

Онтология декомпозиции экспертных источников информации в НИВС "Молекулярная спектроскопия"

Н.А. Лаврентьев¹, А.И. Привезенцев¹

¹Институт оптики атмосферы СО РАН, Томск

e-mail: lnick@iao.ru, remake@iao.ru

Аннотация. В работе описывается онтология результатов декомпозиции экспертных источников информации по отношению к первичным источникам информации. Декомпозиция проводится по полному набору первичных источников информации, относящихся к однотипным решениям задач предметной области. Онтология содержит описание результатов декомпозиции всех опубликованных экспертных источников информации. Декомпозиция проведена по следующим группам опубликованных первичных источников информации: только экспериментальные источники информации, только теоретические источники информации, все источники информации. Кроме этого, декомпозиция проведена по различным диапазонам физической величины. Онтология может использоваться для анализа и выбора экспертных источников информации. В онтологии видно наличие данных, которые не были опубликованы в исследуемом диапазоне физической величины.

Введение

Модель предметной области, положенная в основу систематизации данных в научной информационно-вычислительной системе (НИВС) W@DIS, представляется в виде совокупности решаемых в ней задач, где каждая задача представлена в виде IPO-модели (Input-Processing-Output). Следуя этой модели предметной области, опубликованные числовые данные являются результатом решения определенной предметной задачи. Связка результата решения определенной предметной задачи и библиографической информации о публикации, по которой можно однозначно идентифицировать этот результат, является источником информации [1].

Под первичным источником информации подразумеваются все части опубликованного решения задачи количественной спектроскопии, дополненные базовыми свойствами: названием молекулы, библиографической ссылкой и названием метода решения задачи, при этом кардинальность всех этих базовых свойств равна единице. Составным источником информации называется источник информации, в котором кардинальность любого базового свойства первичного источника информации больше единицы [2].

Экспертные источники информации относятся к классу составных источников информации. Экспертные источники информации представляют собой сами спектральные данные и связанные с ними публикации, в которых авторы предлагают эти составные массивы спектральных данных, отобранные по некоторым критериям, как правило, в явном виде не описанным. Эти массивы, содержащие несколько миллионов значений физических величин, снабжаются библиографией, относящейся почти каждому из значений такого массива. Проверить соответствие библиографии и значений, размещенных в экспертном массиве, не используя компьютерные приложения, практически невозможно.


На практике проверка опубликованности данных важна в силу разных причин. Например, при решении задач спектроскопии важны условия, при которых проводились измерения, или приближения, используемые при вычислениях. Использование библиографических ссылок на публикации, позволяет узнать условия экспериментов или приближения, при которых построены модели, что увеличивает доверие к данным.

Декомпозиция в научной информационно-вычислительной системе W@DIS

Критерий для проверки ограничения опубликования, используемый в спектроскопии, основан на допущении, что проверяемые данные должны отличаться от опубликованных данных не более чем на величину ошибок измерений, характерных для соответствующего диапазона вакуумных волновых чисел [3].

При проведении работ по систематизации спектральных данных по изотопологам воды создана информационная система W@DIS в которой аккумулированы все первичные данные, собранные за период более 70 лет по ряду молекул. Часть данных о вакуумных волновых числах воды была собрана группой IUPAC [4-6] и выделены данные, содержащие не корректные отнесения квантовых чисел. Эти скорректированные решения задач использованы при декомпозиции.

Для решения задачи декомпозиции экспертного массива данных на часть, содержащую опубликованные данные и часть, содержащую неопубликованные данные, в научной информационно-вычислительной системе W@DIS был создан сервис, проводящий такое разложение в соответствии с ограничениями на допустимую величину отклонения сравниваемых величин [3].

 Декомпозиция

| | |
|------------------------------|---|
| 2009_RoGoBaBe_CO2-air | L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc, The HITRAN 2008 molecular spectroscopic database. // Journal of Quantitative Spectroscopy and Radiative Transfer, 2009, Volume 110, Issue 9, Pages 533-572, DOI: 10.1016/j.jqsrt.2009.02.013. |
|------------------------------|---|

Поиск источников информации

| | |
|---|--|
| Выбор задачи | <input type="radio"/> Уровни энергии <input checked="" type="radio"/> Профили линий <input type="radio"/> Переходы |
| Вещество | CO2 |
| Диапазон вакуумных частот (см ⁻¹) | 0 - 30000 |
| Слова для поиска источников данных по контексту, содержащемуся в аннотации или ссылке на публикацию. (Фамилии авторов публикаций, журнал, год публикации, слова из названий статьи) | |

Искать источники информации

Выбрать источник информации

Показать 40 строк от 1 Всего строк 337 << < > >> Настройки

| Выбор | Название Вычисления/Эксперимент | Число записей | Публикация |
|-----------------------|------------------------------------|------------------|---|
| <input type="radio"/> | 1971_ToHuPI_CO2 | 202 | Robert A. Toth, Robert H. Hunt, Earle K. Plyler, Lines intensities of the CO ₂ Σ-Σ bands in the 1.43-1.65 μ region, Journal of Molecular Spectroscopy, 1971, Volume 38, Issue 1, Pages 107-117, DOI: 10.1016/0022-2852(71)90096-8. Annotation |
| <input type="radio"/> | 1971_ToHuPI_CO2 | 248 | Robert A. Toth, Robert H. Hunt, Earle K. Plyler, Lines intensities of the CO ₂ Σ-Σ bands in the 1.43-1.65 μ region, Journal of Molecular Spectroscopy, 1971, Volume 38, Issue 1, Pages 107-117, DOI: 10.1016/0022-2852(71)90096-8. Annotation |

Рис. 1. Интерфейс для выбора источника данных с целью его декомпозиции

Для пользователя процедура декомпозиции массива данных состоит из

последовательного набора действий.

1. Выбор данных для последующей декомпозиции (рис.1).
2. Задание допустимого приближения вакуумных волновых чисел (рис.2).
3. Задание интервала вакуумных волновых чисел (рис. 2).
4. Указание на то по какой части первичных данных (расчетной и/или экспериментальной) проводить декомпозицию переходов с одинаковыми квантовыми числами (рис. 2).
5. Просмотр результатов коррелирующих источников (рис. 3).
6. Просмотр остатка декомпозиции, то есть переходов не найденных в первичных опубликованных источниках информации (рис. 4).

Параметры декомпозиции

| | |
|------------------------------|---|
| 2009_RoGoBaBe_CO2-air | L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc, The HITRAN 2008 molecular spectroscopic database. // Journal of Quantitative Spectroscopy and Radiative Transfer, 2009, Volume 110, Issue 9, Pages 533-572, DOI: 10.1016/j.jqsrt.2009.02.013. |
|------------------------------|---|

| | | |
|--|--|---|
| <input checked="" type="checkbox"/> | Ошибка определения частоты (см ⁻¹) | <input type="text" value="1"/> |
| <input type="checkbox"/> | Ошибка определения интенсивности (%) | <input type="text" value="0"/> |
| <input checked="" type="checkbox"/> | Ограничения на частоту (см ⁻¹) | <input type="text" value="0"/> - <input type="text" value="30000"/> |
| <input checked="" type="checkbox"/> | Декомпозиция по эксперименту | |
| <input type="checkbox"/> | Декомпозиция по расчету | |
| <input type="button" value="Провести декомпозицию"/> | | |

| <input type="button" value="Показать"/> | <input type="text" value="40"/> | строк от <input type="text" value="0"/> | Всего строк 255 | <input type="button" value=">"/> | <input type="button" value=">>"/> | <input type="button" value="Настройки"/> |
|---|---------------------------------|---|---|---|---|--|
| Название Вычисления/Эксперимент | Число записей | Число совпадений | Показать | Публикация | | |
| 1990_JoAu_CO2 | 761 | 413 | <input type="button" value="Показать"/> | <i>J.W.C. Johns and J. Vander Auwera, Absolute intensities in CO₂: The ν₂ fundamental near 15 μm, Journal of Molecular Spectroscopy, 1990, Volume 140, Issue 1, Pages 71-102, DOI: 10.1016/0022-2852(90)90008-E.</i> <input type="button" value="Annotation"/> | | |
| 1987_Johns_b_CO2_300 | 265 | 239 | <input type="button" value="Показать"/> | <i>Johns, J.W.C., High resolution and the accurate measurement of intensities, Mikrochimica Acta, 1987, Volume 93, Issue 1-6, Pages 171-188, DOI: 10.1007/BF01201689.</i> <input type="button" value="Annotation"/> | | |

Рис. 2. Интерфейс для задания условий декомпозиции и представления ее результатов.

При просмотре результатов декомпозиции (рис.2) для каждого первичного источника данных указывается число переходов, удовлетворяющих заданным ограничениям, и в колонке «Показать» размещена ссылка на таблицы, в которых размещены переходы из первичного источника данных, соответствующие указанным в интерфейсе ограничениям.

| Источник информации слева | | | | | | | | | | Источник информации справа | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| 2009_RoGoBaBe_CO2-air | | | | | | | | | | 1990_JoAu_CO2 | | | | | | | | | |
| L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc, The HITRAN 2008 molecular spectroscopic database. // Journal of Quantitative Spectroscopy and Radiative Transfer, 2009, Volume 110, Issue 9, Pages 533-572, DOI: 10.1016/j.jqsrt.2009.02.013. | | | | | | | | | | J.W.C. Johns and J. Vander Auwera, Absolute intensities in CO ₂ : The v ₂ fundamental near 15 μm, Journal of Molecular Spectroscopy, 1990, Volume 140, Issue 1, Pages 71-102, DOI: 10.1016/0022-2852(90)90008-E. | | | | | | | | | |
| Annotation | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | |
|----------|----|----------|---|-----------------|--|--|--|--|--|--|--|--|--|--|--|---|----|-----------|
| Показать | 40 | строк от | 0 | Всего строк 413 | | | | | | | | | | | | > | >> | Настройки |
|----------|----|----------|---|-----------------|--|--|--|--|--|--|--|--|--|--|--|---|----|-----------|

| ω_1 (cm ⁻¹) | | n ₁ ^{up} | n ₂ ^{up} | l ₂ ^{up} | n ₃ ^{up} | r ^{up} | Sym ^{up} | j ^{up} | n ₁ ^{low} | n ₂ ^{low} | l ₂ ^{low} | n ₃ ^{low} | r ^{low} | Sym ^{low} | j ^{low} | ω_2 (cm ⁻¹) | | $\Delta\omega$ (cm ⁻¹) | ΔI (см/мол) % |
|--------------------------------|-----------|------------------------------|------------------------------|------------------------------|------------------------------|-----------------|-------------------|-----------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|------------------|--------------------|------------------|--------------------------------|-----------|------------------------------------|-----------------------|
| 587.311009 | 5.106e-22 | 1 | 0 | 0 | 0 | 2 | e | 38 | 0 | 1 | 1 | 0 | 1 | e | 39 | 587.3113 | 5.376e-22 | 2.9100e-4 | 5.2879e+0 |
| 588.901402 | 6.543e-22 | 1 | 0 | 0 | 0 | 2 | e | 36 | 0 | 1 | 1 | 0 | 1 | e | 37 | 588.9017 | 6.744e-22 | 2.9800e-4 | 3.0720e+0 |
| 590.489672 | 8.235e-22 | 1 | 0 | 0 | 0 | 2 | e | 34 | 0 | 1 | 1 | 0 | 1 | e | 35 | 590.49 | 8.571e-22 | 3.2800e-4 | 4.0801e+0 |

Рис. 3. Интерфейс для просмотра совпадения двух источников информации

| | | | | | | | | | | | | | | | | | | | |
|-----------------------|--|--|--|--|--|--|--|--|--|---|--|--|--|--|--|--|--|--|--|
| 2009_RoGoBaBe_CO2-air | | | | | | | | | | L.S. Rothman, I.E. Gordon, A. Barbe, D.Chris Benner, P.F. Bernath, M. Birk, V. Boudon, L.R. Brown, A. Campargue, J.-P. Champion, K. Chance, L.H. Coudert, V. Dana, V.M. Devi, S. Fally, J.-M. Flaud, R.R. Gamache, A. Goldman, etc, The HITRAN 2008 molecular spectroscopic database. // Journal of Quantitative Spectroscopy and Radiative Transfer, 2009, Volume 110, Issue 9, Pages 533-572, DOI: 10.1016/j.jqsrt.2009.02.013. | | | | | | | | | |
|-----------------------|--|--|--|--|--|--|--|--|--|---|--|--|--|--|--|--|--|--|--|

| | | | | | | | | | | | | | | | | | | |
|----------|----|----------|---|--------------------|--|--|--|--|--|--|--|--|--|--|--|---|----|-----------|
| Показать | 40 | строк от | 0 | Всего строк 110956 | | | | | | | | | | | | > | >> | Настройки |
|----------|----|----------|---|--------------------|--|--|--|--|--|--|--|--|--|--|--|---|----|-----------|

| Волновое число (cm ⁻¹) | Интенсивность (см/мол) | n ₁ ^{up} | n ₂ ^{up} | l ₂ ^{up} | n ₃ ^{up} | r ^{up} | Sym ^{up} | j ^{up} | n ₁ ^{low} | n ₂ ^{low} | l ₂ ^{low} | n ₃ ^{low} | r ^{low} | Sym ^{low} | j ^{low} |
|------------------------------------|------------------------|------------------------------|------------------------------|------------------------------|------------------------------|-----------------|-------------------|-----------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|------------------|--------------------|------------------|
| 352.06615 | 5.13e-30 | 2 | 1 | 1 | 0 | 3 | e | 41 | 2 | 0 | 0 | 0 | 1 | e | 42 |
| 353.59269 | 6.78e-30 | 2 | 1 | 1 | 0 | 3 | e | 39 | 2 | 0 | 0 | 0 | 1 | e | 40 |
| 355.11789 | 8.78e-30 | 2 | 1 | 1 | 0 | 3 | e | 37 | 2 | 0 | 0 | 0 | 1 | e | 38 |
| 356.64223 | 1.12e-29 | 2 | 1 | 1 | 0 | 3 | e | 35 | 2 | 0 | 0 | 0 | 1 | e | 36 |

Рис. 4. Интерфейс для просмотра остатка

На рис. 3 показано сравнение коррелирующих переходов двух источников, а на рис. 4 показана часть экспертного источника информации, не содержащаяся ни в одном из первичных источников информации с указанной точностью.

Техническая реализация программного модуля декомпозиции

Декомпозиция составных источников данных в системе W@DIS производится в несколько шагов, на каждом из которых определённым классом реализуется функциональность одной или нескольких операций с источниками данных и разнообразными критериями выборки и отображения результатов. Список этих классов приводится ниже [7].

Класс `decomposition_sel`, предназначен для выбора источника данных, участвующего в декомпозиции. Класс предоставляет возможность фильтровать источники данных по типу задачи (профили линий и переходы), веществу, диапазону вакуумных волновых чисел и ключевым словам в описании источника.

Класс `decomposition` осуществляет декомпозицию. Он позволяет пользователю задать

параметры декомпозиции, такие как: ошибка определения вакуумных волновых чисел, диапазон изменения вакуумных волновых чисел и тип источников, по которым проводится декомпозиция (экспериментальные и расчетные). Показывает результаты декомпозиции в формате: название источника, число записей в источнике, число совпадений при декомпозиции, ссылка на совпадающие переходы и публикация, к которой привязан источник.

Класс `decomposition_equal` осуществляет табличное представление совпадающих переходов для пары источников. Показывает квантовые числа, вакуумные волновые числа и интенсивность линий для первого и второго источника и их разность.

Класс `decomposition_rem` осуществляет табличный вывод остатка: переходов, содержащихся в раскладываемом источнике, но не имеющих в заданном диапазоне изменения вакуумных волновых чисел совпадений со всеми остальными первичными источниками. Показывает в зависимости от задачи вакуумные волновые числа и интенсивность соответствующего перехода и его квантовые числа.

Базовый класс `vars`, содержащий ассоциативные массивы конфигурационных переменных для всех классов этого модуля.

Онтология декомпозиции экспертных источников информации

Созданную онтологию на языке OWL DL [8] используют для машинной систематизации и интерпретации знаний, относящихся к качеству спектральных данных. Классы прикладной онтологии результатов декомпозиции показаны на рис. 5.

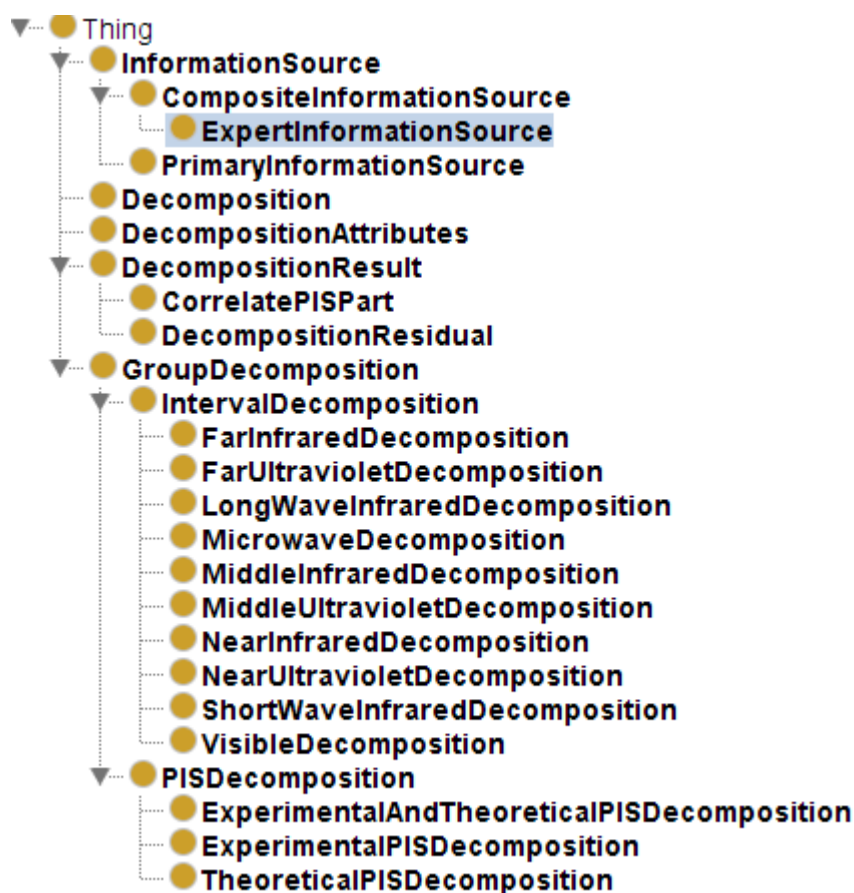


Рис. 5. Классы прикладной онтологии результатов декомпозиции

Объектные свойства прикладной онтологии результатов декомпозиции показаны на рис. 6.

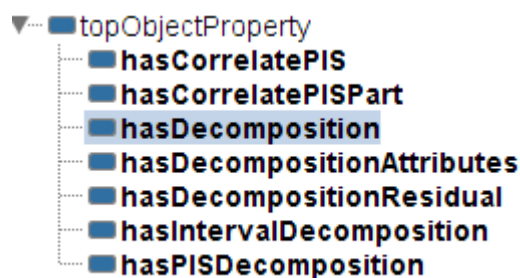


Рис. 6. Объектные свойства прикладной онтологии результатов декомпозиции

Типизированные свойства прикладной онтологии результатов декомпозиции показаны на рис. 7.

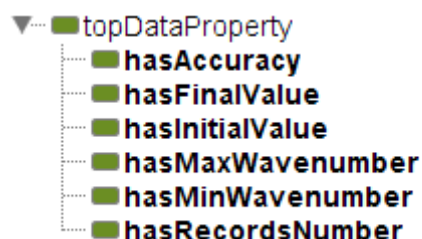


Рис. 7. Типизированные свойства прикладной онтологии результатов декомпозиции

Ключевым понятием в онтологии является источник информации (InformationSource). Источник информации может быть первичным (PrimaryInformationSource, сокращённое название PIS) или составным (CompositeInformationSource). К классу составных источников информации относятся экспертные источники информации (ExpertInformationSource). Экспертный источник информации содержит результаты декомпозиции (hasDecomposition). Декомпозиция источника информации (Decomposition) состоит из трёх групп (hasPISDecomposition) опубликованных первичных источников информации: только экспериментальные массивы данных (ExperimentalPISDecomposition), только теоретические массивы данных (TheoreticalPISDecomposition), все массивы данных (ExperimentalAndTheoreticalPISDecomposition). Каждая из этих трёх групп разбита (hasIntervalDecomposition) на 10 диапазонов декомпозиции (подклассы класса IntervalDecomposition). Каждый диапазон декомпозиции имеет параметры декомпозиции (hasDecompositionAttributes). Параметры декомпозиции (DecompositionAttributes) описывают интервал изменения вакуумных волновых чисел (hasInitialValue, hasFinalValue) и значение допустимого приближения вакуумных волновых чисел (hasAccuracy).

Результат декомпозиции состоит из остатка (DecompositionResidual) и описания частей коррелированных источников информации (CorrelatePISPart), в каждом из 10 диапазонов декомпозиции (подклассах класса IntervalDecomposition) будут свои (hasCorrelatePISPart) описания частей коррелированных источников информации (CorrelatePISPart). Описание части коррелированного источника информации (CorrelatePISPart) описывает диапазон, где экспертный источник информации связан с первичным источником информации. Кардинальность свойства hasCorrelatePISPart может

быть от 0 до бесконечности, так как первичных источников информации по которым раскладывается составной источник информации может быть много. Описание части коррелированного источника информации (CorrelatePISPart) содержит информацию о первичном источнике информации (hasCorrelatePIS), а также описывает диапазон вакуумных волновых чисел в котором произошло совпадение квантовых чисел переходов (hasMinWavnumber, hasMaxWavenumber) и число таких совпадений (hasRecordsNumber).

Результат остатка декомпозиции составных экспертных источников информации (DecompositionResidual) в каждом из 10 диапазонов декомпозиции (подклассах класса IntervalDecomposition) будет свой (hasDecompositionResidual), кроме этого, в каждой из 3 групп декомпозиции (подклассах класса PISDecomposition) результат остатка будет свой. Результат остатка декомпозиции источников информации (DecompositionResidual) описывает диапазон, где экспертный источник информации не связан ни с одним первичным источником информации. Кардинальность свойства hasDecompositionResidual будет всегда 1, так как остаток всегда должен быть указан, даже если он будет нулевым. Результат остатка декомпозиции источников информации (DecompositionResidual) описывает диапазон вакуумных волновых чисел в котором не произошло совпадение квантовых чисел переходов (hasMinWavnumber, hasMaxWavenumber) и число таких пропусков (hasRecordsNumber).

Экземпляры (индивиды) описанных выше классов онтологии результатов декомпозиции содержат полную информацию о всех опубликованных составных экспертных источниках информации.

Заключение

В работе описывается онтология результатов декомпозиции экспертных источников информации по отношению к первичным источникам информации. Декомпозиция проводится по полному набору первичных источников информации, относящихся к однотипным решениям задач предметной области. Онтология содержит описание результатов декомпозиции всех опубликованных экспертных источников информации. Декомпозиция проведена по следующим группам опубликованных первичных источников информации: только экспериментальные источники информации, только теоретические источники информации, все источники информации. Кроме этого, декомпозиция проведена по различным диапазонам физической величины. Онтология может использоваться для анализа и выбора экспертных источников информации. В онтологии видно наличие данных, которые не были опубликованы в исследуемом диапазоне физической величины.

ЛИТЕРАТУРА

[1]. Привезенцев А.И. Концепция создания информационно-вычислительных систем, содержащих опубликованные числовые массивы данных. – Электрон. текстовые дан. – 2010. – Режим доступа: <http://conf.nsc.ru/dicr2010/ru/reportview/29818>. – Доклады XIII Российской конференции "Распределенные информационные и вычислительные ресурсы" (DICR'2010).

[2]. Апанович, З.В., Ахлестин А.Ю., Винокуров П.С., Привезенцев А.И., Фазлиев А.З. Цифровая библиотека научных статей по количественной спектроскопии // Труды XIV Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL'2012) – Переславль-Залесский, 15-18 октября 2012. – Переславль-Залесский: Университет города Переславля, 2012. – С. 257-266.

[3]. Лаврентьев Н.А., Фазлиев А.З., Макогон М.М. Система проверки ограничений существования для молекулярной спектроскопии. – Электрон. текстовые дан. – 2010. – Режим доступа: <http://conf.nsc.ru/dicr2010/ru/reportview/29501>. – Доклады XIII Российской конференции "Распределенные информационные и вычислительные ресурсы"(DICR'2010).

[4]. IUPAC project N 2004-035-1-100 «A database of water transitions from experiment and theory». <http://www.iupac.org/web/ins/2004-035-1-100>.

[5]. J.Tennyson, P.F.Bernath, L.R.Brown, et al., IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part I. Energy Levels and Transition Wavenumbers for H₂¹⁷O and H₂¹⁸O // J. Quant. Spectrosc. Radiat. Transfer, 2009, V. 110, Issue 9, P. 573-596.

[6]. J. Tennyson, P. F. Bernath, L.R. Brown, et al., IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part II: Energy levels and transition wavenumbers for HD¹⁶O, HD¹⁷O, and HD¹⁸O // J. Quant. Spectrosc. Radiat. Transfer, 2010, V. 111, Issue 15, P. 2160-2184.

[7]. Ахлестин А.Ю., Козодоев А.В., Лаврентьев Н.А., Привезенцев А.И., Фазлиев А.З. Базы знаний для описания информационных ресурсов в молекулярной спектроскопии 4. Программное обеспечение ИВС "Молекулярная спектроскопия" [Электронный ресурс] // Электронные библиотеки: Российский научный электронный журнал. – Электрон. журн. – М.: Институт развития информационного общества, 2012. – Т. 15. – № 3. – Режим доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2012/part3/AKLPF>, свободный.

[8]. OWL Web Ontology Language Semantics and Abstract Syntax / edited by P. F. Patel-Schneider, I. Horrocks. – Электрон. текстовые дан. – 2004. – Режим доступа: <http://www.w3.org/TR/owl-semantics/>. – W3C Recommendation.