

УДК 004.89  
DOI: 10.15827/0236-235X.127.512-517

Дата подачи статьи: 21.05.19  
2019. Т. 32. № 3. С. 512–517

## **Разработка базы данных и конвертера для извлечения и анализа специализированных данных, получаемых с медицинского аппарата**

А.П. Еремеев<sup>1</sup>, д.т.н., профессор, зав. кафедрой, [eremeev@aprrmat.ru](mailto:eremeev@aprrmat.ru)  
С.А. Ивлиев<sup>1</sup>, аспирант, [siriusfrk@gmail.com](mailto:siriusfrk@gmail.com)

<sup>1</sup> Национальный исследовательский университет  
«Московский энергетический институт», г. Москва, 111250, Россия

При разработке экспертных систем могут возникнуть затруднения, связанные с используемыми форматами хранения или обмена данными. Возможны ситуации, когда данные хранятся в закрытом формате либо закрытый формат имеют файлы обмена для таких систем. Это затрудняет автоматический анализ данных, поскольку приходится заносить их в экспертную систему вручную. Однако существуют методы, позволяющие преобразовывать данные в удобный для работы формат.

В статье рассматривается анализ двоичных файлов базы данных медицинского аппарата для исследования сложных патологий зрения с целью извлечения из нее данных биофизических исследований для последующего анализа. Поскольку в стандартном программном обеспечении отсутствуют возможности обмена информацией с внешними системами в открытых форматах, требуется разработка дополнительных методов и программных средств для определения физической структуры данных для последующего конвертирования в открытый формат.

Исходными данными для анализа являются информация о данных, хранящихся в базе данных медицинского аппарата, а также общие принципы физического представления данных в компьютерных системах. После определения структуры файлов с данными выполняется разработка конвертера. Выходные файлы конвертера могут быть использованы в дальнейшем при обучении нейронных сетей.

Такой подход позволяет достаточно быстро создавать базу образцов (прецедентов), исключив необходимость ручного переноса данных, и может служить основой для анализа данных в других подобных ситуациях.

**Ключевые слова:** анализ данных, представление данных, БД, закрытый формат, открытые данные, нейронные сети, двоичные файлы.

Важной частью любой системы (в статье рассматривается экспертная система) является подсистема хранения данных. В зависимости от важности и оперативности доступа к данным, значимости самих данных, используемых алгоритмов и методов работы в системе выбираются различные способы организации хранения данных. Это могут быть простые текстовые файлы, представляющие наборы пар вида <имя параметра, значение>, а также структурированные файлы как в текстовом, так и в двоичном виде. Могут использоваться файлы с разметкой типа XML или JSON [1, 2]. Нередки случаи смешанного (гетерогенного) хранения данных, например, заголовки для данных хранятся в одном формате, а сами данные – в другом. Все это обуславливает необходимость наличия (разработки), помимо стандартного ПО для хранения и оперирования данными, специализированного ПО – конвертера, который может быть использован для преобразования данных из одного формата в другой.

Такая ситуация сложилась из-за отсутствия единых подходов к организации хранения данных [1]. Простой пример – метод хранения числовых последовательностей. Если используется БД, то можно хранить числовые последовательности в отдельных таблицах, на что будет расходоваться дополнительная память и создаваться нагрузка на ядро СУБД при работе с ними. Другим подходом является сохранение (запись) числовой последовательности в некоторый массив (строковый или бинарный) в отдельное поле БД. По мере дальнейшего развития технологии хранения данных наблюдается переход к форматам, простым в понимании и содержащим как сами данные, так и их разметку. Легкость понимания обеспечивается наличием разметки в файлах и соглашения о формате представления тех или иных типов данных (строковые, целые, вещественные, логические, дата и время и т.д.) [3–5].

Однако довольно большое количество данных хранится в системах (обычно специализи-

рованных, в частности, медицинского назначения), в которых не уделяется должное внимание процессу хранения и организации данных, что приводит к сложности извлечения и анализа этих данных с применением современных компьютерных средств, в частности, экспертных систем поддержки принятия решений. Проблему можно решить через обращение к производителям соответствующих систем и их ПО, но это потребует довольно значительных временных и финансовых затрат, причем при условии, что разработчик согласится этим заниматься. Более того, нередко ситуации, когда используемый продукт на момент запроса снят с производства и не поддерживается разработчиком.

Альтернативным подходом является решение задачи так называемой обратной разработки формата данных и соответствующей БД с применением специализированного ПО, который будет далее проиллюстрирован на примере медицинского аппарата, используемого для диагностики сложных патологий зрения (патологий сетчатки).

### Описание ПО аппарата медицинской диагностики Tomey

Медицинский аппарат Tomey является компьютеризированной полуавтоматической системой диагностики, предназначенной для биофизических исследований зрения. Он позволяет получить данные о потенциалах, возникающих в глазу при разных световых стимулах, в виде *электроретинограмм* (ЭРГ) – специальных кривых для оценки функционального состояния сетчатки, получаемых при регистрации биопотенциалов, возникающих в ней при световом раздражении [6, 7]. На основании данных ЭРГ специалист по физиологии зрения (Tomey используется в МНИИ глазных болезней им. Г. Гельмгольца) может делать вывод о наличии или отсутствии патологий зрения и стадии заболевания и давать рекомендации врачу-офтальмологу. ПО, поставляемое с Tomey, позволяет осуществлять исследования и хранить данные о них в БД с возможностью вывода графического представления результатов исследований. Пример представления ЭРГ приведен на рисунке 1, где  $a_1$  и  $a_2$  – амплитуда  $a$ -волны;  $b_1$  и  $b_2$  – амплитуда  $b$ -волны;  $D$  – длительность  $b$ -волны;  $L$  – латентный период;  $T_b$  – время кульминации; по оси абсцисс представлена длительность волн ЭРГ, по оси ординат – амплитуда волн. Каждый компонент ЭРГ гене-

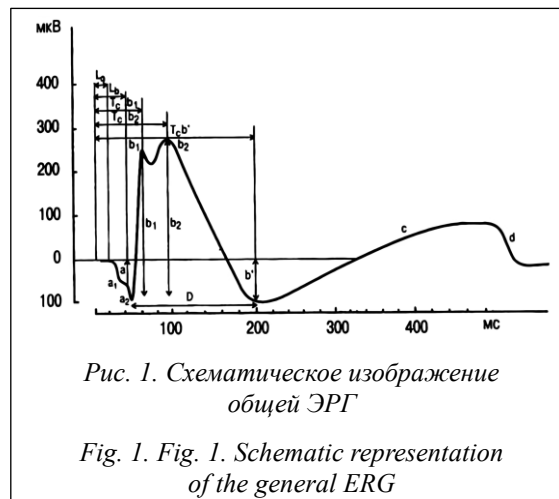


Рис. 1. Схематическое изображение  
общей ЭРГ

Fig. 1. Fig. 1. Schematic representation  
of the general ERG

рируется различными структурами сетчатки.

Аппарат Tomey предоставляет пользователю ряд экранов-окон для анализа ЭРГ (см. <http://www.swsys.ru/uploaded/image/2019-3/2019-3-dop/3.jpg>, <http://www.swsys.ru/uploaded/image/2019-3/2019-3-dop/4.jpg>).

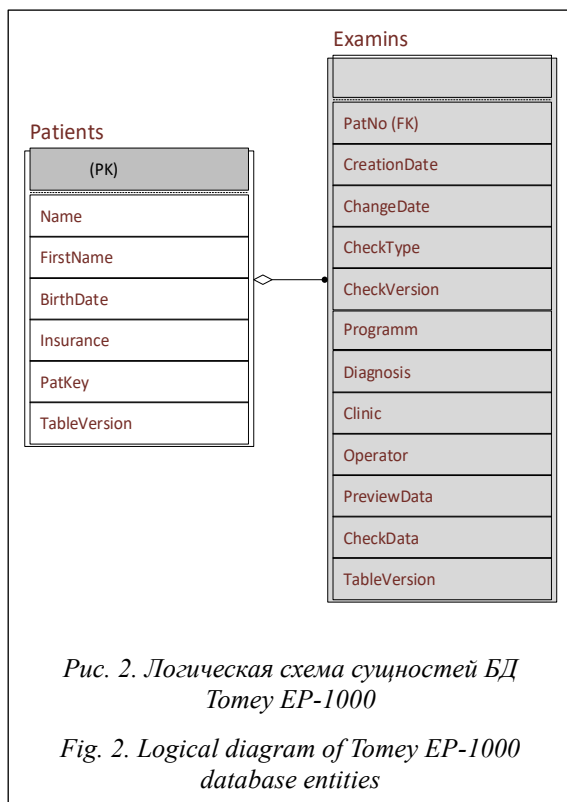
ПО Tomey предоставляет возможности для просмотра и создания отметок на результатах исследования, однако в нем отсутствуют функции для экспорта в форматы, которые были бы пригодны для взаимодействия с другими программными продуктами, например, экспертными системами.

В результате было решено разработать дополнительную БД для аппарата Tomey.

### Описание БД аппарата Tomey

БД Paradox, поставляемая в составе ПО Tomey EP-1000, реализована на ядре Borland Database Engine версии 5.0.1, собственно СУБД написана на языке Delphi. Средства разработки и драйверы чтения для данной БД обновлялись в последний раз в 1996 г., после чего разработка и поддержка данного продукта были прекращены.

БД имеет всего две таблицы – Examins и Patients. На рисунке 2 приведена логическая схема сущностей БД, которой соответствуют таблицы в БД. Ключевой сущностью является пациент, для каждого экземпляра сущности пациента может быть несколько экземпляров сущности исследования. Ключевой атрибут – числовой атрибут PatNo, входящий в состав сущности пациента. Наиболее значимыми являются данные, хранящие непосредственно результаты исследования. Это бинарные объекты (BLOB), хранящиеся в полях PreviewData и CheckData таблицы Examins. Необходимо реа-



лизовать ПО (конвертер) для корректного считывания информации из всех полей, а для этого провести анализ двоичных данных, хранящихся в этих полях [8, 9].

### Анализ двоичных данных

Файловая структура и архитектура БД Tomey достоверно неизвестны, поэтому будем исходить из предположения, что поля, хранящие двоичные данные об исследованиях, на самом деле не содержат эту информацию непосредственно, а ссылаются на некоторые внешние файлы или области внешних файлов. Действительно, в папке с БД есть два файла, имена которых соответствуют описанным выше таблицам: Examins.db и Patients.db. Файлы имеют значительный размер, поэтому будем считать, что они представляют информацию из полей таблицы.

Известно, что бинарные поля включены в состав сущности Examins, поэтому рассмотрим именно этот файл в побайтовом виде. Будем исходить из предположения, что файл разделен на записи, число которых равно числу исследований в БД. Известно, что БД с тестовым набором данных содержит 3 934 записи об исследованиях. Для анализа файла используем редактор файлов 010 Editor.

Просматривая файл, можно заметить, что он заполнен информацией неравномерно: пустые участки чередуются с участками, содержащими некоторую информацию. Предположим, что информативные участки и являются записями, хранящими данные об исследованиях. Поскольку считаем, что файл состоит из записей, выделим для всех информативных участков некоторую последовательность байтов, которая может быть заголовком для всех записей об исследованиях. В результате поиска (см. <http://www.swsys.ru/uploaded/image/2019-3/2019-3-dop/5.jpg>) удастся найти последовательность 0С 51 АВ 67 03 00 02 00 и определить число вхождений данной последовательности в файл. Число вхождений соответствует количеству записей об исследованиях в БД (3 934). Поскольку данная последовательность находится в начале каждой записи, будем считать ее заголовочной. Далее проанализируем байты, расположенные между найденными заголовками, то есть байты, образующие запись.

Будем исходить из предположения, что данные внутри каждой записи также упорядочены и структурированы, а результаты исследования хранятся в виде последовательностей дробных значений, то есть в виде float-значений. Удалось найти некоторый набор байтов, который делит запись на участки 0С 51 АВ 67, а так как значение типа float представляется четырьмя байтами, то это последовательности из четверок байтов. Заметим, что найденный набор не всегда относится к участкам записи, содержащим данные исследования. Поскольку результатом исследования является хорошо детализированный график, количество точек в нем велико, поэтому нас интересуют длинные промежутки между найденными наборами. Удалось найти еще несколько наборов байтов, наличие которых характерно для промежутков большой длины. Это наборы 0С 00 00 00 и 0D 00 00 00. При дальнейшем анализе найденных участков, содержащих последовательности дробных значений, удалось обнаружить, что следующие четыре байта после найденных наборов хранят целое число, отражающее количество идущих за ним дробных значений. В результате получена полная картина структуры бинарного файла, хранящего результаты исследований (рис. 3).

В итоге получен формат хранения бинарных данных результатов исследований, который может быть использован при реализации конвертера. Стоит отметить, что на данных в

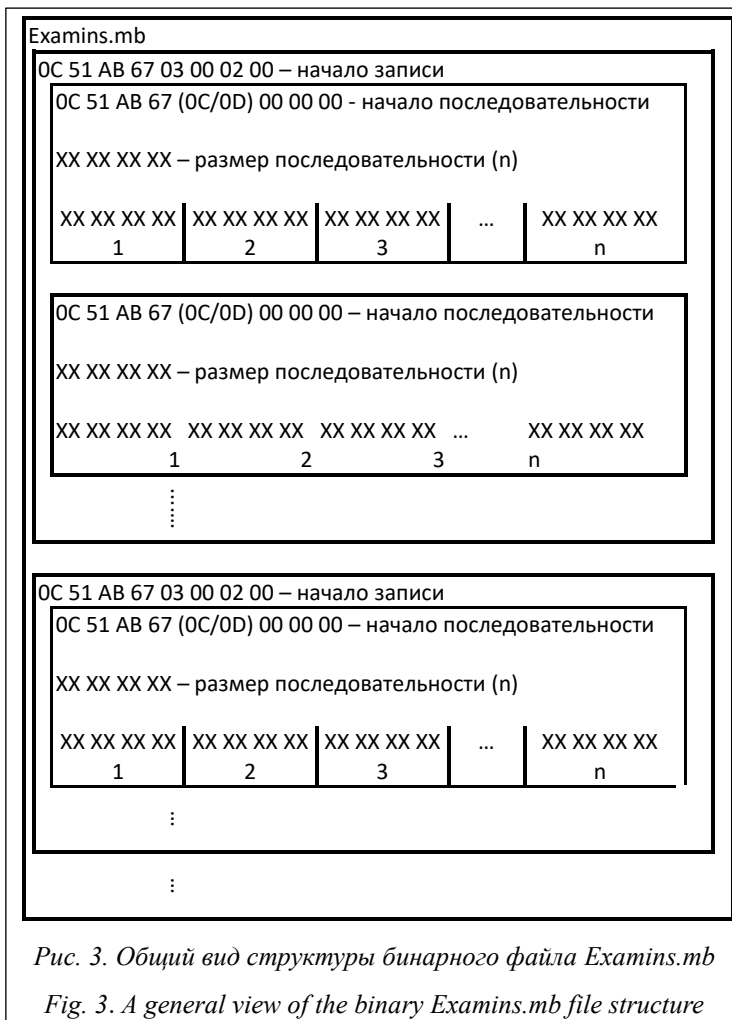


Рис. 3. Общий вид структуры бинарного файла Examins.mb  
 Fig. 3. A general view of the binary Examins.mb file structure

БД Tomeu не было никакой дополнительной защиты, что упростило анализ [10].

### Программная реализация конвертера

Реализованное в виде конвертера ПО позволяет устанавливать подключение к БД Tomeu и считывать всю информацию как с полей таблиц БД, так и из бинарного файла, структура которого была рассмотрена ранее. Имеется также возможность просмотра считанной информации. Весь основной функционал конвертера реализован в главном окне, изображенном на рисунке 4. Используются такие опции, как выпадающее меню, контекстное меню, кнопки и радио-группы.

Главное окно включает три основных блока:

- таблица с перечнем исследований;
- блок с просмотром графика исследования и подробными сведениями об исследовании;

- блок просмотра отладочной информации и переключения вариантов отображения графиков результата исследования.

В главном окне имеется возможность просмотра сведений о пациенте, к которому относится текущее исследование. Для этого необходимо нажать на кнопку Show patient info.

Разработано также меню экспорта для передачи результатов в формате CSV в другие программы, в которых возможно проведение интеллектуального анализа и диагностики с применением методов искусственного интеллекта, в частности, на основе байесовских сетей доверия и нейронных сетей.

### Заключение

В статье рассмотрена проблема хранения большого количества полезных данных в закрытых форматах, что тормозит разработку приложений (экспертных систем, систем поддержки принятия решений и т.д.), требующих обработки большого количества примеров. В качестве одного из методов решения данной проблемы предложен подход на основе разработки дополнительной БД и конвертера (так называемой обратной разработки БД для за-

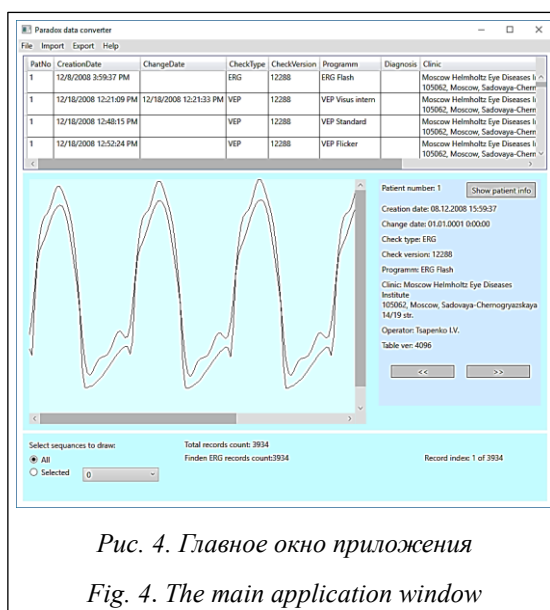


Рис. 4. Главное окно приложения  
 Fig. 4. The main application window

крытого формата данных) на примере аппарата биофизической диагностики зрения Tomeu, позволяющий получить данные в открытом формате для их последующего анализа и диагностирования проблемной ситуации (патоло-

гии зрения). Полученные результаты могут быть использованы для обучения нейронных сетей и для проведения аналогичных операций в других прикладных системах (например, для анализа кардиограмм).

*Работа выполнена при финансовой поддержке РФФИ (проекты №№ 17-07-00553 а, 18-01-00201 а, 18-51-00007 Бел а, 18-29-03088 МК).*

### Литература

1. Cleve A., Gobert M., Meurice L., Maes J., Weber J.H. Understanding database schema evolution: a case study. *Sci Comput Program* 97, 2015, no. P1, pp. 113–121. DOI: 10.1016/j.scico.2013.11.025.
2. Ma Z., Zhao Z., Yan L. Heterogeneous fuzzy XML data integration based on structural and semantic similarities. *Fuzzy Sets and Systems*, 2018, vol. 351, pp. 64–89.
3. Vinoth P. and Sankar P. Encoding of coordination complexes with XML. *J. Mol. Graphics Modell.*, 2017, vol. 76, pp. 242–259. DOI: <http://dx.doi.org/10.1016/j.jmgm.2017.07.009>.
4. Eito-Brun R. Chapter 4 – Databases for XML Data. *XML-based Content Management*. Chandos Publishing, 2018, pp 117–153. DOI: 10.1016/b978-0-08-100204-9.00004-4.
5. Mata C., Oliver A., Lalande A., Walker P., Martí J. On the use of XML in medical imaging web-based applications. *IRBM*, 2017, vol. 38, pp. 3–12. DOI: 10.1016/j.irbm.2016.10.001.
6. Eremeev A., Khasiev R., Tsapenko I., Zueva M. The intelligent decision support system for diagnostics of difficult diseases of vision. *IJ ICP*, 2014, vol. 1, no. 3, pp. 269–279.
7. Еремеев А.П., Ивлиев С.А. Построение онтологии на основе нереляционной базы данных для интеллектуальной системы поддержки принятия решений медицинского назначения // Программные продукты и системы. 2017. № 4. С. 5–14. DOI: 10.15827/0236-235X.030.4.005-014.
8. Панов А.С. Реверсинг и защита программ от взлома. СПб: БХВ-Петербург, 2006. 256 с.
9. Хогланд Г., Мак-Гроу Г. Взлом программного обеспечения: анализ и использование кода; [пер. с англ.]. М.: Вильямс, 2005. 400 с.
10. Altheide C., Carvey H. Chapter 8 – Digital Forensics with Open Source Tools. *File Analysis*, 2011, pp. 169–210. DOI: 10.1016/B978-1-59749-586-8.00001-7.

Software & Systems  
DOI: 10.15827/0236-235X.127.512-517

Received 21.05.19  
2019, vol. 32, no. 3, pp. 512–517

### Development of a database and a converter for retrieval and analysis of specialized data from a medical device

*A.P. Eremeev*<sup>1</sup>, *Dr.Sc.(Engineering), Professor, Head of the Chair, eremeev@appmat.ru*  
*S.A. Ivliev*<sup>1</sup>, *Postgraduate Student, siriusfrk@gmail.com*

<sup>1</sup>*National research University "MPEI", Moscow, 111250, Russian Federation*

**Abstract.** When developing expert systems, there may be difficulties associated with storage or data exchange formats. There may be situations when the data is stored in a proprietary format, or exchange files for such systems have proprietary format. This makes automated data analyze difficult, since they have to be manually entered into an expert system. However, there are methods that allow converting data into an easy-to-use format.

The paper considers the analysis of database binary files of a medical apparatus for studying complex vision impairment in order to extract biophysical studies data for further analysis. Since the standard software does not allow information exchange with external systems in open formats, it is necessary to develop additional methods and software to determine data physical structure for subsequent conversion to an open format.

The initial data for the analysis is information about what data are stored in a medical device database, as well as general principles of physical data representation in computer systems. Converter developing follows determining the structure of data files. Converter output files can be used in further neural network training.

This approach allows quick creating of a database of samples (precedents) eliminating the need for manual data transfer. The proposed approach can further serve as a basis for data analysis in other similar situations.

**Keywords:** data analysis, data presentation, database, closed format, open data, neural networks, binary files.

**Acknowledgments.** The work has been financially supported by RFBR (projects no. 17-07-00553, 18-01-00201, 18-51-00007, 18-29-03088).

### References

1. Cleve A., Gobert M., Meurice L., Maes J., Weber J.H. Understanding database schema evolution: a case study. *Sci. Comput. Program.* 2015, no. 97, pp. 113–121. DOI: 10.1016/j.scico.2013.11.025.
2. Ma Z., Zhao Z., Yan L. Heterogeneous fuzzy XML data integration based on structural and semantic similarities. *Fuzzy Sets and Systems.* 2018, vol. 351, pp. 64–89.
3. Vinoth P., Sankar P. Encoding of coordination complexes with XML. *J. Mol. Graphics Modell.* 2017, no. 76, pp. 242–259. DOI: <http://dx.doi.org/10.1016/j.jmgm.2017.07.009>.
4. Eito-Brun R. *Ch. 4 – Databases for XML Data, XML-based Content Management.* Chandos Publ., 2018, pp. 117–153. DOI:10.1016/b978-0-08-100204-9.00004-4.
5. Mata C., Olivera A., Lalande A., Walker P., Martí J. On the use of xml in medical imaging web-based applications. *IRBM.* 2017, vol. 38, iss. 1, pp. 3–12. DOI: 10.1016/j.irbm.2016.10.001.
6. Ereemeev A., Khasiev R., Tcapenko I., Zueva M. The intelligent decision support system for diagnostics of difficult diseases of vision. *Information Content & Processing.* 2014, vol. 1, no. 3, pp. 269–279.
7. Ereemeev A.P., Ivliev S.A. Ontology design based on non-relational database for intelligent decision support system for medical purposes. *Software & Systems.* 2017, no. 4, pp. 5–14. DOI: 10.15827/0236-235X.030.4.005-014.
8. Panov A.S. *Reversing and Protecting Programs from Hacking.* St. Petersburg, BHV-Peterburg Publ., 2006, 256 p.
9. Hoglund G., McGraw G. *Exploiting Software: How to Break Code.* Addison-Wesley Publ., 2004, 512 p. (Russ. ed.: Moscow, Vilyams Publ., 2005, 400 p.).
10. Altheide C., Carvey H. *Ch. 8 – File Analysis, Digital Forensics with Open Source Tools.* 2011, pp. 169–210. DOI: 10.1016/B978-1-59749-586-8.00001-7.

### Для цитирования

Еремеев А.П., Ивлиев С.А. Разработка базы данных и конвертера для извлечения и анализа специализированных данных, получаемых с медицинского аппарата // Программные продукты и системы. 2019. Т. 32. № 3. С. 512–517. DOI: 10.15827/0236-235X.127.512-517.

### For citation

Ereemeev A.P., Ivliev S.A. Development of a database and a converter for retrieval and analysis of specialized data from a medical device. *Software & Systems.* 2019, vol. 32, no. 3, pp. 512–517 (in Russ.). DOI: 10.15827/0236-235X.127.512-517.