Прикладная онтология для задач молекулярной спектроскопии

© Привезенцев А.И., Фазлиев А.З.
Институт оптики атмосферы СО РАН, Томск
faz@iao.ru

Аннотация

В работе описана прикладная онтология задач, используемая для систематизации информационных ресурсов по молекулярной спектроскопии. Описаны некоторые возможности формирования фактов об уровнях энергии молекулы воды, их машинное отнесения к классам и задача поиска источников данных в рамках концептов, на основе созданной прикладной онтологии.

1 Введение

Молекулярная спектроскопия, как часть оптики, ориентирована на изучение спектров молекул. Результатом измерений являются спектральные функции, значения которых используются для нахождения параметров спектральных линий. Вычислению параметров спектральных линий предшествует нахождение уровней энергии молекулы. На практике расчетные данные помещаются в базы данных и используются для решения задач атмосферной радиации, оптики атмосферы и астрономии

Расчеты выполняемые в молекулярной спектроскопии связаны с двумя типами задач: прямыми и обратными задачами. Входными данными для обратных задач являются либо непосредственно результаты измерений, либо выходные данные, как прямых, так и обратных задачи. Прямые задачи в качестве входных данных используют универсальные константы, либо величины, не относящиеся к предмету исследований в молекулярной спектроскопии (потенциальные функции, мультипольные моменты и т.д.). Последовательность решения прямых задач начинается с задачи нахождения уровней энергии молекулы и заканчивается задачей вычисления спектральных функций. Последовательность решения обратных задач начинается с задачи определения параметров спектральных линий из экспериментальных измерений и заканчивается задачей нахождения уровней энергии молекулы.

Труды 9ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2007, Переславль Залесский, Россия, 2007.

При решении задач обоих типов проводятся вычисления одних и тех же физических величин. Их сравнение позволяет делать выводы о корректности расчетов.

Существующие базы данных (БД) [1, 2] параметров спектральных линий содержат данные о решениях обратных задач. Схемы таких баз данных состоят только из структурных метаданных, содержащих физические величины, характеризующих спектр, погрешностей их определения и ссылок на публикации, в которых содержатся данные. Недостатками существующих схем данных является то, что некоторые характеристики молекул, используемые при решении прямых и обратных задач молекулярной спектроскопии, не включены в схемы данных и форма представления библиографических ссылок не позволяет определять с помощью SQL-запросов к БД принадлежность данных к публикации. Это означает, что невозможно при компьютерной обработке данных установить авторов данных и провести сравнение экспериментальных и расчетных данных.

В нашей работе расширение схемы БД осуществлено дополнением физических величин, относящихся к задачам нахождения уровней энергии молекулы и спектральным функциям. Введение сущности "источник данных" позволило выделить части данных, относящиеся к публикации и связать с ними наборы метаданных. В дальнейшем, предполагается использовать эти источники для формирования экспертных составных источников данных применяя операции реляционной алгебры.

При таком подходе в БД ИВС аккумулируются данные и метаданные, представляющие факты о спектральных свойствах молекул, источниках данных и свойствах данных, отнесенных к источнику данных.

Используемые в работе термины "A-box" и "T-box" определены в искусственном интеллекте при рассмотрении логического вывода. Принято делать различие [6] между терминологическим выводом (T-box) и выводом на множестве утверждений (A-box).

В естественных науках, в частности в молекулярной спектроскопии, большая часть задач связана с процедурными знаниями, основанными на решениях задач предметной области. Целью решения предметной задачи является получение

состояний исследуемой системы. Эти состояния исследуемой системы при представлении знаний рассматриваются как наборы фактов (A-box). Задачи классификации в молекулярной спектроскопии, как правило, сводятся к построению таксономии терминов предметной области, и на практике рассматриваются как вспомогательные задачи. Предполагается, что в задаче классификации концепты представляют интенсионалы предметной области (T-box).

С точки зрения представления знаний эти структурированные данные представляют факты (A-box). В ИВС часть этих фактов была представлена в виде индивидуалов онтологии, формализованной средствами языка OWL DL [3]. Для классификации фактов использовались классы и свойства OWL DL, представленные в виде таксономии (T-box). Таксономии хранятся в файлах и не связаны с БД в которой размещены факты.

Особенностью молекулярной спектроскопии является тот факт, что некоторые таксономии концептов являются динамическими, т.е. подвержены изменениям, или, другими словами, некоторая часть классификации фактов не является устоявшейся. Наиболее частым является изменение терминологии относящейся к тем сущностям, которые связаны с решаемыми в молекулярной спектроскопии задачами. На наш взгляд ментальное расширение и изменения таксономий концептов должно сопровождаться машинным отнесением фактов к классам созданных. Для отнесения фактов к классам в работе использовалась машина вывода Racer [4].

Описание предметной области основано на нескольких таксономиях и наборах фактов для каждой из задач модели молекулярной спектроскопии.

Целью данной работы являлось создание онтологий задач молекулярной спектроскопии используемых для систематизации информационных ресурсов ИВС “Атмосферная спектроскопия” и реализации поиска источников данных, относящихся к разным задачам молекулярной спектроскопии. Данные в эту ИВС поставляются пользователями в виде файлов, содержащих результаты решенных задач и в виде массивов данных, являющихся решениями задач, осуществленными приложениями интегрированными в ИВС.

Прикладные онтологии, связанные с приложениями, относящимися к нашей информационной системе, формализованные таким образом, чтобы с ними можно было работать в реальном масштабе времени (например, машинное отнесение фактов к классам осуществляется за небольшой промежуток времени) созданы, в частности для решения задачи поиска источников данных, характеризующих решения конкретных задач молекулярной спектроскопии. По выделенной цели и выразительности такие онтологии, согласно классификации, называют прикладными онтологиями [5]. Для целей проектирования мы использовали справочную онтологию (reference ontology).

2 Концептуализации и онтология

Известно, что представление данных и знаний в информационно-вычислительной системе основано на концептуализации предметной области. Общее описание назначения концептуализации, сформулированное в [7], хорошо описывается цитатой “формально представленное знание основано на концептуализации: объектах, концептах и других сущностях, которые, предполагается, существуют в некоторой интересующей нас области и отношений между ними. *Концептуализация является абстракцией, упрощающей мир, который мы хотим представить с некоторой целью*. Каждая база знаний, система, основанная на знании или агент уровня знаний фиксирует некоторую концептуализацию, явно или неявно” [7]. В этой же работе было сформулировано определение понятия “концептуализация”.

Определение [7]. *Концептуализация* является парой (**D**, *R*), где **D** – предметная область, а *R* – множество отношений на **D**.

Предполагается, что **D** – множество концептов, а R – экстенсиональные отношения.

Существует другое определение понятия «концептуализация», данное N.Guarino [8]. Guarino отметил, что экстенсиональные отношения отражают конкретную безвариантную предметную область. Фактически же “при представлении знания необходимо сосредоточиться на смысле отношений, независимо от конкретной реализации предметной области” [5]. Такой подход привел к иному определению понятия “концептуализация”.

Определение [8]. *Концептуализация* является триадой C=(**D**, W, **R**), в которой **D** – предметная область, W – возможные реализации предметной области, а **R** – множество концептуальных отношений на **D**.

На наш взгляд последнее определение понятия “концептуализация” явно обозначило различие между представлениями данных и знаний. Различие в трактовке понятия “концептуализация” привело к возникновению термина “онтологическая фиксация”, обобщающего термин “моделирование предметной области”. Описание деталей различия и условий совместимости этих терминов можно найти в монографии [5].

Понятие онтологической фиксации является связью между концептуализацией C, не зависящей от языка, и, онтологией, т.е. логической теорией L, выраженной в соответствие с концептуализацией и интерпретацией словаря этого языка.

Наиболее часто используемое определение онтологии, данное Т.Грубером [9], а именно, “онтология является явной спецификацией концептуализации” в рамках описываемого подхода

трактуется следующим образом. Явная спецификация достигается с помощью логической теории, т.е. множества логических аксиом, выраженных в логическом языке L.

Онтологию можно рассматривать как логическую теорию назначением, которой является обеспечивать намеченный смысл словаря V языка L. Словарь логического языка содержит константы, функции и отношения [5].

Введенный формализм позволяет выбирать онтологии близкие к описанию предметной области. В рамках естественного языка и набора логик, используемого при написании научных работ неявным образом формулируются онтологии разные как по выразительности, так и степени детализации предметной области. На практике для создания автоматизированных ИС используются менее выразительные языки для того, чтобы обработка выражений была ограничена разумным интервалом времени. Использование разных логических языков и словарей приводит к множеству возможных онтологий, требующих классификации. На рис.1 приведена классификация онтологий, описанная в монографии [5].

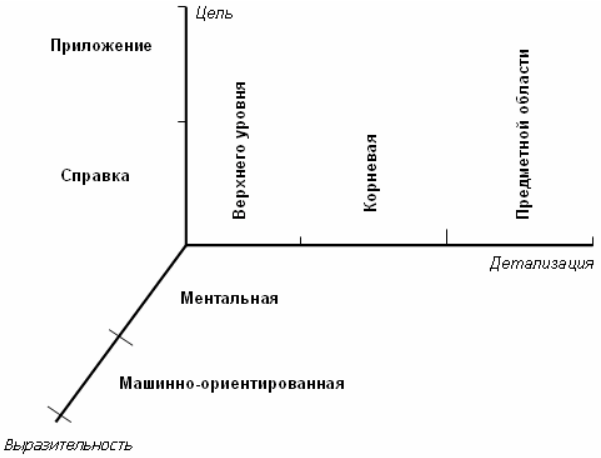


Рис.1. Классификация онтологий по цели, выразительности и уровню детализации.

3 Модель предметной области

Классическая линейная молекулярная спектроскопия, являясь частью физики, имеет все ее характерные особенности, в части проведения измерений и вычислений с последующим сравнением результатов. В спектроскопии единственной экспериментально определимой характеристикой молекул являются спектральные функции.

Основной задачей молекулярной спектроскопии является изучение спектров молекул. Её решение можно представить в виде сети подзадач. Решение подзадач позволяет утверждать о решении основной задачи.

Прямые задачи молекулярной спектроскопии связаны с расчетом из первых принципов фундаментальных характеристик молекул, таких как уровни энергии молекул, частоты перехода, коэффициенты Эйнштейна и т.д. Обратные задачи молекулярной спектроскопии связаны с обработкой данных измерений спектральных функций, что позволяет в дальнейшем при машинной обработке классифицировать их выходные данные как экспериментальные. В сети задач молекулярной спектроскопии существуют связи между прямыми и обратными задачами.

3.1 Прямые задачи молекулярной спектроскопии

К использованным нами при проектировании информационной системы элементарным прямым задачам относятся следующие [10]:

1. **Задача определения физических характеристик изолированной молекулы (Т1)**. Результатом решения задачи являются вычисленные уровни энергии молекулы, волновые функции, которым соответствуют стационарные состояния и интегралы движения, определяющие квантовые числа для уровней энергии.
2. **Задача определения параметров спектральной линии изолированной молекулы (Т2)**. Результатом решения являются частоты переходов (центры линий) и коэффициенты Эйнштейна. Входными данными для задачи являются уровни энергии, волновые функции и квантовые числа.
3. **Задача определения параметров контура спектральной линии (Т3)**. Входными данными являются частоты переходов, волновые функции, коэффициенты Эйнштейна и др. Результатом решения являются вычисленные полуширины, сдвиги, интенсивности, параметры, характеризующие интерференцию спектральных линий, статистические веса.
4. **Задача расчета спектральных функций (Т4)**. Входными спектральными данными являются параметры спектральных линий взаимодействующей молекулы. Рассчитываются коэффициенты поглощения, функция пропускания и т.д. при заданных термодинамических и электромагнитных условиях.
5. **Измерения спектральных функций (Е).** Результатом решения этой задачи значимым для ИВС являются значения спектральных функций и метаданные об условиях проведения эксперимента.

Существенным фактом является то, что эти задачи образуют сеть, определяющую последовательность их решения. Например, для решения задачи Т3 необходимо иметь решение задачи Т2, или, иными словами, входные данные задачи Т3 должны включать в себя выходные данные задачи Т2.

Выделение первых двух классов задач (Т1 и Т2) обусловлено важным физическим фактором, а именно, свойства изолированных молекул не зависят от термодинамических параметров. Задача Т3 позволяет определить параметры спектральных линий одной молекулы при разных

термодинамических условиях и учете столкновений молекул в газе. Задачи T4 и E1 описывают излучающие или поглощательные способности газов.

3.2 Обратные задачи молекулярной спектроскопии

К элементарным обратным задачам относятся

1. **Задача определения параметров спектральной линии взаимодействующей молекулы (ET)**. Входными данными являются измеренные спектральные функции и условия измерения. Результатом решения задачи являются параметры спектральных линий взаимодействующих молекул.
    a. **Подзадача определения центров спектральных линий (ET1.1).** Результатом решения являются частоты переходов.
    b. **Подзадача определения интенсивностей спектральных линий (ET1.2).** Результатом решения являются интенсивности, отнесенные к центрам спектральных линий при заданных термодинамических и электромагнитных условиях.

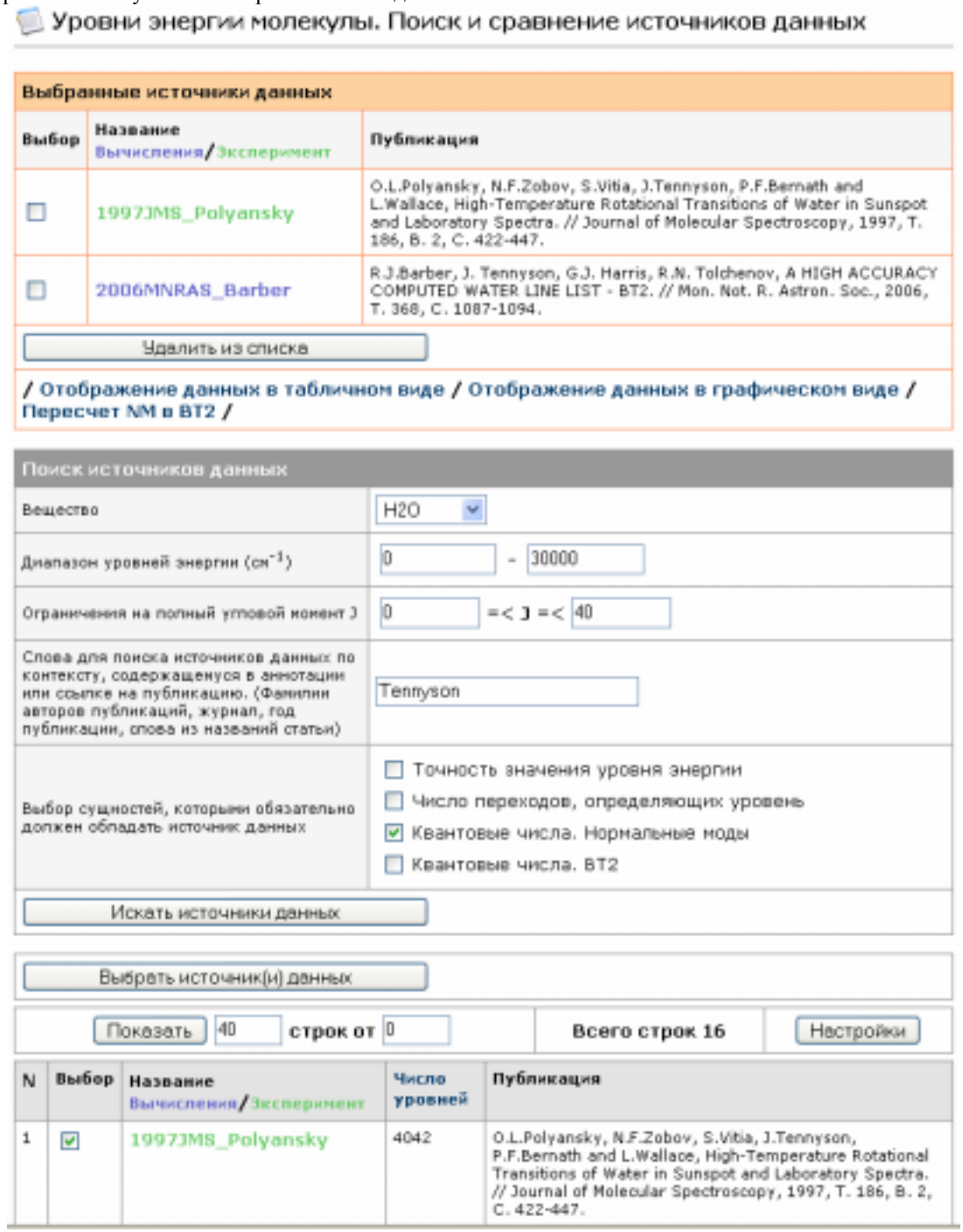


Рис.2. Интерфейс для организации поиска и сравнения расчетных и экспериментальных данных

c. **Подзадача определения полуширин, сдвигов и температурных зависимостей полуширин и сдвигов (ET1.3).** Результатом решения задачи являются значения параметров контура спектральной линии (полуширина линии, обусловленная столкновениями молекул, сдвиг линии, обусловленный давлением, и температурная зависимость полуширины линии)

d. **Подзадача определения параметров смешения линий (ET1.4).**

e. **Подзадача определения коэффициентов Эйнштейна (ET1.5).** Результатом являются коэффициенты Эйнштейна, отнесенные к частотам перехода.

2. **Задача приписывания квантовых чисел спектральным линиям (T5)**. Входными данными являются расчетные спектры с идентифицированными переходами и решения обратных задач **ET**. Результатом является установление связи между частотами перехода и квантовыми числами.
3. **Задача определения уровней энергии изолированной молекулы (T6)**. Результатом является список уровней энергии с приписанными к ним квантовыми числами, погрешности определения уровней энергии и число переходов, использованных для определения значения уровня энергии.

Задачи T4 реализованы в виде приложений в ИВС "Атмосферная спектроскопия" (http://saga.atmos.iao.ru). Остальные задачи, в настоящее время, представлены в виде информационных ресурсов, представляющих результаты решения конкретных задач (T1-T6, E и ET).

4 Модель информационной системы

В качестве модели информационной системы была выбрана модель, включающая в себя три слоя [11]: слой данных и вычислений, информационный слой и слой знаний. Каждый слой имеет свое назначение. Слой данных и вычислений ориентирован на непосредственное решение предметных задач и задач манипуляции данными и основан на концептах и отношениях предметной области. В части работы с данными для этих целей использовалась СУБД MySQL. Информационный слой предназначен для уменьшения величины неявного знания, что достигается машинной процедурной обработкой информации о решенных задачах и описанием результатов решения в терминах метаданных. Основной концепт, используемый при формирования этого уровня - *источник_данных*. Этот слой соответствует наборам утверждений (фактам, A-box) при представлении знаний. Эта задача решается с помощью языка OWL DL. Слой знаний создается на основе онтологий, включающих как факты так и таксономии терминов (T-box) для описания.

В основу модели слоя данных и вычислений, представляющих пользователям как процедурные, так и декларативные знания в области молекулярной спектроскопии, были положены задачи предметной области, описанные в предыдущем параграфе. Проектирование ИВС и программная реализация задач предметной области в виде приложений тесным образом связаны с моделями данных и сетями потока работ [12], характерными для молекулярной спектроскопии.

Как правило, в задачах молекулярной спектроскопии, данные, относящиеся к задаче, описываются моделью структурированных данных, характеризующейся интенсионалами и экстенсионалами этих данных. При реализации частей задачи в виде приложений часть этих данных является для приложений входными и выходными данными.

Уровень данных и вычислений в нашей ИВС представлен приложениями по вычислению спектральных функций и приложениями по манипуляции данными. К числу последних относятся задачи загрузки информационных ресурсов, выборка источников данных по набору атрибутов, табличное и графическое сравнение экспериментальных и расчетных данных.

Загрузка пользовательских данных, в описываемой ИВС, является основным способом наполнения системы новыми элементарными наборами данных. Как экспериментальные, так и расчетные данные, относящиеся к любому из перечисленных выше классов задач, могут попасть в ИВС только в результате их загрузки пользователем или решения соответствующей задачи в информационно-вычислительной системе. При загрузке данных пользователем отсутствует связь загружаемых данных с другими ресурсами, уже находящимися в ИВС. Загружаемые пользователем данные образуют элементарные ресурсы. Как было описано выше, на примере задачи T6, эти ресурсы характеризуются источником данных, содержащим библиографическую ссылку.

Как правило, данные загружаются в ИВС, в виде файлов. Структуры данных, используемые в файлах, могут не соответствовать структурам данных, используемым для хранения ресурсов. 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 [13] форматирования спектральных данных в молекулярной спектроскопии.

Конкретная структура данных, используемая при загрузке, обусловлена задачей молекулярной спектроскопии, решением которой эти данные являются. Структура данных,

используемая для их хранения, может быть иной, и обусловленной задачами которые их используют.

Рассмотрим на примере задачи формирования составных ресурсов механизм изменения структуры ресурсов. После загрузки ресурсы имеют статус персональных ресурсов и доступны только собственнику. Для организации доступа к загруженным пользователем ресурсам используется процедура экспертного отбора. Каждый пользователь по определенной процедуре после экспертной оценк может опубликовать свои элементарные ресурсы, загруженные в ИВС. Статус рекомендованного для публикации ресурса означает невозможность его изменения собственником ресурса. Рекомендованные к публикации ресурсы становятся доступными экспертам. Отобранные экспертами ресурсы приобретают статус опубликованных ресурсов. Они помещаются в хранилище данных и становятся общедоступными.

На основе опубликованных ресурсов все пользователи, в том числе и эксперты, могут формировать составные ресурсы в рамках правил, поддерживаемых в ИВС. Формирование хранилища данных решает проблему непрозрачности процедуры формирования ресурсов, имеющихся в банках данных Hitran и Geisa. Созданные экспертами составные ресурсы могут также использоваться пользователями для их задач. Для параметров спектральных линий по умолчанию выбран формат файла данных используемый в банке данных Hitran.

Структура ресурсов, формируемая экспертами, определяется прикладными задачами, для которых эти ресурсы являются входными данными. Она может не совпадать со структурой, используемой для хранения данных в хранилище данных, содержащем опубликованные ресурсы.

На рис.2 представлен интерфейс для организации поиска источников данных для задач T1 и T6 по названию вещества, диапазону значений уровней энергии молекулы, ограничениям на полный угловой момент молекулы и т.д. Интерфейс не предоставляет пользователю возможности самостоятельной организации запроса к БД. На рис. 3 показан результат графического сравнения расчетных данных с данными, полученными из измерений.

5 Онтологическая фиксация

Выбор модели предметной области в виде сети задач существенно уменьшает число концептов и отношений в нашей концептуализации, по сравнению с неформализованными концептуализациями существующими в представлении молекулярной спектроскопии на естественном языке. Концепт “задача” в ИВС использовался только для задач молекулярной спектроскопии и не применялся к их декомпозиции на приложения с помощью которых эти задачи были реализованы в ИВС. Такая детализация является излишней, т.к. описание конкретной декомпозиции задачи слабо связано с целями, поставленными для задач молекулярной спектроскопии.

Онтология верхнего уровня, предложенная Sowa J.F. [6], использовалась для классификации концептов молекулярной спектроскопии. Ключевыми объектами (независимыми физическими континуантами) в спектроскопии являются вещество и излучение, а процессами (независимыми физическими оккурентами) – поглощение и испускание излучения. Концепты, относящиеся к независимым абстракциям, в этой онтологии относятся к двум классам: Schema и Script. Класс Schema содержат независимые континуанты, а класс Script – независимые оккуренты. В молекулярной спектроскопии к этим классам относятся концепты *уровни энергии*, *переходы между уровнями* и *спектральные функции*.

При кодировании в OWL DL объекты и процессы были представлены классами, а значительная часть экземпляров класса Schema – объектными свойствами и свойствами типов данных. Максимальная кардинальность некоторых свойств, например, для молекулы воды достаточна велика. Так для основного изотопомера воды максимальная кардинальность свойства *иметь_уровень_энергии* превышает 220 000, а свойства *иметь_переход_между_уровнями* – 500 000 000 [14]. По причине такого количества фактов мы отказались от использования этих свойств в прикладной онтологии.

Большая часть классов прикладной онтологии задач построена с помощью ограничений на свойства. Мы избегали построения классов с помощью перечисления индивидуалов по той причине, что известные нам машины вывода не работают с такими конструкциями.

Важную роль при формировании информационного слоя ИВС играет концепт “*источник данных*”. Концепт *источник данных* не относится к молекулярной спектроскопии. Он описывает часть данных, размещенных в БД. Определение этого понятия дано ниже. Оно важно при решении задачи семантического поиска результатов решения задач молекулярной спектроскопии.

Определение. *Источником данных* называется информационный объект, содержащий метаданные о данных, являющихся значениями свойств некоторой вещи и описывающий место нахождения этих данных.

Таксономия источников данных

- Data_Sources
  - Compound_DS
  - Elementary_DS
    - Energy_Levels_EDS
    - Transition_EDS
  - Parts
- DimensionalQuantity
  - Energy_Levels_Md

    - Wavenumbers_Md
- Metadata
    - DSPartName
    - Energy_Levels_Md
    - Input_Data_Md
    - Output_Data_Md
    - Quantum_Numbers_Md
    - Transitions_Md
    - Wavenumbers_Md
- Units

Класс Data_Sources содержит подклассы Compound_DS, Elementary_DS и Parts. Экземплярами расширения класса Elementary_DS являются наборы фактов о физических величинах относящиеся к одной молекуле и одной публикации. Составные источники могут содержать данные о нескольких молекулах, которые опубликованы в нескольких статьях. Семантически значимые части составных и элементарных источников данных могут быть выделены в элементы класса Parts. Примером, такого значимого выделения может служить разбиение уровней энергии молекулы воды на смежные части: уровни энергии орто- и пара-воды.

6 Прикладная онтология задач

Этот раздел посвящен описанию информационного слоя ИВС. В нем рассмотрено построение источника данных для задачи T1. Концептами этой задачи являются физические характеристики, такие как уровень энергии, квантовые числа, потенциальная функция и т.д., которые представляют числовые значения свойств молекулы.

Обязательным атрибутом элементарного источника данных является название молекулы, и библиографическая ссылка. Составной источник данных является композицией элементарных. С каждым элементарным источником данных связаны количественные и качественные метаданные.

В ИВС создание метаданных проводится в два этапа. Порядок формирования метаданных организован так, что пользователь сначала создает шаблон элементарного источника данных и указывает публикацию с которой он связан. Затем пользователь может загрузить в систему данные о решенной задаче и занести качественные метаданные. Возможен и обратный порядок, при котором сначала создаются качественные метаданные.

Количественные метаданные создаются автоматически при загрузке данных в ИВС или при проведении пользователем расчетов с помощью приложений, интегрированных в ИВС. Значениями количественных метаданных являются числа. Количественные метаданные актуализируются автоматически при каждом обновлении предметных данных

Качественными метаданными характеризуют начальные условия задачи и метод ее решения. Они создаются пользователем с помощью форм, а при решении задач с помощью приложений они формируются автоматически. Для каждого типа задач набор качественных метаданных будет индивидуален. Например, для задачи T1, решаемой методом эффективного гамильтониана, в качестве метаданных используются резонансные полиады, центробежные константы, резонансные константы, колебательная энергия полиад и т.д.

На втором этапе имеющееся в ИВС приложение представляет эти данные в виде утверждений, размеченных с помощью спецификации OWL DL. Этот этап заканчивается формированием множества фактов, представленного в виде наборов индивидуалов, являющихся экземплярами расширений классов. Если пользователь не занес качественные метаданные, в этом случае формируется индивидуал со значением количественных метаданных - UNDEFINED. Пользователь имеет возможность дополнить или изменить значения качественных метаданных к тем решениям задач, которые он загрузил в систему.

Работа по созданию слоя знаний состояла в построении классов и таксономии классов с помощью свойства subClassOf. Эта часть работы выполнена с помощью редактора онтологий Protégé [15]. Наиболее употребляемым приемом при создании классов являлось задание ограничений на свойства. Машина вывода Racer, используемая нами для анализа концептов, позволяла делать выводы о том является ли множество, определяемое концептом пустым, находить несогласованные имена концептов в таксономии, определять порождающие и порожденные концепты по отношению к таксономии, проверять согласованность между набором фактов и таксономиями.

Рабочие версии онтологий, используемых в ИВС, можно найти в Интернете по адресам (aaa: substances; DataSources; task; spectra; task_t1; task_t6) и (bbb: T6, T1):

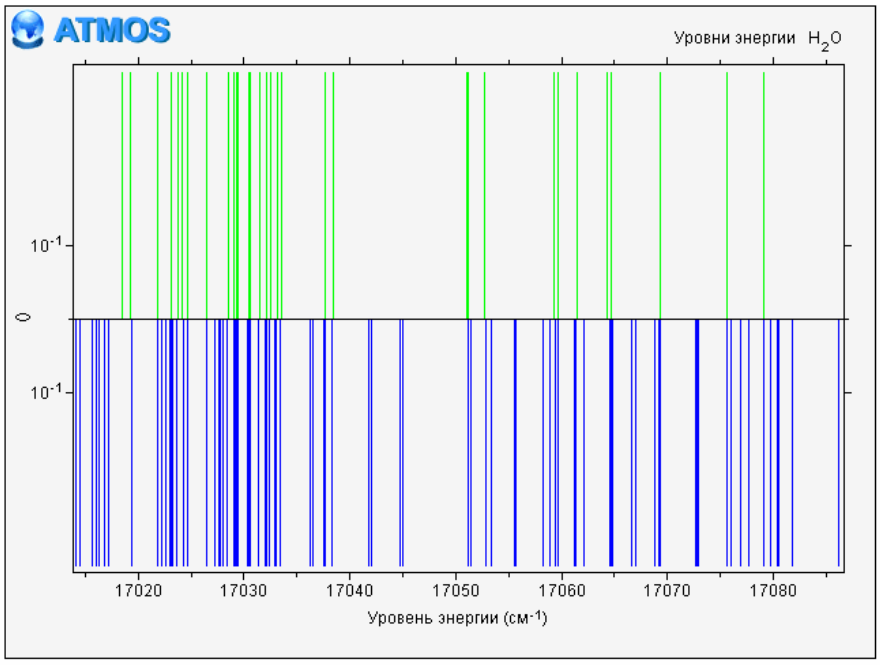
Таксономии

- http://atmos.iao.ru/Ontology3/aaa.owl

Наборы утверждений (фактов)

- http://atmos.iao.ru/Ontology3/bbb.owl

Особенностью работы с онтологиями задач в ИВС является то, что пользователь при решении задачи или загрузки в систему ее решения механически составляют свою собственную онтологию (A-box), содержащую факты, относящиеся к конкретной решенной задаче. Эти факты можно объединять с соответствующими фактами других задач или других пользователей, если позволяют права доступа к базе знаний.



Выбранные источники данных		
N	Название	Публикация
1	2005JMS2_Tolchenov_	Tolchenov R.N., O.Naumenko, N.F.Zobov, S.V.Shirin, O.L.Polyansky, J.Tennyson, M.Carleer, P.-F.Coheur, S.Fally, A.Jenouvrier and A.C.Vandaele, Water vapour line assignments in the 9250-26 000 cm-1 frequency range. // Journal of Molecular Spectroscopy, 2005, v. 233, B. 1, p. 68-76.
2	2006_Zobov_H2_16O	Н.Ф. Зобов, Р. И. Овсянников, С.В. Ширин, О.Л. Полянский, N. Vogt, J. Vogt, Приписывание квантовых чисел теоретическим спектрам H216O, H217O и H218O изотопомеров молекулы воды. // Optics and Spectroscopy, 2006.

Рис.3. Графическое сравнение экспериментальных и расчетных данных

Аннотация (2006_Barber) Расчет/Эксперимент

Вещество	
H2O	
Входные данные	
Потенциальная функция (URL)	UNDEFINED
Массы атомов (URL)	UNDEFINED
Базисные волновые функции (URL)	UNDEFINED

Метод (название и ссылка)

UNDEFINED

Публикация

R.J.Barber, J. Tennyson, G.J. Harris, R.N. Tolchenov, A HIGH ACCURACY COMPUTED WATER LINE LIST - BT2. // Mon. Not. R. Astron. Soc., 2006, T. 368, C. 1087-1094.

A computed list of $H_2{}^{16}O$ infrared transition frequencies and intensities is presented. The list, BT2, was produced using a discrete variable representation two-step approach for solving the rotation–vibration nuclear motions. It is the most complete water line list in existence, comprising over 500 million transitions (65 per cent more than any other list) and it is also the most accurate (over 90 per cent of all known experimental energy levels are within 0.3 cm^{-1} of the BT2 values). Its accuracy has been confirmed by extensive testing against astronomical and laboratory data. The line list has been used to identify individual water lines in a variety of objects including comets, sunspots, a brown dwarf and the nova-like object V838 Mon. Comparison of the observed intensities with those generated by BT2 enables water abundances and temperatures to be derived for these objects. The line list can also be used to provide an opacity for models of the atmospheres of M dwarf stars and assign previously unknown water lines in laboratory spectra.

Выходные данные		
Уровни энергии		
Единица измерения	cm-1	
Минимальное значение	0	
Максимальное значение	29999.840396	
Число уровней энергии	221097	[T]
Квантовые числа		
Тип квантовых чисел	BT2	
Минимальное значение для полного углового момента J	0	
Максимальное значение для полного углового момента J	50	
Число уровней энергии с уникальной идентификацией	221097	[T]
Число уровней энергии с не уникальной идентификацией	0	[T]
Квантовые числа		
Тип квантовых чисел	NormalModes	
Минимальное значение для полного углового момента J	0	
Максимальное значение для полного углового момента J	50	
Число уровней энергии с уникальной идентификацией	28548	[T]
Число уровней энергии с не уникальной идентификацией	0	[T]

Рис.4. Визуализация метаданных. Аннотация решения задачи по определению уровней энергии молекулы воды.

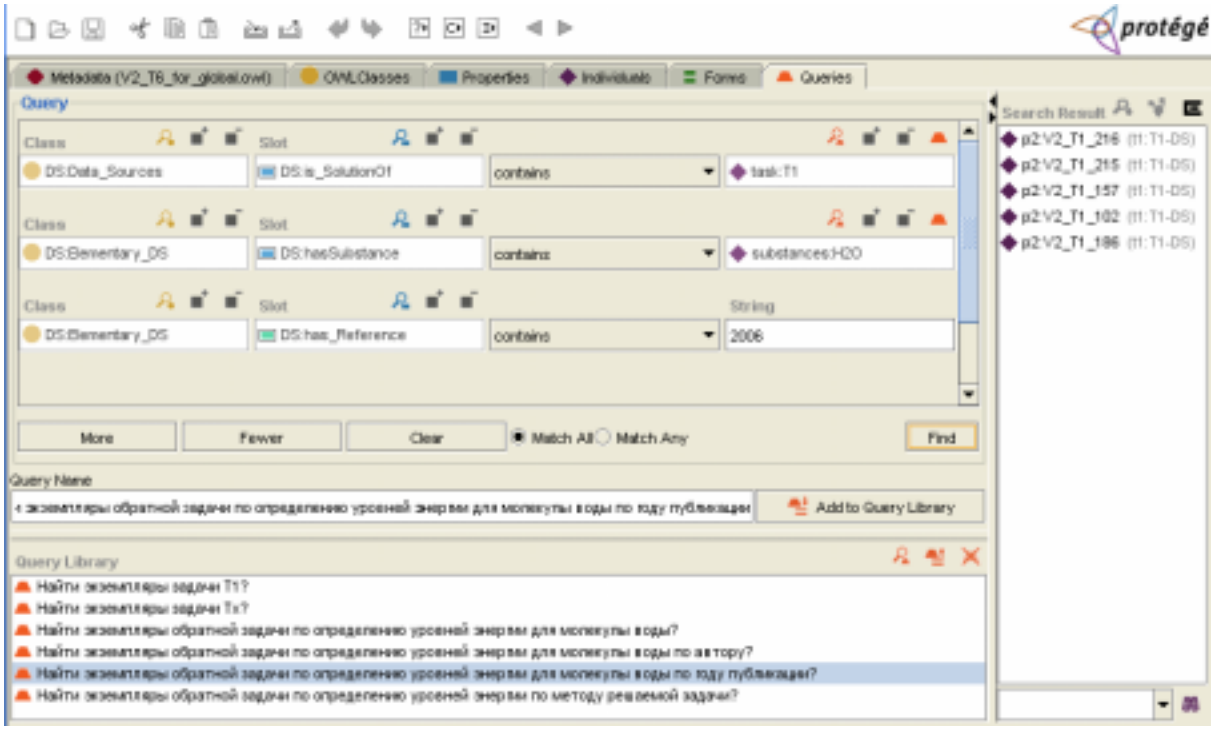


Рис.5. Использование Protégé для организации ответов на вопросы пользователя

Интенсионалами выходных данных задачи T6 являются уровень энергии молекулы, погрешность определения уровня энергии, число переходов, использованных для определения уровня энергии, и квантовые числа, характеризующие уровень энергии. Стоит отметить, что для молекулы воды существует три типа квантовых чисел, в силу сложившихся исторически традиций. Минимальная кардинальность свойства *иметь_уровень_энергии* равна единице, также как минимальная кардинальность свойства *тип_квантовых_чисел*.

Таблица 1. Структура метаданных задачи T1.

Интенсионал	Тип	Назначение
Название вещества	string	H_2O
Потенциальная функция	URI	Ссылка на массив, описывающий потенциальную функцию молекулы
Базисные волновые функции	URI	Ссылка на массив, описывающий набор базисных волновых функций (типичные объемы > 100Гб)
Метод	string	Название метода которым была решена задача
Emin	float	Минимальное значение уровня энергии в массиве данных
Emax	float	Максимальное значение уровня энергии в массиве данных
N	integer	Число уровней энергии
Угловой момент J_{max}	integer	Максимальное значение углового момента
Тип квантовых чисел	string	(нормальные моды, BN2, Schwenke)
Единица измерения	string	Единица измерения уровня энергии

Для работы пользователя в ИВС факты (индивидуалы классов) прикладной онтологий задач визуализируются в виде HTML-страниц. Пример такой визуализации представлен на рис.4. Визуализация необходима при сравнении предметных данных из разных источников, так как при ней явным образом представляются количественные и качественные метаданные для сравниваемых массивов.

В настоящее время в ИВС “Атмосферная спектроскопия” нет приложений, поддерживающих слой знаний. Использование прикладной онтологии возможно на клиентском месте с помощью редактора Protégé. Средства этого редактора позволяют пользователю составлять запросы с помощью конструкции утверждения [16] (субъект, предикат и объект) экземпляры которого можно создавать из концептов прикладных онтологий. На рис.5 показан интерфейс редактора Protégé, представляющий фиксированный список вопросов, относящихся к задаче поиска решений спектроскопических задач T1 и T6.. Выделенный на рис. 5 вопрос разлагается на три вспомогательных вопроса. Пользователь может менять значения объекта в каждом утверждении.

6 Заключение

В работе представлена прикладная онтология задач, созданная с помощью рекомендации W3C OWL, ориентированная на представление и обработку знаний в молекулярной спектроскопии. В нашей работе представлена прикладная онтология задач, специфицированная в рамках OWL DL и ориентированная на решение задачи поиска источников данных в молекулярной

спектроскопии в рамках выполненной нами концептуализации предметной области.

Авторы благодарны Российскому фонду фундаментальных исследований за поддержку работы (гранты 05-07-90196, 06-07-89201) и д.ф.-м.н. Родимовой О.Б. за помощь в составлении типовых вопросов для задачи нахождения уровней энергии молекулы.

Литература

1] Hitran, http://www.hitran.com
2] The GEISA Spectroscopic Database, http://ether.ipsl.jussieu.fr
3] OWL Web Ontology Language Semantics and Abstract Syntax, W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-owl-semantics-20040210/
4] RacerPro version 1.8.1, http://www.racer-systems.com/
5] D.Oberle, Semantic Management of Middleware, Springer, Berlin, 2006, 268p
6] J. Sowa, Knowledge Representation, Brooks/Cole, 2000, 592p.
7] Genesereth, M. R. and Nilsson, N. J. 1987. *Logical Foundation of Artificial Intelligence.* Morgan Kaufmann, Los Altos, California
8] Nicola Guarino, Formal Ontology and Information Systems, Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.
9] T. R. Gruber, Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, Volume 43 , Issue 5-6 Nov./Dec. 1995, p. 907-928.
10] A.D. Bykov, A.Z.Fazliev,A.V. Kozodoev, A.I. Privezentsev, L.N.Sinitsa, M.V.Tonkov, N.N.Filippov, and M.Yu. Tretyakov, Distributed information system on molecular spectroscopy, Proc. of SPIE, International Symposium on High Resolution Molecular Spectroscopy, 2006**, v.** 6580 pp. 65800W.
11] De Roure D., Jennings N., Shadbolt N., A Future e-Science Infrastructure, Report commissioned for EPSRC/DTI Core e-Science Programme, 2001, 78p.
12] J.Dehnert, A methodology for workflow modeling, Dissertation, Berlin, 2003, 200p.
13] R. Lancashire, T. Davies Spectroscopic Data: The Quest for a Universal Format, Chemistry International, Vol. 28, No. 1, 2006, http://www.iupac.org/publications/ci/2006/2801/3_ref5.html
14] R.J.Barber, J. Tennyson, G.J. Harris, R.N. Tolchenov, A high accuracy computed water line list - BT2. // Mon. Not. R. Astron. Soc., 2006, T. 368, C. 1087-1094.
15] Protégé, ontology editor and knowledge-base framework, http://protege.stanford.edu/
16] RDF Semantics, W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-rdf-mt-20040210

Application task ontology for molecular spectroscopy

Fazliev A.Z., Privezetsev A.I.

The description of the application task ontology used for systematization of informational resources in molecular spectroscopy is presented. A few possibilities of formation of molecular energy levels facts, their computer check of consistency with relation to taxonomy of spectroscopic tasks and the search of the data sources of solved spectroscopic tasks on the base of concepts of task ontology are described.