

Опыт построения информационной системы на основе онтологии *информационных ресурсов* количественной предметной области

Фазлиев А.З., Привезенцев А.И., Лаврентьев Н.А.
(ИОА СО РАН)



Содержание

Абстрактная структура информационной системы

(данные, приложения и интерфейсы пользователя)

Слои информационной системы (e-Science)

Опыт построения слоя знаний в ИС с помощью OWL DL.

Tbox. Вычислительные задачи в области молекулярной спектроскопии. Данные и метаданные.

Abox. Данные измерений и вычислений. Открытый или замкнутый мир? Отношения транзитивности и рефлексивности в количественной спектроскопии.

Примеры. Ограничения на значения и существование, среднеквадратические отклонения, упорядочение сущностей по значениям свойств.



Абстрактная структура информационной системы

- Данные
- Приложения
- Интерфейсы

Guarino N. Formal Ontology and Information Systems // Proceedings of FOIS'98. Trento. Italy, 6–8 June 1998. Amsterdam: IOS Press. P. 3–15.



Слои информационной системы (e-Science)

- Данные и вычисления
- Метаданные (информационный слой)
- Онтологии (слой знаний)

De Roure D., Jennings N., Shadbolt N. A Future e-Science Infrastructure // Report commissioned for EPSRC/DTI Core e-Science Programme. 2001. 78 p.



Опыт построения слоя знаний в ИС с помощью **OWL DL**

Предметные области ИС

“**W**ater **@**ccessible **D**istributed **I**nformation **S**ystem”

D = D1 & D2 & D3

D1 – Молекулярная спектроскопия

D2 – Атомы и молекулы

D3 - Публикации



Открытый или замкнутый мир?

Точка зрения пользователя

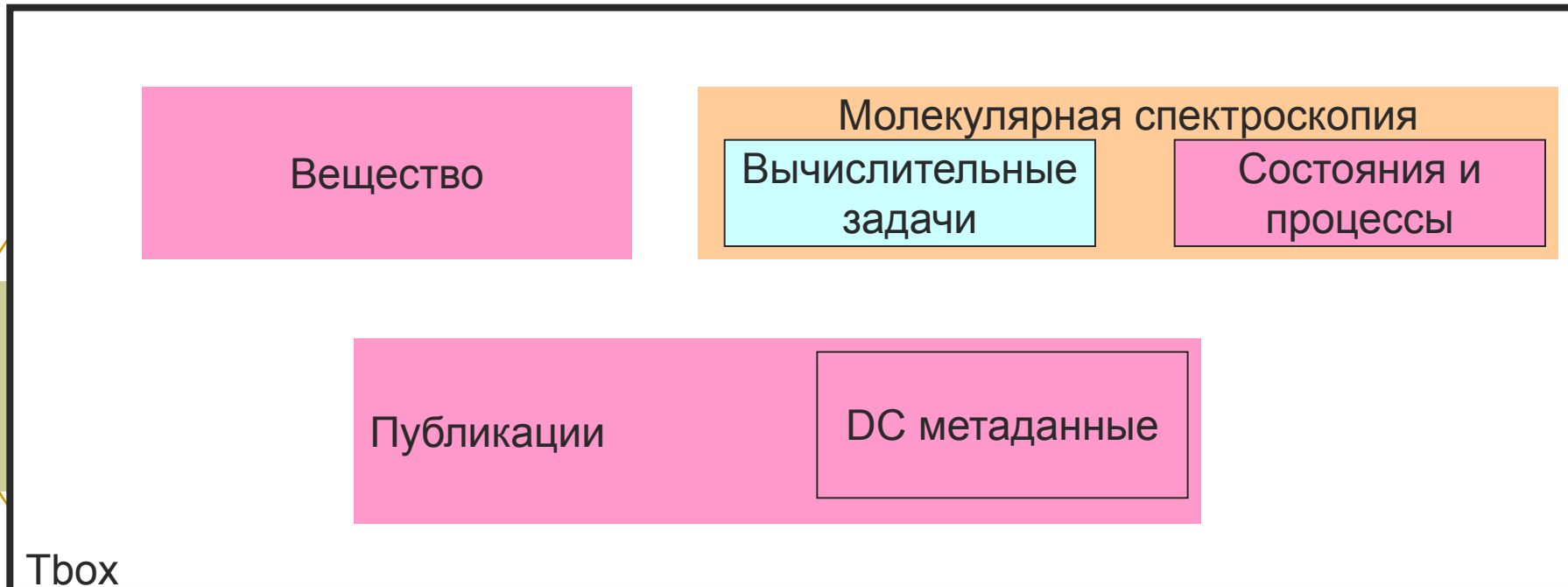
Tbox

- Замкнутое множество понятий ПО
- Расширяемое множество понятий ПО

Abox

- Возможность дополнения фактов (измерения и вычисления)
- Фиксация числа фактов





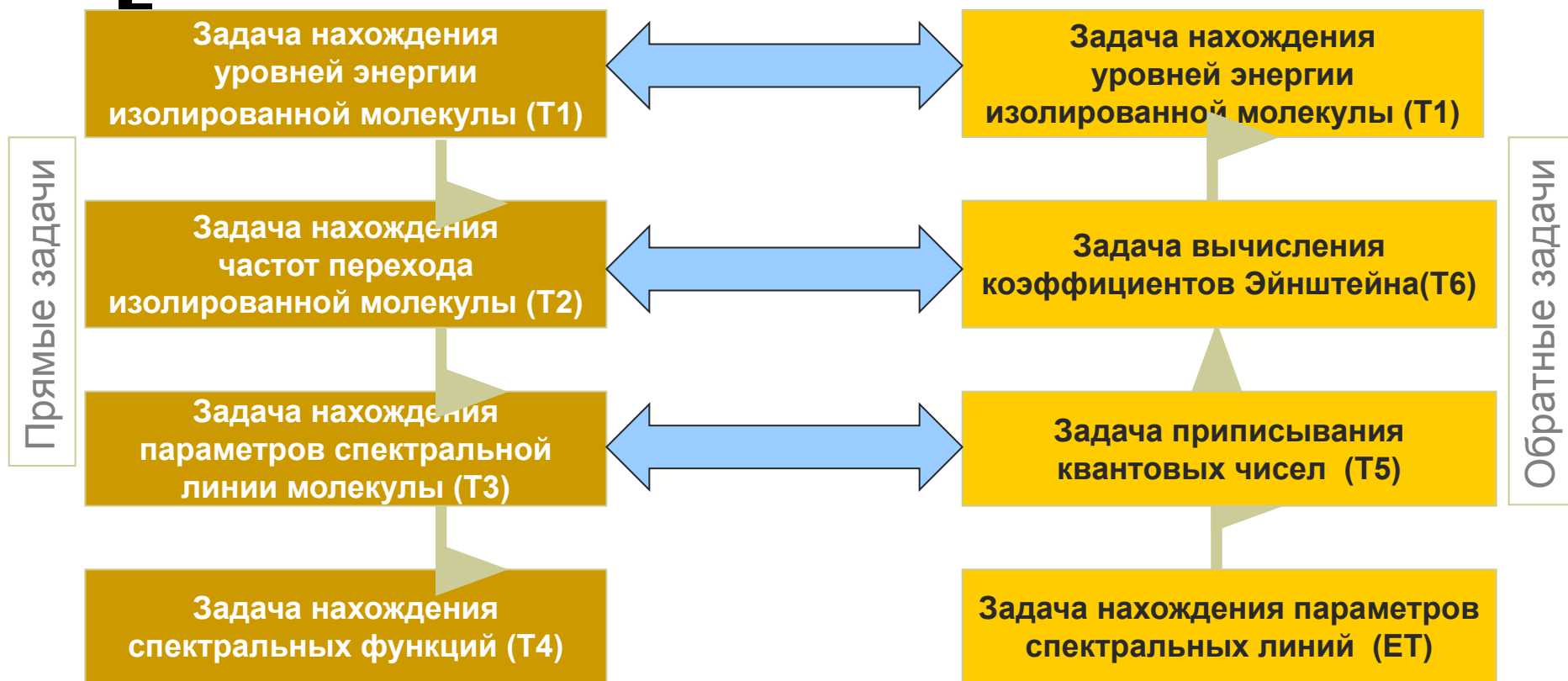
Модель молекулярной спектроскопии

(0-ое приближение)



Модель молекулярной спектроскопии

(1-ое приближение)



Стандартные отклонения



Абох. Данные измерений и вычислений.

Индивид «information source V4_T5_279_1998_ToBr_H2_170-H2O»

V4_T5_279_1998_ToBr_H2_170-H2O

isSolutionOf T5

hasMethod UNDEFINED

hasSubstance H2_17O

hasOutputData_MD V4_T5_279-T5_OutputData_MD

hasReference **Toth R.A., Brown L.R., Self-broadened widths and frequency shifts of water vapor lines between 590 and 2400 cm⁻¹. // Journal of Quantitative Spectroscopy and Radiation Transfer, 1998, v.59, p.529-562.**

V4_T5_279-T5_OutputData_MD

hasBroadeningSubstance_MD

V4_T5_279_BroadeningSubstances_MD_for_H2O

hasWavenumbers_MD V3_T5_279_Wavenumbers_MD

hasTransitions_MD

V4_T5_279_Transitions_MD_for_NormalModes

hasPhysicalCondition_MD

V4_T5_279_PhysicalCondition_MD

hasIntensity_MD V3_T5_279_Intensity_MD

V4_T5_279_BroadeningSubstances_MD_for_H2O

hasBroadeningSubstance H2O

hasHalfwidth_MD V3_T5_279_Halfwidth_MD_for_H2O

hasPressure_MD

V4_T5_279_PressureValue_MD_for_H2O

hasPressureDependence_MD

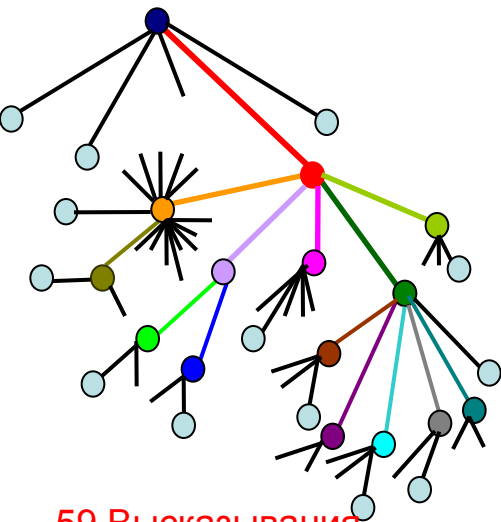
V4_T5_279_PressureDependence_MD_for_H2O

hasShift_MD V3_T5_279_Shift_MD_for_H2O

hasTemperatureDependence_MD

V4_T5_279_TemperatureDependence_MD_for_H2O

Субъектно-предикатная структура



59 Высказывания

V4_T5_279_PhysicalCondition_MD

hasTemperature_MD

V4_T5_279_TemperatureValue_MD

hasPressure_MD

V4_T5_279_PressureValue_MD

V4_T5_279_PressureValue_MD

hasUnit atm

hasFloatValue 1

V4_T5_279_TemperatureValue_MD

hasUnit K

hasFloatValue 296

H2O

hasSymmetryGroup C2v

hasPhysicalState SingleMolecule

V4_T5_279_Intensity_MD

hasUnit cm-1_molecule

hasUncertainty false

isPresented false

V4_T5_279_Wavenumbers_MD

hasUnit cm-1

hasUncertainty false

hasMaxWavenumber 2010.911865

hasMinWavenumber 1315.606567

hasNumberOfWavenumbers 142

V4_T5_279_Transitions_MD_for_NormalModes

hasSpectralBand

V4_T5_279_for_NormalModes_v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW_SpectralBand

hasQuantumNumbersType NormalModes

hasTotalMaxAngularMomentum 12 hasTotalMinAngularMomentum 0

hasNumberOfInvalidTransitions 0 hasNumberOfValidWater-C2V-Transitions 142

hasNumberOfRejectedTransitions 0 hasNumberOfUnassignedTransitions 0

hasNumberOfValidTransitions 142 hasNumberOfUniqueTransitions 142

hasNumberOfValidIdentifications 142 hasNumberOfInvalidWaterTransitions 0

hasNumberOfInvalidWater-C2V-Transitions 0 hasNumberOfInvalidIdentifications 0

hasNumberOfNonuniqueTransitions 0

V4_T5_279_Halfwidth_MD_for_H2O

hasUnit cm-1_atm-1

hasUncertainty true

isPresented true

V4_T5_279_PressureValue_MD_for_H2O

hasUnit atm

hasFloatValue 1

V4_T5_279_PressureDependence_MD_for_H2O

hasUncertainty false

isPresented false

V4_T5_279_Shift_MD_for_H2O

hasUnit cm-1_atm-1

hasUncertainty true

isPresented true

V4_T5_279_TemperatureDependence_MD_for_H2O

hasUncertainty false

isPresented false

V4_T5_279_for_NormalModes_v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW_SpectralBand

hasBandType

v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW

hasNumberOfSpectralBands 1

Индивид «information source 1998_ToBr_H2_170-H2O»

V4_T5_279_1998_ToBr_H2_170-H2O

isSolutionOf T5 hasMethod UNDEFINED hasSubstance H2_170 hasOutputData_MD V3_T5_279-T5_OutputData_MD
hasReference Toth R.A., Brown L.R., Self-broadened widths and frequency shifts of water vapor lines between 590 and 2400 cm⁻¹. // Journal of Quantitative Spectroscopy and Radiation Transfer, 1998, v.59, p.529-562.

V4_T5_279-T5_OutputData_MD

hasBroadeningSubstance_MD
V4_T5_279_BroadeningSubstances_MD_for_H2O
hasWavenumbers_MD V3_T5_279_Wavenumbers_MD
hasTransitions_MD
V4_T5_279_Transitions_MD_for_NormalModes
hasPhysicalCondition_MD
V4_T5_279_PhysicalCondition_MD
hasIntensity_MD V3_T5_279_Intensity_MD

V4_T5_279_BroadeningSubstances_MD_for_H2O

hasBroadeningSubstance H2O
hasHalfwidth_MD V3_T5_279_Halfwidth_MD_for_H2O
hasPressure_MD V3_T5_279_PressureValue_MD_for_H2O
hasPressureDependence_MD
V4_T5_279_PressureDependence_MD_for_H2O
hasShift_MD V3_T5_279_Shift_MD_for_H2O
hasTemperatureDependence_MD
V4_T5_279_TemperatureDependence_MD_for_H2O

V4_T5_279_PhysicalCondition_MD

hasTemperature_MD
V4_T5_279_TemperatureValue_MD
hasPressure_MD
V4_T5_279_PressureValue_MD

H2O

hasSymmetry Group C2v
hasPhysicalstate SingleMolecule

V4_T5_279_Halfwidth_MD_for_H2O

hasUnit cm-1_atm-1
hasUncertainty true
isPresented true

V4_T5_279_PressureValue_MD

hasUnit atm
hasFloatValue 1

V4_T5_279_Intensity_MD

hasUnit cm-1_molecule
hasUncertainty false
isPresented false

V4_T5_279_Shift_MD_for_H2O

hasUnit cm-1_atm-1
hasUncertainty true
isPresented true

V4_T5_279_TemperatureValue_MD

hasUnit K
hasFloatValue 296

V4_T5_279_Wavenumbers_MD

hasUnit cm-1
hasUncertainty false
hasMaxWavenumber 2010.911865
hasMinWavenumber 1315.606567
hasNumberOfWavenumbers 142

V4_T5_279_PressureDependence_MD_for_H2O

hasUncertainty false
isPresented false

V4_T5_279_TemperatureDependence_MD_for_H2O

hasUncertainty false
isPresented false

V4_T5_279_Transitions_MD_for_NormalModes

hasSpectralBand
V4_T5_279_for_NormalModes_v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW_SpectralBand
hasQuantumNumbersType NormalModes
hasTotalMaxAngularMomentum 12 hasTotalMinAngularMomentum 0
hasNumberOfInvalidTransitions 0 hasNumberOfValidWater-C2V-Transitions 142
hasNumberOfRejectedTransitions 0 hasNumberOfUnassignedTransitions 0
hasNumberOfValidTransitions 142 hasNumberOfUniqueTransitions 142
hasNumberOfValidIdentifications 142 hasNumberOfInvalidWater-Transitions 0
hasNumberOfInvalidWater-C2V-Transitions 0 hasNumberOfInvalidIdentifications 0
hasNumberOfNonuniqueTransitions 0

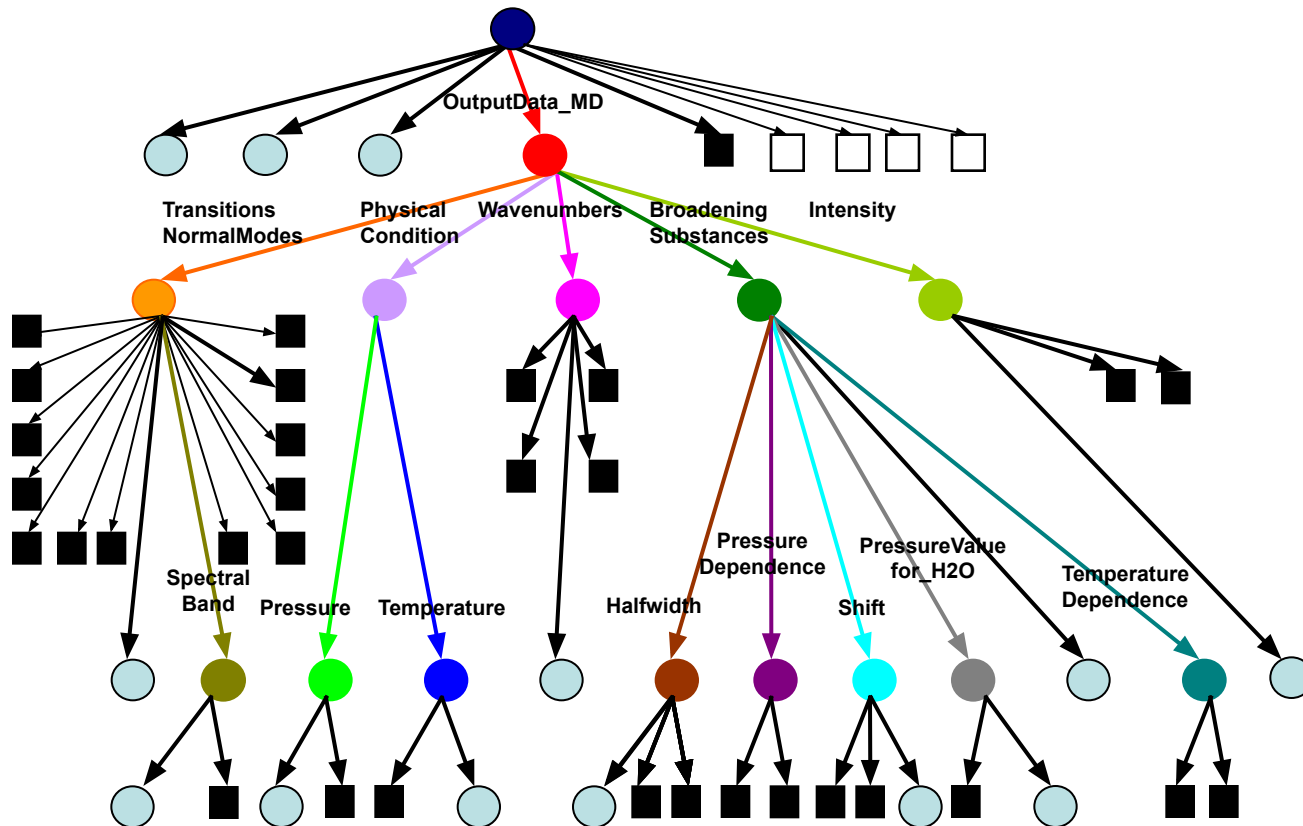
V4_T5_279_PressureValue_MD_for_H2O

hasUnit atm
hasFloatValue 1

V4_T5_279_for_NormalModes_v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW_SpectralBand
hasBandType
v1UP_v2UP_v3UP_v1LOW_v2LOW_v3LOW
hasNumberOfSpectralBands 1

Субъектно-предикатная структура «information source 1998_ToBr_H2_170-H2O»

1998_ToBr_H2_170-H2O



61 высказывания

Количественная предметная область



N_T – число высказываний в $Tbox$

N_A – число высказываний в $Abox$ (M - факты предметной области)

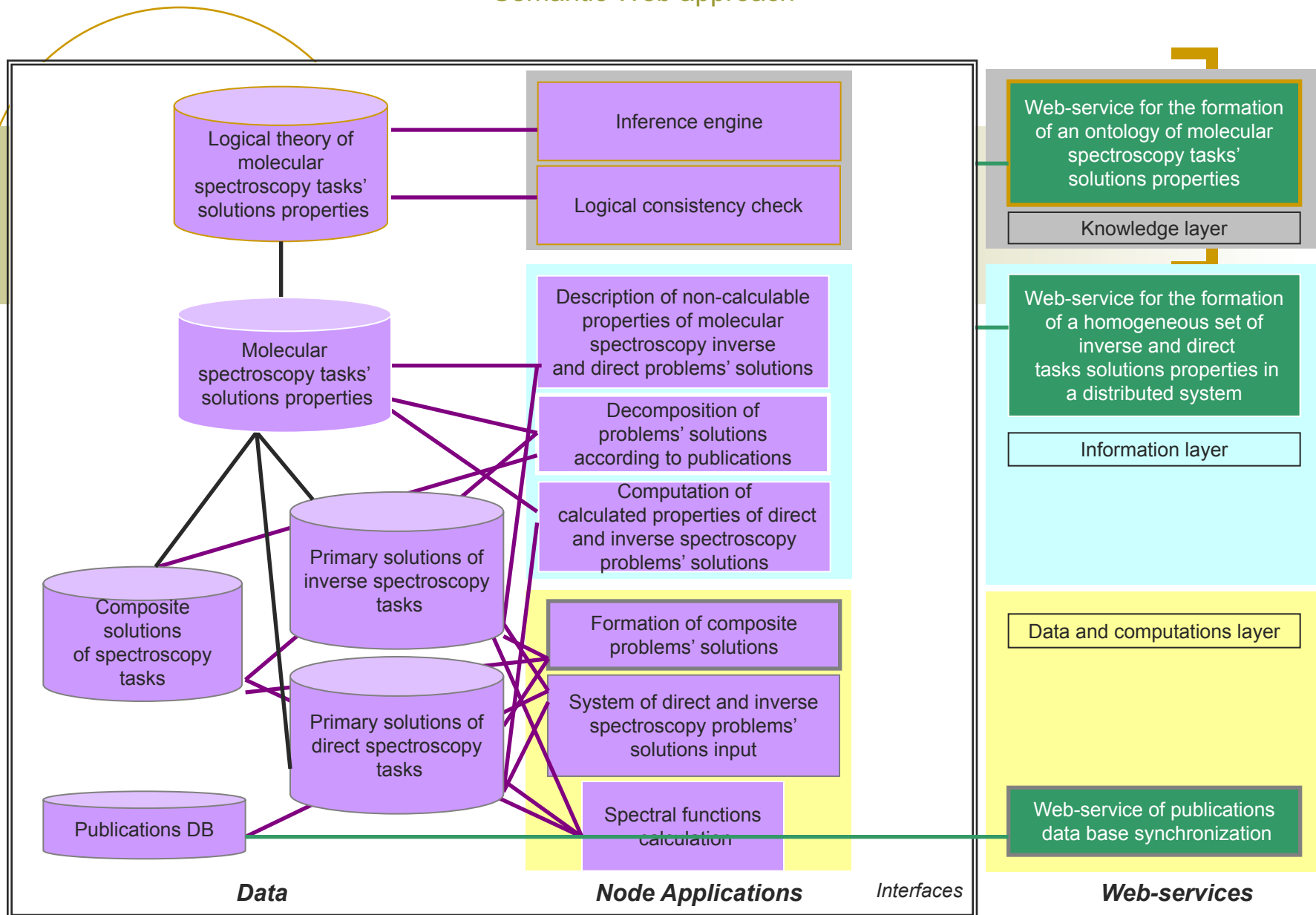
$N_{A(P)}$ – число высказываний в $Abox$ (P - факты предметной области, $P \subseteq M$)

$$N_T \ll N_{A(P)} \ll N_A$$



Архитектура распределенной информационной системы для молекулярной спектроскопии

Semantic Web approach



Иерархия математических теорий



FIGURE 2.14 Обобщение иерархии теорий

Sowa J.F. Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

594 p.



Отношения транзитивности и рефлексивности в количественной спектроскопии

- Транзитивность

$$a=b \ \& \ b=c \ \Rightarrow \ a=c$$

- Рефлексивность

$$a = a$$

- Симметрия

$$a=b \ \Rightarrow \ b=a$$

- Антисимметрия

$$a = < b \ \& \ b = < a \ \Rightarrow \ a = b$$



Примеры. Ограничения на значения и существование

Ограничения в задачах достоверности данных в молекулярной спектроскопии

Формальные ограничения

Тип данных –

квантовые числа – натуральные числа,
интенсивность, столкновительная полуширина, вакуумные волновые числа, уровни энергии – положительные действительные числа,

Интервалы изменения –

$0 < \text{волновые числа} < 45000 \text{ см}^{-1}$, $10^{-17} \text{ см/mol} < \text{интенсивность} < 10^{-30} \text{ см/mol}$

Правила отбора -

нормальные моды - $k_a + k_c = J$ or $J+1$,
точные квантовые числа – $J < 60$, $0 < s < 5$,

Ограничения опубликования

Решение задачи должно быть опубликованным информационным ресурсом

Неформальные ограничения

Точки зрения экспертов



Примеры. Сводка публикаций. Ограничения на значения

H ₂ O		H ₂ S											
Вещь	Вещь	T1		T2		T3		T5		T6		T7	
		HN	CN	HN	CN	HN	CN	HN	CN	HN	CN	HN	CN
H ₂ O	CO ₂	-	-	19 (13)	4 (3)	14 (12)	1 (0)	52 (48)	-	47 (41)	-	-	-
H ¹⁸ OH	H ₂ S	-	-	12 (6)	4 (3)	6 (6)	1 (1)	22 (21)	-	26 (23)	-	-	-
H ¹⁷ OH	HDS	-	-	12 (11)	4 (4)	5 (5)	1 (1)	14 (13)	-	21 (21)	-	-	-
HOD	D ₂ S	-	-	10 (9)	4 (4)	5 (5)	1 (1)	11 (10)	-	16 (15)	-	-	-
D ₂ O	H ³⁴ S	-	-	3 (3)	2 (2)	-	-	6 (6)	-	8 (8)	-	-	-
H ¹⁸ OD	H ³³ S	-	-	1 (1)	-	1 (1)	-	5 (5)	-	5 (5)	-	-	-
H ¹⁷ OD	HD ³⁴	-	-	6 (4)	2 (2)	1 (1)	-	8 (7)	-	8 (7)	-	-	-
D ₂ ¹⁷ O	Итого	-	-	1 (0)	-	-	-	-	-	1 (0)	-	-	-
D ₂ ¹⁸ O	Триады	-	-	1 (1)	-	-	-	-	-	3 (3)	-	-	-
HTO	A-Bo:	-	-	4 (3)	-	-	-	3 (3)	-	7 (6)	-	-	-
Итого	A-Bo:	-	-	2 (1)	-	-	-	-	-	3 (3)	-	-	-
		-	-	-	-	1 (1)	-	1 (1)	-	1 (1)	-	-	-
		-	-	-	-	-	-	-	-	1 (1)	-	-	-
		-	-	75 (56)	22 (20)	35 (33)	4 (3)	133 (125)	-	161 (148)	-	-	-

Триады	T1	T7	T2	T6	T3	T5	Итого
A-Box	-	-	4064	8463	2461	10658	25646
A-Box (корр.)	-	-	425758	290139	290139	1015036	1015036



Примеры. Декомпозиция. Ограничения на существование

Сравнение массивов спектральных данных Hitran и GEISA в диапазоне 0-50 см⁻¹

	0-10 см ⁻¹		10-30 см ⁻¹		30-50 см ⁻¹	
	HITRAN	GEISA	HITRAN	GEISA	HITRAN	GEISA
H₂¹⁷O		<u>0/1</u> 0%		<u>0/10</u> 0%	9/25 36%	9/29 36%
		<u>0/1</u> 0%		<u>0/10</u> 0%	<u>0/25</u> 0%	<u>0/29</u> 0%
H₂¹⁸O	28/31 90%	<u>0/1</u> 0%	<u>0/15</u> 0%	<u>0/15</u> 0%	17/33 52%	17/33 52%
	8/31 26%	<u>0/1</u> 0%	<u>0/15</u> 0%	<u>0/15</u> 0%	<u>0/33</u> 0%	<u>0/33</u> 0%
HDO	20/65 31%	20/66 30%	39/76 51%	20/76 26%	60/76 79%	60/76 79%
	<u>0/65</u> 0%	<u>0/66</u> 0%	<u>0/76</u> 0%	<u>0/76</u> 0%	<u>2/76</u> 3%	<u>2/76</u> 3%
HD¹⁸O	26/53 49%		52/61 85%		28/45 62%	
	8/53 15%		3/61 5%		0/45 0%	

Точность, с которой проводилось сравнение данных измерений задана шрифтом.

Подчеркнутые цифры означают, что сравнение проведено с точностью 0.01 см⁻¹, цифры курсивом - с точностью *0.1* см⁻¹, и выделенные жирным шрифтом – с точностью **1** см⁻¹.



Декомпозиция источников данных

Сравнение массивов спектральных данных Hitran и GEISA в диапазоне 50-20000 см⁻¹

	50-200 см ⁻¹		200-10000 см ⁻¹		10000-20000 см ⁻¹	
	HITRAN	GEISA	HITRAN	GEISA	HITRAN	GEISA
H₂¹⁷O	147/279 53%	147/279 53%	1309/4762 28%	1383/4828 29%	118/1916 6%	9/325 3%
	<i>0/279</i> 0%	<i>0/279</i> 0%	669/4762 14%	<i>119/4828</i> 2%	639/1916 33%	107/325 33%
H₂¹⁸O	205/342 60%	205/342 60%	1529/6889 22%	1380/7109 20%	164/2443 7%	23/863 3%
	2/342 1%	<i>0/342</i> 1%	302/6889 4%	317/7109 4%	208/2443 9%	49/863 6%
HDO	370/599 62%	370/599 62%	1674/8825 19%	2980/11045 27%	8/3483 0%	8/119 7%
	2/599 0%	0/599 0%	5/8825 0%	27/11045 0%	94/3483 3%	0/119 0%
HD¹⁷O			7/175 4%	7/175 4%		
			0/175 0%	0/175 0%		
HD¹⁸O	378/544 69%		434/908 48%	195/679 29%		
	16/544 3%		221/908 24%	221/679 33%		

Точность, с которой проводилось сравнение данных измерений задана шрифтом.

Подчеркнутые цифры означают, что сравнение проведено с точностью 0.01 см⁻¹, цифры курсивом - с точностью *0.1* см⁻¹, и выделенные жирным шрифтом – с точностью **1** см⁻¹.



Вопросы

