образ представляет собой численную аппроксимацию изображенной спектральной функции. Публикации в формате PDF для последующего извлечения графики отбираются экспертами в предметной области.

Систематизация спектральных характеристик и спектральных функций слабосвязанных молекулярных комплексов, изображенных на рисунках в научных публикациях, результаты которой представлены в форме информационной системы (ИС), еще далека до завершения. Наиболее удачной попыткой систематизации является работа [1] по коллекции сечений поглощения. В этой работе описана коллекция, часть которой представляет дискретные значения сечений поглощения молекул и молекулярных комплексов. Эти значения занесены в базу данных (БД) наряду с метаданными, характеризующими каждый набор точек графика, соответствующий определенным термодинамическим условиям. Оцифровка данных и метаданные позволили включить ранее не доступные для компьютерной обработки данные в научный оборот. В нашей работе [2] по систематизации сечений поглощения был сделан следующий шаг: число метаданных было

расширено, и с каждым набором метаданных связан индивид онтологии сечений поглощения. Таксономия онтологии представляла собой аналог каталога информационных ресурсов.

### **Типизация научной графики и изображений в ИС W@DIS**

Научные графики и изображения, встречающиеся в публикациях, посвященных слабосвязанным молекулярным комплексам в спектроскопии, относятся к ограниченному числу типов графиков и изображений. Некоторые из этих типов использованы в данной работе. Для определенности в докладе даны определения ключевых терминов, используемых ниже.

*Научный график* – одна или несколько кривых, изображающих зависимость физической величины от других (другой) физических величин в одной и той же системе координат, опубликованные в научных изданиях. Научные графики разделены на три группы: элементарные научные графики и составные одностатейные или многостатейные научные графики, содержащие результаты, полученные в одной или нескольких публикациях. Элементарный научный график содержит графическое изображение зависимости одной физической величины от другой физической величины в одной системе координат. Составной научный график содержит изображение нескольких зависимостей физической величины от одной и той же физической величины, при разных условиях (например, термодинамических).

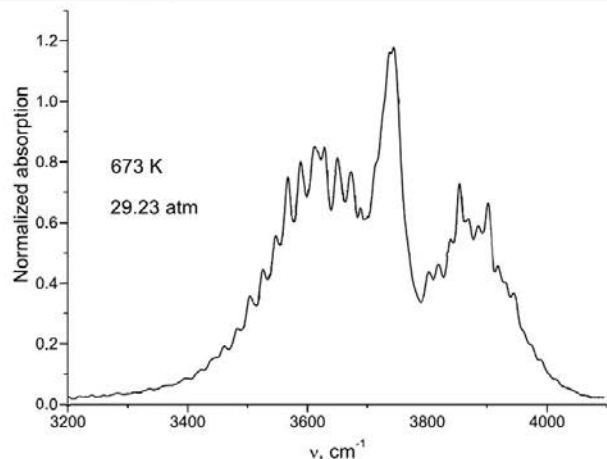
*Изображение* – представление субъектов, характерных для спектроскопии (молекул, комплексов, диаграмм характеристик и т.д.). Изображения также разделены на три группы: элементарные изображения, содержащие один субъект или составные изображения.

Каждый научный график или изображение является *элементарным рисунком*.

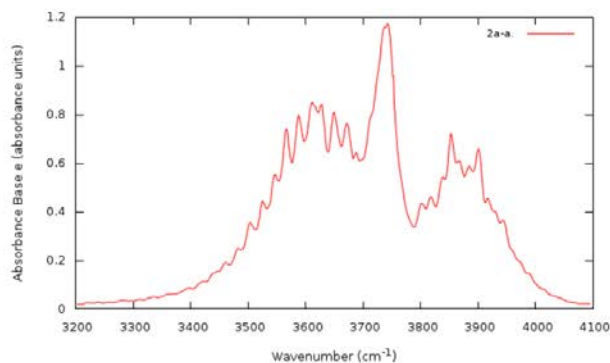
*Составной рисунок* – набор элементарных рисунков в научной публикации, относящиеся к одному рисунку публикации.

### **Программная реализация системы работы с графической информацией**

Систематизация результатов исследований спектральных данных, представленных в графическом виде, необходима для решения нескольких задач: формирования доступной для исследователей коллекции цифровых данных, характеризующих многочисленные научные графики, создания инструмента для сравнения физических величин, полученных при разных термодинамических условиях, создания онтологии спектральных данных о слабосвязанных молекулярных комплексах, позволяющей поисковым агентам получать информацию о цифровом прототипе научной графики по спектроскопии.



Water vapor absorption in the OH stretching fundamental. Resulting spectral profile. 2005 - A.A. Vigasin, A.I. Pavlyuchko, Y. Jin, S. Ikawa (2a-a) H<sub>2</sub>O-H<sub>2</sub>O T=673 K P=29.23 atm



Библиографическая ссылка

A.A. Vigasin, A.I. Pavlyuchko, Y. Jin, S. Ikawa, Density evolution of absorption bandshapes in the water vapor OH-stretching fundamental and overtone: evidence for molecular aggregation, Journal of Molecular Structure, 2005, Volume 742, Issue 1-3, Pages 173-181, DOI: 10.1016/j.molstruc.2004.12.060, http://dx.doi.org/10.1016/j.molstruc.2004.12.060.

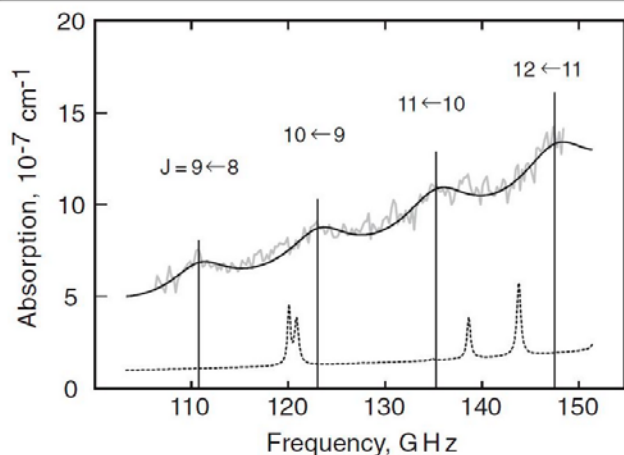
Annotation

Рисунок 2a-a. Water vapor absorption in the OH stretching fundamental. Resulting spectral profile.

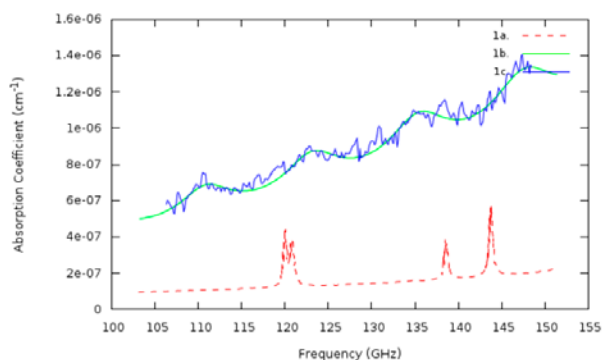
Графические характеристики

Физическая величина по оси абсцисс (X) (Размерность) (Число графиков)	Волновое число (см <sup>-1</sup> )	Физическая величина по оси ординат (Y) (Размерность) (Число графиков)	Поглощение по базе e (единицы поглощения)
Физические характеристики			
Вещество	H <sub>2</sub> O-H <sub>2</sub> O	Физическая величина	Absorbance
Температура	673 K	Метод измерения (решения)	
Давление	29.23 atm	Тип данных	Теоретические

Рис. 1. Представление элементарного графика для поглощения димера воды [3]



Observed water vapor spectrum after monomer contribution subtraction (broken line), calculated water monomer spectrum (smooth dotted line), and model dimer... 2013 - M.Yu. Tretyakov, E. A. Serov, M. A. Koshelev, V.V. Parshin, and A. F. Krupnov H<sub>2</sub>O-H<sub>2</sub>O T=296 K P=0.0171 atm



Библиографическая ссылка

M.Yu. Tretyakov, E. A. Serov, M. A. Koshelev, V.V. Parshin, and A. F. Krupnov, Water Dimer Rotationally Resolved Millimeter-Wave Spectrum Observation at Room Temperature, Physical Review Letters, 2013, Volume 110, Issue 9, Pages 093001, DOI: 10.1103/PhysRevLett.110.093001, http://link.aps.org/doi/10.1103/PhysRevLett.110.093001.

Annotation

Рисунок 1. Observed water vapor spectrum after monomer contribution subtraction (broken line), calculated water monomer spectrum (smooth dotted line), and model dimer absorption (smooth solid line). Vertical bars show positions of the J + 1 ← J, K = 0, E1-type dimer rotational lines.

Элементарные графики

- Рисунок 1a
- Рисунок 1b
- Рисунок 1c

Рис. 2. Составной график, характеризующий коэффициент поглощения водяного пара [4]

Последовательность действий при создании коллекции спектральных данных о комплексах воды включает в себя:

1. Выделение элементарных графиков из составных и создание набора координат, характеризующих элементарный график (цифровой образ);
2. Импорт элементарных графиков (рис. 1) в ИС (оригинального графика и его цифрового образа). Построение метаданных, характеризующих элементарный график;
3. Построение опубликованных сложных графиков (рис. 2) из элементарных графиков.
4. Построение произвольных композиций элементарных графиков.

Заметим, что первые три действия приводят к созданию публичных ресурсов и выполняются администратором системы. Последнее действие производится исследователем и его результат может быть экспортирован и в не сохраняется в системе.

Загрузка элементарных графиков спектральной функции

Физические характеристики		Графические характеристики	
Вещество	H <sub>2</sub> O-H <sub>2</sub> O	Единицы измерения по оси X	Волновое число (см <sup>-1</sup> )
Температура	400 К	Единицы измерения по оси Y	Коэффициент поглощения
Давление	2 atm	Масштаб по оси X	lin
Спектральная функция	Absorption Coefficient	Масштаб по оси Y	lin
Метод		Стиль графика	Lines
Тип данных	Теоретические	Номер рисунка в публикации	2a
Подпись к графику			
Библиографическая ссылка			
Поиск по публикации	Введите ключевое слово или автора публикации, в которой представлен график Varanasi <input type="text"/> <input type="button" value="Найти статью по ключевому слову"/>		
Выбранная статья	Varanasi P., Chou S., Penner S.S., Absorption coefficients for water vapor in the 600-1000 cm <sup>-1</sup> region, Journal of Quantitative Spectroscopy and Radiative Transfer, 1968, T. 8, Страницы 1537-1541.		
Комментарий			
Файл изображения	Обзор... 1968_VaChPe_Fig	Файл с данными	Обзор... 1968_VaChPe_Fig
<input type="button" value="Загрузить"/>			

Рис. 3. Интерфейс для описания метаданных при импорте простого графика и соответствующего ему массива данных

Базовой структурой данных для системы является элементарный график. Он хранится как совокупность абсцисс и ординат для соответствующего набора данных и сопутствующих им метаданных. На рис. 3 набор метаданных для каждого графика включает в себя физические величины, такие как: молекулярный комплекс, участвующий в физическом процессе, показанном на графике, температура и давление, соответствующие процессу, тип данных — экспериментальные или теоретические, спектральная функция и метод (измерения или вычисления), величины по осям X и Y, и их единицы

измерения; а также вспомогательные метаданные, включающие в себя: стиль графика — кривая, которая может быть представлена несколькими способами, набор точек или стиков; линейные или логарифмические масштабы по осям абсцисс и ординат, подпись и комментарий к графику, библиографическая ссылка на статью, из которой взят график, и номер рисунка в этой публикации. С каждым элементарным графиком связано отсканированное изображение из публикации-источника, что позволяет сравнить оригинальный рисунок с графиком, построенным системой автоматически, используя загруженный набор координат. В свою очередь, объединяя элементарные графики из одной публикации, можно получить составной график, для которого метаданными служат библиографическая ссылка и дополнительная подпись уже к составному графику как целому.

### **Формирование составных графиков**

Интерфейс поиска и сравнения позволяет найти уже загруженные графики по широкому набору критериев, таких как физические величины по обеим осям с соответствующими единицами измерений, вещество, температура и давление; либо других физических или вспомогательных метаданных. В результате поиска мы получаем наборы данных из различных публикаций, которые затем можно объединить в одном пространстве координат для последующего сравнения. Этот процесс можно разделить на три шага:

1. первоначальный поиск элементарных графиков по набору параметров (метаданным);
2. выбор из множества найденных источников данных подмножества, для которого необходимо произвести сравнение;
3. непосредственно графическое сравнение выбранных источников данных.

На приведённом ниже рис. 4 показан интерфейс поиска. Программный модуль позволяет искать источники данных, удовлетворяющие следующим параметрам:

- Единицы измерения по осям X и Y (обязательные параметры поиска)
- Вещество, которому соответствует источник данных
- Диапазон температур (от минимальной до максимальной)
- Диапазон давлений
- Тип данных в источнике: теоретический или экспериментальный
- Спектральная функция источника
- Метод получения
- Ключевые слова для библиографической ссылки на публикацию, из которой взят рисунок, включая фамилии авторов статьи.
- Номер рисунка в публикации
- Текстовый фрагмент подписи к графику.



**Поиск и сравнение графиков**

**Графические характеристики**

Единицы измерения по оси X: Длина волны (нм) (295)  Единицы измерения по оси Y: Сечение поглощения (см<sup>-1</sup>)

**Физические характеристики**

Вещество: H<sub>2</sub>O (51)  Спектральная функция: Ø (19)

Температура: 0 - 500 K  Метод: Ø (501)

Давление: 0 - 10 atm  Тип данных: Экспериментальные (44)

**Библиографическая ссылка**

Поиск по публикации: Введите ключевое слово или автора публикации, в которой представлен график: Lewis

Номер рисунка в публикации:   Подпись к графику:

Найти графики

Найдено графиков: 3 Уникальных статей: 1

Сравнить выбранные графики

#	Изображение	Подпись к графику	Физ. характеристики (Вещество, Температура, Давление)	Единицы измерения (по осям Y, X)	Библиографическая ссылка	Операции
<input checked="" type="checkbox"/>		1a. 1983_LeVaCa_235_H2O	H <sub>2</sub> O T = 235 K P = 1 atm	Y. Сечение поглощения (см <sup>-2</sup> ) X. Длина волны (нм)	B.R. Lewis, I.M. Vardavas, and J.H. Carver, The aeronomic dissociation of water vapor by solar H Lyman $\alpha$ radiation, Journal of Geophysical Research, 1983, Volume 88, Issue 6, Pages 4935-4940, DOI: 10.1029/JA088IA06p04935. Annotation	Изменить Удалить
<input type="checkbox"/>		1b. 1983_LeVaCa_292_H2O	H <sub>2</sub> O T = 235 K P = 1 atm	Y. Сечение поглощения (см <sup>-2</sup> ) X. Длина волны (нм)		Изменить Удалить
<input checked="" type="checkbox"/>		1c. 1983_LeVaCa_367_H2O	H <sub>2</sub> O T = 367 K P = 1 atm	Y. Сечение поглощения (см <sup>-2</sup> ) X. Длина волны (нм)		Изменить Удалить
#	Изображение	Подпись к графику	Физ. характеристики (Вещество, Температура, Давление)	Единицы измерения (по осям Y, X)	Библиографическая ссылка	Операции

Сравнить выбранные графики

Рис. 4. Интерфейс поиска элементарных графиков, удовлетворяющих заданным критериям

Выбрав необходимый набор параметров, по которому системе следует провести поиск, необходимо поставить галочку перед соответствующим параметром в табличном интерфейсе и задать или выбрать значение параметра, которому должны соответствовать источники данных. Если значение параметра можно выбрать из списка, рядом с этим значением в круглых скобках указано количество загруженных в систему источников данных, соответствующих данному значению параметра.

На рис. 4 показан интерфейс для поиска по следующим критериям: единицы измерения по осям X и Y должны соответствовать значениям “Длина волны (нм)” и “Сечение поглощения (см<sup>-1</sup>)”, источник данных должен содержать рисунок для вещества H<sub>2</sub>O, данные должны быть получены экспериментально, и рисунок должен быть взят из статьи за авторством Lewis. Также на рисунке можно увидеть, что в систему загружено 295 источников с единицами измерения “Длина волны (нм)” и “Сечение поглощения (см<sup>-1</sup>атм<sup>-1</sup>)”, 51 источник для молекулы воды и 445 источников с экспериментальными данными. Всем критериям поиска удовлетворяют три элементарных графика, показанные внизу рисунка. Там же проставлены 2 галочки из 3 (для рисунков 1a и 1c соответственно), поэтому в результате на конечном графике композиции (рис. 5) показаны 2 кривые, соответствующие отмеченным рисункам.

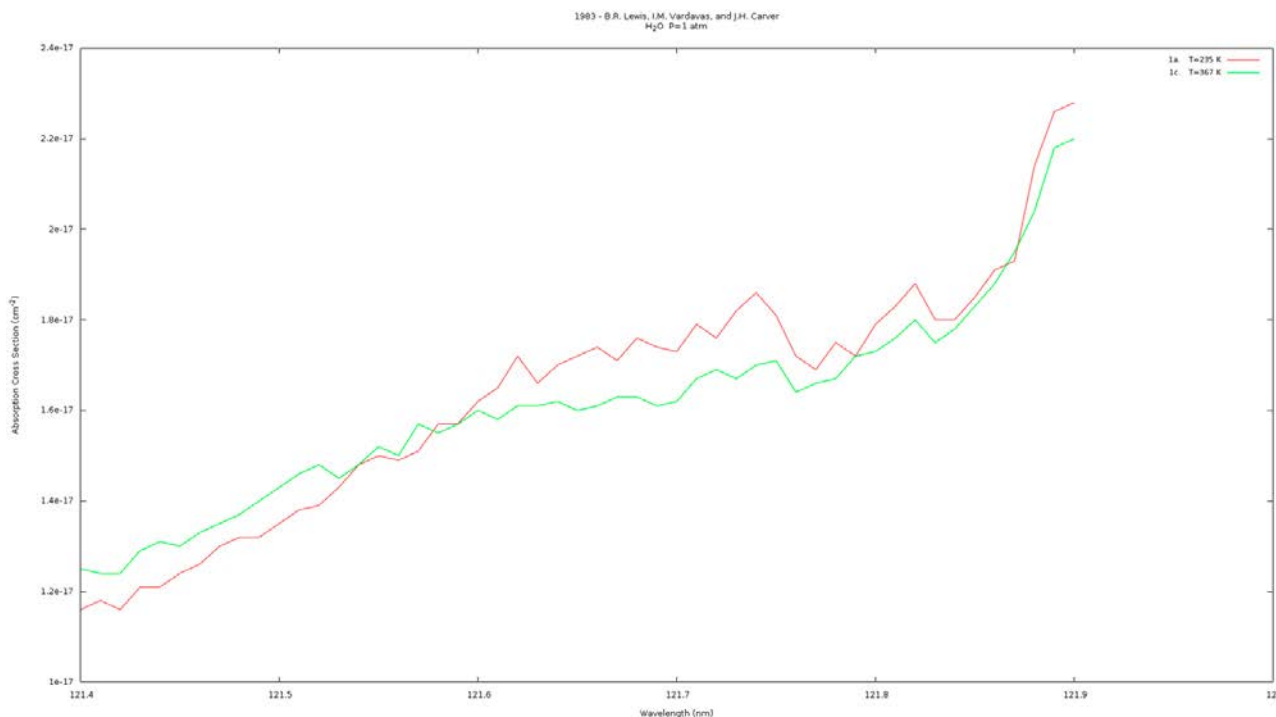


Рис. 5. Композиция выбранных простых графиков

### Технические особенности системы GrafOnto

Для построения научных графиков по загруженным наборам данных, характеризующих спектральные функции, используется пакет научной графики *gnuplot*, управляемый скриптами, динамически создаваемыми системой в зависимости от требований отображения одиночных или составных графиков различных типов. Промежуточное программное обеспечение ИС W@DIS [5] описано в нашей работе [6].

Система GrafOnto использует для функционирования три различных базы данных (рис. 6): базы веществ, библиографии и непосредственно базу для хранения данных элементарных и составных графиков. База данных веществ содержит таблицы молекул и молекулярных комплексов, атомов, изотопов и ионов, а также связывающую их таблицу молекулярного состава. Библиографическая база разделена на три части: таблицы библиографических ссылок, журналов и издателей. Основная функциональность системы GrafOnto обеспечивается базой данных научной графики, содержащей таблицы элементарных и составных графиков, а также таблицу абсцисс и ординат, привязанную к элементарным графикам. Оригинальные изображения GrafOnto хранит в файловой системе в форматах JPG, PNG или TIF.



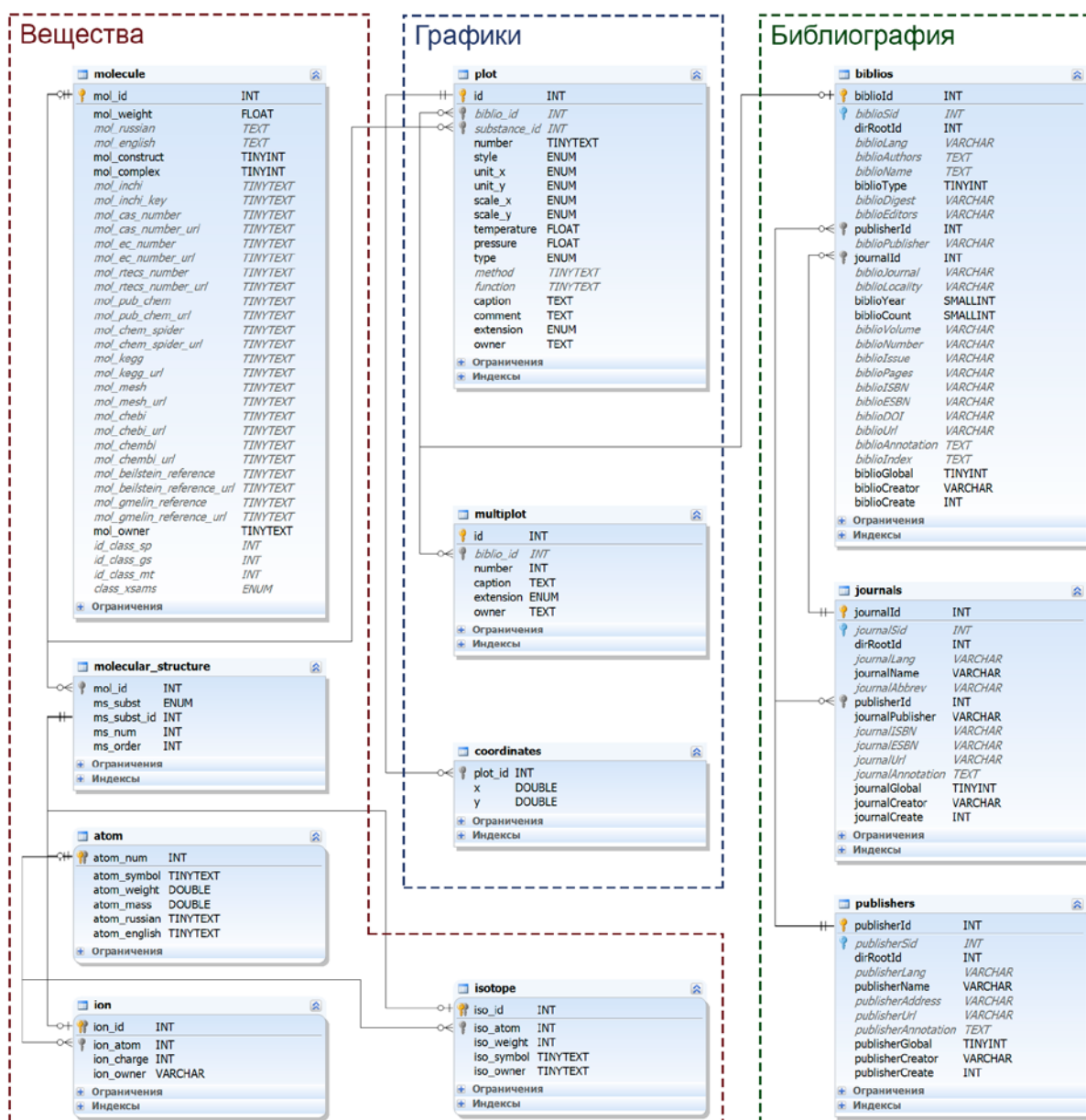


Рис. 6. Схема базы данных системы GrafOnto

## Заключение

Для систематизации опубликованных графиков решен ряд технических задач: проведена оцифровка графиков спектральных функций, импортированы данные в ИС, представляющие спектральные функции разных типов, построено приложение для графического представления спектральных функций стандартными средствами.

Дальнейшее развитие системы связано с созданием составных рисунков из элементарных и составных графиков и изображений. В разработке находится инструментарий для формирования составных графиков и изображений, состоящих из элементарных графиков и изображений, взятых из разных публикаций. В настоящее время в системе GrafOnto пользователю позволяет конструировать составные графики и рисунки, только если все составляющие

их элементарные графики принадлежат одной публикации. Кроме того, поскольку в публикациях встречаются рисунки, на которых графики привязаны не к одному веществу, а к некоторой их смеси в различных пропорциях, планируется изменить интерфейсы таким образом, чтобы появилась возможность задавать смеси веществ (как в процентном соотношении, так и с указанием только качественного, а не количественного состава) в качестве метаданных.

Система GrafOnto размещена в глобальной сети Интернет по адресу <http://wadis.saga.iao.ru/complexes/plots/>.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 07-13-00411.

### Литература

1. Keller-Rudek, H., Moortgat, G. K., Sander, R., and Sörensen, R. The MPI-Mainz UV/VIS spectral atlas of gaseous molecules of atmospheric interest, *Earth System Science Data*, 5, 365–373, (2013) doi:10.5281/zenodo.6951
2. Yu. Voronina, N. Lavrentiev, A. Privezentsev, A. Fazliev, K. Firsov, Representation of Absorption Cross-Sections in Information System W@DIS, Proc. of SPIE XX-th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, 2014, Vol. 9292, 929209
3. A.A. Vigasin, A.I. Pavlyuchko, Y. Jin, S. Ikawa, Density evolution of absorption bandshapes in the water vapor OH-stretching fundamental and overtone: evidence for molecular aggregation, *Journal of Molecular Structure*, 2005, Volume 742, Issue 1-3, Pages 173-181, doi: 10.1016/j.molstruc.2004.12.060
4. M.Yu. Tretyakov, E.A. Serov, M.A. Koshelev, V.V. Parshin, and A. F. Krupnov, Water Dimer Rotationally Resolved Millimeter-Wave Spectrum Observation at Room Temperature, *Physical Review Letters*, 2013, Volume 110, Issue 9, Pages 093001, doi: 10.1103/PhysRevLett.110.093001
5. Информационная система W@DIS; <http://wadis.saga.iao.ru>
6. Akhlyostin A.Yu., Kozodoev A.V., Lavrentiev N.A., Privezentsev A.I., Fazliev A.Z. Computed knowledge base for description of information resources of molecular spectroscopy. 4. Software, *Russian Digital Library Journal*, 2012, v.15,n.3, <http://elbib.ru/index.phtml?page=elbib/rus/journal/2012/part3/AKLPF>
7. , <http://elbib.ru/index.phtml?page=elbib/rus/journal/2012/part3/AKLPF>