

Ч. А. Найданов¹, Д. Е. Пальчунов^{1,2}, П. А. Сазонова¹

¹Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия

²Институт математики им. С. Л. Соболева СО РАН
пр. Акад. Коптюга, 4, Новосибирск, 630090, Россия

naydanov.fit@ya.ru, palch@math.nsc.ru, psazonova@gmail.com

ТЕОРЕТИКО-МОДЕЛЬНЫЕ МЕТОДЫ ИНТЕГРАЦИИ ЗНАНИЙ, ИЗВЛЕЧЕННЫХ ИЗ МЕДИЦИНСКИХ ДОКУМЕНТОВ*

Статья посвящена методам представления знаний, интеграции знаний, извлеченных из разных источников, и порождения новых знаний. Подход основан на четырехуровневой модели представления знаний. Интеграция знаний, извлеченных из медицинских документов, и порождение новых оценочных знаний производится при помощи онтологической модели предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника». Онтологическая модель состоит из онтологии, общих теоретических знаний, эмпирических знаний (прецедентов) и оценочных вероятностных знаний. Разработана программная система, предназначенная для предотвращения рисков возникновения критических состояний пациентов.

Ключевые слова: онтология, онтологическая модель, представление знаний, порождение знаний, теоретико-модельные методы, атомарная диаграмма, нечеткая модель, прецедентная модель, деформации позвоночника, дегенеративные заболевания позвоночника, критические состояния.

Введение

Статья посвящена проблемам представления знаний, разработке методов интеграции знаний, извлеченных из разных источников, и порождения новых знаний. Целью исследования является построение онтологической модели предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника» на основе интеграции знаний, представленных в различных медицинских документах: Международной классификации болезней, справочниках по лекарствам, нормативных документах Минздрава, монографиях и статьях по данной области медицины, историях болезней пациентов.

Для интеграции знаний, извлеченных из различных текстов естественного языка, предложена четырехуровневая модель представления знаний. Рассматриваются уровень онтологических знаний (онтология предметной области), уровень общих теоретических знаний, уровень частных эмпирических знаний (прецедентов предметной области) и уровень оценочных, вероятностных знаний. Методы интеграции знаний основаны на теоретико-модельном подходе к разработке онтологий и онтологических моделей предметных областей [1–6].

На основе четырехуровневой модели представления знаний разработана онтологическая модель предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника». В качестве эмпирических знаний, прецедентов предметной области, рассматриваются истории болезней пациентов. За счет обработки историй болезней пациентов порожд-

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 14-07-00903_a.

даются оценочные знания, предназначенные, в частности, для выдачи рекомендаций врачам с целью предотвращения рисков возникновения у пациентов критических состояний.

Теоретико-модельный подход к разработке онтологий и онтологических моделей

В данном разделе мы рассмотрим формальный, теоретико-модельный подход к определению понятия онтологической модели. Этот подход является развитием теоретико-модельного подхода к разработке онтологий предметных областей [1–6].

Теоретико-модельный подход к формализации онтологии предметной области основан на разделении знаний о предметной области на три типа: логические, аналитические и синтетические [1–3]. Это разделение восходит к классификации истинности суждений, предложенной Р. Карнапом [7]. Логически истинные предложения истинны всегда, независимо от смысла входящих в них понятий. Истинность аналитических предложений зависит только от смысла входящих в них понятий и не зависит от состояния, законов и свойств реального мира. Истинность синтетических предложений определяется состоянием и свойствами реального мира.

Мы рассматриваем классификацию Карнапа одновременно с логической и теоретико-модельной точек зрения. С логической точки зрения мы говорим о предложениях и теориях, аксиоматизируемых этими предложениями. С теоретико-модельной точки зрения мы имеем дело с моделями, то есть с алгебраическими системами, на которых истинны эти теории. Для удобства здесь мы будем считать, что логические предложения являются аналитическими. Множество аналитических предложений вместе с множеством синтетических предложений образует теорию предметной области. Таким образом, множество логических предложений является подмножеством множества аналитических предложений, а множество аналитических предложений в свою очередь является подмножеством теории предметной области.

Это соответствует обратному включению классов моделей: класс моделей теории предметной области (множества аналитических и синтетических предложений) содержится в классе моделей аналитической теории предметной области (онтологии предметной области). Класс моделей аналитической теории предметной области содержится в классе моделей логической теории предметной области, то есть в классе моделей данной сигнатуры.

Определение [2]. Формальной онтологией предметной области SD назовем пару $O = \langle A, \sigma \rangle$, где σ – множество ключевых понятий предметной области, и A – множество аналитических предложений, описывающих смысл этих ключевых понятий. Множество T предложений, которые являются верными в каждом примере данной предметной области, будем называть теорией предметной области SD .

Заметим, что на практике нам никогда не известно все множество аналитических предложений, вследствие чего приходится иметь дело только с конечной частью онтологического знания о ключевых понятиях данной предметной области. Поэтому в реальных приложениях в качестве онтологии предметной области рассматривается пара $O_0 = \langle A_0, \sigma \rangle$, где $A_0 \subseteq A$, A_0 – конечное подмножество множества всех аналитических предложений. Таким образом, A_0 дает определенную спецификацию смысла ключевых понятий данной предметной области.

Онтологическую модель, как и онтологию, мы рассматриваем с точки зрения теоретико-модельного подхода к формализации естественного языка [2–6; 8]. При этом мы объединяем синтаксический и семантический подходы [5]: в рамках семантического подхода мы рассматриваем модели и классы моделей, в рамках синтаксического подхода – множества формул и теории. Кроме того, для формализации неточного и оценочного знания мы используем нечеткие модели.

Для интеграции синтаксического и семантического подходов в качестве базового объекта мы рассматриваем атомарные диаграммы моделей (более точно, фрагменты атомарных диаграмм моделей) [5]. С одной стороны, фрагмент атомарной диаграммы модели является

множеством предложений, то есть синтаксическим объектом. С другой стороны, атомарная диаграмма модели определяет эту модель с точностью до изоморфизма. Поэтому, работая с атомарными диаграммами моделей, мы фактически работаем с классами моделей, то есть полностью определяем семантику.

Для целей данной работы нам достаточно ограничиться случаем, когда сигнатура σ рассматриваемых моделей состоит только из символов предикатов и констант.

Модель $\mathfrak{A} = \langle A; \sigma \rangle$ сигнатуры σ нам удобно рассматривать в расширенной сигнатуре $\sigma_A = \sigma \cup \{c_a \mid a \in A\}$, обогащенной новыми константами – именами всех элементов модели \mathfrak{A} . Это означает, что $c_a \notin \sigma$ при $a \in A$. Через \mathfrak{A}_A мы обозначаем обогащение модели \mathfrak{A} до сигнатуры σ_A : обеднение модели \mathfrak{A}_A до сигнатуры σ совпадает с \mathfrak{A} . Кроме того, выполняется $c_a^{\mathfrak{A}_A} = a$ при $a \in A$.

$S(\sigma_A)$ обозначает множество предложений сигнатуры σ_A , т. е. формул без свободных переменных этой сигнатуры.

Предложение φ сигнатуры σ_A называется атомарным, если φ имеет один из следующих видов: $\varphi = (c_1 = c_2)$, $\varphi = \neg(c_1 = c_2)$, $\varphi = P(c_1, \dots, c_n)$ или $\varphi = \neg P(c_1, \dots, c_n)$, где $P, c_1, \dots, c_n \in \sigma_A$.

Атомарной диаграммой модели \mathfrak{A} называется множество предложений

$$AD(\mathfrak{A}) = \{ \varphi \in S(\sigma_A) \mid \mathfrak{A}_A \models \varphi \text{ и предложение } \varphi \text{ – атомарное} \}.$$

Конечное подмножество атомарной диаграммы $AD(\mathfrak{A})$ модели \mathfrak{A} мы будем называть *конечным фрагментом атомарной диаграммы* модели \mathfrak{A} .

Использование в качестве базового объекта конечных фрагментов атомарных диаграмм моделей позволяет нам решить две проблемы. Во-первых, таким образом мы избавляемся от бесконечности. Действительно, рассматриваемые нами алгебраические системы, как правило, бесконечны; любая теория также бесконечна, так как множество тождественно истинных предложений уже бесконечно. А фрагменты атомарных диаграмм, с которыми мы имеем дело, как правило, конечны. Таким образом, мы конечными множествами предложений аппроксимируем бесконечные модели и теории.

Во-вторых, использование фрагментов атомарных диаграмм моделей позволяет нам решить еще более важную и сложную проблему – необходимость работать с моделями разной сигнатуры. В классической теории моделей при рассмотрении аксиоматизируемых классов всегда имеют дело с классами моделей одной и той же сигнатуры. С другой стороны, в практических приложениях возникает необходимость работать с классами моделей разной сигнатуры (см., например, [9]; там это называется моделями разных типов). Более того, часто приходится иметь дело с ситуацией, когда сигнатура рассматриваемой модели заранее не известна. Такая ситуация возникает, например, когда мы извлекаем знания из текстов естественного языка: сигнатура алгебраической системы, которую мы строим, может быть в любой момент времени расширена. В частности, если мы интегрируем знания, извлеченные из разных документов, мы должны объединять сигнатуры, то есть множества понятий, представленных в этих документах. Кроме того, при возникновении омонимии – когда одно и то же понятие в разных текстах имеет разный смысл, мы должны дублировать сигнатурные символы, соответствующие этим понятиям.

Эта проблема легко решается при использовании атомарных диаграмм моделей: мы просто объединяем несколько фрагментов в один. При этом необходимо только следить, чтобы полученный фрагмент был непротиворечив, в частности, отслеживать омонимию, согласовывать онтологии, которые относятся к объединяемым фрагментам. Здесь следует отметить, что фрагменты атомарных диаграмм моделей являются множествами бескванторных предложений, поэтому проблема непротиворечивости в данном случае будет алгоритмически разрешимой.

Онтологическая модель состоит из онтологии и описания предметной области. В определенном смысле можно было бы считать эти два компонента аналитической теорией и теорией данной предметной области. Однако, как было указано выше, проблема состоит в том, что

в полном, законченном виде мы не имеем ни аналитическую теорию, ни теорию предметной области. У нас нет не только полного формального представления всех аналитических предложений, но отсутствует даже сколько-нибудь полная спецификация на естественном языке определений ключевых понятий предметной области. Поэтому, вместо аналитической теории предметной области мы используем онтологию предметной области как некоторую спецификацию смысла ключевых понятий предметной области. Отметим, что проблема явного извлечения из текстов естественного языка онтологической информации, в том числе явных и неявных определений понятий, является крайне сложной и важной [4].

Точно так же, мы не имеем полной теории предметной области. Вместо этого у нас есть определенная информация о ней – определенный набор предложений, которые мы считаем истинными в предметной области. Кроме этой информации мы имеем данные о конкретных прецедентах предметной области. Исходя из знаний о прецедентах мы, во-первых, можем утверждать ложность утверждений, которые являются ложными хотя бы на одном прецеденте. То есть, мы можем гарантировать, что некоторые предложения (опровергнутые на одном из прецедентов) не входят в теорию предметной области. Во-вторых, знания о прецедентах мы можем экстраполировать – формулировать гипотезы, которые могут оказаться истинными на всей предметной области. И, наконец, на основе анализа информации о прецедентах мы можем формулировать оценочные, вероятностные утверждения, которые могут иметь достаточно большую практическую ценность.

Для интеграции знаний, извлеченных из разных источников, разных медицинских документов, мы используем четырехуровневую модель представления знаний, реализованную в виде онтологической модели предметной области. Онтологическая модель состоит из онтологии и еще трех уровней представления знаний о предметной области: общих (теоретических) знаний, эмпирических знаний и оценочных (вероятностных) и знаний (рис. 1).



Рис. 1. Четырехуровневое представление знаний в онтологической модели

1. Онтология предметной области содержит описание, спецификацию ключевых понятий, на языке которых описывается данная предметная область. В случае необходимости, в частности, при возможности омонимии (один и тот же термин в разных контекстах имеет разный смысл), целесообразно рассматривать не одну онтологию, а иерархию онтологий. Каждая

из онтологий в данном случае будет иметь свой набор ключевых понятий (т. е. свою сигнатуру) и множество определений, задающих смысл ключевых понятий.

Собственно описание предметной области содержит еще три уровня представления знаний:

2. Универсальные, общие утверждения – законы и постулаты предметной области. Это знания о предметной области: общих принципах, законах и закономерностях и пр., которые на данный момент времени считаются полностью достоверными. Универсальные утверждения являются известной нам частью теории предметной области. Эти знания являются синтетическими в отличие от аналитических знаний, представленных в онтологии. Синтетические знания, в отличие от аналитических, не вытекают из смысла употребляемых терминов, их истинность зависит от реального мира. Поэтому эти знания могут изменяться (т. е. истинные утверждения становятся ложными, а ложные утверждения могут становиться истинными) даже в том случае, когда смысл употребляемых терминов остался полностью неизменным. Эти утверждения носят общий, универсальный характер. В отличие от эмпирических данных они описывают не отдельную ситуацию, а предметную область в целом, являются истинными для всех ситуаций. Универсальные, общие утверждения о предметной области формально записываются универсальными предложениями (\forall -предложениями) логики предикатов первого порядка. Для того, чтобы произвольные утверждения представить в виде \forall -предложений мы совершаем операцию, аналогичную скульемизации (от *Skolemization*; в некоторых источниках пишут *сколемизация*): добавляем в сигнатуру дополнительные символы – новые константы и отношения и таким образом избавляемся от кванторов существования \exists .

3. Эмпирические данные содержат описание конкретных ситуаций, прецедентов предметной области. Формальное описание каждого прецедента представляется в виде фрагмента атомарной диаграммы алгебраической системы. При этом, кроме сигнатурных символов, соответствующих реальным понятиям предметной области (смысл которых специфицирован онтологией), вводятся дополнительные, служебные сигнатурные символы [5]. Эти символы обозначают неизвестные объекты, действия и отношения, то есть это специальные символы констант и предикатов.

4. Вероятностные и оценочные знания. Эти знания либо берутся из внешних источников, либо порождаются на основе онтологической модели. Знания о прецедентах представляются в виде прецедентной модели, которая является булевозначной моделью [10, 11]. В результате фазификации этой булевозначной модели [11, 12] получается нечеткая модель, в которой значениями истинности формул являются числа из интервала $[0; 1]$. Таким образом, на основе анализа имеющихся в онтологической модели эмпирических данных и сопоставления их с универсальными знаниями и онтологическими (аналитическими) знаниями, также представленными в онтологической модели, порождаются вероятностные и оценочные знания и закономерности. Ниже, в параграфе 5 описаны три алгоритма порождения оценочных знаний на основе анализа эмпирической информации, представленной в онтологической модели. Формально вероятностные и оценочные знания представляются в виде множества предложений $\Gamma \subseteq S(\sigma_A)$ и отображения $\mu: \Gamma \rightarrow [0, 1]$, ставящего в соответствие каждому предложению $\varphi \in \Gamma$ его нечеткое значение истинности $\mu(\varphi) \in [0, 1]$ (см. [11; 12]).

Онтологический подход к моделированию предметных областей в медицине

Группой специалистов под руководством А. С. Клещева был разработан подход к моделированию предметных областей в медицине [13–15], основанный на применении онтологий. В центре внимания находится построение структуры онтологических моделей.

Модель предметной области должна отвечать следующим требованиям [13]: термины баз знаний должны быть понятны специалистам предметной области; знания должны быть полезны в течение всего времени эксплуатации системы; должна существовать возможность пополнять базы знаний, а также должна существовать возможность автоматического накопления результатов верификации принятых решений.

Наибольшая работа по автоматизации предметных областей была проделана А. С. Клещевым и его группой исследователей в области медицины. Была разработана онтологическая модель предметной области медицинской диагностики, которая состоит из следующих частей:

- 1) термины, описывающие действительность предметной области;
- 2) связи между терминами;
- 3) функциональная часть, содержащая множество функций, которые на основании введенных данных выдают конкретную информацию из онтологии.

Действительность в медицинской диагностике рассматривается как множество ситуаций, каждая из которых соответствует диагностическому случаю (пациенту) [14; 15]. А термины действительности – это термины, используемые для описания таких ситуаций; ограничения действительности – ограничения на значения терминов действительности.

Процессы, происходящие в организме пациента, условно подразделяются на внешние и внутренние. Первые представлены наблюдаемыми признаками, последние протекающими заболеваниями.

Значения наблюдаемых признаков в некоторый момент времени определяются заболеваниями, которыми болен пациент, его анатомо-физиологическими особенностями и произошедшими с ним событиями. Эти зависимости описываются в виде причинно-следственных связей. Причинно-следственная связь «клиническое проявление» отражает зависимость между заболеванием, анатомо-физиологическими особенностями, произошедшими событиями с одной стороны, и наблюдаемым признаком с другой. Причинно-следственные связи «нормальная реакция» и «реакция на воздействие событий» отражают зависимость между анатомо-физиологическими особенностями, произошедшими событиями, с одной стороны, и наблюдаемым признаком у здорового человека, с другой. Наблюдаемый признак не может описываться одновременно несколькими причинно-следственными связями. Если у наблюдаемого признака несколько возможных причин, то среди них выбирается наиболее приоритетная. Приоритет причинно-следственных связей задается экспертом.

Работа системы происходит следующим образом: система медицинской диагностики строит множество гипотез о том, что у пациента присутствует некоторое заболевание. Для каждой гипотезы проверяется, что все значения наблюдаемых клинических признаков являются следствием этого заболевания. Решение представляет собой множество истинных гипотез и множество соответствующих причинно-следственных связей, описывающих все наблюдаемые признаки.

Данный подход был применен к конкретным направлениям медицины: была создана онтология для работы в области офтальмологии, структурированы общие методы медицинской диагностики, разработаны база знаний лекарственных средств и база знаний заболеваний [14–18].

Онтологическая модель предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника»

В данном параграфе описана структура онтологической модели, которая позволяет формализовать предметную область на четырех уровнях представления знаний, изложенных выше. Особую важность для решаемых задач имеют уровень эмпирических знаний (прецедентов) и уровень оценочных (вероятностных) знаний.

Описание онтологической модели

Разработанная онтологическая модель описывает предметную область «Деформации позвоночника и дегенеративные заболевания позвоночника». Она формализует ключевые понятия предметной области, отношения между ключевыми понятиями предметной области, описания явлений и процессов, происходящих в предметной области. При помощи булево-значных моделей формализуются прецеденты предметной области; при помощи нечетких моделей формализуются оценочные знания о предметной области.

Программная реализация онтологической модели состоит из следующих частей:

- 1) онтология;
- 2) база общих знаний;
- 3) база прецедентов;
- 4) база вероятностных знаний;
- 5) функциональная часть.

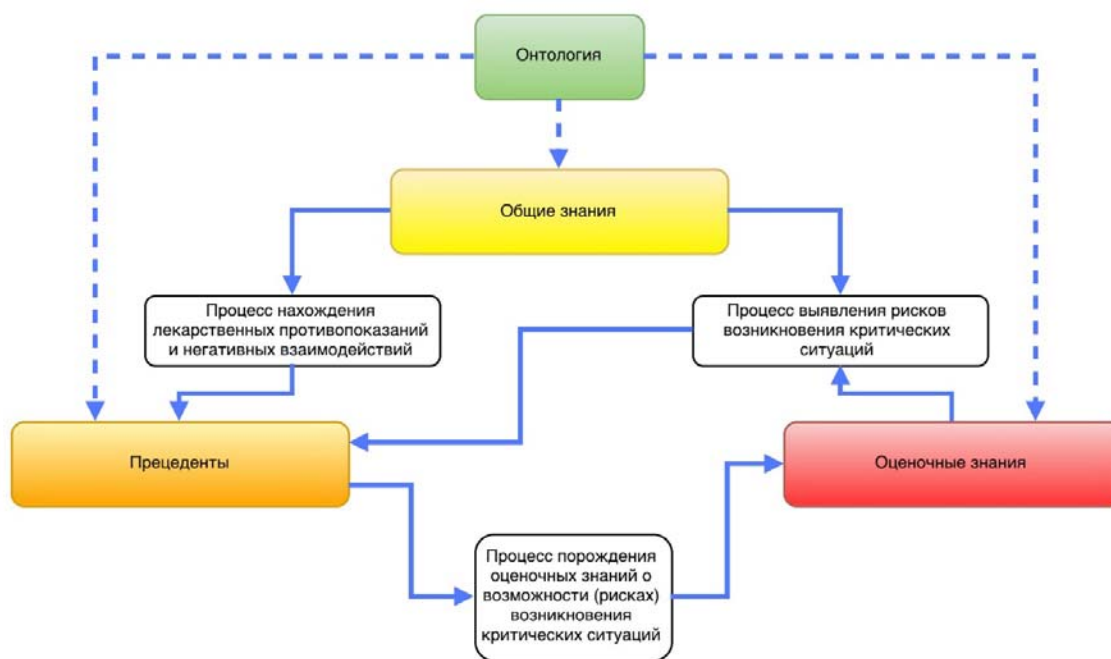


Рис. 2. Структура онтологической модели

Онтология

Разработка онтологий в настоящее время играет огромную роль в исследовании и моделировании предметных областей [19]. Онтология является основой разрабатываемой нами онтологической модели.

Онтология предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника» состоит из нескольких частей – «подонтологий». Перечислим наиболее важные из них.

Онтология наблюдений описывает виды анализов, обследований, их возможные значения. Онтология симптомов содержит термины разновидностей симптомов заболеваний.

Важное значение имеет онтология критических состояний и осложнений. С каждым из критических состояний, представленных в онтологии, связаны: а) условия на анатомо-физиологические особенности; б) список анализов, обследований и симптомов, входящих в клиническую картину соответствующего критического состояния. Эти условия представлены в онтологической модели.

В онтологии лекарственных препаратов описываются разновидности лекарств; каждому термину-лекарству в онтологической модели сопоставлены условия на анатомо-физиологические особенности и противопоказания.

База общих знаний

В базе общих знаний описаны клинические проявления заболеваний, критических состояний и осложнений. Клинические проявления представлены в виде причинно-следственных

связей, где причинами являются заболевания, критические состояния и осложнения, а следствиями – результаты анализов, обследований, наблюдаемые симптомы и прочее.

Также в базе общих знаний описаны негативные взаимодействия лекарственных препаратов. Каждое негативное взаимодействие представлено в виде причинно-следственной связи, где причинной является введение пациенту несколько различных действующих веществ, а следствием негативное взаимодействие.

База прецедентов

База прецедентов содержит истории болезней реальных пациентов. Для описания историй болезни используются формальные прецеденты. Формальный прецедент представляется в виде конечного фрагмента атомарной диаграммы алгебраической системы – множества логических предложений, описывающих факты из истории болезни пациента. Например, факт о том, что у пациента наблюдалось повышение температуры, будет записан в следующем виде: Наблюдаться(ТемператураТела, Повышение).

Все понятия, используемые в предложениях, представлены в онтологии.

База оценочных знаний

В базе оценочных знаний хранятся знания, полученные из множества прецедентов с помощью алгоритмов, реализованных в функциональной части. Оценочные знания состоят из множества гипотез о наличии или отсутствии критических состояний, осложнений или заболеваний у пациента.

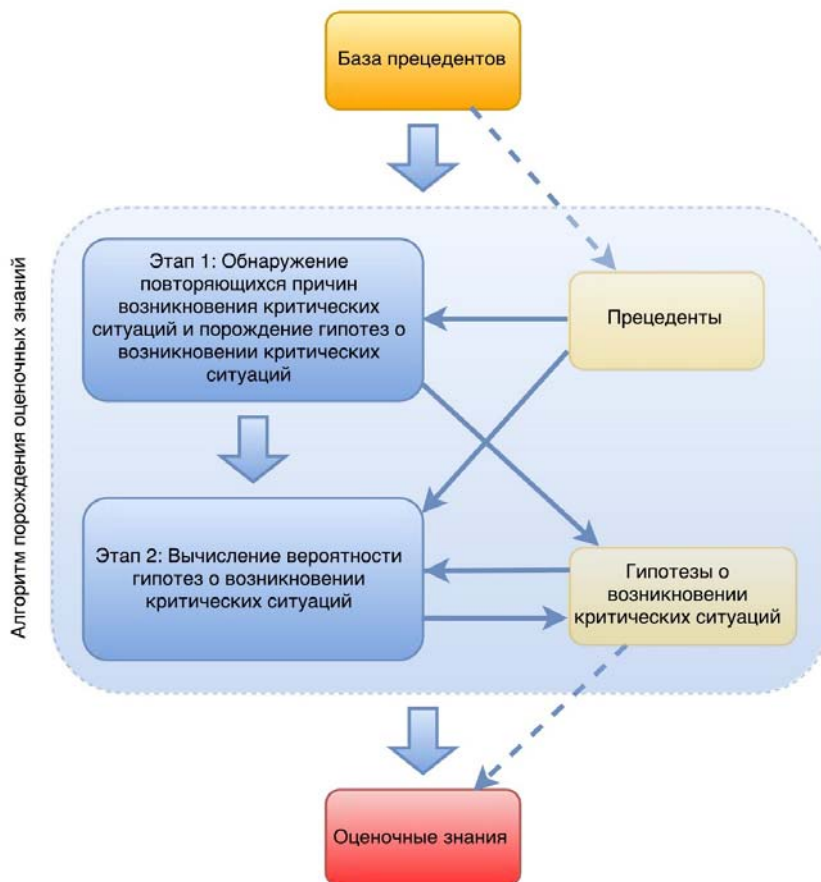


Рис. 3. Пополнение базы вероятностных знаний

Функциональная часть

Функциональная часть онтологической модели реализует алгоритмы порождения новых знаний. Работа алгоритмов основана на интеграции уже имеющихся знаний, представленных в разных уровнях онтологической модели. Функциональная часть содержит модули диагностики критических состояний, нахождения лекарственных противопоказаний и негативных взаимодействий, порождения оценочных знаний. Одной из основных задач функциональной части является своевременное определение и предотвращение возникновения критической ситуации.

В целом, работу функциональной части можно описать следующим образом. Для базы прецедентов, в которой хранятся записи из медицинских карт пациентов, применяется алгоритм порождения оценочных знаний о возможности (рисках) возникновения критического состояния. В результате получается набор гипотез о возникновении критических состояний, которые составляют базу оценочных знаний.

По запросу врача программная система для конкретного пациента выдает список лекарственных противопоказаний, список потенциально возможных критических ситуаций у пациента и список необходимых дополнительных обследований. Для этого используются знания из онтологии, из базы общих знаний онтологической модели, а также сведения из медицинской карты пациента. Результаты формируются при помощи алгоритма диагностики и алгоритма нахождения лекарственных противопоказаний и негативных взаимодействий.

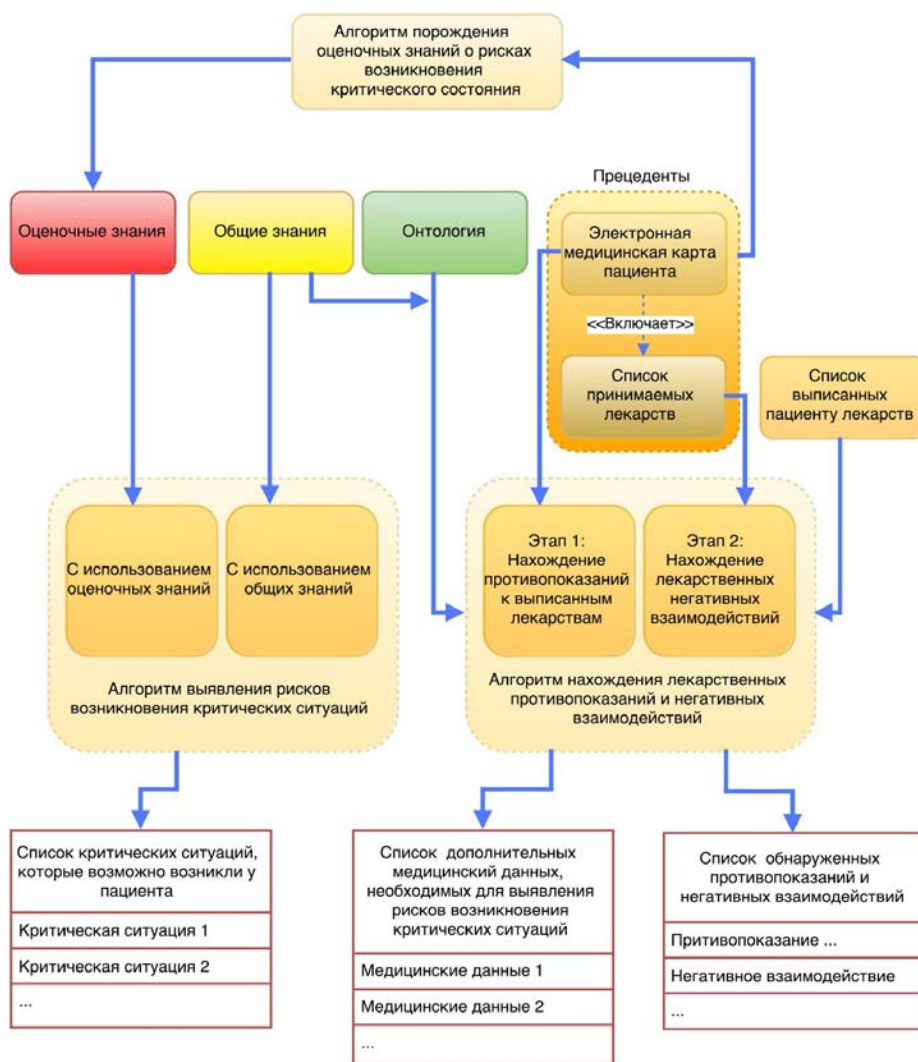


Рис. 4. Функциональная часть

1. *Алгоритм порождения оценочных знаний о возможности (рисках) возникновения критического состояния.* Алгоритм работает с базой прецедентов, где представлены истории болезней пациентов. В частности, в базе прецедентов имеется информация об уже возникших критических состояниях пациентов. В результате работы алгоритм пополняет базу оценочных знаний.

2. *Алгоритм процесса диагностики.* Данный алгоритм позволяет определить набор возможных критических ситуаций, которые могут возникнуть у конкретного пациента. В случае недостатка данных для определения такого набора, будут рекомендованы дополнительные обследования для пациента. Разработанный нами алгоритм основан на методах, представленных в [19].

На вход алгоритм принимает электронную медицинскую карту (ЭМК) конкретного пациента. ЭМК содержит информацию об анамнезе, обследованиях, результатах анализов.

Алгоритм работает в два этапа. На первом этапе предположения строятся на основе базы общих знаний онтологической модели. На втором этапе используются сведения из базы оценочных знаний. Для каждого критического состояния из базы извлекаются вероятностные гипотезы, наиболее подходящие для данного пациента. Далее вычисляется числовая оценка, которая интерпретируется как наличие либо отсутствие критической ситуации (более точно, как *степень* ее наличия или отсутствия).

3. *Алгоритм нахождения лекарственных противопоказаний и негативных взаимодействий.* Согласно статистике, осложнения от лекарственных средств являются одним из самых часто возникающих причин возникновения критических ситуаций, а впоследствии и смертности. Осложнения от лекарственных средств может вызвать введение пациенту противопоказанного ему препарата, а также совместное использование препаратов, в состав которых входят негативно взаимодействующие между собой вещества.

Чтобы минимизировать вероятность возникновения таких ситуаций, применяется алгоритм нахождения лекарственных противопоказаний и негативных взаимодействий. Он осуществляет проверку лекарств, назначаемых врачом, на совместимость и на наличие противопоказаний для пациента.

Результаты работы алгоритмов

С помощью алгоритма порождения оценочных знаний из медицинских данных, представленных в историях болезней пациентов (которые рассматриваются как прецеденты данной предметной области), были сформулированы вероятностные гипотезы о наличии или отсутствии неврологических и воспалительных осложнений. Порожденные гипотезы были добавлены в оценочные знания MedOntoModel и использовались в дальнейшей работе системы.

При помощи алгоритма выявления рисков возникновения критических ситуаций у части пациентов были обнаружены осложнения и возможность возникновения критических состояний. Программная система MedOntoModel выдала соответствующие рекомендации для врачей. При выявлении рисков возникновения критических ситуаций использовались как знания из базы общих знаний, так и оценочные знания, полученные из анализа прецедентов (историй болезней).

У части пациентов система MedOntoModel нашла лекарственные противопоказания и негативные взаимодействия в списке выписанных препаратов. Для каждого случая система выдала предупреждение с названиями конфликтующих препаратов.

Все «обнаруженные» системой критические ситуации были подтверждены наличием критических ситуаций в реальных историях болезни соответствующих пациентов.

Заключение

В настоящей работе предложена четырехуровневая модель представления знаний. Четыре уровня представления знаний это:

1) онтологические знания об определениях и смысле ключевых понятий предметной области, представленные в онтологии;

- 2) теоретические знания, которые являются общими для данной предметной области, истинными на всех ее экземплярах;
- 3) эмпирические знания, представленные в виде прецедентов предметной области;
- 4) оценочные, вероятностные знания.

На основе предложенной четырехуровневой модели представления знаний описаны методы разработки онтологий и онтологических моделей предметных областей. Построена онтологическая модель предметной области «Деформации позвоночника и дегенеративные заболевания позвоночника».

Онтологическая модель реализована при помощи программной системы MedOntoModel. Программная система включает в себя модуль порождения оценочных знаний, модуль выявления рисков возникновения критических ситуаций и модуль нахождения лекарственных противопоказаний и негативных взаимодействий.

Программная система MedOntoModel была протестирована на реальных (деперсонифицированных) медицинских данных пациентов ННИИТО им. Я. Л. Цивьяна.

Список литературы

1. Пальчунов Д. Е. Решение задачи поиска информации на основе онтологий // Бизнес-информатика. 2008. № 1. С. 3–13.
2. Пальчунов Д. Е. Моделирование мышления и формализация рефлексии. II: Онтологии и формализации понятий // Философия науки. 2008. № 2 (37). С. 62–99.
3. Palchunov D. E. Virtual catalog: the ontology-based technology for information retrieval // Knowledge Processing and Data Analysis. LNAI 6581. Berlin; Heidelberg: Springer-Verlag, 2011. P. 164–183.
4. Пальчунов Д. Е., Степанов П. А. Применение теоретико-модельных методов извлечения онтологических знаний в предметной области информационной безопасности // Программная инженерия. 2013. № 11. С. 8–16.
5. Махасоева О. Г., Пальчунов Д. Е. Автоматизированные методы построения атомарной диаграммы модели по тексту естественного языка // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, вып. 2. С. 64–73.
6. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, вып. 3. С. 34–47.
7. Carnap R. Meaning and Necessity. A Study in Semantics and Modal Logic. Chicago: Chicago University Press, 1956. 220 p.
8. Pal'chunov D. E. Algebraische Beschreibung der Bedeutung von Äußerungen der natürlichen Sprache // Zelger, Josef / Maier, Martin (Hrsg.): GABEK. Verarbeitung und Darstellung von Wissen. Innsbruck-Wien. STUDIENVerlag, 1999. S. 310–326.
9. Клещев А. С., Артемьева И. Л. Математические модели онтологий предметных областей. Ч. 1–3 // НТИ. Серия 2. 2001. № 2. С. 20–27; № 3. С. 19–29; № 4. С. 10–15.
10. Pal'chunov D. E., Yakhyaeva G. E. Interval Fuzzy Algebraic Systems // Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference'05. World Scientific Publishers, 2006. P. 191–202.
11. Пальчунов Д. Е., Яхьяева Г. Э. Нечеткие алгебраические системы // Вестн. Новосиб. гос. ун-та. Серия: Математика, механика, информатика. 2010. Т. 10, вып. 3. С. 75–92.
12. Пальчунов Д. Е., Яхьяева Г. Э. Нечеткие логики и теория нечетких моделей // Алгебра и логика. 2015. Т. 54, № 1. С. 109–118.
13. Клещев А. С., Шалфеева Е. А. Содержание системного анализа при автоматизации интеллектуальной деятельности на уровне отрасли // Материалы IV Междунар. науч.-техн. конф. «Open Semantic Technologies for Intelligent Systems (OSTIS-2014)». Минск, 2014. С. 285–290.
14. Клещев А. С., Москаленко Ф. М., Черняховская М. Ю. Модель онтологии предметной области «Медицинская диагностика». Ч. 1: Неформальное описание и определение базовых терминов // НТИ. Серия 2. 2005. № 12. С. 1–7.

15. Клещев А. С., Москаленко Ф. М., Черняховская М. Ю. Модель онтологии предметной области «Медицинская диагностика». Ч. 2: Формальное описание причинно-следственных связей, причин значений признаков и причин заболеваний // НТИ. Серия 2. 2006. № 2. С. 19–30.

16. Грибова В. В., Петряева М. В., Федорищев Л. А., Черняховская М. Ю. Формализация методов исследования в офтальмологии для компьютерных диагностических тренажеров // Материалы III Междунар. науч.-практ. конф. «Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине». СПб.: Изд-во Политехн. ун-та, 2012. Т. 2. С. 191–195.

17. Грибова В. В., Окунь Д. Б., Черняховская М. Ю. Онтология и модель онтологии предметной области «Медикаментозное лечение» // Информатика и системы управления. 2015. № 2 (44). С. 70–79.

18. Черняховская М. Ю. Формирование баз наблюдений на основе онтологии медицины // Информатика и системы управления. 2009. № 4 (22). С. 198–200.

19. Staab S., Studer R. (eds.) The Handbook on Ontologies in Information Systems. Springer Verlag, 2003. 811 p.

Материал поступил в редколлегию 20.05.2015

C. A. Naydanov¹, D. E. Palchunov^{1,2}, P. A. Sazonova¹

¹ Novosibirsk State University

2 Pirogova Str., 630090, Novosibirsk, Russia

² Sobolev Institute of Mathematics SB RAS

4 Acad. Koptug avenue, 630090, Novosibirsk, Russia

naydanov.fit@ya.ru, palch@math.nsc.ru, psazonova@gmail.com

MODEL-THEORETIC METHODS OF INTEGRATION OF KNOWLEDGE EXTRACTED FROM MEDICAL DOCUMENTS

The paper is devoted to methods of knowledge representation, integration of knowledge extracted from different texts and methods of new knowledge generation. This approach is based on four-level model of knowledge representation. Integration of knowledge extracted from medical documents and generation of estimated knowledge are done with the help of the ontological model of the domain “Spinal deformity and degenerative diseases of the spine”. The ontological model consists of ontology, general theoretical knowledge, empirical knowledge (precedents), and probabilistic knowledge. The software system for generating recommendations to prevent and reduce the risk of the patient’s critical condition has been developed.

Keywords: ontology, ontology model, knowledge representation, generation of knowledge, model-theoretic methods, atomic diagram, fuzzy model, precedent model, spinal deformity, degenerative diseases of the spine, critical conditions.

References

1. Palchunov D.E. The solution of the problem of information retrieval based on ontologies. // *Bisnes-informatika*, No. 1, 2008, p. 3–13 (in Russian).

2. Palchunov D.E. Modeling of reasoning and formalization of reflection II: Ontologies and formalization of concepts. // *Filosofiya nauki*, No. 2 (37), 2008, p. 62-99 (in Russian).

3. Palchunov D.E. Virtual catalog: the ontology-based technology for information retrieval // *Knowledge Processing and Data Analysis*. LNAI 6581. Springer-Verlag Berlin Heidelberg, 2011. Pp. 164–183.

4. Palchunov D.E., Stepanov P.A. The use of model-theoretic methods for extracting ontological knowledge in the domain of information security // *Programnaya ingeneriya*, No. 11, 2013, p. 8-16. (in Russian).
5. Makhasoeva O.G., Palchunov D.E. Semi-automatic methods of a construction of the atomic diagrams from natural language texts // *Vestnik NGU*, series: *Informacionnye tehnologii*, Vol. 12, No. 2, 2013, p. 64–73. (in Russian).
6. Derevyanko D.V., Palchunov D.E. Formal methods of development of the question-answering system on natural language // *Vestnik NGU*, series: *Informacionnye tehnologii*, Vol. 12, No. 3, 2013, p. 34–47 (in Russian).
7. Carnap R. *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: Chicago University Press, 1956. 220 p.
8. Pal'chunov D. E. Algebraische Beschreibung der Bedeutung von Äußerungen der natürlichen Sprache // *Zelger, Josef/Maier, Martin (Hrsg.): GABEK. Verarbeitung und Darstellung von Wissen*. Innsbruck-Wien. STUDIENVerlag, 1999. Pp. 310–326.
9. Kleshev A.S., Artemeva I.L. Mathematical models of subject domain ontologies. Parts 1-3 // *Nauchno-tehnicheskaya informatsiya. Seriya 2*. 2001, No. 2, p. 20-27, No. 3, p.19-29, No. 4, p. 10-15. (In Russian).
10. Pal'chunov D.E., Yakhyaeva G.E. Interval Fuzzy Algebraic Systems // *Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference'05*. World Scientific Publishers, 2006. Pp. 191-202.
11. Pal'chunov D.E., Yakhyaeva G.E. Fuzzy algebraic systems // *Vestnik NGU. Seriya: Matematika, mexanika, informatika*, vol. 10, no. 3, 2010, p. 75-92 (in Russian).
12. Pal'chunov D.E., Yakhyaeva G.E. Fuzzy logics and fuzzy model theory // *Algebra and Logic*, vol. 54, no. 1, 2015, p. 74-80.
13. Kleshev A.S., Shalfeyeva E.A. System analysis contents for intelligent activity automation at branch level // *Proceedings of the IV Conf. "Open Semantic Technologies for Intelligent Systems 2014 (OSTIS-2014)"*. Minsk: BSUIR, 2014. Pp. 285–290 (In Russian).
14. Kleshev A.S., Moskalenko F.M., Chernyakhovskaya M.Yu. Ontology model of domain "Medical diagnostics". Part 1. Informal description and definition of basic terms // *Nauchno-tehnicheskaya informatsiya. Seriya 2*. 2005. Vol. 12. Pp. 1–7 (In Russian).
15. Kleshev A.S., Moskalenko F.M., Chernyakhovskaya M.Yu. Ontology model of domain "Medical diagnostics". Part 2. Formal description of causal links, reasons of feature values and reasons of diseases // *Nauchno-tehnicheskaya informatsiya. Seriya 2*. 2006 Vol. 2. Pp. 19–30 (In Russian).
16. Gribova V.V., Petryaeva M.V., Fedorischev L.A., Chernyakhovskaya M.Yu. Formalization of Examination Method in Ophthalmology for Computer Diagnostic Simulators // *Proceedings of the III International scientific-practical conference "High Technologies, Basic and Applied Researches in Physiology and Medicine"*. St. Petersburg: Polytechnical University Publishing House, 2012. Vol. 2. Pp. 191-195 (In Russian).
17. Gribova V.V., Okun D.B., Chernyakhovskaya M.Yu. Ontology and ontology model of domain "Medication treatment" // *Informatika i sistemy upravleniya*. 2015. Vol. 2 (44). Pp. 70-79 (In Russian).
18. Chernyakhovskaya M.Yu. Formation of observation database based on medicine ontology // *Informatika i sistemy upravleniya*. 2009. Vol. 4 (22). Pp. 198-200 (In Russian).
19. Staab S., Studer R. (eds.) *The Handbook on Ontologies in Information Systems*. Springer Verlag, 2003. 811 p.