

Д. Е. Пальчунов<sup>1,2</sup>, Г. Э. Яхьяева<sup>1</sup>, О. В. Ясинская<sup>1</sup>

<sup>1</sup>Новосибирский государственный университет  
ул. Пирогова, 2, Новосибирск, 630090, Россия

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

palch@math.nsc.ru, gul\_nara@mail.ru, yasinskaya.olga@gmail.com

## ПРИМЕНЕНИЕ ТЕОРЕТИКО-МОДЕЛЬНЫХ МЕТОДОВ И ОНТОЛОГИЧЕСКОГО МОДЕЛИРОВАНИЯ ДЛЯ АВТОМАТИЗАЦИИ ДИАГНОСТИРОВАНИЯ ЗАБОЛЕВАНИЙ\*

Статья посвящена разработке автоматизированных методов порождения знаний о возможном диагнозе пациента на основе анализа историй болезни других пациентов. Эти методы основаны на теоретико-модельном подходе к формализации предметной области. На основе интеграции знаний, извлечённых из историй болезней, строится онтологическая модель данной предметной области. Формализация оценочных (нечетких) суждений производится на языке теории нечетких моделей. Для получения формульного описания диагнозов пациентов используется методология анализа формальных понятий.

*Ключевые слова:* онтология, онтологическая модель, прецедентная модель, нечеткая модель, формальный контекст, формальное понятие, диагностирование заболеваний.

### Введение

Успешное лечение больного в значительной степени зависит от быстрого и правильного диагностирования. Для исключения ошибок в диагностике и своевременного назначения лечения необходима последовательность в проведении обследования больного. На сегодняшний день существует множество различных технологий диагностики. Однако врачи (особенно поликлинического звена, а также работающие в небольших городах и районных больницах) ежедневно сталкиваются с проблемой постановки диагноза, определения необходимых диагностических процедур и консультаций смежных специалистов. Врачу необходимо на основании клинических и лабораторных исследований определить предварительный (рабочий) диагноз, на основании которого направить пациента на дополнительные и, зачастую, дорогостоящие исследования.

Поэтому критически важно и экономически обоснованно разработать такую программную систему, которая позволяла врачам на основе статистических данных определять предварительный диагноз и максимально быстро получать информацию о необходимости проведения тех или иных дополнительных диагностических процедур для постановки заключительного диагноза и выбора оптимальной тактики лечения.

---

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

В данной работе мы описываем методологию разработки программной системы, помогающей лечащему врачу пристраивать индивидуальный план диагностирования пациента. Данная методология основана на статистической обработке историй болезней пациентов, прошедших полный цикл диагностирования и лечения конкретного заболевания. В разработанной методологии для представления знаний, извлечённых из различных текстов на естественном языке (историй болезней), используется прецедентный подход к представлению знаний [1; 2]. Он основан на разработанном теоретико-модельном подходе к формализации онтологий предметных областей [3]. Для обработки представленных в системе знаний используется методология анализа формальных понятий [4; 5].

### **Теоретико-модельная формализация предметной области**

Одним из важнейших направлений инженерии знаний является представление знаний (knowledge representation). Еще в середине 70-х годов прошлого столетия появились первые прикладные интеллектуальные системы, использующие различные способы представления знаний – экспертные системы. Экспертные системы являются программными системами, имитирующими принятие решений квалифицированным специалистом в данной предметной области. Однако постепенно пришло понимание необходимости моделирования не эксперта, а непосредственно самой предметной области.

С теоретико-модельной точки зрения онтологическое моделирование предметной области и построение теории предметной области содержит следующие этапы [3].

1. Определение множества ключевых понятий предметной области. Множество ключевых понятий мы рассматриваем как сигнатуру онтологии и онтологической модели. На этом этапе мы получаем логическую теорию данной предметной области, т. е. теорию класса всех моделей рассматриваемой сигнатуры.

2. Описание смысла ключевых понятий предметной области (наиболее просто – в виде глоссария, наиболее полно – в виде множества явных и неявных определений ключевых понятий). На этом этапе мы получаем онтологию данной предметной области; в идеале – аналитическую теорию предметной области, т. е. теорию класса всех моделей, представляющих все мыслимые случаи этой предметной области.

3. Задание аксиом, эмпирических постулатов, то есть, предложений, которые должны быть истинны на всей предметной области. На этом шаге мы получаем в большей или меньшей степени полное описание всех реальных примеров предметной области, т. е. на этом этапе строится теория предметной области.

Кроме онтологии и теории предметной области онтологическая модель содержит множество конкретных прецедентов предметной области, а также нечёткую теорию предметной области – множество предложений, истинность которых оценивается числом из промежутка  $[0; 1]$ .

4. Формальное представление множества реальных прецедентов предметной области. Безусловно, множество имеющихся прецедентов не покрывает все возможные ситуации, оно представляет лишь то частичное знание о конкретных примерах предметной области, которое нам уже известно. На этом этапе строится булевозначная прецедентная модель предметной области.

5. Порождение оценочных, нечётких знаний о предметной области на основе анализа множества имеющихся прецедентов. На этом этапе строится нечёткая модель предметной области, или класс нечетких моделей, представляемый обобщённой нечёткой моделью предметной области.

Далее мы более подробно рассмотрим построение онтологической модели предметной области диагностики заболеваний позвоночника.

#### *Сигнатура предметной области*

В данной работе рассматривается предметная область, содержащая сведения о пациентах с заболеваниями только из одного фиксированного класса заболеваний (например, класс заболеваний позвоночника, представленных в МКБ-10 [6]). Мы не рассматриваем механизмы

взаимодействия пациентов друг с другом и с медицинским персоналом. Следовательно, в формализации данной предметной области используются только понятия-свойства, описывающие конкретных пациентов. Все понятия делятся на три класса:

- $\mathbb{P}$  – класс первичных диагностических процедур (например, различные клинические или лабораторные исследования). Понятия этого класса используются для постановки предварительного (рабочего) диагноза пациента;
- $\mathbb{Q}$  – класс дополнительных диагностических процедур (например, набор инструментальных исследований). Понятия этого класса используются для уточнения рабочего диагноза и постановки заключительного диагноза;
- $\mathbb{D}$  – класс различных диагнозов, рассматриваемых в рамках данного класса заболеваний.

Таким образом, мы рассматриваем сигнатуру, состоящую из конечного числа одноместных предикатных символов. Далее, сигнатуру предметной области  $\Delta$  будем обозначать через  $\sigma_\Delta$ . Как указано выше,  $\sigma_\Delta = \mathbb{P} \cup \mathbb{Q} \cup \mathbb{D}$ . Предикаты  $P(x) \in \sigma_\Delta$  будем называть *сигнатурными предикатами*.

Через  $S(\sigma_\Delta)$  будем обозначать множество всех одноместных бескванторных формул сигнатуры  $\sigma_\Delta$ . Заметим, что каждая формула  $\varphi(x) \in S(\sigma_\Delta)$  является булевой комбинацией сигнатурных предикатов.

#### *Онтология предметной области*

Для описания онтологии предметной области задается конечное множество аксиом  $Ax(\Delta) \subseteq S(\sigma_\Delta)$ . В данной статье мы рассматриваем три вида аксиом: аксиомы общего-частного, аксиомы исключения и аксиомы полноты.

*Аксиомы общего-частного.* Иерархическая упорядоченность некоторых понятий должна быть отображена аксиоматически (например, иерархия диагнозов, приведенная в МКБ-10). Схема таких аксиом следующая:

$$(P_1(x) \rightarrow P_2(x)).$$

Например: «Если *M48.0 Спинальный стеноз*, то *M48 Стеноз*».

*Аксиомы исключения.* Некоторые понятия являются взаимоисключающими (например, различные результаты лабораторных исследований). Схема таких аксиом следующая:

$$(P_1(x) \rightarrow \neg P_2(x)).$$

Например: «Если *Уровень гемоглобина в крови превышает норму*, то не верно, что *Уровень гемоглобина в крови в норме*».

*Аксиома полноты.* Для описания предметной области мы рассматриваем истории болезней пациентов прошедших полный цикл диагностирования, для которых установлен окончательный диагноз и проведено соответствующее лечение. При этом мы считаем, что в каждой рассматриваемой истории болезни отражено хотя бы по одному признаку из каждого из трех рассматриваемых классов. Таким образом, мы имеем следующую аксиому полноты:

$$\left( \left( \bigvee_{P \in \mathbb{P}} P(x) \right) \& \left( \bigvee_{P \in \mathbb{Q}} P(x) \right) \& \left( \bigvee_{P \in \mathbb{D}} P(x) \right) \right).$$

Упорядоченная пара  $\langle \sigma_\Delta, Ax(\Delta) \rangle$  образует *онтологию* предметной области  $\Delta$ . Эта онтология является подмножеством *аналитической теории* предметной области  $\Delta$ .

#### *Онтологическая модель предметной области*

Рассмотрим теперь конечное множество  $\{e_1, \dots, e_n\}$  историй болезней пациентов. История болезни  $e_i$  – это полуструктурированный текст, написанный на естественном языке. Заметим, что в каждой истории болезни однозначно указаны все проведенные диагностические

исследования, их результаты, а также заключительный диагноз для данного пациента. Следовательно, для каждой истории болезни  $e_i$  мы можем описать множество признаков (сигнатурных предикатов), истинных на  $e_i$ . Таким образом, по каждой истории болезни  $e_i$  мы строим одноэлементную модель  $e_i = \langle \{e_i\}, \sigma_\Delta \rangle$ , которую будем называть *прецедентом* предметной области  $\Delta$ .

Для построения онтологической модели нам необходимо рассмотреть конечное множество  $\mathbb{E} = \{e_1, \dots, e_m\}$  прецедентов предметной области. Заметим, что мы не исключаем случая, когда для некоторых  $i, j \in \{1, \dots, m\}$  ( $i \neq j$ ) модели  $e_i$  и  $e_j$  будут изоморфны. Тем не менее мы будем учитывать эти модели как формализации различных прецедентов.

Необходимым условием, накладываемым на прецеденты, является истинность на них аналитической теории предметной области. С теоретико-модельной точки зрения это, в частности, означает, что на каждой модели  $e_i$  должно быть истинным множество аксиом  $\mathcal{A}_x(\Delta)$ , т. е.  $e_i \models \mathcal{A}_x(\Delta)$ .

Однако, не каждая одноэлементная модель аналитической теории, является прецедентом предметной области. Выбор тех или иных моделей в качестве прецедентов и задает структуру *онтологической модели*  $\mathfrak{A}_\Delta = \langle \mathbb{E}, \sigma_\Delta \rangle$ . В онтологической модели  $\mathfrak{A}_\Delta$  для каждого сигнатурного предиката  $P(x) \in \sigma_\Delta$  и для каждого прецедента  $e_i \in \mathbb{E}$  имеем  $\mathfrak{A}_\Delta \models P(e_i)$  тогда и только тогда, когда  $e_i \models P(e_i)$ .

Заметим, что если формула  $\varphi(x) \in \mathcal{S}(\sigma_\Delta)$  принадлежит множеству аксиом предметной области  $\Delta$  (т. е.  $\varphi(x) \in \mathcal{A}_x(\Delta)$ ), то для любого  $e_i \in \mathbb{E}$  имеет место  $\mathfrak{A}_\Delta \models \varphi(e_i)$ . Таким образом, на онтологической модели является истинной онтология данной предметной области, более того, онтологическая модель принадлежит классу моделей аналитической теории предметной области.

#### Формализация оценочных знаний

Для решения задач статистической обработки данных нам понадобятся *прецедентная* и *нечеткая* модели рассматриваемой предметной области [7, 8]. Эти модели мы будем строить на основе онтологической модели.

Для дальнейшего рассмотрения нам понадобится понятие прецедентной модели, которая является частным случаем булевозначной модели.

**ОПРЕДЕЛЕНИЕ 1** [7]. Пусть  $\mathbb{B}$  – полная булева алгебра и  $\tau: \mathcal{S}(\sigma_A) \rightarrow \mathbb{B}$ . Тогда упорядоченная тройка  $\mathfrak{A}_\tau = \langle A, \sigma, \tau \rangle$  называется **булевозначной моделью**, если истинностная функция  $\tau$  замкнута относительно логических операций.

**ОПРЕДЕЛЕНИЕ 2** [7]. Упорядоченную тройку  $\mathfrak{A}_E = \langle \{a\}, \sigma_\Delta, \tau_E \rangle$  назовем **прецедентной моделью** предметной области  $\Delta$ , порожденной онтологической моделью  $\mathfrak{A}_\Delta = \langle \mathbb{E}, \sigma_\Delta \rangle$ , если для любой формулы  $\varphi(x) \in \mathcal{S}(\sigma_\Delta)$  имеем  $\tau_E(\varphi(a)) = \{e \in E \mid \mathfrak{A}_\Delta \models \varphi(e)\}$ .

В прецедентной модели каждой формуле ставится в соответствие множество прецедентов, для которых она является истинной.

Таким образом, по множеству прецедентов  $\mathbb{E}$  мы определяем булевозначную модель  $\mathfrak{A}_E$ . В этой булевозначной модели каждому предложению ставится в соответствие элемент булевой алгебры  $\rho(\mathbb{E})$ .

Данное описание основано на следующем результате.

**ТЕОРЕМА О ДВОЙСТВЕННОСТИ** [7]. Пусть  $\mathbb{B}$  – полная атомная булева алгебра,  $\mathfrak{A}_\mathbb{B}$  – булевозначная модель,  $\mathbb{E} = \{\mathfrak{A}_b \mid b \in \text{At}(\mathbb{B})\}$  и  $\mathfrak{A}_E$  – прецедентная модель. Тогда  $\mathfrak{A}_\mathbb{B} \cong \mathfrak{A}_E$ .

В большинстве методик статистической обработки данных используются объективные и / или субъективные вероятности. Под объективной вероятностью понимается относительная частота появления какого-либо события в общем объеме наблюдений или отношение числа благоприятных исходов к общему количеству наблюдений. Под субъективной вероятностью имеется в виду мера уверенности некоторого эксперта или группы экспертов в том, что данное событие в действительности будет иметь место.

В рассматриваемом подходе, для описания объективных вероятностей используется понятие нечеткой модели [7].

ОПРЕДЕЛЕНИЕ 3. Упорядоченную тройку  $\mathfrak{A}_{\mu_E} \equiv \langle \{a\}, \sigma_{\Delta}, \mu_E \rangle$  назовем **нечеткой моделью** предметной области  $\Delta$ , порожденной онтологической моделью  $\mathfrak{A}_{\Delta} = \langle \mathbb{E}, \sigma_{\Delta} \rangle$ , если

для любой формулы  $\varphi(x) \in S(\sigma_{\Delta})$  имеем  $\mu_E(\varphi) = \frac{\|\{e \in E \mid \mathfrak{A}_{\Delta} \models \varphi(e)\}\|}{\|E\|}$ .

Значениями истинности предложений (понятий) в нечеткой модели являются числа из интервала  $[0,1]$ , которые отражают объективную вероятность наличия того или иного понятия у случайно выбранного прецедента. Более подробное описание свойств прецедентных и нечетких моделей можно найти в работах [7; 2; 9; 10].

### Алгоритмы обработки знаний

#### Формальное описание предварительных диагнозов

Для формального описания предварительных диагнозов была использована техника анализа формальных понятий (АФП). Анализ формальных понятий – это прикладная ветвь алгебраической теории решеток. На сегодняшний день АФП считается одним из мощнейших методов интеллектуального анализа данных. Более подробно с этим направлением можно ознакомиться в работах [4; 11; 12].

Центральным понятием АФП является понятие *формального контекста*. С теоретико-модельной точки зрения формальный контекст задается классом моделей  $K \subseteq K(\sigma)$  фиксированной сигнатуры  $\sigma$  и множеством предложений  $S \subseteq S(\sigma)$  этой же сигнатуры и является упорядоченной тройкой  $\langle K, S, \models \rangle$  [5].

В данной работе будем рассматривать формальный контекст  $\mathcal{K}_{\Delta} = \langle \mathbb{E}, \sigma_{\Delta}, \models \rangle$ , порожденный онтологической моделью  $\mathfrak{A}_{\Delta}$ .

Пусть  $\mathfrak{A}_E$  и  $\mathfrak{A}_{\mu_E}$  – прецедентная и нечеткая модели, порожденные онтологической моделью  $\mathfrak{A}_{\Delta}$ . Тогда пара множеств  $(A, B)$ , таких что  $A \subseteq \mathbb{E}$ ,  $B \subseteq \sigma_{\Delta}$ , является *формальным понятием* контекста  $\mathcal{K}_{\Delta}$ , если выполняются следующие условия:

- 1)  $\mu_E(\&_{\varphi(x) \in B} \varphi(a)) > \mu_E(\psi(a) \& (\&_{\varphi(x) \in B} \varphi(a)))$ , для любого  $\psi(x) \in S(\sigma_{\Delta})/B$ ;
- 2)  $A = \tau_E(\&_{\varphi(x) \in B} \varphi(a))$ ,

где  $\{a\}$  – основное множество нечеткой модели  $\mathfrak{A}_{\mu_E}$ .

Множество  $B$  называется *содержанием* формального понятия  $(A, B)$ . Для удобства изложения мы вместо множества  $B$  будем называть содержанием формального понятия  $(A, B)$  формулу  $\&_{\varphi(x) \in B} \varphi(x)$ .

Формальное понятие  $(A_1, B_1)$  называется *более общим понятием*, чем понятие  $(A_2, B_2)$  (и обозначается  $(A_1, B_1) \supseteq (A_2, B_2)$ ) если  $A_2 \subseteq A_1$ . Заметим, что если  $(A_1, B_1) \supseteq (A_2, B_2)$ , то  $B_1 \subseteq B_2$ .

Для постановки предварительного диагноза мы будем использовать формальный контекст  $\mathcal{K}_{\mathbb{P}} = (\mathbb{E}, \mathbb{P}, \models)$ , где  $\mathbb{P} \subset \sigma_{\Delta}$  – класс первичных диагностических процедур. Заметим, что формальный контекст  $\mathcal{K}_{\mathbb{P}}$  является подконтекстом контекста  $\mathcal{K}_{\Delta}$ .

Формальное понятие  $(A, B)$  контекста  $\mathcal{K}_{\mathbb{P}}$  будем называть *положительной гипотезой* для диагноза  $D(x) \in \mathbb{D}$ , если выполняется условие

$$\mu_E \left( \&_{\varphi(x) \in B} \varphi(a) \rightarrow D(a) \right) \geq \alpha,$$

где  $0 < \alpha \leq 1$ , т. е. здесь  $\alpha$  – коэффициент допустимой погрешности. Этот коэффициент определяется эмпирически, в зависимости от специфики рассматриваемого класса заболеваний. Он отражает допустимую долю прецедентов (историй болезни), в которых рабочий диагноз не был впоследствии подтвержден.

Рассмотрим множество  $G(D)$  всех положительных гипотез для диагноза  $D$ . Заметим, что может случиться так, что для некоторого диагноза  $D(x) \in \mathbb{D}$  множество  $G(D)$  является пустым. Это означает, что класс  $\mathbb{P}$  первичных диагностических процедур является недостаточно информативным для определения рабочего диагноза  $D(x)$ . Если таких диагнозов много, то необходимо произвести пересмотр класс  $\mathbb{P}$  первичных диагностических процедур на уровне определения сигнатуры предметной области.

Допустим, что множество  $G(D)$  не пусто. Определим множество  $G_{\max}(D) \subseteq G(D)$  максимальных положительных гипотез для диагноза  $D(x)$ , т. е. таких гипотез, что для любого понятия  $(A, B) \in G_{\max}(D)$  не существует более общего понятия, принадлежащего множеству  $G(D)$ . Тогда формулу

$$F_D(x) = \bigvee_{(A, B) \in G_{\max}(D)} \&_{\varphi(x) \in B} \varphi(x),$$

будем называть *формульным описанием* диагноза  $D(x)$ .

Коэффициент

$$k_D = \frac{\mu_E(F_D(x) \& D(a))}{\mu_E(D(a))}$$

отражает степень полноты формульного описания диагноза  $D(x)$ . Очевидно, что  $0 < k_D \leq 1$ . При этом выполнение условия  $k_D = 1$  означает, что в формульном описании  $F_D(x)$  учтены все прецеденты, на которых предикат  $D(x)$  истинен.

#### *Алгоритм определения рабочего диагноза*

Рассмотрим пациента  $Pat$ . Предположим, что проведено частичное диагностирование этого пациента, и необходимо установить предварительный диагноз. Это означает, что имеется информация об истинности некоторых, но возможно не всех, предикатов из множества  $\mathbb{P}$ . Обозначим через  $True(Pat)$  множество сигнатурных предикатов из множества  $\mathbb{P}$ , истинность которых для пациента  $Pat$  известна. Обозначим через  $Th(Pat)$  замыкание относительно выводимости множества  $True(Pat)$  (т. е. теорию, порожденную множеством  $True(Pat)$ ). По теории  $Th(Pat)$  будем строить *модель пациента*  $Pat$ .

**ОПРЕДЕЛЕНИЕ 4.** Упорядоченную тройку  $\mathfrak{A}_{Pat} = \langle \{Pat\}, \sigma_{\Delta}, \eta_{Pat} \rangle$  будем называть *нечеткой моделью пациента*  $Pat$ , если для любой формулы  $\varphi(x) \in S(\sigma_{\Delta})$  истинностная функция  $\eta_{Pat}$  определяется следующим образом:

$$\eta_{Pat}(\varphi(Pat)) = \begin{cases} 1, & \varphi(x) \in Th(Pat); \\ 0, & \neg\varphi(x) \in Th(Pat); \\ [0,1] & \text{в противном случае.} \end{cases}$$

Модель  $\mathfrak{A}_{Pat}$  является обобщенной нечеткой моделью сигнатуры  $\sigma_{\Delta}$ . Формальное определение и описание свойств таких моделей можно найти в работах [7; 9].

Далее, для определения предварительного диагноза (или нескольких предварительных диагнозов) нам необходимо проверять истинность на модели  $\mathfrak{A}_{Pat}$  формульных описаний  $F_D(x)$  всех диагнозов  $D$  из множества  $\mathbb{D}$ . Диагнозы, для которых выполняется условие  $\eta_{Pat}(F_D(Pat)) = 1$ , объявляются предварительными диагнозами для данного пациента.

Однако может сложиться ситуация, когда рабочие диагнозы не определены, т. е. для любого  $D(x) \in \mathbb{D}$  мы имеем  $\eta_{Pat}(F_D(Pat)) \neq 1$ . Тогда если найдется хотя бы один диагноз  $D(x)$  такой, что  $\eta_{Pat}(F_D(Pat)) = [0,1]$ , то система предлагает произвести дополнительное обследование пациента. Если же складывается ситуация, когда для любого  $D(x) \in \mathbb{D}$  имеем  $\eta_{Pat}(F_D(Pat)) = 0$ , то мы имеем дело с нестандартной ситуацией, т. е. диагностировать такого пациента при помощи разработанной системы невозможно.

#### *Алгоритм назначения дополнительного диагностирования*

Допустим, что в ходе первичного обследования пациента  $Pat$  были выдвинуты предварительные диагнозы  $D_1(x), \dots, D_k(x) \in \mathbb{D}$ . Дальнейшей задачей системы является подбор наиболее оптимального набора дополнительных диагностических процедур для постановки заключительного диагноза.

Из множества прецедентов  $\mathbb{E}$  выберем подмножество прецедентов  $\mathbb{E}'$ , для которых был определен хотя бы один из диагнозов  $D_1(x), \dots, D_k(x)$ , т. е.

$$\mathbb{E}' = \{e \in \mathbb{E} \mid \mathfrak{A}_{\Delta} \models D_1(e) \vee \dots \vee \mathfrak{A}_{\Delta} \models D_k(e)\}.$$

Рассмотрим формальный контекст  $\mathcal{K}_{\mathbb{Q}} = (\mathbb{E}', \mathbb{Q}, \models)$ , являющийся подконтекстом контекста  $\mathcal{K}_{\Delta}$ . Содержанием этого контекста является класс  $\mathbb{Q}$  дополнительных диагностических процедур. В этом контексте будем искать наибольшее формальное понятие  $(A, B)$  по отношению  $\sqsupseteq$  – «быть более общим понятием». Если содержание  $B$  этого понятия не пусто, то оно является искомым набором инструментальных средств исследования с учетом рабочих диагнозов  $D_1(x), \dots, D_k(x)$ .

При этом может оказаться, что наибольшее понятие  $(A, B)$  имеет пустое содержание  $B$ . Поскольку рассматриваемое множество формальных понятий образует конечную решетку, существует конечное множество максимальных формальных понятий среди формальных понятий с непустым содержанием. Пусть  $\{(A_1, B_1), \dots, (A_l, B_l)\}$  – множество таких максимальных формальных понятий с непустым содержанием для контекста  $\mathcal{K}_{\mathbb{Q}}$ . Тогда система предлагает альтернативные решения:  $B_1, \dots, B_l$ .

Несложно проверить, что множество понятий  $\{(A_1, B_1), \dots, (A_l, B_l)\}$  обладает следующими свойствами:

- 1)  $A_1 \cup \dots \cup A_l = \mathbb{E}'$ ;

- 2) для любого формального понятия  $(A, B)$  контекста  $\mathcal{K}_{\mathbb{Q}}$  найдется такое  $i = 1, \dots, l$ , что  $B_i \subseteq B$ .

Таким образом, предлагая набор альтернативных решений  $B_1, \dots, B_l$ , мы, с одной стороны, обеспечиваем охват всех рассматриваемых прецедентов, а с другой стороны, минимизируем количество дополнительных диагностических процедур, необходимых для постановки заключительного диагноза.

#### *Апробация разработанных алгоритмов*

Разработанные методы реализованы в программной системе «Diagnostic Panel». Программная система апробирована на предметной области «деформации позвоночника и дегенеративные заболевания позвоночника».

Онтологическая модель является ядром программы и реализуется в базе данных MS SQL Server 2014. Для описания историй болезни в базе данных было создано 12 таблиц и организованы связи между таблицами. Они описывают 7 категорий рассматриваемых признаков в историях болезни – Пол, Возрастная категория, Диагноз, Жалобы, Первичный осмотр, Анализы и Инструментальные средства. Каждая из этих категорий признаков, применяемых для описания историй болезни, представлена в базе данных в виде отдельной таблицы, связанной с главной таблицей, описывающей историю болезни с помощью связей по внешнему ключу («один ко многим») и вспомогательных таблиц («многие ко многим»). Каждая история болезни в базе характеризуется наличием определенных признаков из каждой категории.

На основании проведенных клинических и лабораторных исследований пациента с заболеваниями позвоночника система выдает множество предварительных (рабочих) диагнозов для данного пациента. На основании имеющихся предварительных диагнозов, система помогает пользователю (лечащему врачу) выбрать минимальный необходимый набор инструментальных средств диагностирования пациента с заболеванием позвоночника с целью определения заключительного диагноза.

#### **Заключение**

В статье разработаны автоматизированные методы порождения знаний о возможном диагнозе пациента на основе анализа историй болезней других пациентов. Данные методы основаны на применении теоретико-модельного подхода к разработке онтологий, теории нечетких моделей и анализа формальных понятий.

Для формализации знаний о предметной области используется множество историй болезней пациентов, т. е. множество полуструктурированных текстов, написанных на естественном языке. По каждой истории болезни строится одноэлементная алгебраическая система – формальный прецедент предметной области. На основе множества всех прецедентов строится онтологическая модель рассматриваемой предметной области.

По онтологической модели предметной области определяется формальный контекст. В построенном формальном контексте выделяются формальные понятия, подтверждающие диагнозы пациентов. Исходя из этого, строятся формульные описания диагнозов. На основе оценочных знаний о предметной области определяется степень полноты формульного описания каждого диагноза.

Далее строится нечеткая модель пациента, прошедшего частичное обследование. На этой модели проверяется истинность формульных описаний различных диагнозов, формируется множество предварительных диагнозов для данного пациента. Рассматривается формальный контекст диагнозов, являющийся подконтекстом формального контекста предметной области. В этом формальном контексте ищется множество максимальных формальных понятий с непустым содержанием. Содержания этих понятий являются альтернативными наборами необходимых дополнительных диагностических процедур.

Разработанные методы были апробированы на деперсонифицированных историях болезни реальных пациентов.

Мы выражаем признательность сотрудникам Новосибирского НИИТО им. Я. Л. Цивьяна, любезно предоставившим нам необходимую медицинскую информацию.



**Список литературы**

1. *Pulchunov D., Yakhyaeva G.* Interval fuzzy algebraic systems // Proceedings of the Asian Logic Conference. 2005. P. 23–37.
2. *Пальчунов Д. Е., Яхьяева Г. Э.* Нечеткие логики и теория нечетких моделей // Алгебра и логика. 2015. Т. 54, № 1. С. 109–118.
3. *Пальчунов Д. Е.* Моделирование мышления и формализация рефлексии. Ч. 2: Онтологии и формализации понятий // Философия науки. 2008. Т. 37, № 2. С. 62–99.
4. *Ganter B., Wille R.* Formal Concept Analysis: Mathematical Foundations. Heidelberg: Springer, 1999.
5. *Pal'chunov D.* Lattices of Relatively Axiomatizable Classes // ICFCA. 2007. Vol. LNAI 4390. P. 221–239.
6. Международная статистическая классификация болезней и проблем, связанных со здоровьем. 10-й пересмотр. Женева, 1995.
7. *Пальчунов Д. Е., Яхьяева Г. Э.* Нечеткие алгебраические системы // Вестн. Новосиб. гос. ун-та. Серия: Математика, механика, информатика. 2010. Т. 10, № 3. С. 75–92.
8. *Yakhyaeva G.* Fuzzy model truth values // Proc. of the 6<sup>th</sup> International Conference Aplimat. Bratislava, 2007. P. 423–431.
9. *Яхьяева Г. Э., Ясинская О. В.* Методы согласования знаний по компьютерной безопасности, извлеченных из различных документов // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, № 3. P. 63–73.
10. *Yakhyaeva G.* Logic of Fuzzifications // Proc. of the 4<sup>th</sup> Indian International Conference on Artificial Intelligence (ICAI-09). Tumkur, India, 2009. P. 222–239.
11. *Gartner S., Ruhroth T., Burger J., Schneider K., Jurjens J.* Maintaining requirements for long-living software systems by incorporating security knowledge // IEEE 22<sup>nd</sup> International Requirements Engineering Conference. 2014. P. 103–112.
12. *Priss U.* Formal Concept Analysis in Information Science // Annual Review of Information Science and Technology. 2006. Vol. 40. P. 521–543.

*Материал поступил в редколлегию 10.06.2015*

**D. E. Palchunov, G. E. Yakhyaeva, O. V. Yasinskaya**

*Novosibirsk State University,  
2 Pirogov Str., Novosibirsk, 630090, Russia*

*Sobolev Institute of Mathematics SB RAS  
4 Acad. Koptyug avenue, Novosibirsk, 630090, Russia*

*palch@math.nsc.ru, gul\_nara@mail.ru, yasinskaya.olga@gmail.com*

**APPLICATION OF MODEL-THEORETIC METHODS  
AND ONTOLOGICAL MODELING TO AUTOMATE THE DIAGNOSIS OF DISEASES**

The article is devoted to developing automated methods of generating knowledge about possible diagnosis of the patient based on the analysis of clinical records of other patients. These methods are based on the model-theoretic approach to the formalization of the domain. An ontological model of the domain is constructed on the base of integration of knowledge extracted from clinical records. Formalization of estimated statements is described in the language of the fuzzy model theory. We use methodology of the formal concept analysis to obtain formular descriptions of diagnoses of patients.

*Keywords:* ontology, ontological model, case-based model, fuzzy model, formal context, formal concept, diagnosis of diseases.

## References

1. Pal'chunov D.E., Yakhyaeva G.E. Interval fuzzy algebraic systems. *Proceedings of the Asian Logic Conference*, pp. 23–37, 2005.
2. Pal'chunov D.E., Yakhyaeva G.E. Fuzzy logic and the theory of fuzzy models. *Algebra and Logic*, vol. 54, no. 1, 2015, pp. 74–80.
3. Palchunov D.E. Simulation of thinking and formalization of reflection: II. Ontologies and formalization of concepts. *Filosofiya nauki*, 37(2), pp. 62–99, 2008 (in Russian).
4. Ganter B., Wille R. *Formal Concept Analysis: Mathematical Foundations*. Heidelberg: Springer, 1999.
5. Pal'chunov D.E. Lattices of Relatively Axiomatizable Classes. *ICFCA 2007*, LNAI 4390, p. 221–239, 2007.
6. The international statistical classification of the diseases and problems connected with health; 10th revision. *Medecina, Geneva*, 1995.
7. Palchunov D.E., Yakhyaeva G.E. Fuzzy algebraic systems. *Vestnik NGU. Seriya: Matematika, mexanica, informatica*. 2010. V.10, № 3. p. 75–92.
8. Yakhyaeva G. Fuzzy model truth values. *Proceedings of the 6-th International Conference Aplimat*, February 6-9, 2007, Bratislava, Slovak Republic, p. 423–431.
9. Yakhyaeva G., Yasinskaya O. Methods of coordination of knowledge on computer security, taken from various documents. *Vestnik NGU. Seriya: Informacionnie texnologii*, V. 11, № 3, pp. 63–73, 2013.
10. Yakhyaeva G. Logic of Fuzzifications. *Proceedings of the 4th Indian International Conference on Artificial Intelligence (IICAI-09)*, Tumkur, India, pp. 222–239.
11. Gartner S., Ruhroth T., Burger J., Schneider K., Jurjens J. Maintaining requirements for long-living software systems by incorporating security knowledge. *IEEE 22nd International Requirements Engineering Conference*, pp. 103–112, 2014.
12. Priss U. Formal Concept Analysis in Information Science. *Annual Review of Information Science and Technology*, т. 40, pp. 521–543, 2006.