УДК 004.056 Вестник УрФО № 2(44) / 2022, с. 49–56

Манжосов А.В., Болодурина И.П.

DOI: 10.14529/secur220207

МЕТОД АВТОМАТИЗИРОВАННОГО ПОСТРОЕНИЯ ГРАФА ЗНАНИЙ СВЯЗНОСТИ ФОРМАЛЬНЫХ МОДЕЛЕЙ НОРМ И ТРЕБОВАНИЙ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В последние десятилетия объем накопленной человечеством информации увеличился невероятно. Люди придумывают всё более оптимальные способы для систематизации, совместного использования и анализа большого объема данных с помощью традиционных алгоритмов и структур данных, но при этом они не предусматривают анализ естественного языка и зачастую не используют семантические связи.

Исходя из этого, назрела необходимость в таком анализе и представлении информации, которые позволяли бы с одной стороны хранить огромное количество объектов и отношений между ними, а с другой – предоставляло высокоскоростной доступ большому количеству пользователей к хранящимся данным, и, кроме того, сохраняло семантику. Одной из самых эффективных структур данных, позволяющей решать подобные задачи, являются графы и базы знаний, которые относительно недавно появились и стали предметом исследований в последние годы.

Ключевые слова: *граф знаний, онтология, анализ, нормативно-правовой акт, информационная безопасность.*

AUTOMATED CONSTRUCTION OF THE KNOWLEDGE GRAPH OF REFERENCE TO FORMAL MODELS OF NORMS AND TREATMENTS IN THE FIELD OF INFORMATION SECURITY METHOD

The amount of information accumulated by mankind has increased incredibly in recent decades. People are coming up with ever better ways to organize, share, and analyze large amounts of data using algorithms and data structures.

There is a need for such an analysis and presentation of information that would on the one hand, allow to store a huge number of objects and relationships between them, and on the other hand, provide high-speed access to a large number of users to the stored data, and, in addition, preserve the semantics. One of the most effective data structures that allows to solve such problems are graphs and knowledge bases, which have appeared relatively recently and have become the subject of research in recent years.

Keywords: *knowledge graph, ontology, analysis, normative act, information security.*

Введение

В этой статье рассматриваются все этапы реализации графа знаний, а также проблемы, с которыми, возможно, придется столкнуться при создании собственного экземпляра данной абстракции. Помимо этого, рассмотрены методы создания интерактивного визуального представления информации для ее эффективного хранения в графе, а также практические шаги по его реализации и использованию.

Разработку метода автоматизированного построения графа знаний связности формальных моделей норм и требований в области информационной безопасности предлагается разделить на четыре основных этапа.

Во-первых, анализ разработок в области визуализации нормативно-правовой базы, а также использования графовой модели в сфере информационной безопасности.

Во-вторых, исследование способов формального представления норм права и требований стандартов для использования компьютерных технологий в области права.

В-третьих, анализ норм права на предмет возможности создания их машиночитаемого представления и автоматизированного способа пополнения базы знаний.

В-четвертых, разработка инструментов для работы экспертов в области информационной безопасности над формальными моделями норм и требований.

Анализ исследований по способам визуализации нормативно-правовой базы, а также использованию графовой модели в сфере информационной безопасности

Авторы статьи [1] предлагают анализ существующих подходов представления знаний в виде графов. Исследование включает в себя извлечение отношений из текста, методы встраивания отношений для создания ссылок, обзор существующих графов знаний,

сравнение хранилищ графов знаний, импорт графов знаний. Статья полезна для первичного анализа работы с графами знаний, определения концепции построения, хранения и импорта графов знаний. При этом отмечается, что подходы, предложенные в статье, применимы для любой области знаний, т.е. универсальны.

Основная работа в статье [2] разделена на части. Обсуждается создание базы знаний по кибербезопасности в соответствии с трехэтапной процедурой и предлагается структура для создания базы знаний:

– во-первых, извлечение информации путем сбора и анализа структурированных и неструктурированных данных;

– во-вторых, построение онтологии в соответствии с полученной информацией компьютерных атак;

– в-третьих, метод машинного обучения для извлечения объектов, связанных с компьютерной атакой.

Данное исследование [3] описывает создание онтологии по информационной безопасности в рамках прецедентного подхода к описанию компьютерных угроз. Разрабатываемая онтология достаточно полно описывает предметную область компьютерных угроз и является концептуальной. Авторы предложили способ создания на основе онтологии базы данных основных известных прецедентов компьютерных атак и нарушений информационной безопасности, с их описанием, методами идентификации и способами устранения последствий.

В статье [4] авторы предлагают новую общую структуру защиты промышленной управляющей сети на основе данных. С целью повышения качества анализа отношения сущностей в данных о кибербезопасности, авторами предлагается прототип новой модели извлечения отношений кибербезопасности ResPCNN-ATT.

Авторами [5] разработана модель информационной безопасности баз знаний, включающая следующие составляющие: функции защиты информации в базах знаний; методы защиты баз знаний, технико-экономические показатели методов защиты информации в базах знаний.

По сравнению с представленными работами, в статье предлагается решение нескольких задач:

– исследование способов формального представления норм права и требований стандартов для использования компьютерных технологий в области права;

– разработка формального представления предметной области;

– анализ норм права, на предмет возможности создания их машиночитаемого представления и автоматизированного способа пополнения базы знаний;

– разработка инструментов для работы экспертов в области информационной безопасности над формальными моделями норм и требований.

Описание способа формального представления норм права и требований стандартов для использования компьютерных технологий в области права

Наиболее часто применяемым инструментом при визуализации реляционных типов данных является графовая модель, которая также применяется и в других направлениях визуализации. Модель представляет не только данные, но и отношения между данными. Графовая модель является структурой данных, состоящей из связанных сущностей. Представление информации в виде графа знаний не ново, но в последнее время оно приобрело популярность благодаря его использованию в приложениях искусственного интеллекта. Ниже представлены основные термины и определения, использующиеся в статье.

Граф знаний – это семантическая сеть, в которой хранятся сущности и отношения между сущностями в виде графа. Преимущество использования графа знаний – это возможность обработки большего количества запросов, чем у традиционных методов хранения. Граф знаний – гибкая форма хранения и визуализации данных, которые легко обновить.

Ключевое слово — особо важное, общепонятное, емкое и показательное слово в тексте, набор которых может дать высокоуровневое описание его содержания.

Нормативно-правовой акт – концепт, соответствующий официальному документу, изданному в установленном порядке органом государственной власти, органом местного самоуправления или должностным лицом, и содержащий правовые нормы.

Онтология – классы, свойства классов, правила, которые в совокупности отражают формальную концепцию предметной области; попытка всеобъемлющей и подробной формализации некоторой области знаний с помощью концептуальной схемы.

Отношение – связь между сущностями, являющаяся ребром графа знаний.

Ссылка – связь, ведущая от фрагмента текста одного нормативно-правового акта к другому нормативно-правовому акту.

Сущность – нормы права, являющиеся узлами (вершинами) графа знаний.

Тип отношения – способ указать, что означает отношение между сущностями, семантику соответствующего отношения.

Сегодня термины, применяемые при построении баз и графов знаний, употребляются в различных контекстах. Моделирование и формальное представление схемы данных в виде баз и графов знаний обеспечивают гораздо большие возможности, чем традиционные базы данных или объектно-ориентированный подход.

Разработка формального представления предметной области

В этом разделе представлено формальное представление онтологии графа знаний нормативно-правовой базы в сфере информационной безопасности.

Согласно определению, формализация онтологии – процесс выражения концептуализации предметной области в соответствии с парадигмой представления знаний, предлагаемой языком моделирования.

Формализация онтологии проведена в соответствии с условиями:

– онтология является одним из инструментов, необходимых для моделирования предметной области;

– онтология содержит перечень ключевых слов данной предметной области;

– знание о смысле ключевых слов, представленное онтологией, должно быть очевидным для любого эксперта в данной предметной области.

Цель онтологии – повысить количество и качество описываемых общих свойств предметной области, не зависящих от ее конкретных реализаций.

Формальной онтологией предметной области *EA* называется пара <*e, a*>, где *e* – множество ключевых слов предметной области, а *a* – множество аналитических предложений, отношений, ссылок, описывающее смысл данных ключевых понятий.

Онтология предметной области включает в себя:

– словарь ключевых понятий, использующихся в данной предметной области;

– совокупность отношений, обеспечивающих корректную интерпретацию понятий и их правильное использование.

Представление знаний в виде онтологий применяется ради семантической интеграции информационных ресурсов, корректной интерпретации содержания или названия текстовых документов, представленных с помощью естественного языка. На основе онтологий разрабатываются базы знаний и графы знаний.

Поэтому всё множество *S* предложений, которые являются верными в предметной области *EA*, будем называть теорией предметной области *EA*.

Тогда, если *E*, *A* – формальная онтология предметной области *EA*, а *S* – теория предметной области *EA*. Тогда каждый элемент *E* является элементом *S*.

Примем во внимание, что если некоторое утверждение базы данных не является истинным, то утверждение принимается ложным. В то время, как для базы знаний в этом случае такое утверждение является ни истинным, ни ложным. Это свойство существенно влияет на то, какие факты считаются логически следующими из заданной базы знаний, и на понятие логического следования в эти факты базы знаний.

Анализ норм права, на предмет возможности создания их машиночитаемого представления и автоматизированного способа пополнения базы знаний

Законодательная база Российской Федерации содержит множество нормативно-правовых актов в сфере информационной безопасности. Анализ нормативно-правовых актов свободного доступа позволит сформировать базу знаний. Алгоритм анализа нормативно-правового акта представлен в виде алгоритма на рисунке 1.

Описание основных этапов алгоритма представлены ниже.

1. Добавление нормативно-правовых актов.

На данном этапе происходит добавление пользователем нормативно-правовых актов, которые в дальнейшем будут считаться нормативно-правовой базой, и на основании которых предлагается строить граф знаний. Добавление НПА происходит посредством выбора директории с необходимым набором документов текстового формата.

2. Обработка заголовков НПА.

В загруженных документах определяются их заголовки и выносятся в отдельный мас-

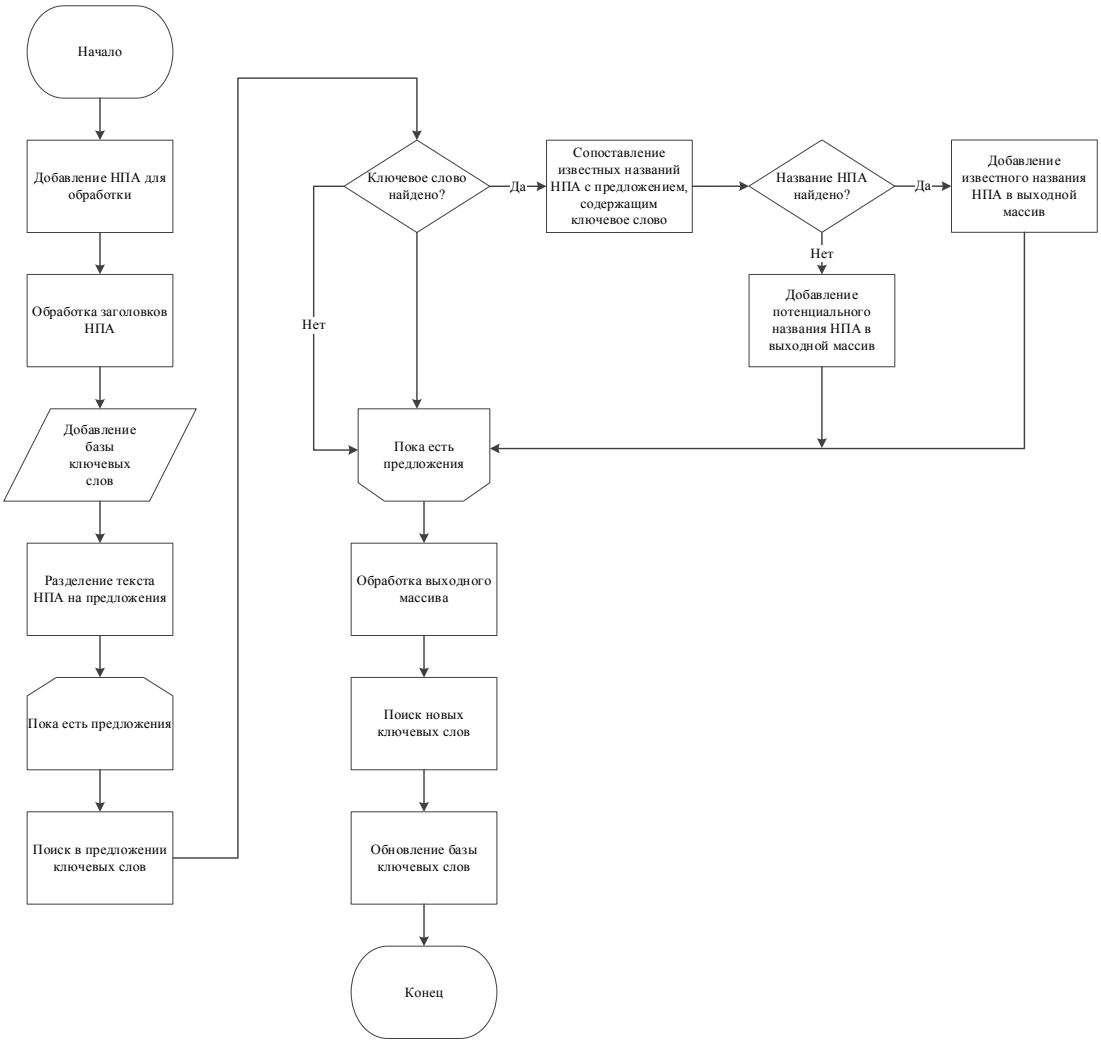

Рис. 1. Алгоритм анализа нормативно-правового акта

сив. При этом подразумевается, что файлы, найденные в каталоге, могут иметь, а могут не иметь соответствующие шаблону наименования норм законодательства названия (%тип документа% %издатель% от %дата% %номер% %название%). Набор ключевых слов определяется исходя из шаблонов наименования норм законодательства в полуавтоматическом режиме.

3. Разделение текста НПА на предложения.

В данном цикле осуществляется основная работа по определению связей между НПА следующим образом: осуществляется поиск по всем предложениям на предмет явного содержания в предложении или предложений, потенциально содержащих ключевые слова заголовков НПА.

4. Обработка выходного массива.

Массив полученных из предложений названий НПА (ссылок на документы) соотносится с заголовками из текущего множества загруженных НПА. Следующим шагом выявляются совпадения между полученной ссылкой и заголовком с помощью ключевых слов. Найденные ключевые слова или потенциальные ключевые слова добавляются в массив, который в последствии редактируется.

5. Поиск новых ключевых слов.

Пользователю предлагается анализировать потенциальные ключевые слова на предмет добавления их в базу ключевых слов для последующего использования при обработке НПА. Результатом анализа пользователя на этом этапе является Обновленная база ключевых слов.

На практике извлечение сущностей заключается в извлечении из текста нормативно-правового акта ссылок на другие нормативно-правовые акты. Процесс построения

графа знаний состоит из извлечения сущностей и их взаимоувязки отношениями с базой нормативно-правовых актов через онтологию. Это сформирует граф знаний и предоставит возможность экспертам в области ИБ проводить рассуждения о связности документов. Разработанную онтологию также можно связать с внешними знаниями и онтологиями.

Основной подход к извлечению отношений основан на техниках обработки естественного языка, включая тегирование части речи, синтаксический анализ, распознавание именованных сущностей.

Онтологии представляются в виде в графов знаний, чтобы визуализировать взаимосвязанные отношения между сущностями области знания. Графы знаний – наилучшие средства представления сущностей и отношений между ними. Онтология же состоит из набора классов с атрибутами и отношениями.

База знаний представляет собой семантический граф знаний, описывающий семантику источников информации. Таким образом, получившаяся онтология предлагает для аналитиков и профессионалов в данной области инструмент для анализа. Граф знаний является результатом связывания области знания и модели представления данных, а именно онтологии. В нашем случае граф знаний является связностью нормативно-правовых актов, извлеченных в специальную онтологию.

Разработка инструментов для работы экспертов в области информационной безопасности над формальными моделями норм и требований

Реализацией предложенных алгоритмов и формализованных представлений является разработанная программа для ЭВМ [6]. Программа является инструментом для работы экспертов в области информационной безопасности над формальными моделями норм и требований в области законодательства РФ. На рисунке 2 представлен интерфейс программы для ЭВМ.

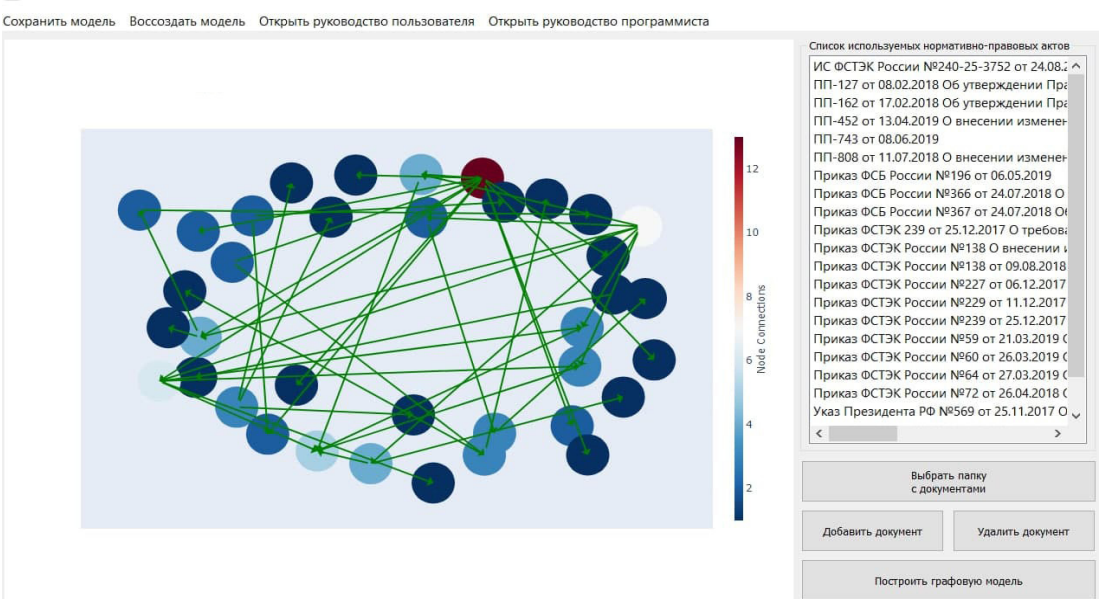

Рис. 2. Интерфейс программы для ЭВМ

Интерфейс состоит из:

– элементов управления;

– зоны управления загруженными нормативно-правовыми актами;

– области визуализации отношений сущностей нормативно-правовой базы.

Элементы управления позволяют пользователю определить входные данные для работы программы, корректировать состав НПА, ознакомиться с руководствами пользователя и администратора, сохранить или восстановить граф знаний.

Зона управления загруженными НПА отражает состав документов, выбранных пользователем в качестве входных данных. Перечень документов, отображаемых в этой зоне, имеет функционал интерактивного взаимодействия с областью визуализации отношений сущностей нормативно-правовой базы.

Область визуализации отношений сущ-

ностей нормативно-правовой базы является основным полем работы пользователя с программой для ЭВМ. Область визуализации отношений отражает связь всех обрабатываемых пользователем НПА. Визуализация отношений реализована в виде графа знаний и является интерактивной. Пользователю доступен функционал перемещения вершин графа, взаимодействия с вершинами и ребрами графа для получения дополнительной информации. Визуально количество упоминаний документа в других НПА отличается цветом. В области визуализации отношений представлена шкала, демонстрирующая соотношение количества упоминаний документа в других НПА (число связных ребер графа знаний с конкретной вершиной) и цветового оттенка.

Целью применения средства автоматизированного построения нормативно-правовой базы является снижение трудоемкости задачи анализа требований нормативно-правовой базы. Анализ требований нормативно-правовой базы проводится аналитиками информационной безопасности периодически.

Заключение

Представленные разработки предлагают метод автоматизированной визуализации отношений между нормативно-правовыми актами посредством построения графа знаний, что позволяет ускорить анализ информации специалистом в сфере ИБ. Основным преимуществом предложенного метода является автоматизированный способ пополнения базы ключевых слов, а также скорость анализа данных, поскольку автоматизированная реализация способна в десятки раз сократить время, затрачиваемое на ручной поиск, анализ и систематизацию информации, изложенной в нормативно-правовой базе. Программная реализация предложенного метода зарегистрирована в государственном фонде электронных ресурсов. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-01065, а также гранта Президента Российской Федерации для государственной поддержки ведущих научных школ Российской Федерации (НШ-2502.2020.9).

Литература

1. Гурин В. С., Костров Е. В., Гавриленко Ю. Ю., Саада Д. Ф., Ильюшин Е. А., Чижов И. В. Представление знаний в виде графа: основные технологии и подходы // Современные информационные технологии и ИТ-образование. 2019. Т. 15, № 4. С. 932-944. DOI: 10.25559/SITITO.15.201904.932-944.

2. A Practical Approach to Constructing a Knowledge Graph for Cybersecurity Yan Jia, Yulu Qi, Huaijun Shang, Rong Jiang, Aiping Li https://doi.org/10.1016/j.eng.2018.01.004.

3. Мирзагитов А. А., Пальчунов Д. Е. Методы разработки онтологии по информационной безопасности, основанные на прецедентном подходе // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, вып. 3. С. 37–46.

4. Guowei Shen, Wanling Wang, Qilin Mu, Yanhong Pu, Ya Qin, and Miao Yu Data-Driven Cybersecurity Knowledge Graph Construction for Industrial Control System Security doi.org/10.1155/2020/8883696.

5. Рихтер Т.В., Абрамова И.В. Разработка модели информационной безопасности баз знаний. Физико-математическое образование. 2020. Выпуск 1(23). С. 106-110. DOI 10.31110/2413-1571-2020-023-1-017.

6. Свидетельство 2021666949 «Автоматизированное построение графовой модели нормативно-правовой базы»: программа для ЭВМ / А.В. Манжосов, И.П. Болодурина, В.Д. Родионов, М.С. Гнамм (RU); правообладатель А.В. Манжосов. заявл. от 13.09.2021. опуб. 21.10.2021.

References

1. Gurin V. S., Kostrov Ye. V., Gavrilenko YU. YU., Saada D. F., Il'yushin Ye. A., Chizhov I. V. Predstavleniye znaniy v vide grafa: osnovnyye tekhnologii i podkhody // Sovremennyye informatsionnyye tekhnologii i IT-obrazovaniye. 2019. T. 15, № 4. S. 932-944. DOI: 10.25559/SITITO.15.201904.932-944.

2. A Practical Approach to Constructing a Knowledge Graph for Cybersecurity Yan Jia, Yulu Qi, Huaijun Shang, Rong Jiang, Aiping Li https://doi.org/10.1016/j.eng.2018.01.004.

3. Mirzagitov A. A., Pal'chunov D. Ye. Metody razrabotki ontologii po informatsionnoy bezopasnosti, osnovannyye na pretsedentnom podkhode // Vestn. Novosib. gos. un-ta. Seriya: Informatsionnyye tekhnologii. 2013. T. 11, vyp. 3. S. 37–46.

4. Guowei Shen, Wanling Wang, Qilin Mu, Yanhong Pu, Ya Qin, and Miao Yu Data-Driven Cybersecurity Knowledge Graph Construction for Industrial Control System Security doi.org/10.1155/2020/8883696.

5. Rikhter T.V., Abramova I.V. Razrabotka modeli informatsionnoy bezopasnosti baz znaniy. Fiziko-matematicheskoye obrazovaniye. 2020. Vypusk 1(23). S. 106-110. DOI 10.31110/2413-1571-2020-023-1-017.

6. Svidetel'stvo 2021666949 «Avtomatizirovannoye postroyeniye grafovoy modeli normativno-pravovoy bazy»: programma dlya EVM / A.V. Manzhosov, I.P. Bolodurina, V.D. Rodionov, M.S. Gnamm (RU); pravoobladatel' A.V. Manzhosov. zayavl. ot 13.09.2021. opub. 21.10.2021.

МАНЖОСОВ Артём Владимирович, аспирант кафедры прикладной математики Факультета математики и информационных технологий Федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет». 460018, Оренбург, пр. Победы, 13. E-mail: a.v.manzhosov@gmail.com

БОЛОДУРИНА Ирина Павловна, профессор, доктор технических наук, заведующий кафедрой прикладной математики Факультета математики и информационных технологий Федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет». 460018, Оренбург, пр. Победы, 13. E-mail: ipbolodurina@yandex.ru

MANZHOSOV Artyom Vladimirovich, post-graduate student of the Department of Applied Mathematics, Faculty of Mathematics and Information Technologies, Federal State Budgetary Educational Institution of Higher Education "Orenburg State University". 460018, Orenburg, Pobedy Ave., 13. E-mail: a.v.manzhosov@gmail.com, +79228133818.

BOLODURINA Irina Pavlovna, Professor, Doctor of Technical Sciences, Head of the Department of Applied Mathematics, Faculty of Mathematics and Information Technologies, Federal State Budgetary Educational Institution of Higher Education "Orenburg State University". 460018, Orenburg, Pobeda ave., 13. E-mail: ipbolodurina@yandex.ru