

УДК 007:519.816
DOI: 10.15827/0236-235X.120.739-744

Дата подачи статьи: 18.07.17
2017. Т. 30. № 4. С. 739–744

ПОСТРОЕНИЕ ОНТОЛОГИИ НА ОСНОВЕ НЕРЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ МЕДИЦИНСКОГО НАЗНАЧЕНИЯ

А.П. Еремеев, д.т.н., профессор, зав. кафедрой, eremeev@appmat.ru;

С.А. Ивлиев, аспирант, siriusfrk@gmail.com

*(Национальный исследовательский университет «Московский энергетический институт»,
ул. Красноказарменная, 14, г. Москва, 111250, Россия)*

По мере наращивания объемов данных в системах здравоохранения становится возможным использование комплексных методов программной обработки данных для поддержки принятия решений в сложных проблемных ситуациях. Ввиду сильной гетерогенности среды (разные формы отчетности, форматы файлов, итеративный процесс работы с пациентом) требуется создание гибкой системы, которая позволит эффективно работать в описанных условиях.

Рассматривается процесс построения и использования онтологии на основе нереляционной БД для интеллектуальной системы поддержки принятия решений для исследования и диагностики сложных патологий зрения и для проведения интеллектуального анализа данных в гетерогенной среде. Также рассматривается построение программного интерфейса для работы с такой онтологией. Создаваемая онтология ориентирована на хранение как данных об обследованиях пациентов, проводимых врачом, так и заключений по ним. Нереляционная БД позволяет в таком случае более эффективно оперировать данными в гетерогенных средах, примером которых являются медицинские исследования.

Показаны основные преимущества нереляционных БД по сравнению с традиционными реляционными: более удобное оперирование данными в рамках заданной онтологии, ее расширение и дополнение, извлечение требуемых данных по запросу. Описана программная реализация предложенного подхода и проиллюстрирована ее работа.

По итогам проведенной работы предложены дальнейшие пути развития в данной области, а именно: обобщение полученных результатов на решение других схожих задач медицинской диагностики, а также подключение новых методов, таких как семантический поиск по историям болезней, в реализуемую интеллектуальную систему поддержки принятия решений.

Ключевые слова: интеллектуальная система, поддержка принятия решений, нереляционная БД, онтология.

Часто интеллектуальные системы поддержки принятия решений (ИСППР), в частности для медицинских приложений, должны быть ориентированы на функционирование в гетерогенной среде, поскольку, как правило, постоянно происходит взаимодействие разработчиков или пользователей и экспертов. Кроме того, сама предметная область подвержена изменениям в результате как появления новых знаний, так и внедрения нового оборудования для исследований и диагностики, что может, в свою очередь, привести к изменению формата данных, с которыми работает система. Также отметим, что в различных медицинских учреждениях могут быть использованы разные форматы отчетности, технологии подготовки специалистов и другие особенности. В результате может возникнуть ситуация, когда уже действующая ИСППР не будет обладать достаточной гибкостью и адаптируемостью для эффективного функционирования в новых условиях. Одним из подходов к решению данной проблемы является разработка более абстрактных онтологий предметной области, что, однако, повышает накладные расходы и приводит к усложнению модификаций системы.

Отметим, что в последние годы в связи с резким увеличением обрабатываемых объемов данных и их сложности началось широкое развитие технологии хранения данных в нереляционных БД (НБД), которые предоставляют более высокую гибкость при работе в сложных средах, а также простоту модификации на самом нижнем уровне – уровне хра-

нения данных по сравнению с традиционными реляционными БД (РБД) [1, 2].

Другой важной задачей является обработка поступаемых в ИСППР данных. Поскольку они могут быть представлены в разной форме (форматах), можно применять различные методы и алгоритмы, которые подразумевают как извлечение знаний из данных (Data mining), так и их анализ с выводом некоторого заключения (Knowledge discovery). В таком случае для эффективной работы требуются также создание и хранение соответствующей онтологии предметной области, что достаточно эффективно может быть реализовано посредством НБД. Далее для обработки данных и поиска решения (диагноза, рекомендации по лечению) могут быть применены методы искусственного интеллекта на основе нейронных сетей и генетических алгоритмов, метод опорных векторов, а при наличии статистических данных – известные теоретико-вероятностные методы, в частности, на основе байесовских сетей доверия и теории свидетельств Демпстера–Шеффера и др.

Рассмотрим проблему построения онтологии на основе НБД в плане ее использования в ИСППР для диагностики сложных патологий зрения [3, 4].

БД для задачи хранения медицинских данных

Любой системе, связанной с обработкой данных, требуется система их хранения [5]. Поскольку

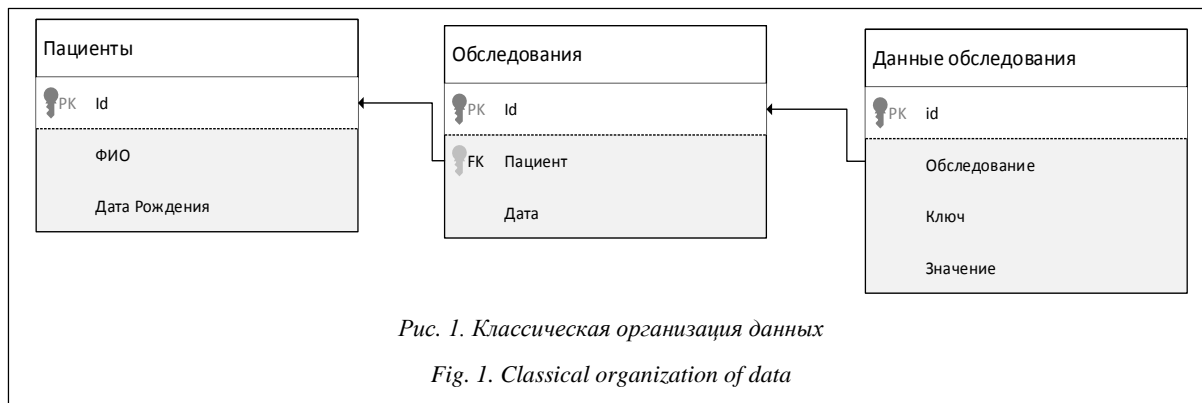


Рис. 1. Классическая организация данных
Fig. 1. Classical organization of data

проектируемая экспертная система является ИСППР, следует ожидать наличия высокой гетерогенности в представляемых данных. В связи с этим сравним обычную реляционную модель хранения данных и нереляционную.

В качестве примера используем историю болезни, содержащую следующие данные:

- результаты обследований специалистами-медиками, записанные на естественном языке;
- результаты анализов, которые могут представлять собой ряды данных, кортежи (наименование, значение), фотоматериалы;
- результаты обследований, записанные в виде опросников;
- назначения препаратов и процедур.

Помимо распределенности данных во времени (которая всегда сложно обрабатывается в традиционных РБД и требует реализации дополнительных структур для работы с темпоральными данными [6, 7]), имеются данные, представленные на естественном языке, в виде временных рядов и изображений, и, возможно, данные (знания), полученные в рамках другой модели.

В случае выбора для реализации РБД потребуется создание ряда таблиц и связей между ними, причем по мере приближения модели данных к уровню их физического хранения будет теряться гибкость, а по мере отдаления усложняться работа на уровне представления данных.

Продемонстрируем сказанное на следующих примерах.

1. Пусть РБД содержит в себе следующие таблицы: Пациенты, Обследования, Данные обследования. В первой таблице хранится общая информация о пациентах, во второй – связи пациентов и обследований, а также данные о проведении обследования (время, место, тип), в третьей – собственно данные исследований. Можно сразу заметить, что третья таблица будет перегружена данными вне зависимости от способа их размещения, при этом, помимо деградации скорости работы ввиду большого количества записей, потребуются создание промежуточных таблиц и сущностей для извлечения знаний из данных, что будет приводить, в свою очередь, к замедлению скорости работы системы в це-

лом. На рисунках 1 и 2 приведено схематичное сравнение подходов.

2. Пусть в РБД хранятся исключительно знания, полученные в результате обработки данных, которые были представлены, то есть по сути имеем базу знаний (БЗ). Однако в данном случае теряется возможность повторного исследования данных при получении каких-либо новых сведений, так как собственно данные не обрабатываются для постоянного использования.

3. Пусть в БД (БЗ) хранятся как данные, так и получаемые из них знания. Тогда количество таблиц будет расти по мере использования новых типов знаний. При этом качество доступа к сырым (необработанным) данным будет по-прежнему низким, поскольку проводится разделение изначальной понятной сущности (обследования) на искусственные подсущности.

Также стоит выделить сложности с масштабируемостью, переносимостью и обработкой больших массивов данных в обычных РБД [1, 8].

В качестве альтернативы традиционным РБД предлагается использование НБД. В отличие от концепции ACID (atomicity, consistency, isolation, durability – неделимость, согласованность, изоли-

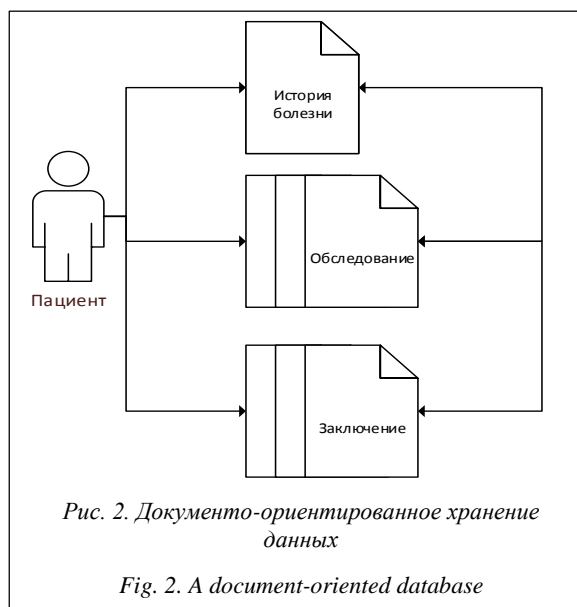


Рис. 2. Документно-ориентированное хранение данных
Fig. 2. A document-oriented database

рованность, устойчивость) в НБД используется концепция BASE:

BA (basic availability) – базовая доступность, то есть любой запрос к БД будет завершен (успешно или безуспешно) без ошибок;

S (soft state) – гибкое состояние, то есть состояние системы может изменяться со временем даже без внесения новых данных для достижения согласованности данных в базе;

E (eventual consistency) – конечная согласованность, то есть данные могут быть несогласованными некоторое время, но при этом стремятся к согласованному состоянию и в итоге приходят к нему [9].

Таким образом, НБД можно использовать для решения задач, связанных с гетерогенной средой, поскольку они исключают требования к необходимости хранения четких схем данных. Важным достоинством НБД является также возможность хранения документов как единых сущностей, что позволяет эффективно организовывать разные виды обработки данных и поиска в знаниях (например, используя алгоритм MapReduce [10]), а также создавать новые формы хранения знаний и получения их из данных путем дополнения уже имеющихся документов и установления связей между ними. Сказанное означает, что каждый из элементов истории болезни можно интерпретировать как некоторый документ и формировать на его основе новые производные документы, проводить выборку среди таких документов для отбора тех из них, которые лучше всего согласуются с имеющимися и планируемыми результатами, то есть сформировать соответствующую онтологию.

Онтология представления истории болезни в НБД

В работе [2] показан пример использования онтологий для представления геномных данных для хранения в граф-ориентированной НБД. В рассматриваемом случае предлагается использовать документные НБД в сочетании с онтологией. Опишем основные элементы предлагаемой онтологии.

Понятия (классы):

1) *обследование* (O) – проведение некоторого обследования, выраженного на основе некоторого документа определенной или неопределенной структуры; обследование может содержать в себе разные экземпляры типа «исследование ЭРГ», «фотография сетчатки» и др.;

2) *назначение* (N) – итоговый результат, выданный экспертом на основании рассмотрения данных одного или нескольких обследований; назначение может содержать в себе разные экземпляры, такие как «назначение лекарств», «назначение обследования у другого врача» и т.д.;

3) *знание* (Z) – результат получения знаний из данных автоматическими методами либо в резуль-

тате формирования при помощи опросников и других методов прямого получения знаний от экспертов; знание состоит из экземпляров разного рода, например: «опросник офтальмолога», «опросник невропатолога», «результат анализа нейросетью», «результат вывода методом Демпстера–Шеффера» и др.;

4) *документ порождения* (DR) – документ, порождаемый при применении генерации новых документов и связанный с одним из отношений порождения.

Отношения:

1) *порождения* (R_{gen}) – порождение Z из множества O, Z, N ;

2) *обобщения* (R_{int}) – порождение Z из множества Z ;

3) *заклучения* (R_{sum}) – порождение N из множества Z на основании экспертного заключения.

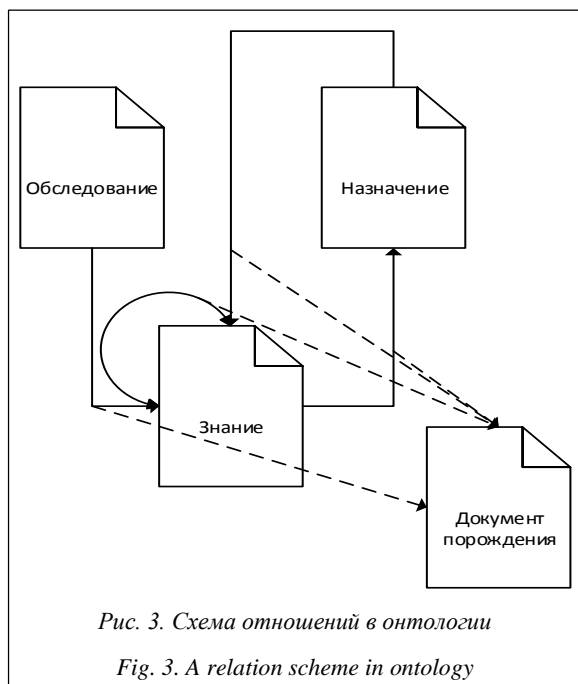
Предложенная онтология позволяет описать следующий процесс вывода нового назначения N из имеющегося по следующей формуле: $Z(t+1) = R_{int}(R_{gen}(O(t) \times N(t) \times Z(t))^2)$, $N(t+1) = R_{sum}(Z(t+1))$, где $N(t+1)$ – новое заключение; $Z(t+1)$ – обобщение имеющихся знаний на момент построения вывода; $O(t)$ – множество всех обследований на момент построения; $N(t)$ – множество всех назначений на момент построения; $Z(t)$ – множество всех имеющихся знаний на момент построения; $(O(t) \times N(t) \times Z(t))^2$ – множество всех подмножеств декартова произведения, полученных на данный момент из обследований, назначений и знаний, то есть исходные данные для повторной операции извлечения знаний из данных и знаний, полученных на данный момент.

Онтология является упрощенной базовой моделью для организации хранения знаний. Поскольку каждая генерация происходит посредством реализации методов и алгоритмов с применением отношения порождения R_{gen} , при такой операции генерируется некоторый документ DR , который позволяет проводить ретроспективные изучения последовательностей выводов.

На рисунке 3 представлена схема отношений в онтологии.

Следует отметить, что все понятия могут храниться в НБД в открытом и расширяемом видах, как формализованном, так и нет. Кроме того, НБД открыта для занесения в нее новых понятий.

В итоге можно сделать вывод, что использование предложенной онтологии позволяет формализовать процессы работы с НБД и расширить тем самым ее возможности для применения в ИСППР в сложных проблемных ситуациях, используя средства НБД, позволяющие объединить онтологию с БД, сделав ее расширяемой на низшем уровне. На рисунке 4 в качестве иллюстрации приведена последовательность сбора данных о пациенте и генерации назначений. Каждая стрелка отображает отношение порождения, создающее документ DR для



последующего анализа. Такое гармоничное представление открывает большие возможности для обработки данных посредством машинного обучения.

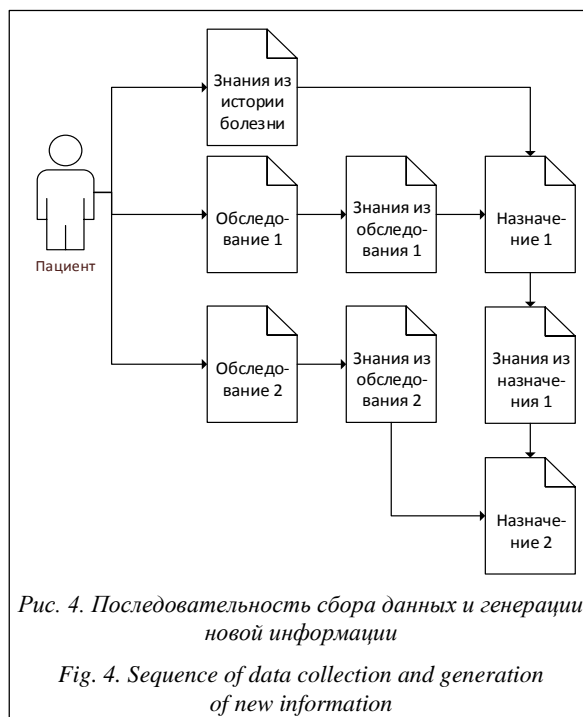
Методы машинного обучения для анализа данных и извлечения знаний

Предложенная модель позволяет интегрировать различные модели извлечения знаний и подготовки итоговых решений [3, 4, 11]. При этом для разных обследований могут использоваться разные методы получения знаний. В частности, при анализе *электроретинограмм* (ЭРГ) [4] могут быть использованы *нейронные сети* (НС), при помощи которых решается задача классификации с применением правила вывода вида $RD(\text{классификация на основе НС}) \leftarrow (O(\text{ЭРГ}), R_{gen}, Z(\text{возможное заболевание на основе НС}))$.

Также может быть использован метод на основе *имитационных моделей* (ИМ) [11] с правилом вида $RD(\text{классификация на основе ИМ}) \leftarrow (O(\text{ЭРГ}), R_{gen}, Z(\text{возможное заболевание на основе ИМ}))$.

При формировании опросников, которые позволяют получить знания (Z), возможно порождение новых знаний, которые окажутся более качественными и понятными для эксперта [3], на основе правил вида $RD(\text{вывод на основе метода Демстера-Шеффера (МД-Ш)}) \leftarrow (Z(\text{опросник 1}), Z(\text{опросник 2})), R_{int}, Z(\text{возможное заболевание опросник}))$.

Общее заключение может быть сформировано на основе интегрированного правила вывода: $RD(\text{итоговый вывод назначения}) \leftarrow Z(\text{возможное заболевание на основе НС}), Z(\text{возможное заболевание на основе ИМ}), Z(\text{возможное заболевание на основе МД-Ш}), R_{sum} N(\text{назначение})$.



Как видно из приведенных правил, в них не учитываются возможность использования данных, полученных во времени, а также изменение их динамики в результате тех или иных назначений. Для этих целей могут быть использованы методы прогнозирования, в частности:

- на основе НС, когда НС просто анализирует N параметров [12];
 - на основе кластеризации и выделения признаков (метод SVM, случайные леса [13] и др.).
- Следует отметить, что построенная модель не исключает, а способствует использованию методов семантического поиска и анализа среди заключений, сделанных специалистами (врачами как лицами, принимающими решения, или экспертами-физиологами), что может позволить получить дополнительные знания (см. пример в [14]).

Практическая реализация

В рамках предложенной модели был разработан прототип ИСППР для агрегации знаний и проведения исследований с целью диагностирования сложных патологий зрения с применением описанных ранее методов. Программная система реализована на базе фреймворка Flask и НБД MongoDB и выполняет следующие базовые функции.

1. Занесение данных о результатах биофизических исследований (например, данных, полученных при снятии ЭРГ).

2. Добавление дополнительных данных к данным исследований (в случае, если каким-то другим прибором производится необходимое дополнительное измерение, соотносящееся с исследованием).

3. Занесение данных текстового характера для результатов медицинских обследований.

4. Выполнение дифференциальной диагностики заболеваний на основе вейвлет-преобразований и нейросетевого подхода [4].

5. Вывод заключений на основании полученных результатов диагностики.

В дальнейшем планируется расширить возможности прототипа посредством использования теоретико-вероятностных методов (байесовских сетей доверия, метода Демпстера–Шефера, методов на основе нечеткой логики) [3, 11], а также методов с применением семантического поиска.

Заключение

В результате проведенного исследования разработана модель на основе НБД, позволяющая хранить и обрабатывать с применением методов машинного обучения данные, используемые для медицинской диагностики. Показаны преимущества модели на основе НБД и ее отличия от классического представления данных в традиционных РБД, в частности, возможности хранения онтологий вместе с самими данными, а также гибкость предлагаемой модели. Дано описание методов обработки данных и получения на их основе знаний, проведен анализ методов и их комбинаций (интегрирования) с целью получения более качественных итоговых заключений и с учетом специфики поступающей информации. К достоинствам предложенного подхода и реализованного на его основе прототипа ИСППР следует отнести также то, что отсутствие привязки к фактической структуре БД позволило сделать модель более универсальной и применить не только к диагностике патологий зрения на основе ЭРГ, но и для анализа кардиограмм на основе данных ЭКГ. Исследования проводятся на кафедре прикладной математики НИУ «МЭИ» совместно с Московским НИИ глазных болезней им. Гельмгольца в плане создания ИСППР для диагностики сложных патологий зрения и обучения молодых специалистов (ординаторов) офтальмологов.

Работа выполнена при финансовой поддержке РФФИ (проект № 17-07-00553) и РФФИ-БФФИ (проект № 16-51-00058).

Литература

1. Zohreh Goli-malekabady, Mohammad kazem Akbari-fatidahi, Morteza Sargozaei-javan. An effective model for store and retrieve big health data in cloud computing. *Computer Methods and Programs in Biomedicine*. 2016, vol. 132, pp. 75–82.
2. Naresh Kumar Gundla, Zhengxin Chen. Creating NoSQL Biological Databases with Ontologies for Query Relaxation. *Procedia Comp. Sc.*, 2016, vol. 91, pp. 460–469.
3. Еремеев А.П., Хазиев Р.Р., Зуева М.В., Цапенко И.В. Прототип диагностической системы поддержки принятия решений на основе интеграции байесовских сетей доверия и метода Демпстера–Шефера // Программные продукты и системы. 2013. № 1. С. 11–16.
4. Еремеев А.П., Ивлиев С.А. Анализ и диагностика сложных патологий зрения на основе вейвлет-преобразований и нейросетевого подхода // Интегрированные модели и мягкие вычисления в искусственном интеллекте: сб. науч. тр. VIII Международ. науч.-технич. конф. (Коломна, 18–20 мая 2015 г.). М.: Физматлит, 2015. Т. 2. С. 589–595.
5. Джарратано Д., Райли Г. Экспертные системы: принципы разработки и программирование; [пер. с англ.]. М.: Вильямс, 2007. С. 120–200.
6. Еремеев А.П., Куриленко И.Е., Смирнова А.Е. Разработка темпорального расширения методов рассуждений на основе прецедентов // IS&IT'11: тр. конгресса по интелект. сист. и информ. технол. М.: Физматлит, 2011. Т. 1. С. 50–59.
7. Еремеев А.П. Логика ветвящегося времени и ее применение в интеллектуальных системах поддержки принятия решений // КИИ-2006: сб. тр. 10-й национальной конф. по искусствен. интел. М.: Физматлит, 2006. Т. 3. С. 746–754
8. Ken Ka-Yin Leea, Wai-Choi Tangb,1, Kup-Sze Choia. Alternatives to relational database: Comparison of NoSQL and XML approaches for clinical data storage. *Computer methods and programs in biomedicine*. 2013, pp. 99–109.
9. Antonios Makris, Konstantinos Tserpes, Vassiliki Andronikou, Dimosthenis Anagnostopoulos. A classification of NoSQL data stores based on key design characteristics. *Procedia Computer Science*, 2016, vol. 97, pp. 94–103
10. Ralf Lämmel. Google's MapReduce programming model – Revisited. *Science of Computer Programming*. 2008, vol. 70, iss. 1, pp. 1–30.
11. Анисимов Д.Н., Вершинин Д.В., Колосов О.С., Зуева М.В., Цапенко И.В. Диагностика текущего состояния динамических объектов и систем сложной структуры методами нечеткой логики с использованием имитационных моделей // Искусственный интеллект и принятие решений. 2012. № 3. С. 39–50.
12. Ясницкий Л.Н., Думлер А.А., Богданов К.В., Полещук А.Н., Черепанов Ф.М., Макурина Т.В., Чугайнов С.В. Диагностика и прогнозирование течения заболеваний сердечно-сосудистой системы на основе нейронных сетей // Медицинская техника. 2013. № 3. С. 42–44.
13. Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda. Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Jour*. 2017, vol. 15, pp. 104–116.
14. Jay Urbain. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. *Jour. of Biomedical Informatics*. 2015, vol. 58, Supplement, pp. 143–149.

ONTOLOGY DESIGN BASED ON NON-RELATIONAL DATABASE FOR INTELLIGENT DECISION SUPPORT SYSTEM FOR MEDICAL PURPOSES

*A.P. Eremeev*¹, *Dr.Sc. (Engineering), Professor, Head of Chair, eremeev@appmat.ru*

*S.A. Ivliev*¹, *Postgraduate Student, siriusfrk@gmail.com*

¹ *National Research University "MPEI", Krasnokazarmennaya St. 14, Moscow, 111250, Russian Federation*

Abstract. As the volume of data in healthcare systems increases, it becomes possible to use complex methods of software data processing to support decision-making in complex problem situations. Since the environment has a strong heterogeneity (different form of reporting, file formats, iterative process of working with the patient), it is required to create a flexible system that will provide effective work under the conditions described.

The paper represents building and using ontologies based on a non-relational database for intelligent decision support systems in order to investigate and diagnose complex pathologies. It also describes constructing a software interface to work with this ontology. The created ontology is oriented both to the storage of data on patient examinations conducted by a doctor, and medical assessment reports. In this case, a non-relational database allows operating data in heterogeneous environments (e.g. medical research) more efficiently. The paper discusses the main advantages of non-relational databases compared to traditional relational databases, such as more convenient data handling within a given ontology, its expanding and supplementing, extracting the required data upon a request. The paper also describes software implementation of the proposed approach and illustrates its work.

Based on the results of the research, the authors suggest ways of further development in this area. They are: generalization of the obtained results to solve other similar problems of medical diagnostics, as well as using new methods, such as the semantic search against case records, in the implemented intelligent decision support system.

Keywords: intelligent system, decision support, non-relational database, ontology.

Acknowledgements. The work has been financially supported by RFBR (project no. 17-07-00553) and by RFBR-BFBR (project no. 16-51-00058).

References

1. Zohreh Goli-malekabady, Mohammad kazem Akbari-fatidahi, Morteza Sargozaei-javan. An effective model for store and retrieve big health data in cloud computing. *Computer Methods and Programs in Biomedicine*. 2016, vol. 132, pp. 75–82.
2. Naresh Kumar Gundla, Zhengxin Chen. Creating NoSQL Biological Databases with Ontologies for Query Relaxation. *Procedia Computer Science*. 2016, vol. 91, pp. 460–469.
3. Ereemeev A.P., Haziev R.R., Zueva M.V., Tsapenko I.V. Prototype of the diagnostic decision support system on the basis of integration bayesian belief networks and the Dempster–Shafer method. *Programmnye produkty i sistemy* [Software & Systems]. 2013, no. 1, pp. 11–16 (in Russ.).
4. Ereemeev A.P., Ivliev S.A. Analysis and diagnosis of complex vision pathologies based on wavelet transformations and the neural network approach. *Sb. nauch. tr. VIII Mezhdunar. nauch.-tekhnich. konf. "Integrirovannyye modeli i myagkie vychisleniya v iskusstvennom intellekte"* [Proc. 8th Int. Scientific and Tech. Conf. on Integrated Models and Soft Calculations in Artificial Intelligence]. Moscow, Fizmatlit Publ., 2015, vol. 2, pp. 589–595 (in Russ.).
5. Giarratano J.C., Riley G.D. *Expert Systems: Principles and Programming*. 4th ed. Course Technology Publ., 2004, 288 p. (Russ. ed.: Moscow, Vilyams Publ., 2007, 1152 p.).
6. Ereemeev A.P., Kurilenko I.E., Smirnova A.E. Development of temporal expansion of reasoning methods based on precedents. *Tr. kongressa po intellektualnym sistemam i informatsionnym tekhnologiyam (IS&IT'11)* [Proc. Congr. on Intelligence. Syst. and Inform. Technology]. Moscow, Fizmatlit Publ., 2011, vol. 1, pp. 50–59 (in Russ.).
7. Ereemeev A.P. The branching-time logic and its application in intelligent decision support systems. *Sb. tr. 10 nats. konf. po iskusstvennomu intellektu s mezhdunar. uchastiem KII-2006* [Proc. 10th National Conf. on Artificial Intelligence KII-2006]. Moscow, Fizmatlit Publ., 2006. vol. 3, pp. 746–754 (in Russ.).
8. Ken Ka-Yin Leea, Wai-Choi Tangb, Kup-Sze Choia. Alternatives to relational database: Comparison of NoSQL and XML approaches for clinical data storage. *Computer Methods and Programs in Biomedicine*. 2013, pp. 99–109.
9. Antonios Makris, Konstantinos Tserpes, Vassiliki Andronikou, Dimosthenis Anagnostopoulos. A Classification of NoSQL Data Stores Based on Key Design Characteristics. *Procedia Computer Science*. 2016, vol. 97, pp. 94–103
10. Ralf Lämmel. Google's MapReduce programming model – Revisited. *Science of Computer Programming*. 2008, vol. 70, iss. 1, pp. 1–30.
11. Anisimov D.N., Vershinin D.V., Kolosov O.S., Zueva M.V., Capenko I.V. Diagnostics of complex structure dynamic objects and systems current state by fuzzy logic methods using simulation model parameters. *Iskusstvenny intellekt i prinyatie resheny* [Artificial intelligence and decision-making]. 2012, no. 3, pp. 39–50 (in Russ.).
12. Yasnitsky L.N., Dumler A.A., Bogdanov K.V., Poleshchuk A.N., Cherepanov F.M., Makurina T.V., Chugaynov S.V. Diagnosis and prognosis of a cardiovascular diseases course based on neural networks. *Meditinskaya tekhnika* [Medical Technology]. 2013, no. 3, pp. 42–44 (in Russ.).
13. Kavakiotis I., Tsave O., Salifoglou A., Maglaveras N., Vlahavas I., Chouvarda I. Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Jour.* 2017, vol. 15, pp. 104–116.
14. Urbain J. Mining heart disease risk factors in clinical text with named entity recognition and distributional semantic models. *Jour. of Biomedical Informatics*. 2015, vol. 58, Supplement, pp. 143–149.