

УДК 614; 614.2; 614:33

**Е.А. Берсенева***Национальный НИИ общественного здоровья им. Н.А. Семашко,  
г. Москва, Российская Федерация***А.П. Суходолов***Байкальский государственный университет,  
г. Иркутск, Российская Федерация***А.А. Седов***Национальный НИИ общественного здоровья им. Н.А. Семашко,  
г. Москва, Российская Федерация***Д.Ю. Михайлов***Национальный НИИ общественного здоровья им. Н.А. Семашко,  
г. Москва, Российская Федерация***В.Г. Кудрина***Российская медицинская академия непрерывного профессионального образования,  
г. Москва, Российская Федерация***Т.В. Андреева***Российская медицинская академия непрерывного профессионального образования,  
г. Москва, Российская Федерация***М.Г. Спасенникова***Национальный НИИ общественного здоровья им. Н.А. Семашко,  
г. Москва, Российская Федерация*

## **ЦИФРОВАЯ МЕДИЦИНА: ИНФОРМАЦИОННО-ТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ ПРИМЕНЕНИЯ ЛЕКСИЧЕСКОГО АНАЛИЗА ПРИ ФОРМАЛИЗАЦИИ КОДИРОВАНИЯ ДИАГНОЗОВ**

**АННОТАЦИЯ.** В статье приводятся результаты анализа существующей практики кодирования диагнозов, обсуждаются недостатки существующей модели, приводящие, в конечном итоге, к невозможности грамотного построения систем мониторинга заболеваемости и смертности. Обобщаются разработанные и предлагаемые для реализации подходы к созданию модуля лексического анализа в рамках автоматизированных информационных систем поддержки кодирования. Создаваемый модуль лексического анализа в рамках системы автоматизированного кодирования пригоден как к применению непосредственно в момент кодирования, так и постфактум, актуален (при определенной технологической доработке) для любой версии Международной классификации болезней. Эффективность внедрения принципов создания и функционирования системы автоматизированной поддержки кодирования по Международной классификации болезней с модулем лексического анализа при поддержке трансформации клинического диагноза в формализованный код определяется тем, что при использовании такой системы повышается качество формирования статистики заболеваемости и смертности в РФ, что должно позволить перевести на новый качественный уровень планирование медицинской помощи, а также оснащенность ресурсами.

**КЛЮЧЕВЫЕ СЛОВА.** Информационные технологии, комплексные автоматизированные информационные системы медицинских организаций, медицинские информационные системы, качество медицинской документации, качество кодирования диагноза, лексический анализ.

**ИНФОРМАЦИЯ О СТАТЬЕ.** Дата поступления 9 сентября 2019 г.; дата принятия к печати 2 декабря 2019 г.; дата онлайн-размещения 29 декабря 2019 г.

© Берсенева Е.А., Суходолов А.П., Седов А.А., Михайлов Д.Ю., Кудрина В.Г., Андреева Т.В., Спасенникова М.Г., 2019

# **Baikal Research Journal**

электронный научный журнал Байкальского государственного университета

**E.A. Berseneva***N.A. Semashko National Research Institute of Public Health,  
Moscow, Russian Federation***A.P. Sukhodolov***Baikal State University,  
Irkutsk, Russia***A.A. Sedov***N.A. Semashko National Research Institute of Public Health,  
Moscow, Russian Federation***D.Yu. Mikhailov***N.A. Semashko National Research Institute of Public Health,  
Moscow, Russian Federation***V.G. Kudrina***Russian Medical Academy of Continuing Vocational Education,  
Moscow, Russian Federation***T.V. Andreeva***Russian Medical Academy of Continuing Vocational Education,  
Moscow, Russian Federation***M.G. Spasennikova***N.A. Semashko National Research Institute of Public Health,  
Moscow, Russian Federation*

## DIGITAL MEDICINE: INFORMATION AND TECHNOLOGICAL BASES FOR USING LEXICAL ANALYSIS IN FORMALIZATION OF DIAGNOSIS CODING

**ABSTRACT.** The article brings forth the results of analysis of the existing practice of coding diagnoses, discusses the drawbacks of the existing model leading, in the long run, to impossibility of literate building of a monitoring system for morbidity rate and death rate. It substantiates the approaches, developed and proposed for implementation, to creating a lexical analysis module in terms of automated information systems of supporting the coding. The lexical analysis module under formation in terms of the automated coding system fits both for usage immediately at the moment of coding and post-factum and is topical (with certain technological improvement) for any version of the International Classification of Diseases. The efficiency of implementing the principles of creating and functioning of the system of the automated coding support in terms of the International Classification of Diseases with the lexical analysis module with support of transforming the clinical diagnosis into a formalized code is determined by the fact that usage of such a system increases the quality of forming the statistics of morbidity and death rates in the Russian Federation, which must allow to transfer the planning of medical aid and resource infrastructure to a new qualitative level.

**KEYWORDS.** Information technologies, complex automated systems of medical organizations, medical information systems, quality of medical documentation, quality of coding diagnosis, lexical analysis.

**ARTICLE INFO.** Received September 9, 2019; accepted December 2, 2019, available online December 29, 2019.

Анализ существующей практики кодирования и возможностей использования информационных систем на основе лексического анализа, обоснование использования современных методов информационных технологий.

Методологический инструментарий, используемый для решения поставленных в настоящем исследовании задач, включает общенаучные и специальные методы: изучение и обобщение опыта, логический метод, анализ и синтез, абстраги-

рование, математическое моделирование, метод вывода на основе аналогии, метод формализации, экспертный метод, системный анализ, функциональная, объектно-ориентированная и алгоритмическая декомпозиции, активное и пассивное наблюдение, статистический анализ.

Формирование цифровой экономики — вопрос национальной безопасности Российской Федерации. Правительство Российской Федерации занимается цифровым преобразованием таких сфер, как здравоохранение, образование, промышленность, сельское хозяйство, строительство, городское хозяйство, транспортную и энергетическую инфраструктуры, финансовые услуги и др.

Кодирование диагнозов по международной классификации болезней 10-го пересмотра является важной задачей в повседневной деятельности врача, так как результаты кодирования, особенно при кодировании диагнозов в свидетельстве о смерти оказывают обширное влияние на процессы организации здравоохранения. Более того, корректный мониторинг заболеваемости и смертности в РФ невозможен без решения задачи правильного кодирования диагнозов. Несмотря на это, научно не изучены фундаментальные процессы, приводящие врача к тому или иному заключению, не выработаны методические подходы, позволяющие унифицировать процесс кодирования и проверку правильности получаемого результата.

В ходе ранее выполненной в Национальном научно-исследовательском институте общественного здоровья имени Н.А. Семашко» НИР 0528-2018-006 «Научно обоснование организации информационной поддержки кодирования по международной классификации болезней 10-го пересмотра (МКБ-10, версия 2014–2016)» была выявлена универсальная особенность процесса трансформации клинического диагноза (поставленного и записанного в традициях русской медицинской школы) в статистический код: отсутствие как фундаментально-научного, так и методического инструментария поддержки такой трансформации. В рамках данной работы был создан справочник клинических состояний, включающий 10 800 позиций, а также создана таблица взаимосвязей с типом связи «один ко многим» для диагнозов МКБ-10 и справочника клинических состояний. Однако при использовании такой таблицы необходимым является переход от клинического диагноза, сформулированного в традициях русской медицинской школы к определению клинического состояния. Этот переход, осуществляемый в нашей стране врачом, является весьма субъективным и может привести к ошибкам в кодировании заболеваемости и смертности.

Создание автоматизированного инструментария трансформации клинического диагноза в статистический код позволит произвести структурные изменения описанного процесса, без которых невозможно осуществить его «цифровую трансформацию», поскольку основной сутью такой трансформации является не перевод медицинских данных в электронную форму как таковой, а обеспечение данного процесса алгоритмической базой, позволяющей повысить качество медицинских данных, обрабатываемых в электронном виде, по сравнению с общепринятыми на данный момент методами их обработки.

При этом особо следует отметить, что на сегодняшний день, на практике не проводилась формализованная оценка влияния процесса кодирования на другие процессы в сфере здравоохранения, не применялись методы математической лингвистики, новые методы математической статистики, не разрабатывались ни математический аппарат, ни алгоритмы, ни, тем более методические рекомендации, имеющие специфическую нацеленность на унификацию, формализацию и автоматизацию процесса кодирования диагноза по МКБ-10.

Опыт, накопленный на сегодняшний момент специалистами Национального научно-исследовательского института общественного здоровья имени Н.А. Се-

машко» [1, с. 49–53; 2, с. 13–22; 3, с. 297–301; 4], а также степень развития математического аппарата, который заложен в основу проводимой в настоящее время работы, позволят обеспечить практический результат, позволяющий обеспечить цифровую трансформацию этого важного процесса, так востребованную на сегодняшний день.

Математический аппарат лексического анализа весьма развит, и постепенно, лексический анализ становится неотъемлемой частью множества информационных систем [5, с. 6–11; 6, с. 179–185; 7, с. 4–8; 8, с. 138–145; 9, с. 93–102; 10, с. 67–74; 11, с. 1–25].

Одним из перспективных направлений применения систем лексического анализа является их использование при переводе в формализованный код клинического формулирования диагнозов. В данном контексте под этим понятием подразумевается любой диагноз (предварительный, окончательный), для любого заболевания (основное, сопутствующее, осложнение, конкурирующее заболевание). Точное кодирование диагнозов с использованием международной классификации болезней является важной задачей. Сложившаяся ныне практика выбора кода по МКБ-10 подразумевает жесткое размежевание понятий «диагноз для статистического учета», и «клиническая формулировка диагноза». При этом фактическое содержание клинического диагноза зачастую весьма существенно отличается от статистического, указанного в учетных формах.

Подталкивают клиницистов к этому прежде всего отсутствие практических механизмов работы с третьим томом МКБ-10, который может быть использован для кодирования клинического диагноза. Указанная практика проиллюстрирована на рис. 1.

Предлагаемый нами подход состоит в автоматизации самого процесса выбора кода по МКБ-10 с использованием формулировок тома № 3 МКБ-10. При этом роль информационной системы состоит в как в реализации интерфейсного решения, пригодного для повседневного использования, так и в реализации требуемого алгоритмического аппарата. Общий принцип, лежащий в основе системы, проиллюстрирован на рис.2.

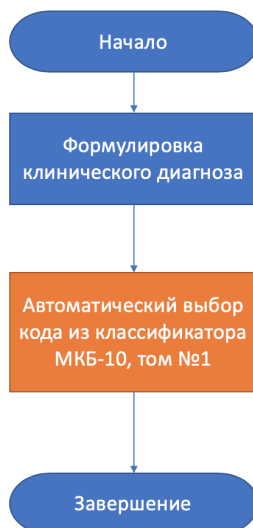
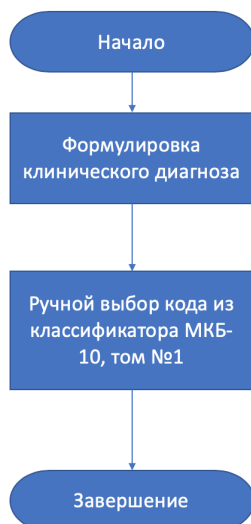


Рис. 1. Схематичный алгоритм формирования диагноза в сложившейся практике



*Рис. 2. Принципиальное изменение схемы кодирования диагноза (обобщенный сценарий)*

В ходе нашего исследования решено было создавать информационную систему, реализующая указанный подход. В практической реализации процедура автоматического выбора в текущей реализации системы подразумевает непрерывное взаимодействие пользователя и системы.

На начальном этапе исследования с использованием эмпирических знаний было спланировано и произведено наблюдение. Результаты наблюдения были подробно описаны, что позволило выделить элементы содержания процесса, формирующие этот процесс объекты, формализовать задачи процесса кодирования. Это дало возможность применить универсальные методы научного познания — анализ полученных результатов, абстрагирование (в данном случае, конечно, прежде всего от конкретных нозологий и особенностей формулирования клинических диагнозов для них, с целью выделения значимых общих свойств) и последующее обобщение свойств выявленных объектов.

Результатами этой части работы явилось полное формализованное описание процесса кодирования, акцент внимания при этом был сделан на элементах процесса, позволяющих преобразовать его таким образом, чтобы задача кодирования была решена оптимальным образом.

На этом этапе потребовалось расширить предмет исследования, проведя предварительный, во многом оценочный анализ способов использования получаемых статистических кодов. Это было сделано с целью выявить пути практического применения получаемых результатов, что, с одной стороны не являлось задачей или предметом исследования, но, с другой стороны позволяло оценить применимость результатов исследования на практике. Более четко практическая востребованность этого решения проявилась на последующем этапе моделирования, который позволил сформировать методологическую часть работы, открывающую путь к практическим результатам. Таким образом методы научного дискурса были дополнены практической составляющей, которая, по сути, представляет собой методологию оптимизации процесса кодирования. Реализация этой методологии требует не только выполнения научной работы, но и реализации ряда как админи-

стративных, так и технологических мероприятий, опирающихся на научную составляющую работы, как на платформу для обеспечения трансформации процесса практического решения задач кодирования. Хотя научный дискурс не подразумевает решения практических задач, в данной работе получение практического результата непосредственно следовало из полученных научных заключений, что послужило основой для решения применить практические методы решения задач информатизации. Это были как совершенно рутинные методы обеспечения эффективного реляционного хранения данных, разработка пользовательского интерфейса, обеспечение технологически безопасного доступа к данным их изменения, так и такие наукоемкие методы, как лексический и семантический анализ. Так же отдельно была выполнена трудоемкая работа по структуризации и сохранении в базе данных содержания третьего тома МКБ-10.

Таким образом, на этапе практической реализации удалось добиться существенного снижения роли так называемого "человеческого фактора" на этапе преобразования клинического диагноза, который, конечно, формируется врачом на основе его практического опыта, теоретических знаний, и уникальных особенностей конкретного клинического случая, в диагноз статистический, который является прежде всего формальным статистическим инструментом, позволяющим оценивать структуру заболеваемости в популяции.

Метод кодирования, реализованный в системе, предполагает следующий порядок взаимодействия с пользователем:

1. Определяется ведущий термин, который вводится в соответствующее поле интерфейса информационной системы.

2. Система осуществляет подбор всех записей по тому № 3 МКБ-10, содержащих данный ведущий термин. При этом не рассматриваются элементы формулировок диагноза, содержащие данный термин в качестве уточняющего.

3. Все найденные элементы предлагаются пользователю для выбора.

4. После выбора ведущего термина фиксируется соответствующий ему код МКБ-10, и далее предлагается уточнить клиническую формулировку диагноза. Последовательное уточнение происходит до тех пор, пока имеются уточняющие записи в томе № 3 МКБ-10.

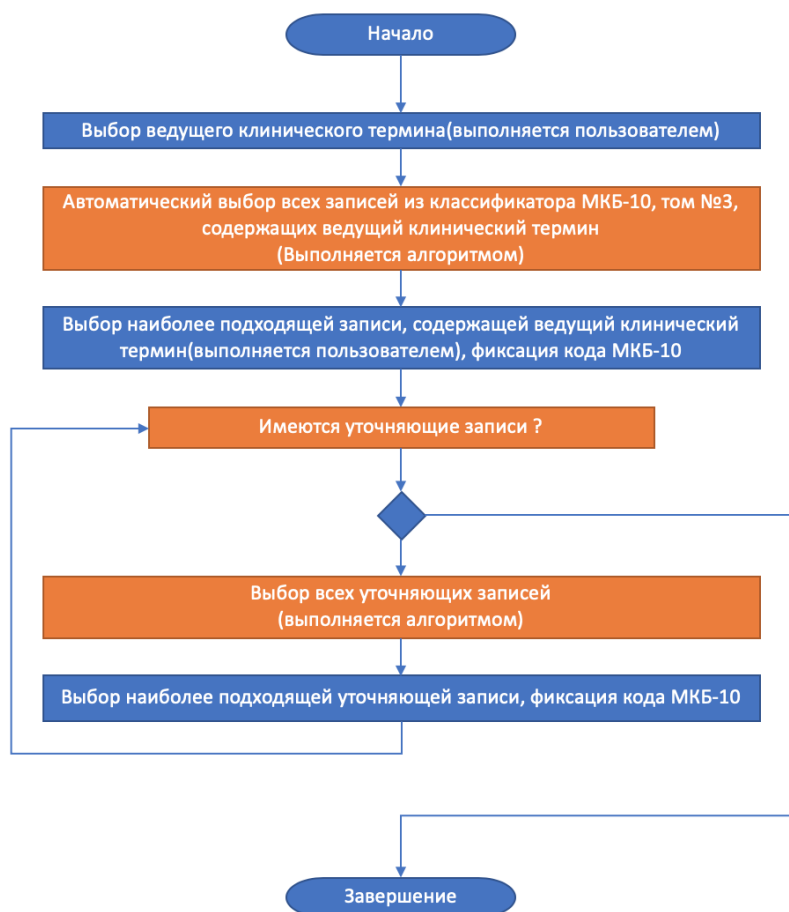
Следует заметить, что в случаях, когда промежуточное уточнение делает возможным выбор кода диагноза МКБ-10, отличного от выбранного ранее, этот выбор фиксируется, и предъявляется пользователю, что делает возможным последующий анализ процесса принятия решения о выборе кода, и, в случае необходимости внесение изменения начиная с любого шага уточнения без необходимости начинать кодирование с начала.

Предлагаемый сценарий использования проиллюстрирован на рис. 3.

При создании автоматизированной системы поддержки кодирования на основе лексического анализа принято решение использовать следующие технологические решения (рис. 4):

- Операционная система семейства Linux;
- База данных FireBird;
- Сервер среднего слоя GlassFish;
- Сервер веб-интерфейса Apache;
- Алгоритмизация производится на языке Scala, с последующим превращением в Java-апплет;
- В качестве интерфейсного(клиентского) решения используется любой браузер на основе ядра «Хромиум». Взаимодействие между клиентской частью системы и сервером реализуется по сервис-ориентированной модели, путем вызова поименованных сервисов.





*Рис. 3. Порядок работы системы при выборе кода с использованием системы лексического анализа*

Для создания информационной система кодирование диагнозов и оформления свидетельств о смерти на основе международных классификаторов (автоматизированной системы поддержки кодирования по МКБ-10) с использованием лексического анализа были выбраны следующие технические характеристики:

- Создание на основе свободно распространяемого ПО (Linux, FireBird, Glasfish) — в свете тенденции по замене проприетарного программного обеспечения;
- Трехзвенная архитектура — в соответствии с требованиями концепции информатизации здравоохранения;
- Территориально распределенная база данных — для возможности развертывания в нескольких регионах одновременно;
- Работа обособленных подразделений в отдельных базах данных с единой точкой входа — что позволяет контролировать процесс на уровне органов управления здравоохранения;
- Web-интерфейс — избавляет от необходимости устанавливать программное обеспечение на рабочие места пользователей;
- Наличие средств агрегации любых имеющихся в системе данных — используется для построения отчетов произвольного формата;

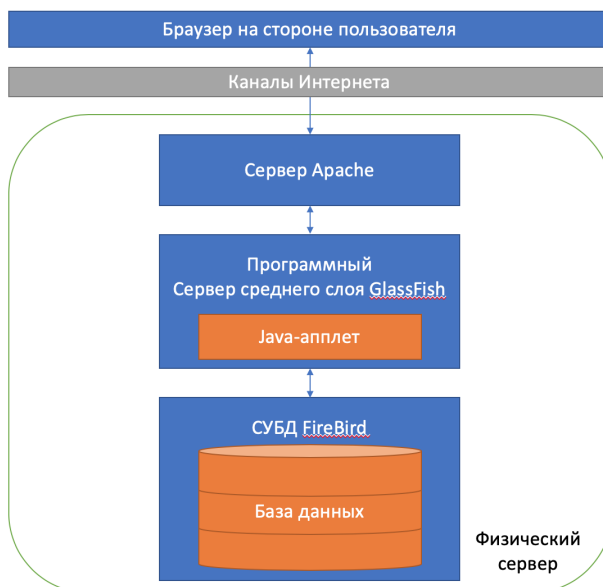


Рис. 4. Принципиальная схема технологических решений

- Поддержка средств визуализации агрегатов;
- Собственная система обмена сообщениями — используется для передачи уведомлений организационного характера;
- Механизмы лексического анализа — используются для обеспечения эффективной работы с классификаторами.

Следует отметить, что при проектировании использован передовой международный опыт информатизации здравоохранения, зафиксированный в отечественных стандартах.

Для создания автоматизированной информационной системы поддержки кодирования на основе лексического анализа нами предлагается использовать облачную модель. Преимущества облачной модели, особенно для сегментов государственной и муниципальной информатизации, делают ее основным вариантом тиражирования типовых программных решений в сфере электронного правительства и информационного общества в России. Кроме того, облачная модель открывает ранее невиданные возможности для разработчиков решений, снимая барьеры в организации сбыта решений и позволяя сосредоточиться на функциональных возможностях и качестве решений.

#### Список использованной литературы

1. Берсенева Е.А. Автоматизированный лексический контроль как средство повышения качества медицинских документов / Е.А. Берсенева, А.А. Седов // Менеджер здравоохранения. — 2014. — № 2. — С. 49–53.
2. Берсенева Е.А. Этапы создания автоматизированной системы лексического анализа медицинских документов (АИС «Элекс») / Е.А. Берсенева, А.А. Седов // Бюллетень Национального НИИ общественного здоровья им. Н.А. Семашко. — 2015. — Вып. 6. — С. 13–22.
3. Берсенева Е.А. Автоматизированный лексический контроль медицинских документов — обязательный элемент информатизации медицинских организаций / Е.А. Берсенева, А.А. Седов // Проблемы социальной гигиены, здравоохранения и истории медицины. — 2016. — Т. 24, № 5. — С. 297–301.



4. Берсенева Е.А. Лексический анализ в здравоохранении / Е.А. Берсенева, Р.Т. Тайрова. — Москва : Светлица, 2018. — 178 с.
5. Система принятия решений в программе льготного лекарственного обеспечения / Р.У. Хабриев, А.П. Суходолов, Б.А. Спасенников, Л.Ю. Безмельницына, Д.О. Мешков. — DOI 10.17150/2500-2759.2018.28(1).6-11 // Известия Байкальского государственного университета. — 2018. — Т. 28, № 1. — С. 6–11.
6. Оценка необходимого объема финансирования лекарственной терапии отдельных заболеваний / Р.У. Хабриев, А.П. Суходолов, Л.Ю. Безмельницына, Б.А. Спасенников, Д.О. Мешков, С.Н. Черкасов. — DOI 10.17150/2500-2759.2018.28(2).179-185 // Известия Байкальского государственного университета. — 2018. — Т. 28, № 2. — С. 179–185.
7. Смертность от внешних причин у лиц группы риска / Р.У. Хабриев, С.В. Кулакова, Л.Ф. Пертли, Б.А. Спасенников // Проблемы социальной гигиены, здравоохранения и истории медицины. — 2019. — Т. 27, № 1. — С. 4–8.
8. Оптимизация тарифной политики в сфере обязательного медицинского страхования в субъекте Российской Федерации / А.П. Суходолов, Л.И. Меньшикова, Н.Н. Ясько, О.А. Ларюшкина, А.Л. Санников, Б.А. Спасенников. — DOI 10.17150/2500-2759.2019.29(1).138-145 // Известия Байкальского государственного университета. — 2019. — Т. 29, № 1. — С. 138–145.
9. Мбайкоджи Э. Метод автоматической классификации коротких текстовых сообщений. / Э. Мбайкоджи, А.А. Драль, И.В. Соченков // Информационные технологии и вычислительные системы. — 2012. — № 3. — С. 93–102.
10. Браславский П.И. Сравнение пяти методов извлечения терминов произвольной длины / П.И. Браславский, Е.А. Соколов // Компьютерная лингвистика и интеллектуальные технологии : материалы ежегод. междунар. конф. «Диалог» (Бекасово, 4–8 июня 2008 г.). — Москва, 2008. — Вып. 7 (14). — С. 67–74.
11. Колесникова И.А. Научно-педагогические параметры оценки качества стратегического документа в области образования / И.А. Колесникова, Е.В. Титова // Непрерывное образование: XXI век. — 2015. — № 1 (9). — С. 1–25.

### References

1. Berseneva E.A., Sedov A.A. Automated Lexical Control as a Medical Document's Quality Control Improvement Instrument. *Menedzher zdravookhraneniya = Manager of Health Care*, 2014, no. 2, pp. 49–53. (In Russian).
2. Berseneva E.A., Sedov A.A. Stages of the Automated System Lexical Analysis of Medical Documents (Ais «Alex»). *Byulleten natsionalnogo NII obshchestvennogo zdorovya imeni N.A. Semashko = N.A. Semashko National SRI of Public Health Bulletin*, 2015, iss. 6, pp. 13–22. (In Russian).
3. Berseneva E.A., Sedov A.A. The Automatized Lexical Control of Medical Documents as an Obligatory Component of Informatization of Medical Organizations. *Problemy sotsial'noi gigieny, zdravookhraneniya i istorii meditsiny = Problems of Social Hygiene, Public Health and History of Medicine*, 2016, vol. 24, no. 5, pp. 297–301. (In Russian).
4. Berseneva E.A., Tairova R.T. *Leksicheskii analiz v zdravookhranении* [Lexical Analysis Public Health]. Moscow, Svetlitsa Publ., 2018. 178 p.
5. Khabriev R.U., Sukhodolov A.P., Spasennikov B.A., Bezmelnitsyna L.Y., Meshkov D.O. The DecisionMaking System in the Program of Subsidized Pharmaceutical Provision. *Izvestiya Baykal'skogo gosudarstvennogo universiteta = Bulletin of Baikal State University*, 2018, vol. 28, no. 1, pp. 6–11. DOI: 10.17150/2500-2759.2018.28(1).6-11. (In Russian).
6. Khabriev R.U., Sukhodolov A.P., Bezmelnitsyna L.Y., Spasennikov B.A., Meshkov D.O., Cherkasov S.N. Evaluation of the Necessary Amount of Financing of Drug Therapy for Particular Diseases. *Izvestiya Baykal'skogo gosudarstvennogo universiteta = Bulletin of Baikal State University*, 2018, vol. 28, no. 2, pp. 179–185. DOI: 10.17150/2500-2759.2018.28(2).179-185. (In Russian).
7. Khabriev R.U., Kulakova S.V., Pertli L.F., Spasennikov B.A. The Mortality from External Causes in Individuals of Risk Group. *Problemy sotsial'noi gigieny, zdravookhraneniya i istorii meditsiny = Problems of Social Hygiene, Public Health and History of Medicine*, 2019, vol. 27, no. 1, pp. 4–8. (In Russian).

8. Sukhodolov A.P., Menshikova L.I., Yasko N.N., Laryushkina O.A., Sannikov A.L., Spasennikov B.A. Optimization of Tariff Policy in the Area of Compulsory Medical Insurance in the Subject of the Russian Federation. *Izvestiya Baikal'skogo gosudarstvennogo universiteta = Bulletin of Baikal State University*, 2019, vol. 29, no. 1, pp. 138–145. DOI: 10.17150/2500-2759.2019.29(1).138-145. (In Russian).

9. Mbaykodzhi E., Dral A.A., Sochenkov I.V. Short Text Messages Classification Method. *Informatsionnye tekhnologii i vychislitel'nye sistemy = Journal of Information Technologies and Computing Systems*, 2012, no. 3, pp. 93–102. (In Russian).

10. Braslavskii P.I., Sokolov E.A. Comparison of Five Methods for Variable Length Term Extraction. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Materialy ezhegodnoi mezhdunarodnoi konferentsii (Bekasovo, 4–8 iyunya 2008 g.)* [Computer Linguistics and Intellectual Technologies. Materials of Annual International Conference, Bekasovo, June 4–8, 2008]. Moscow, 2008, vol. 7 (14), pp. 67–74. (In Russian).

11. Kolesnikova I.A., Titova E.V. Scientific and Pedagogical Quality Parameters of Strategic Documents in the Field of Education. *Nepreryvnoe obrazovanie: XXI vek = Lifelong Education: The XXI Century*, 2015, no. 1 (9), pp. 1–25. (In Russian).

### Информация об авторах

*Берсенева Евгения Александровна* — доктор медицинских наук, руководитель Центра высшего и дополнительного профессионального образования, Национальный научно-исследовательский институт общественного здоровья им. Н.А. Семашко, г. Москва, Российская Федерация, e-mail: eaberseneva@gmail.com.

*Суходолов Александр Петрович* — доктор экономических наук, профессор, проректор по науке, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: science@bgu.ru.

*Седов Александр Анатольевич* — младший научный сотрудник, Национальный научно-исследовательский институт общественного здоровья им. Н.А. Семашко, г. Москва, Российская Федерация, e-mail: gd@seber.ru.

*Михайлов Дмитрий Юрьевич* — кандидат медицинских наук, докторант, Национальный научно-исследовательский институт общественного здоровья им. Н.А. Семашко, г. Москва, Российская Федерация, e-mail: mdudoc@mail.ru.

*Кудрина Валентина Григорьевна* — доктор медицинских наук, профессор, заслуженный врач Российской Федерации, заведующий кафедрой, кафедра медицинской статистики и цифрового здравоохранения, Российская медицинская академия непрерывного профессионального образования, г. Москва, Российская Федерация, e-mail: kudrinu@mail.ru.

*Андреева Татьяна Вадимовна* — кандидат педагогических наук, доцент, кафедра медицинской статистики и цифрового здравоохранения, Российская медицинская академия непрерывного профессионального образования, г. Москва, Российская Федерация, e-mail: tvandreeva@rambler.ru.

*Спасенникова Марина Геннадьевна* — кандидат медицинских наук, доцент, ведущий научный сотрудник, Национальный научно-исследовательский институт общественного здоровья им. Н.А. Семашко, г. Москва, Российская Федерация, e-mail: mspasennikova@gmail.com.

### Authors

*Evgeniya A. Berseneva* — Doctor habil. in Medicine, Head of Centre for Higher and Additional Professional Education, N.A. Semashko National Research Institute of Public Health, Moscow, Russian Federation, e-mail: eaberseneva@gmail.com.

*Alexander P. Sukhodolov* — Doctor habil. in Economics, Professor, Vice Rector for Research, Baikal State University, Irkutsk, Russian Federation, e-mail: science@bgu.ru.

*Alexander A. Sedov* — Junior Researcher, N.A. Semashko National Research Institute of Public Health, Moscow, Russian Federation, e-mail: gd@seber.ru.

*Dmitry Yu. Mikhailov* — Ph.D. in Medicine, Doctoral Student, N.A. Semashko National Research Institute of Public Health, Moscow, Russian Federation, e-mail: mdudoc@mail.ru.

*Valentina G. Kudrina* — Doctor habil. in Medicine, Professor, Honored Doctor of the Russian Federation, Head of Chair of Medical Statistics and Digital Public Health, Russian Medical Academy of Continuing Vocational Education, Moscow, Russian Federation, e-mail: kudrinu@mail.ru.

*Tatyana V. Andreeva* — Ph.D. in Pedagogy, Associate Professor, Russian Medical Academy of Continuing Vocational Education, Moscow, Russian Federation, e-mail: [tvandreeva@rambler.ru](mailto:tvandreeva@rambler.ru).

*Marina G. Spasennikova* — Ph.D. in Medicine, Associate Professor, Leading Researcher, N.A. Semashko National Research Institute of Public Health, Moscow, Russian Federation, e-mail: [mspasennikova@gmail.com](mailto:mspasennikova@gmail.com).

### Для цитирования

Цифровая медицина: информационно-технологические основы применения лексического анализа при формализации кодирования диагнозов / Е.А. Берсенева, А.П. Суходолов, А.А. Седов, Д.Ю. Михайлов, В.Г. Кудрина, Т.В. Андреева, М.Г. Спасенникова // *Baikal Research Journal*. — 2019. — Т. 10, № 4. — DOI: 10.17150/2411-6262.2019.10(4).8.

### For Citation

Berseneva E.A., Sukhodolov A.P., Sedov A.A., Mikhailov D.Yu., Kudrina V.G., Andreeva T.V., Spasennikova M.G. Digital Medicine: Information and Technological Bases for Using Lexical Analysis in Formalization of Diagnosis Coding. *Baikal Research Journal*, 2019, vol. 10, no. 4. DOI: 10.17150/2411-6262.2019.10(4).8.