

УДК 004.891.3

doi:10.18720/SPBPU/2/id23-84

*Лещева Ирина Анатольевна*¹,
старший преподаватель;
*Благосклонов Николай Алексеевич*²,
младший научный сотрудник

ПРИМЕНЕНИЕ ОНТОЛОГИЧЕСКОГО ПОДХОДА К ДИАГНОСТИКЕ ОРФАННЫХ ЗАБОЛЕВАНИЙ

¹ Россия, Санкт-Петербург, Институт «Высшая школа менеджмента»,
Санкт-Петербургский государственный университет, leshcheva@gsom.spbu.ru;

² Россия, Москва, Институт проблем искусственного интеллекта,
Федеральный исследовательский центр «Информатика и управление»
Российской академии наук, nblagosklonov@gmail.com

Аннотация. В работе описаны принципы создания базы знаний онтологического типа для диагностики орфанных заболеваний на долабораторном этапе. Предлагаемый подход основывается на модальности и факторах уверенности для сроков манифестации и выраженности признаков для каждой клинической формы в определенные возрастные периоды. Для реализации предложенного подхода был создан прототип онтологической базы знаний и подробно описано ее наполнение с использованием разработанного ранее метода МЕТЕОР.

Ключевые слова: экспертная диагностическая система, база знаний онтологического типа, наполнение онтологий.

*Irina A. Leshcheva*¹,
Senior lecturer;
*Nikolay A. Blagosklonov*²,
Junior Researcher

APPLICATION OF THE ONTOLOGICAL APPROACH TO THE DIAGNOSIS OF ORPHAN DISEASES

¹ Saint Petersburg State University, St. Petersburg, Russia,
leshcheva@gsom.spbu.ru;

² Russian Artificial Intelligence Research Institute, Federal Research Center
“Computer Science and Control” of the Russian Academy of Sciences,
Moscow, Russia, nblagosklonov@gmail.com

Abstract. The paper describes the principles of creating an ontological-type knowledge base for diagnosing orphan diseases at the pre-laboratory stage. The proposed approach is based on the modality and confidence factors for the timing of manifestation and the severity of signs for each clinical form in certain age periods. To implement the proposed approach, a prototype of the ontological knowledge base was created, and its population was described in detail using the previously developed METEOR method.

Keywords: expert diagnostic system, ontological knowledge base, ontology population.

Введение

Орфанные или редкие заболевания генетической природы — это патология, встречающаяся с низкой частотой в популяции, отличающаяся прогрессирующим течением и без лечения приводящая к смерти или инвалидизации больного, тогда как своевременно начатое лечение способно предупредить или затормозить дальнейшее прогрессирование патологии.

Диагностика орфанных заболеваний вызывает у врачей серьезные трудности ввиду необходимости учета разнообразных проявлений заболеваний на разных этапах ее развития и ввиду высокой стоимости биохимических и молекулярно-генетических исследований. Все это определяет актуальность разработки системы поддержки принятия решений для диагностики редких наследственных болезней в раннем возрасте.

Онтологический подход широко используется в различных медицинских интеллектуальных системах и приложениях [1, 2, 6]. Для диагностики орфанных (редких) заболеваний предложено использовать медицинские онтологии [3, 4]. Анализ предметной области и работа с экспертами позволили предложить подход, в котором диагностирование осуществляется на основе шкал модальности признаков с их коэффициентами и формул для комплексной оценки совокупности факторов уверенности (выраженности и манифестации). Для реализации этого подхода была создана база знаний онтологического типа *ontomorf*. Одним из трудоемких шагов создания онтологической базы знаний является ее наполнение. Для наполнения созданной онтологии использовался метод METEOR, описанный в [7].

1. Анализ предметной области

Источником данных для проведенного анализа служили интервью с экспертами и научные статьи, в которых описаны различные клинические случаи. На первом этапе работы над диагностической базой знаний были рассмотрены 15 заболеваний. По результатам анализа были выделены 22 признака (симптома) и 4 возрастных группы пациентов:

- (1) до 1 года;
- (2) 1–3 года;
- (3) 4–6 лет;
- (4) старше 6 лет.

Далее для каждой возрастной группы были определены коэффициенты модальности, манифестации и выраженности каждого признака.

Модальность (M_{ik} , где i — номер признака, а k — номер возрастного периода) характеризует релевантность признаков заболевания в каждом возрастном периоде. В соответствии с мнением экспертов для каждого заболевания признаки были разделены на главные, необходимые и второстепенные, и для каждой группы признаков были определены коэффициенты. При невозможности наличия признака в данном возрастном диапазоне модальность считалась нулевой ($M_{ik} = 0$).

Манифестация (n_{ij} , где i — номер признака, а j — номер возрастного периода) характеризует меру доверия (уверенность) экспертов в том, что данный признак определённого заболевания манифестирует (обнаруживается) именно в данном возрасте (возрастной группе). Т. е. для каждого возрастного периода экспертами было определено свое значение манифестации, причем сумма этих значений для одного заболевания по всем возрастным группам не превышает 1. Суммарный фактор уверенности в манифестации m_{ik} рассчитывался по формуле:

$$m_{ik} = \sum_{j=1}^k n_{ij}, \quad (1)$$

где i — номер признака, а k — номер возрастного периода.

Выраженность (s_{ik} , где i — номер признака, а k — номер возрастного периода) характеризует уверенность экспертов в том, что данный признак встречается в конкретной возрастной группе с определенной степенью выраженности. Изменение выраженности по возрастным периодам косвенно указывает на скорость развития заболевания (симптомов).

Для получения комплексной количественной оценки признака в определенном возрастном периоде использовалась следующая формула:

$$P_{ik} = M_{ik} * m_{ik} * s_{ik}, \quad (2)$$

где i — номер признака,

k — номер возрастного периода,

P_{ik} — количественная оценка признака заболевания (симптома) для указанного возрастного периода,

M_{ik} — модальность признака для этого возрастного периода,

m_{ik} — суммарный фактор уверенности манифестации для возрастного периода k ,

s_{ik} — фактор уверенности выраженности для этого же возрастного периода.

Для получения интегрированной количественной оценки суммы признаков для каждого диагностируемого случая использовалась формула:

$$I = \sum_{i=1}^n P_{ik}, \quad (3)$$

где n — количество признаков,

k — номер возрастного периода, в который попадает возраст данного пациента,

I — интегрированная оценка признаков,

P_{ik} — количественная оценка признака заболевания i для возрастного периода k .

2. Онтология ontomorf

Для реализации предложенного подхода была создана онтология ontomorf, основные элементы которой представлены на рисунке 1.

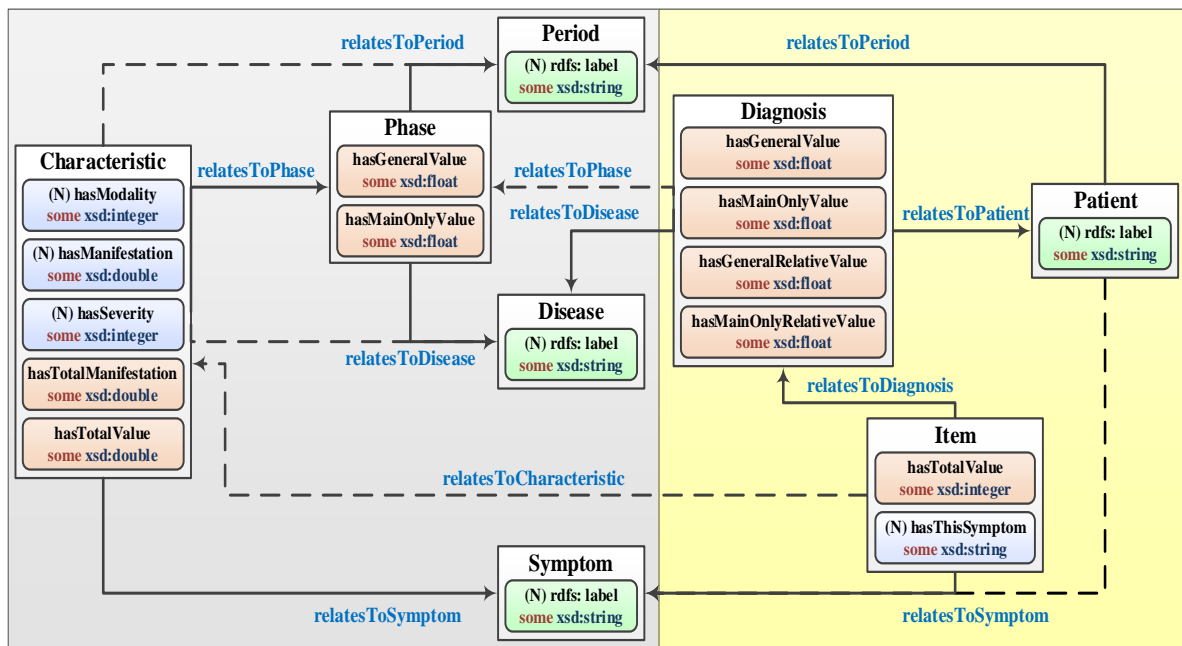


Рис. 1. Структура онтологии диагностики орфанных заболеваний ontomorf

В левой части — классы, в которых содержится «постоянная» информация базы знаний. Предполагается, что в процессе работы базы знаний эта информация не меняется или меняется очень редко:

– Класс **Disease** характеризует заболевания, для диагностики которых предназначена сформированная база знаний.

- Класс **Symptom** служит для описания признаков заболеваний диагностируемой группы.
- Класс **Period** определяет возрастные периоды.
- Класс **Characteristic**: каждый элемент класса соответствует одному признаку (симптому) определенного заболевания в определенный период и содержит значения модальности (M_{ik}), манифестации (n_{ik}) и выраженности (s_{ik}) этого признака для этого заболевания в этого период. Также элементы этого класса используются для расчетов суммарного фактора уверенности манифестации (m_{ik}), комплексной количественной оценки признака (P_{ik}) и суммарных значений элементов класса **Phase**.
- Класс **Phase**: содержит суммарные (референсные) значения для каждого заболевания в определенный период.

Правая часть содержит «переменную» информацию о пациентах:

- Класс **Patient** содержит идентификационные данные пациента, возраст, симптомы.
- Класс **Diagnosis**: в этом классе для каждого пациента создаются экземпляры по одному на каждое заболевание. Содержит суммарные значения по каждому заболеванию для определенного пациента. Эти значения сравниваются с референсными значениями из соответствующего экземпляра класса **Phase** для формирования гипотез о диагнозе.
- Класс **Item**: вспомогательный класс для вычисления значений экземпляров класса **Diagnosis**.

Элементы, подсвеченные розовым, и отношения, показанные пунктирными стрелками, определяются на основе другой имеющейся информации. Для этого заданы 24 правила вывода.

3. Наполнение онтологии *ontomorf*

Для наполнения базы знаний использовался метод МЕТЕОР [6]:

- на этапе создания «базовой» (постоянной) части базы знаний (однократно);
- на этапе эксплуатации для внесения информации о пациентах.

Алгоритм метода МЕТЕОР состоит из 5 шагов и далее будет проиллюстрирован на примере наполнения «базовой» части базы знаний.

Шаг 1: Идентификация источников данных. Данные для наполнения «базовой» части базы знаний содержались в двух Excel-файлах *symptoms.xlsx* и *modalities.xlsx*. Структура файла *symptoms.xlsx* показана на рисунке 2.

	Гурлер								Г-Шейе							
	до 1 года		1-3 года		4-6 лет		старше 6 лет		до 1 года		1-3 года		4-6 лет		старше 6 лет	
Задержка роста	0,7	3	0,1	5	0,1	7	0,1	10	0	0	0,7	7	0,1	7	0,1	8
Короткая шея	0,1	2	0,1	3	0,3	7	0,4	9	0,1	3	0,1	3	0,2	6	0,2	6
Макроцефалия	0,7	2	0,1	4	0,1	5	0,1	8	0,4	2	0,1	3	0,1	4	0,1	4

Рис. 2. Пример структуры файла с коэффициентами манифестации и выраженности

В первом столбце содержатся все признаки диагностируемых заболеваний, а далее в блоках для каждого заболевания и для каждой возрастной группы указаны значения коэффициентов факторов уверенности манифестации и выраженности.

Модальности заданы иерархическим списком следующего вида (см. рисунок 3).

Мукополисахаридоз 1Н (Гурлер)	
до 1 года	
	(https://rarediseases.info.nih.gov/diseases/12559/mucopolysaccharidosis-type-1h)
	Главные признаки (80%-99% больных имеют эти симптомы):
	Короткая шея
	Грубые черты лица
	Гепатомегалия
	Спленомегалия
	Грыжи
	Необходимые признаки (30%-79% больных имеют эти симптомы):
	Задержка роста
	Макроглоссия
	Снижение слуха
	Помутнение роговицы
	Второстепенные признаки (менее 30% больных имеют эти симптомы):
	Макроцефалия
	Скафоцефалия
	Гипертония

Рис. 3. Пример структуры файла с модальностями признаков

На первом шаге данные были сведены в единую нормализованную таблицу. При этом понадобилась дополнительная таблица, в которой текстовым названиям модальностей сопоставлялись числовые значения (Главные – 5, Необходимые – 4, Второстепенные – 2).

Полученная таблица содержит 6 столбцов:

- Disease: названия диагностируемых заболеваний.
- Period: принадлежность к возрастной группе.
- Symptom: названия признаков заболеваний.
- Manifestation: значения манифестации (n_{ik}).
- Severity: значения выраженности (s_{ik}).
- Modality: значения модальности (M_{ik}).

Шаг 2: Спецификация источников данных. На этом шаге была создана вспомогательная онтология `ontomorf_struct`, импортирующая онтологию источников данных `meteor`, схема которой представлена на рисунке 4. Далее был создан экземпляр класса **Workbook**, у которого заданы атрибуты `hasPath` (путь к файлу из шага 1) и `hasName` (имя файла из шага 1).

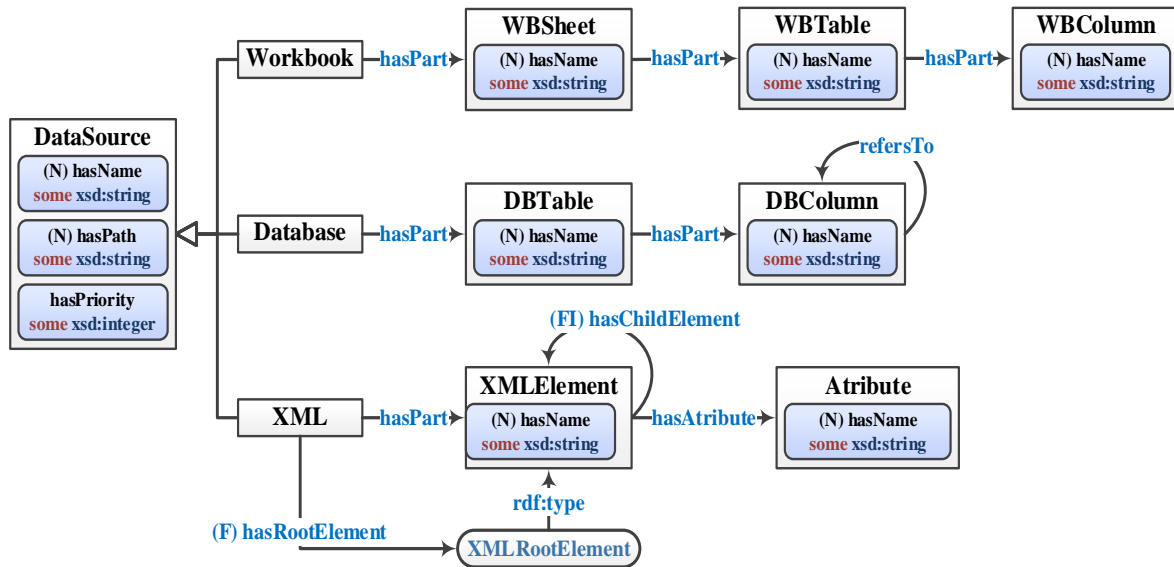


Рис. 4. Основные элементы онтологии источников данных `meteor`

Шаг 3: Извлечение структуры данных. С помощью модуля извлечения структуры данных онтология `ontomorf_struct` была наполнена экземплярами, описывающими структуру файла из шага 1.

Шаг 4: Задание правил отображения. На этом этапе онтология `ontomorf_struct` была импортирована в онтологию `ontomorf`, а затем заданы правила отображения:

- Для класса `Period`:
 - `getInstancesFrom Period`
- Для класса `Disease`:
 - `getInstancesFrom Disease`
- Для класса `Symptom`:
 - `getInstancesFrom Symptom`
- Для класса `Phase`:
 - `combineInstancesFrom Disease useObjectProperty relatesToDisease`
 - `combineInstancesFrom Period useObjectProperty relatesToPeriod`
- Для класса `Characteristic`:
 - `combineInstancesFrom Disease useObjectProperty relatesToDisease`

- combineInstancesFrom Period useObjectProperty relatesToPeriod
- combineInstancesFrom Symptom useObjectProperty relatesToSymptom
- getValuesFrom Manifestation
- getValuesFrom Modality
- getValuesFrom Severity

Для описания правил отображения используются аннотационные свойства, которые назначаются классам онтологии предметной области. Для создания экземпляров классов используются аннотационные свойства *getInstancesFrom* и *combineInstancesFrom*. Для описания связывания экземпляров отношением (с помощью объектного свойства) служит свойство *makeReferenceTo*. Ссылка на объектное свойство задается с помощью вложенного аннотационного свойства *useObjectProperty*. Для задания атрибутов экземпляров класса используется свойство *getValuesFrom*. Ссылка на свойство задается с помощью вложенного аннотационного свойства *useDataProperty*.

Шаг 5: Наполнение онтологии. Выполняется автоматически с помощью модуля наполнения в несколько итераций. На первой итерации создаются экземпляры классов **Period**, **Disease** и **Symptom**, которые аннотированы с помощью свойства *getInstancesFrom*. На второй итерации создаются экземпляры классов **Phase** и **Characteristic**, аннотированные свойством *combineInstancesFrom*, и отношения между созданными на этой итерации экземплярами и экземплярами классов, указанных в качестве значений свойств *combineInstancesFrom*. Объектное свойство задается с помощью аннотационного свойства *useObjectProperty*. На третьей итерации заполняются значения свойств **Manifestation**, **Modality** и **Severity**, описанные аннотационным свойством *getValuesFrom*.

После наполнения онтологии была запущена машина логического вывода, с помощью которой построены недостающие отношения и вычислены необходимые для процесса диагностирования значения. Все полученные результаты вывода были экспортированы в онтологию, так как процесс вывода занял значительное время, а, в данном случае, наполнялась «базовая» часть онтологии, т. е. та часть, регулярное изменение которой не предполагается.

Далее тот же алгоритм был применен для загрузки в полученную базу знаний информации о 15 тестовых пациентах, а затем запущена машина логического вывода для постановки диагноза. Пример результата по одному пациенту приведен в таблице 1.

Пример вывода для одного пациента

Пациент	Синдром	a	b	c	d
Пример [5] Пол: женский Возраст: 3 месяца Признаки: • кардиопатия • тугоподвижность крупных суставов • гепатомегалия • спленомегалия • грубые черты лица • кифосколиоз Диагноз: Мукополисахаридоз I типа (синдром Гурлер)	Гурлер	38,0	40%	38,0	46%
	Гурлер-Шейе	2,5	13%	2,5	28%
	М-Лами быстро прогр	0,0	0%	0,0	0%
	М-Лами медл прогр	0,0	0%	0,0	0%
	Морк-А (быстро-прогр)	0,0	0%	0,0	0%
	Морк-А (медленно-прогр)	0,0	0%	0,0	0%
	Морк-В	0,0	0%	0,0	0%
	Слая	3,0	32%	3,0	86%
	С-Ф D	0,0	0%	0,0	0%
	С-Ф А	0,0	0%	0,0	0%
	С-Ф В	0,0	0%	0,0	0%
	С-Ф С	0,0	0%	0,0	0%
	Хантер-Г	1,8	13%	0,0	0%
	Хантер-Л	0,0	0%	0,0	0%
Шейе	0,0	0%	0,0	0%	

Заключение

В работе был описан созданный прототип базы знаний, которая может стать ядром системы поддержки принятия решений для диагностики орфанных заболеваний. В предложенной базе знаний учитывается возрастная динамика прогрессирующих наследственных болезней в сочетании с количественной экспертной оценкой их проявления. Использование онтологической системы позволяет выдвинуть ряд гипотез, чтобы в дальнейшем провести направленное лабораторное обследование для уточнения окончательного диагноза.

Также необходимо отметить, что метод наполнения онтологий МЕТЕОР продемонстрировал высокую эффективность не только на этапе наполнения «постоянной» части базы знаний, но и, в особенности, при многократной загрузке информации о пациентах.

Для оценки результатов диагностики были взяты реальные истории болезни пациентов из отечественных и зарубежных открытых источников с верифицированными диагнозами. Диагностика с помощью сформированной базы знаний продемонстрировала приемлемые результаты для долабораторного этапа идентификации заболевания.

Список литературы

1. Грибова В.В., Петряева М.В., Окунь Д.Б., Шалфеева Е.А. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. – 2018. – Т. 8. – № 1 (27). – С. 58–73.

2. Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Онтология и модель онтологии предметной области «Медицинская диагностика». – Владивосток: Изд-во ИАПУ ДВО РАН, 2005.

3. Кобринский Б.А. Триединство факторов уверенности в задачах медицинской диагностики // Искусственный интеллект и принятие решений. – 2018. – № 2. – С. 62-72.

4. Кобринский Б.А., Благодосклон Н.А., Демикова Н.С., Грибова В.В., Шалфеева Е.А., Петряева М.В. Возможности применения онтологического подхода к диагностике орфанных заболеваний // Семнадцатая Национальная конференция по искусственному интеллекту с международным участием. КИИ-2019. Сборник научных трудов. В 2 т. – Ульяновск: УлГТУ, 2019. – Т.2. – С. 227.

5. Намазова-Баранова Л.С., Вашакмадзе Н.Д., Бабайкина М.А., Басаргина Е.Н., Журкова Н.В., Геворкян А.К., Кузенкова Л.М., Подклетнова Т.В., Жердев К.В., Челпаченко О.Б., Дегтярёва Т.Д. Эффективность современных методов лечения пациентов с мукополисахаридозом I типа // Педиатрическая фармакология. – 2014. – Т. 11. – № 6. – С. 76–79.

6. Ivanović M., Budimac Z. An overview of ontologies and data resources in medical domains // Expert Systems with Applications. – 2014. – Vol. 41. – No 11. – Pp. 5158–5166.

7. Leshcheva I., Begler A. A method of semi-automated ontology population from multiple semi-structured data sources // Journal of Information Science. – 2022. – Vol. 48. – No 2. – Pp. 223–236.