



УДК 004.067

Д. А. Коробов, М. В. Лапаев
Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики (Университет ИТМО)

А. И. Водяхо, Н. А. Жукова
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Модели представления данных в области медицины

Рассматривается многоуровневая информационная модель обработки и анализа данных для адаптивных информационных систем доменного уровня. Модель позволяет описывать объекты реального мира по имеющимся о них общим, оперативным и историческим данным. Модель имеет 3 уровня абстракции – абстрактная концептуальная модель предметной области обработки и анализа данных, доменные модели для отдельных предметных областей, объектные модели, определяющие реализации конечных систем. В статье рассматриваются вопросы применения модели для решения задач анализа медицинских данных о состоянии пациентов. Формируемые модели объектов реализуются в виде онтологий. Модели апробированы в составе медицинской информационной системы, разрабатываемой для ФГБУ «СЗФМИЦ им. В. А. Алмазова».

Информационные модели, обработка и анализ медицинских данных, глубинное обучение, семантические технологии

В сфере информационных технологий в настоящее время наблюдаются две тенденции. Первая связана с желанием наделить все окружающие вещи некоторого рода искусственным интеллектом, а вторая – с созданием семантической сети (Semantic Network – SN). Первое направление довольно быстро развивается в рамках концепции интернета вещей (Internet of Things – IoT), второе – в рамках концепции семантической паутины (Semantic Web – SW).

Понятие интернета вещей возникло в 1999 г. и применялось для определения способов взаимодействия физических объектов между собой и с окружающим миром [1]. По мере распространения беспроводных сетей и развития методов межмашинного взаимодействия концепция непрерывно развивалась, включая в состав новые идеи и технологии. Значительное количество проектов, основанных на IoT, реализовано для области медицины. В некоторых клиниках широкое распространение получили бейджи, позволяющие следить за перемещением персонала больницы и пациентов. Это позволяет эффективным образом организовать пространство клиники и деятель-

ность большинства структурных подразделений. Например, определить, где лучше разместить шкафчик с медицинскими инструментами. Родные, пришедшие навестить больных, с помощью электронного табло могут узнать, где находится пациент – в палате, в операционной, в коридоре или где-либо еще. Также существуют специальные датчики, которые внедряются в контейнеры с жидким мылом. Они контролируют расход мыла и время, которое врач тратит на то, чтобы вымыть руки. В палатах внедряются устройства, которые позволяют удаленно отслеживать состояние тяжелобольных. Тем самым, один врач, находясь на своем рабочем месте, может на экране дисплея отслеживать состояние сразу десятков пациентов. При отсутствии таких приборов в каждой палате пришлось бы дежурить отдельному врачу [2].

SW работает на основе обычной всемирной паутины (WWW) и использует протокол HTTP для передачи информации. Если всемирная паутина предназначена для хранения информации, понятной и доступной человеку, то SW предоставляет удобный доступ к информации для вычислительных устройств. SW строится на основе

онтологий [3]. Наличие онтологического описания позволяет компьютерам понимать информацию, занесенную в семантическую паутину, а также выполнять над ней операции, включая логический вывод. В медицине применение онтологий позволяет, например, по записям врача и результатам объективных анализов предоставлять информацию о состоянии больного, найти в связанных онтологических источниках данных описание болезни, которой соответствуют наблюдаемые у пациента показатели, предоставить по болезни детальную спецификацию; подготовить сводную информацию по пациенту, которая может учитываться при постановке диагноза. Вся информация, формируемая медицинскими системами, носит рекомендательный характер. В роли потребителей в основном выступают лечащие врачи и сотрудники управленческих подразделений.

Достигнутые успехи в реализации концепции IoT и широкое распространение семантической паутины позволяют выйти на новый уровень их использования в прикладных предметных областях. Новый этап развития требует интеграции IoT и SW и обеспечения дальнейшего совместного использования. Задача интеграции сводится, в первую очередь, к задаче обработки и анализа данных, собираемых от объектов реального мира с использованием технологий IoT, их представлению в виде, потребляемом семантической паутиной.

Решать проблему обработки и анализа данных можно различными средствами. В одних случаях достаточно применения стандартных методов математической статистики, например для определения наличия и степени отклонений результатов тестов от нормы или определения наличия корреляций между значениями различных параметров, измеряемых при проведении диагностических процедур. В случаях, когда требуется сформировать заключение на основе информации, которую практически невозможно обработать в ручном режиме из-за ее больших объемов, необходимости учета значительного числа влияющих факторов и т. д., применяются, как правило, методы машинного обучения. Например, методы поиска ассоциативных зависимостей позволяют выявить влияние погодных условий и генетических предрасположенностей на протекание болезней.

При решении задачи интеграции IoT и SW алгоритмы машинного обучения рассматриваются в контексте одного из современных направлений развития искусственного интеллекта – глубинного обучения (Deep Learning – DL). Алгоритмы глубинного обучения моделируют высокоуровне-

вые абстракции данных, используя архитектуры, состоящие из множества нелинейных трансформаций [4]. В ходе обработки данных, получаемых от объектов реального мира, средства глубинного обучения позволяют построить их информационные модели, которые содержат информацию, необходимую и достаточную для решения прикладных задач. Поддержка формируемыми моделями стандартов W3C обеспечивает согласованность моделей с информационным пространством семантической паутины.

Для построения моделей объектов реального мира (МРМ) и их поэтапного наполнения средствами систем обработки и анализа данных предложена и разработана иерархическая информационная модель обработки и анализа данных (МОАД). Модель определяет этапы обработки данных с использованием статистических алгоритмов и методов машинного обучения, а также способы формализованного представления результатов каждого из этапов. МОАД может быть описана на трех уровнях абстракции. Модель верхнего уровня содержит описания понятий и процессов в общем виде, а по мере снижения уровня абстракции модель детализируется. В состав МОАД входят концептуальная модель, доменная модель, объектные модели.

1. *Концептуальная модель* является абстрактной задачей-ориентированной моделью. Модель предназначена для решения прикладных задач обработки и анализа данных, включая результаты измерений параметров объектов. Состав информации и методы работы с данными, которые определены в модели, описаны на уровне, позволяющем после детализации применять их в различных предметных областях.

2. *Доменная модель* представляет собой модель предметной области, в которой планируется решать прикладные задачи. Модель строится на основе концептуальной модели, которая детализируется и расширяется за счет данных, информации и знаний о предметной области.

3. *Объектная модель* является результатом уточнения и наполнения доменной модели информацией о контекстах решения прикладных задач. Модели определяют реализации конечных систем. Они могут быть описаны в формате OWL/RDF, представлены в виде ER-диаграмм отношений реляционной базы данных или в виде программных структур/классов/объектов.

Информационная модель обработки и анализа измерений (рис. 1) позволяет описать исследуемый

объект ИО на основе получаемых от объекта данных D . Сбор данных осуществляется системой сбора и предобработки данных ССПД. Для их последующей обработки и анализа используется СИОАД (система интеллектуальной обработки и анализа данных). Каждый элемент процесса, вовлеченный в сбор и обработку данных, описывается в виде модели. Модели на рис. 1 представлены в блоке «Модели СОАД». Модели систем обработки и анализа данных включают две составляющие – статическую (СМСИОАД) и динамическую (ДМСИОАД). Модели объектов строятся на основе обобщенных моделей, размещенных в блоке «Обобщенные модели». Названия моделей совпадают с названиями моделируемых объектов, дополненными префиксами «М» и «ОМ» для моделей и обобщенных моделей соответственно. Модели объектов формируются при проектировании систем, обобщенные модели определяются для СИОАД априорно. Блок обработки и анализа данных определяет состав и способы применения моделей, методов и алгоритмов для решения прикладных задач. Модели обработки и анализа данных рассматриваются на уровне предметной области обработки и анализа данных (модель предметной области обработки и анализа данных, МПО ОиАД), уровне СИОД (обобщенная модель обработки и анализа данных, ОМ ОиАД), уровне задач обработки и анализа, решаемых для реальных объектов предметной области (модели обработки и анализа данных исследуемых объектов, М ОиАД ИО). Для возможности учета контекста функционирования объектов в состав информационной модели обработки и анализа измерений включен блок моделей, описывающих среду («Внешняя среда» – ВС). Аналогично объектам

реального мира внешняя среда описывается в виде модели (модель внешней среды – МВС), построенной на основе обобщенной модели (обобщенная модель внешней среды – ОМВС).

Применение модели обработки и анализа данных можно проиллюстрировать на примере одного из вариантов использования медицинской информационной системы, разработанной для Исследовательского центра им. В. А. Алмазова [5]. В результате проведения диагностических мероприятий приборами регистрируются объективные показания параметров, отражающих состояние пациента. Врач в процессе ведения больного ведет текстовые записи, например дневник, эпикризы. Все результаты объективных анализов, а также экспертные текстовые описания фиксируются в электронной карточке больного с разбиением на периоды, соответствующие этапам протекания болезни.

С точки зрения структуры модели обработки и анализа данных сбор информации о состоянии пациентов можно рассмотреть как процесс получения данных об объектах исследования. Объективные и экспертные данные передаются от систем сбора данных в систему обработки и анализа. СИОАД в результате выполнения преобразований над получаемыми данными формирует информационную модель, отражающую состояние и ход лечения пациентов по доступной оперативной, исторической и общей информации (информационная модель объектов области медицины – ИММ). Структура модели показана на рис. 2, описание трансформаций приведено в таблице. С позиций глубинного обучения ИММ – высокоуровневая абстракция данных, а преобразования, выполняемые СИОАД, – нелинейные трансформации.

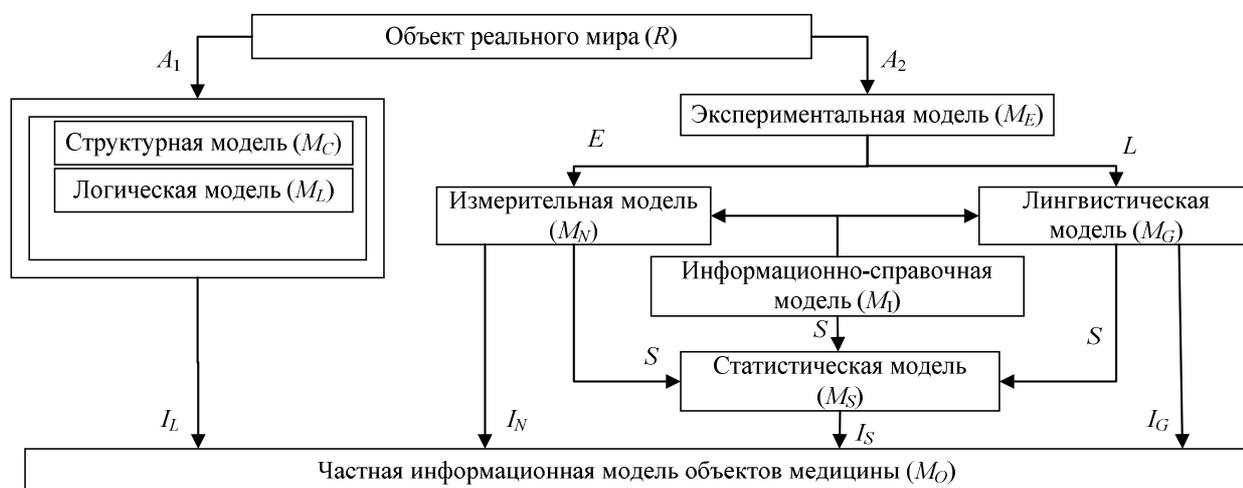


Рис. 2

Символ	Название трансформации	Начальная точка	Конечная точка (модель)	Описание модели
A_1	Структурная и логическая абстракция	Объект реального мира	Структурная модель (M_C)	Описание медицинских учреждений (диагностические средства, лабораторная база, профильные специалисты и т. д.) в части, относящейся к лечению пациента в медицинском учреждении
			Логическая модель (M_L)	Описание хода развития болезни
A_2	Экспериментальная абстракция	Объект реального мира	Экспериментальная модель (M_E)	Описание условий лечения пациента и влияющих факторов
E	Эксперимент	Экспериментальная модель	Измерительная модель (M_N)	Сбор объективных данных о пациенте
L	Обработка текстов	Экспериментальная модель	Лингвистическая модель (M_G)	Обработка и анализ экспертной текстовой информации
S	Статистическая обработка	Измерительная, лингвистическая модели	Статистическая модель (M_S)	Обработка и анализ данных с использованием статистических методов
I_L	Структурно-логическая интерпретация	Структурно-логическая модель	Частная информационная модель объектов медицины (M_O)	Представление объективных и субъективных данных в виде, необходимом для решения задачи в заданном контексте (особенности пациента, пользователь системы (врач/специалист управленческого подразделения))
I_N	Измерительная интерпретация	Измерительная модель		
I_G	Лингвистическая интерпретация	Лингвистическая модель		

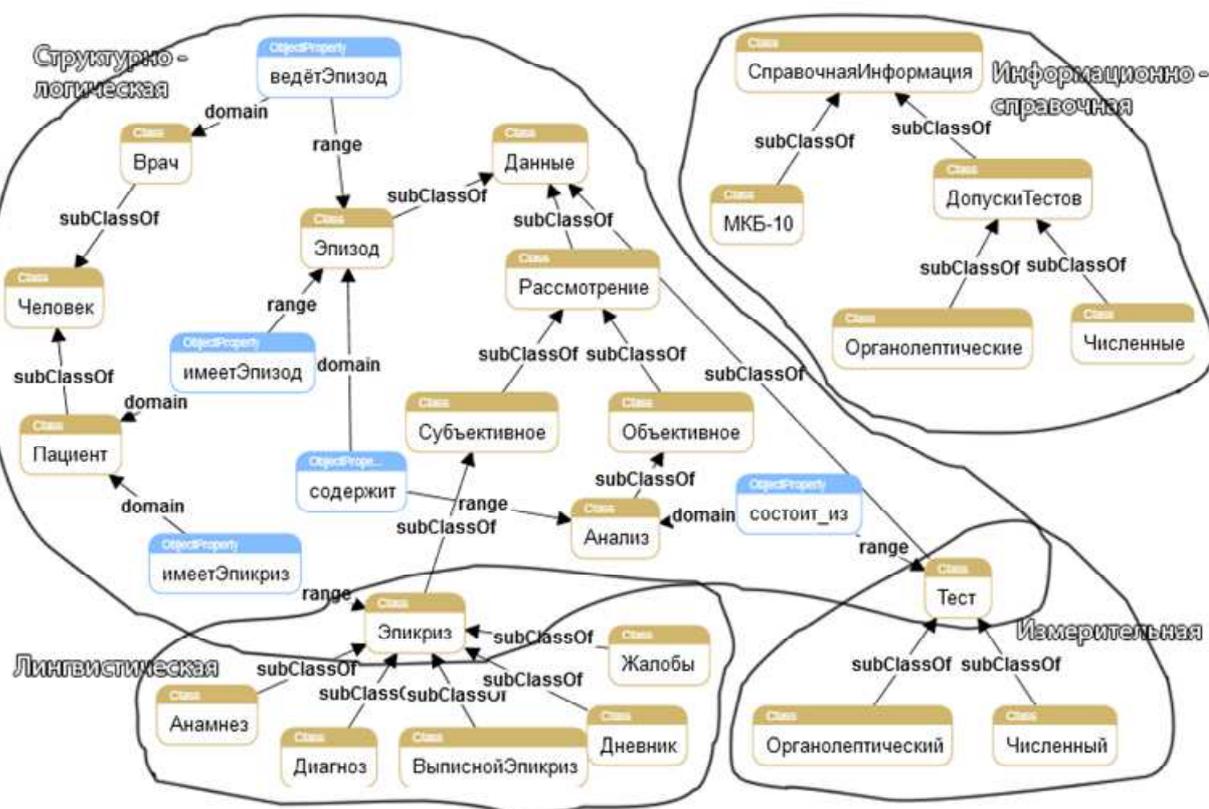


Рис. 3

По предложенной ранее информационной модели объектов области медицины была построена онтология для СИОАД центра им. В. А. Алмазова (рис. 3). Сущности сгруппированы в соответствии с моделями, к которым они принадлежат в ИММ.

К измерительной модели относятся результаты тестов, которые проводятся в рамках обследования. К лингвистической – текстовая информация и результаты ее обработки. Структурная и

логическая модели на рисунке объединены. К структурной составляющей относятся общие сущности, используемые при описании процессов работы с пациентами: пациенты, врачи, анализы, эпикризы. Сущности эпизод и данные определяют логику описания и ведения истории болезни. Вся справочная информация, такая, как классификатор болезней и допусковые интервалы значений параметров для тестов, относится к информационно-справочной модели.

```

<!--
http://www.semanticweb.org/smda/ontologies/2016/2/smda-ontology-rus#имеетЭпикриз
-->
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/smda/ontologies/2016/2/smda-ontology-rus#имеетЭпикриз">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/smda/ontologies/2016/2/smda-ontology-rus#Пациент"/>
<rdfs:range
rdf:resource="http://www.semanticweb.org/smda/ontologies/2016/2/smda-ontology-rus#Эпикриз"/>
</owl:ObjectProperty>

```

Рис. 4

На рис. 4 приведен фрагмент описания онтологии в формате rdf [3]. В качестве примера рассматривается одно из объектных свойств. Остальные свойства описываются в аналогичном виде.

На основе онтологии спроектирована и реализована модель хранения данных в объектной базе данных (рис. 5). Частично алгоритмы обработки и анализа данных взаимодействуют непосредственно с базой данных. Работа алгоритмов с базой данных обусловлена тем, что традиционно медицинские информационные системы строятся на основе объектных баз данных. Для многократного использования алгоритмов требуется обеспечить наличие источника данных в виде объектной базы данных.

В статье показана возможность использования разработанной авторами концептуальной модели обработки и анализа изменений для построения информационных моделей объектов области медицины. Модель реализована и используется в системе семантической обработки и анализа медицинских данных, которая в настоящее время находится в стадии активной разработки. Система создается в рамках исследовательского проекта, выполняемого в интересах центра им. В. А. Алмазова [5]. Актив-

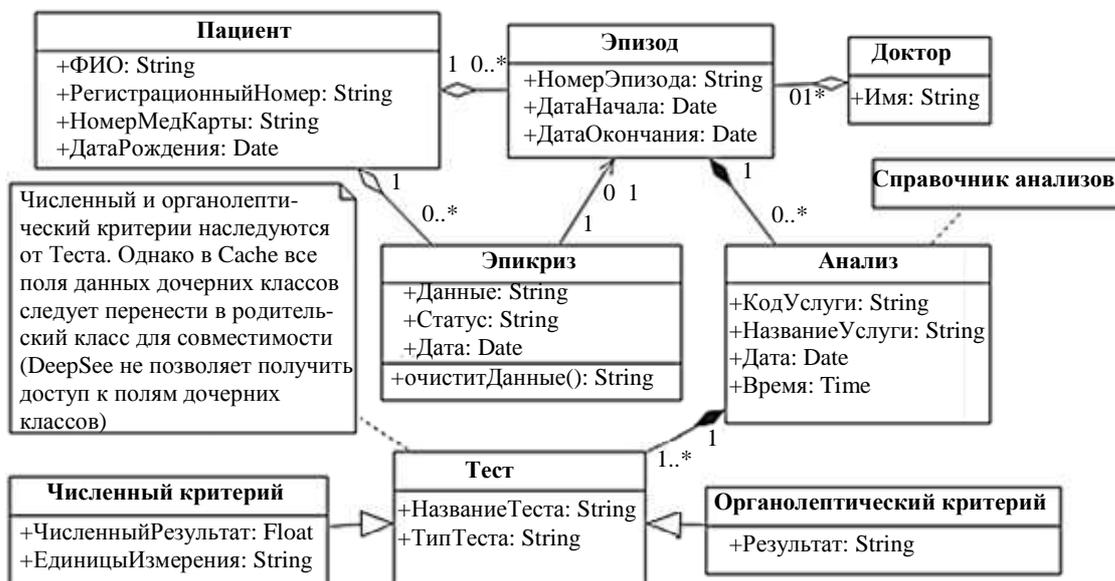


Рис. 5

Субъективная информация (эпикризы и другие записи врача) и объективная информация (результаты анализов: численные и органолептические) выступают в роли данных, получаемых в результате эксперимента (диагностических и лечебных мероприятий).

ное участие в работе над системой принимают специалисты Центра. В проекте также принимают участие другие исследовательские лаборатории и представители промышленных предприятий, в том числе InterSystems [6].

СПИСОК ЛИТЕРАТУРЫ

1. The Internet of Things. URL: <http://www.theinternetofthings.eu/>.
2. Как интернет вещей меняет современную медицину. URL: <https://meduza.io/feature/2015/12/14/bolnitsy-bez-vrachey-i-umnoe-mylo>.
3. W3C. SemanticWeb. URL: <http://www.w3.org/standards/semanticweb/>.
4. Deep learning applications and challenges in big data analytics / M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. WaldEmail, E. Muharemagic // J. of Big Data. 2015. № 2. URL: <http://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7>.
5. Медицинский центр им. В. А. Алмазова. URL: <http://www.almazovcentre.ru/>.
6. InterSystems. URL: <http://www.intersystems.com>

D. A. Korobov, M. V. Lapaev
ITMO University (Saint Petersburg National Research University
of Information Technologies, Mechanics and Optics)

A. I. Vodyaho, N. A. Zhukova
Saint Petersburg Electrotechnical University «LETI»

DATA MODELS FOR MEDICAL DOMAIN

A multilevel information model for adaptive information systems for data processing and analysis is discussed. The model is to be used for building of real world object models taking into account operational and historical data. The model has 3 levels of abstraction: abstract conceptual model, model of subject domain and models of target system. In the article problems of usage of this model in medicine domain are discussed. The models are realized as ontologies. Suggested approach was tested while building the system of medical data processing for Almazov Center in Saint Petersburg

Information models, medical data processing and analysis, deep learning, semantic technologies

УДК 681.3

М. Ф. Галимуллин, Е. Л. Калишенко,
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Н. А. Рапоткин
АО «Информационно-телекоммуникационные технологии» (Санкт-Петербург)

Анализ производительности стратегий синхронизации потоков в структурах данных, основанных на flat-combining

Рассматриваются некоторые сценарии использования конкурентных структур данных, показывающие повышение производительности при увеличении времени работы одного потока, которому остальные потоки делегируют свои задачи. Данный подход получил название flat-combining (FC) [1]. Представлены несколько разработанных стратегий синхронизации, описаны их преимущества и область применения.

Конкурентные структуры данных, анализ производительности, flat-combining, многопоточность

Идея подхода flat-combining (на примере стека) заключается в следующем (рис. 1): со стеком связан мьютекс и список анонсов (publication list) размером, пропорциональным количеству потоков, работающих со стеком. Каждый поток при первом обращении к стеку добавляет в список

анонсов свою запись (I на рис. 1). Если потоку необходимо выполнить операцию над контейнером, то он публикует в своей записи запрос – операцию, например push или pop для стека, и ее аргументы – и пытается захватить мьютекс. Если мьютекс захвачен, поток становится комбайнером