

ВЕСТНИК НОВОСИБИРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

Научный журнал
Основан в ноябре 1999 года

Серия: Информационные технологии

2020. Том 18, № 1

СОДЕРЖАНИЕ

<i>Демин В. О., Пищик Б. Н.</i> Моделирование обмена транзакционными данными в корпоративных системах	5
<i>Евменов Н. Д., Воронин В. В., Дядьков П. Г., Позднякова О. А.</i> Оценка точности определения параметров погребенных археологических объектов по данным разноточной магнитной съемки на основе численного моделирования	16
<i>Курганов Д. В.</i> Об одном методе классификации нефтяного месторождения с использованием комплекса геолого-промысловых данных машинного обучения	27
<i>Лапин В. Н.</i> Модель распространения трещины вдоль гидроизоляции скважины	36
<i>Лукашов О. Ю., Костромин Д. А.</i> Технология построения локально детализированных разностных сеток для расчета зон поражения при взрывах газа и пыли в угольных шахтах	50
<i>Нечта И. В.</i> Построение хрупкого цифрового водяного знака, применяемого коллективом авторов в исполняемых файлах	65
<i>Соловьев Ф. Н.</i> Автоматическая обработка текстов на основе платформы ТХМ с учетом анализа структурных единиц текста	74
<i>Щеглов Б. О., Безуленко Н. И., Атащиков С. А., Щеглова С. Н.</i> Виртуальный атлас персонифицированной анатомии человека «SkiaAtlas» и возможности его применения	83
Информация для авторов	94

VESTNIK

NOVOSIBIRSK STATE UNIVERSITY

Scientific Journal
Since 1999, November
In Russian

Series: Information Technologies

2020. Volume 18, № 1

CONTENTS

<i>Demish V. O., Pischik B. N.</i> Modeling of Transaction Data Exchange in Enterprise Systems	5
<i>Eymenov N. D., Voronin V. V., Dyadkov P. G., Pozdnyakova O. A.</i> Estimation of Accuracy of Determination of Parameters of the Buried Archaeological Objects According to the Data of Different-Altitude Magnetic Survey on the Basis of Numerical Modeling	16
<i>Kurganov D. V.</i> Oil Reservoir Classification by Geological and Production Data Using Un-supervised Machine Learning Algorithm	27
<i>Lapin V. N.</i> Model of Fracture Propagation in Hydraulic Isolation of the Wellbore	36
<i>Lukashov O. Yu., Kostromin D. A.</i> Technology of Building Locally Detailed Differential Grids for the Calculation of Affected Areas at Gas and Dust Explosions in Coal Mines	50
<i>Nechta I. V.</i> Construction Fragile Digital Watermark Used by a Team of Authors in Executable Files	65
<i>Soloviev F. N.</i> Embedding Additional Natural Language Processing Tools into the TXM Platform	74
<i>Shcheglov B. O., Bezulenko N. I., Atashchikov S. A., Scheglova S. N.</i> Virtual Atlas of Per-sonified Human Anatomy “SkiaAtlas” and the Possibility of Its Application	83
Instructions to Contributors	94

Editor in Chief M. M. Lavrentiev

Vice-Editor A. V. Avdeev

Executive Secretary D. P. Iksanova

Editorial Board of the Series

- I. V. Bychkov*, professor, academician (Irkutsk), *B. M. Glinsky*, professor (Novosibirsk)
A. N. Gorban, professor (Lester, GB), *E. P. Gordov*, professor (Tomsk)
B. S. Dobronets, professor (Krasnoyarsk), *A. M. Elizarov*, professor (Kazan)
G. N. Erokhin, professor (Kaliningrad), *A. I. Kamyshnikov*, professor (Khanty-Mansiysk)
G. P. Karev, professor (Maryland, USA), *N. A. Kolchanov*, professor, academician (Novosibirsk)
M. M. Lavrentjev, professor (Novosibirsk), *V. E. Malyshkin*, professor (Novosibirsk)
N. N. Mirenkov, professor (Aizu, Japan), *N. M. Oskorbin*, professor (Barnaul)
D. E. Palchunov, professor (Novosibirsk), *T. Pizansky*, professor (Ljubljana, Slovenia)
V. P. Potapov, professor (Kemerovo), *O. I. Potaturkin*, professor (Novosibirsk)
V. A. Serebryakov, professor (Moscow), *A. V. Starchenko*, professor (Tomsk)
S. I. Smagin, professor, corresponding member of RAS (Khabarovsk)
D. A. Tusupov, professor (Astana, Kazakhstan)
V. V. Shajdurov, professor, corresponding member of RAS (Krasnoyarsk)
Yu. I. Shokin, professor, academician (Novosibirsk)

*The journal is published quarterly in Russian since 1999
by Novosibirsk State University Press*

The address for correspondence

Faculty of Information Technologies, Novosibirsk State University

1 Pirogov Street, Novosibirsk, 630090, Russia

Tel. +7 (383) 363 42 46

E-mail address: inftech@vestnik.nsu.ru

On-line version: <http://elibrary.ru>

Моделирование обмена транзакционными данными в корпоративных системах

В. О. Демиш, Б. Н. Пищик

*Новосибирский государственный университет
Новосибирск, Россия*

*Институт вычислительных технологий СО РАН
Новосибирск, Россия*

Аннотация

Анализируются проблемы проектирования корпоративных систем для автоматизации предприятий с несколькими разделами и видами учета. Решается частная задача взаимодействия оперативного и бухгалтерского контуров учета на примере работы с давальческим сырьем. Обозначена необходимость создания модели взаимодействия блоков учета, обеспечивающих корректность работы с транзакциями уровня бизнес-логики. Предложена соответствующая модель, позволяющая спроектировать интеграционное решение на базе нескольких подсистем с описанием обмена транзакционными данными.

Ключевые слова

разработка корпоративных систем, интеграция данных, автоматизация предприятий

Для цитирования

Демиш В. О., Пищик Б. Н. Моделирование обмена транзакционными данными в корпоративных системах // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 5–15. DOI 10.25205/1818-7900-2020-18-1-5-15

Modeling of Transaction Data Exchange in Enterprise Systems

V. O. Demish, B. N. Pischik

*Novosibirsk State University
Novosibirsk, Russian Federation*

*Institute of Computational Technologies SB RAS
Novosibirsk, Russian Federation*

Abstract

This paper analyzes the problems of designing enterprise systems for solving problems with several sections and types of accounting. The particular problem of the interaction of the operational and accounting circuits is solved by the example of work with tolling raw materials.

The necessity of creating a model for the interaction of different subsystems that ensure the correctness of work with transactions at the level of business logic is indicated. An appropriate model is proposed that allows design integration solutions based on several subsystems with a description of the exchange of transactional data.

Keywords

enterprise systems design, data integration, enterprise automation

For citation

Demish V. O., Pischik B. N. Modeling of Transaction Data Exchange in Enterprise Systems. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 5–15. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-5-15

Введение

Вне зависимости от направления деятельности (торговля, производство, оказание услуг и др.), организациям необходимо вести бухгалтерский и налоговый учет. Для отражения информации о хозяйственной деятельности экономического субъекта в Российской Федерации оба вида учета закреплены законодательно в Федеральном законе № 402-ФЗ и налоговом кодексе. Общие положения налогового учета изложены в статье 313 НК РФ.

С другой стороны, руководителям компаний-холдингов требуется вести корпоративный учет, рассматривая несколько юридических лиц как единое целое. При этом в любом случае требуется бухгалтерская и налоговая отчетность по каждому юридическому лицу. Так появляется еще один вид учета – управленческий [1].

Одна и та же операция при отражении в разных видах учета требует разной степени детализации. Например, фиксация данных о приобретенных товарах может быть отражена так:

- БУ (бухгалтерский учет): общее количество + общая сумма;
- НУ (налоговый учет): общая сумма + сумма НДС;
- УУ (управленческий учет): детальное количество, цены, размещение по складам.

Для обеспечения естественного требования единовременной фиксации (без дублирования ввода данных) каждой операции данные всех видов учета должны быть взаимосвязаны.

Работы в области экономических наук ориентированы на выстраивание наиболее эффективного учета на предприятиях. Это порождает различные требования к средствам автоматизации, которые должны быть предусмотрены разработчиками программного обеспечения.

В тексте работы будем использовать следующие термины предметной области.

Раздел учета – выделенная функциональная часть программы, которая предназначена для решения определенного фиксированного круга задач. Например: Продажи, Закупки, Производство, Кадры и т. п.

Вид учета (в литературе можно также встретить термин «контур учета») – замкнутая область данных, в которой отражаются операции предприятия. Например: оперативный, бухгалтерский, налоговый контуры.

Возможные сочетания разделов и видов учета на предприятиях:

Раздел учета / Вид учета	БУ	НУ	УУ
Планирование	–	–	+
CRM	–	–	+
Продажи	+	+	+
Закупки	–	–	++
Склад	+	–	++
Кадры	+	–	+
Заработная плата	+	+	+
Регламентированная отчетность	+	+	–
...	...	...	...

Состояние вида учета в сочетании с разделом учета в данной таблице обозначено различными знаками:

- не ведется;
- ведется с минимальной детализацией;
- ++ ведется с подробной детализацией.

Деление крайне условно и разным предприятиям будут соответствовать разные комбинации подобных вариантов.

В рамках автоматизации хозяйственной деятельности организации важно отметить, что каждая ячейка рассмотренной таблицы (назовем ее блоком учета) – это отдельная предмет-

ная область, которая пересекается с другими, но при этом может быть автоматизирована посредством отдельной информационной системы. В общем случае для полного покрытия всех проблем автоматизации потребуется интеграционное решение, в котором решены задачи всех видов и разделов учета.

В данной работе рассматриваются проблемы реализации бизнес-логики (в соответствии с терминологией М. Фаулера в [2]) в корпоративных системах автоматизации хозяйственной деятельности предприятий, основанных на интеграции подсистем, работающих с различными блоками учета.

Данные, с которыми осуществляется работа на уровне бизнес-логики, удобно разделить на четыре группы [3]:

- метаданные – описание структур данных;
- НСИ (нормативно-справочная информация) – справочники, классификаторы и др. (практически не изменяются во все время жизни информационной системы);
- мастер-данные – это условно-постоянные данные, определяющие состав автоматизируемой предметной области и основу для описания бизнес-логики приложения;
- транзакционные (или операционные) данные отображают факты выполнения бизнес-операций в системе (примеры: выполнение платежа, поступление товаров, начисление заработной платы).

Полнота интеграционного решения рассматриваемой проблемы заключается в обеспечении единого справочника НСИ и мастер-данных для всех информационных систем, а также в корректном отражении операционных данных в каждой из них.

В работе рассмотрено решение такой проблемы на примере, включающем несколько блоков учета.

Пример сквозного учета

Один из содержательных примеров взаимодействия нескольких блоков учета – переработка давальческого сырья.

Для построения примера достаточно рассмотреть две организации (А и Б), занимающиеся торговлей. При этом в организации А есть блок производства, а организация Б самостоятельного производства не имеет, но использует схему переработки сырья на стороне с помощью организации А.

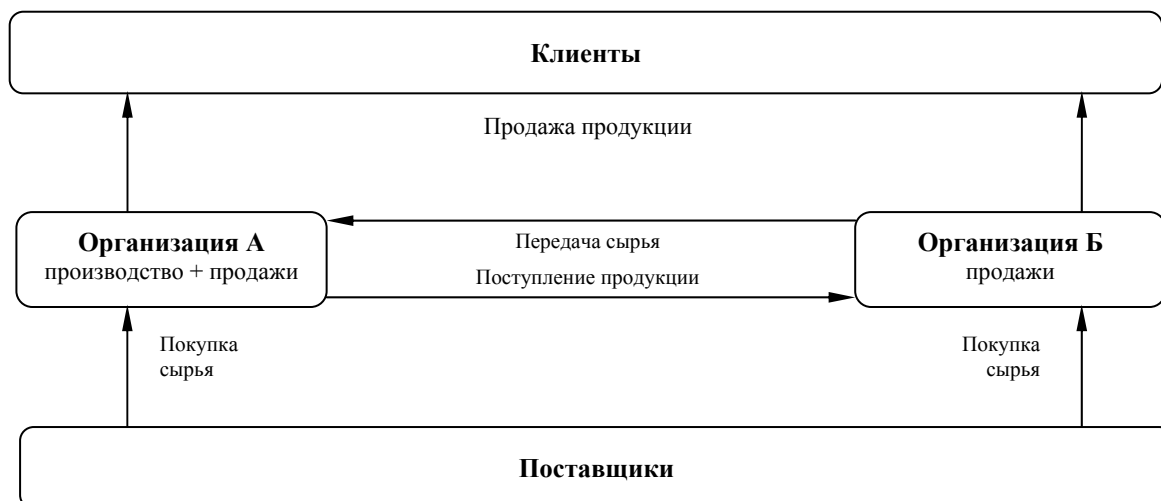


Рис. 1. Схема работы с давальческим сырьем
Fig. 1. Illustration of work with tolling raw materials

Такая схема работы – это один из видов *оптимизации налоговой нагрузки* [4], т. е. имеется экономически обоснованная необходимость автоматизации такой схемы работы. Пример показателен тем, что подсистемы бухгалтерского и управленческого учета рассматривают в нем бизнес-процессы по-разному. Для бухгалтерского учета важно четкое определение «хозяина» сырья и готовой продукции. В том случае, если организация А производит продукцию для организации Б, эта продукция будет учтена на так называемых забалансовых счетах и не будет числиться на балансе организации А. Но для руководителей отдела производства в данном случае важнее будет актуальный остаток произведенной продукции.

Также возможен сценарий, когда на этапе производства продукции из сырья еще не ясно, какой организацией она будет в дальнейшем реализована. Иначе говоря, организация А может произвести продукцию из собственного сырья, но в дальнейшем эту продукцию будет выгоднее реализовать от лица организации Б. И тогда нужно будет оформить производство в организации А через давальческую схему, а не из собственного сырья.

Какой вариант автоматизации такого бизнес-процесса будет наиболее оптимальным? Как предусмотреть возможность изменения данных «задним числом»? Как реализовать возможность однократного ввода данных лишь в одну из систем, если данные должны быть зафиксированы и в других системах?

Полные ответы требуют детального рассмотрения. В настоящей работе вместо этого предлагается провести обобщение примера до более широкого класса проблем и провести его анализ вне конкретной интерпретации.

Похоже, это один из тех примеров бизнес-логики, о которых упоминал М. Фаулер [2], называя их нелогичными по своей природе. Обилие таких примеров также присутствует в предметной области автоматизации учебных заведений [5].

Реализация примера

В рамках работы реализовано решение приведенного примера для конфигураций «ERP 2.4»¹ и «Бухгалтерия 3»² на платформе 1С: Предприятие. Выбранные инструменты актуальны для предприятий РФ ввиду распространенности продуктов 1С в сфере автоматизации предприятий.

Конфигурация ERP – это флагманское решение от 1С, позиционируемое как инструмент для автоматизации *всех контуров учета* предприятий.

В свою очередь, конфигурация «Бухгалтерия» автоматизирует только бухгалтерский учет. Это делает ее удобной для сдачи бухгалтерской отчетности, поскольку отражение всех операций на предприятии происходит с минимальной степенью детализации (необходимой для формирования отчетности).

Ключевая идея решения – использовать ERP в качестве основной информационной системы, но для ведения бухгалтерского учета использовать только конфигурацию «Бухгалтерия». В этом случае целесообразно внесение информации только в ERP с последующим обменом данными с системой для ведения бухгалтерии. Для этого реализованы следующие блоки:

- модуль переноса прочих операций, не относящихся к производству (закупки, продажи, кадры и расчет заработной платы);
- модуль итогового формирования объемов производства в организациях за заданный период;
- модуль переноса итоговых операций из ERP в Бухгалтерию.

¹ 1С: ERP Управление предприятием 2. URL: <http://v8.1c.ru/erp/> (дата обращения 17.11.2019).

² «1С: Бухгалтерия 8» – доступно и всерьез. URL: <http://v8.1c.ru/buhv8/> (дата обращения 17.11.2019).

Перенос прочих операций

Разработанный в 1С универсальный формат обмена данными в XML ³ позволяет переносить объекты данных 1С между различными конфигурациями. Но это решает задачу лишь частично, поскольку стандартные возможности формата не включают все блоки учета.

Еще один универсальный инструмент от 1С – конфигурация «Конвертация данных» ⁴ позволяет самостоятельно настраивать правила обмена между различными базами. С его помощью было настроено соответствие между объектами подсистемы расчета заработной платы в ERP и конфигурацией «Бухгалтерия».

Модуль итогового формирования объемов производства

Алгоритм формирования обобщенных итогов производства состоит из двух фаз. Первая – отбор реализованной продукции за период (месяц). Результатом этой фазы работы является таблица «Продажи за период» (Организация, Номенклатура, Количество, Продажи по давальческой схеме). Значение последней колонки истинно, если реализация была не от организации-производителя.

Вторая фаза – определение нужного объема сырья. Для этого требуется наличие действующих спецификаций производства на всю реализованную продукцию (количество сырья и других расходов на единицу продукции). Алгоритм должен учитывать многоэтапное производство. В таком случае на выпускаемые в процессе полуфабрикаты должна быть собственная спецификация. В результате работы второй фазы алгоритма на каждую пару (Организация, Продажи по давальческой схеме) формируется обобщенная таблица произведенной продукции и необходимого сырья для ее производства. Сохранив эту информацию в базе ERP, останется осуществить заключительный шаг – перенести ее в контур бухгалтерского учета.

Модуль переноса итоговых операций производства

Отображение информации по объему произведенной продукции и списанного сырья в блок бухгалтерского учета заключается в оформлении набора документов.

Для стороны давальца:

- передача на переработку;
- поступление из переработки.

Для переработчика:

- поступление сырья в переработку;
- передача давальческого сырья в производство;
- выпуск продукции из давальческого сырья;
- реализация услуг по переработке;
- передача продукции давальцу.

Вместе с этим для организации-производителя, которая выступает в качестве переработчика, также формируются документы для собственного производства:

- передача сырья в производство;
- выпуск собственной продукции.

³ Универсальный механизм обмена данными. URL: http://v8.1c.ru/overview/Term_000000314.htm (дата обращения 17.11.2019).

⁴ Организация конвертации данных и обмена данными с помощью конфигурации «Конвертация данных 2.0». URL: <https://its.1c.ru/db/metod8dev/content/2943/hdoc> (дата обращения 17.11.2019).

Для реализации такого отображения были созданы правила обмена данными в конфигурации «Конвертация данных». Получившееся решение представляет собой отображение подраздела производства из блока управленческого учета в контур бухгалтерского учета.

Периодичность обмена данными различна для разных компонентов. Основная часть данных (не относящаяся к работе с давальческим сырьем) выгружается из ERP в базу бухгалтерии по ходу добавления / изменения. А данные по давальческому сырью – только по итогам месяца.

Выбор платформы для решения задачи в пользу 1С позволяет снять ряд вопросов по соблюдению законодательных требований к бухгалтерскому учету, поскольку они решены в 1С, и больше внимания уделить проработке архитектуры предлагаемого решения.

Обобщение

Ориентируясь на приведенный пример, можно обобщить ситуации такого плана до более широкого класса задач, в которых несколько информационных систем реализуют определенный набор бизнес-операций и при этом осуществляют обмен следующими видами данных:

- НСИ;
- мастер-данные;
- транзакционные данные.

Данные могут различаться как по структуре, так и по степени детализации.

Отражение бизнес-операции в отдельных системах может включать неопределенность, которая разрешается спустя некоторое время. Так, объем продаж за месяц из приведенного примера становится полностью определенным только по истечении месяца. При этом в других системах этой неопределенности может не возникать ввиду различий отражения операций хозяйственной деятельности на уровне бизнес-логики.

Информация, необходимая для разрешения неопределенности в одной системе, может находиться в другой. Эта информация может быть представлена разными способами. Например, в системах, построенных на реляционных СУБД, это будут отношения и кортежи (таблицы и строки). Также это могут быть классы и объекты (если речь идет об объектно-ориентированном программировании). Можно воспользоваться подходом 1С и работать в контексте предметно-ориентированного программирования [6]. Такая парадигма позволяет оперировать предопределенными классами и объектами, характерными для сферы автоматизации учета на предприятиях. Например: Документ, Справочник, План счетов, Регистр бухгалтерии.

В рамках этой работы все эти варианты допустимы. Конечная реализация модели данных не столь важна – достаточно ограничиться описанием элементов данных как наборов атрибутов.

Моделирование решения

Транзакционные данные состоят из ссылок на НСИ и мастер-данные, а также свойств с атомарными значениями (числа, строки, даты). Например, данные о продаже продукции будут содержать ссылки на клиента, товар, дату операции, цену и др. Чтобы не было путаницы с понятием транзакций в СУБД, следует отметить, что все рассуждения касаются исключительно слоя бизнес-логики, но не слоя источников данных (где осуществляется взаимодействие с СУБД).

Фиксация транзакционных данных в систему может быть:

- *завершенная* – имеющая на момент фиксации все необходимые для отражения в учете данные;

- *частичная* – не имеющая на момент фиксации все необходимые для отражения в учете данные. Недостающие данные заполняются значениями по умолчанию либо остаются в исходном состоянии – не фиксируются.

Условие однократного ввода данных в одну из систем предполагает обязательный обмен данными между блоками учета. Таким образом, отображение блоков учета друг в (на) друга – это средство достижения согласованности данных в системе, если под согласованностью в данном случае понимать завершенность всех транзакций.

В связи с этим требуется разобраться в некоторых вопросах.

- Как определить, находится ли сейчас система в согласованном состоянии?
- Что требуется для достижения актуального состояния?
- Какими могут быть отображения блоков учета?

Рассмотрим построение модели распространения транзакции на примере двух информационных систем S и S' . Пусть все отражаемые в системах бизнес-операции (далее операции) делятся на три группы:

- 1) операции, отражаемые только в S ;
- 2) операции, отражаемые только в S' ;
- 3) операции, изначально заносящие данные в S , но требующие отражения в том числе и в S' .

Все операции, относящиеся к первой и второй группам, завершаются сразу же после внесения их в S или S' соответственно. Третья группа вместе с отражением в S предполагает последующее отображение транзакционных данных в S' .

Для каждой операции A определен набор типов транзакций $T_1, T_2, \dots, T_n$ из S и $T'_1, T'_2, \dots, T'_m$ из S' , которые должны быть зафиксированы в каждой из систем. Для операций из первой группы $m = 0$, а из второй $n = 0$ соответственно.

Данные, которые фиксируются операцией A , в свою очередь, проецируются на НСИ и мастер-данные обеих систем. В рамках настоящей статьи предполагается, что синхронизация НСИ и мастер-данных осуществляется каким-то из возможных способов [3].

Рассмотрим, каким образом могут быть отображены транзакционные данные одной системы на данные другой системы.

Условимся, что все отображения делятся на инъекцию, сюръекцию и биекцию в соответствии с принятыми в математике определениями [7]. Если отображение является инъективным (в том числе биективным), то можно построить обратное отображение и, как следствие, использовать в качестве исходных данных любую из систем (а не только S). Иными словами, наличие инъективных отображений позволяет выбирать, в какую систему вносить данные в первую очередь, а в какую они должны отображаться. Это может быть полезным с практической точки зрения, например, исходя из удобства работы пользователей с какой-то конкретной системой.

Для решения рассмотренного примера «работа с давальческим сырьем» необходимы как минимум отображения вида

$$T_1, T_2, \dots, T_n \rightarrow T'_1, T'_2, \dots, T'_m, \quad (1)$$

$$(T_1, T_2, \dots, T_n) \Delta t \rightarrow T'_1, T'_2, \dots, T'_m, \quad (2)$$

где (1) включает преобразование транзакционных данных системы S в соответствующие данные системы S' посредством сопоставления их свойств и применения функциональных преобразований, доступных в системах.

Преобразования вида (2) более сложные и схожи с агрегатными функциями из языка SQL [8]. Они предполагают вычисление транзакций в S' , основываясь на определенном временном промежутке Δt в системе S . Как вычисление итоговых сумм продаж за месяц в рассмотренном примере. При этом для разных операций соответствующий временной промежуток

может быть разным в зависимости от требований на уровне бизнес-логики (день, месяц, квартал и т. д.).

Несмотря на существенные отличия, преобразование вида (2) можно свести к (1), если ввести в системе S новый «обобщающий» тип транзакции, зависящий от других транзакций за определенный период времени:

$$T_{n+1} = (T_1, T_2, \dots, T_n) \Delta t,$$

т. е. итоговая цепочка отображения данных из S в S' будет выглядеть так:

$$(T_1, T_2, \dots, T_n) \Delta t \rightarrow T_{n+1} \rightarrow T'_1, T'_2, \dots, T'_m.$$

Данные фиктивной транзакции T_{n+1} будут сохранены в S . В результате такого построения любая операция A может быть отображена в S' из S , когда данные за требуемый временной промежуток Δt будут внесены. До этих пор A будет находиться в очереди частично завершенных транзакций, ожидающих внесения недостающих данных в S . В зависимости от типа транзакционных данных можно разными способами реализовать выгрузку данных в S' , например, по требованию пользователя (нажатием на кнопку), по заданному расписанию или с привязкой к заданным событиям в S .

Таким образом, *описание отображений транзакционных данных* между блоками учета позволяет определить корректность отражения операций предприятия в информационной системе.

Во-первых, дает возможность ответить на вопрос: «Можно ли единожды ввести операцию и впоследствии отразить ее на все блоки?».

Во-вторых, дает возможность иметь всю необходимую информацию о состоянии системы в каждый момент времени, а именно какие транзакционные данные уже отражены в системе в целом, какие находятся в стадии частичной фиксации, когда они будут отражены и что для этого необходимо.

Например, если операция A должна быть отражена в системе S транзакциями $T_1, T_2, \dots, T_n, T_{n+1}, T_{n+2}, \dots, T_{n+k}$, а в S' транзакциями $T'_1, T'_2, \dots, T'_m, T'_{m+1}, T'_{m+2}, \dots, T'_{m+q}$, то отображение данных из S в S' может выглядеть как на рис. 2.

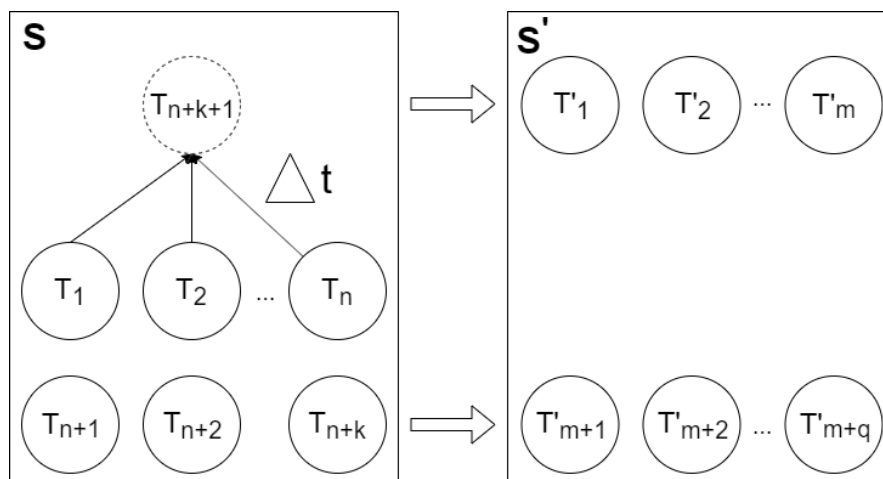


Рис. 2. Отображения транзакционных данных системы из S в S' на примере одной операции
Fig. 2. Example of transactional data export from S in S' for single operation

В этом случае часть транзакций в S' (это $T'_1, T'_2, \dots, T'_m$) будут зафиксированы только по прошествии времени Δt , в то время как $T'_{m+1}, T'_{m+2}, \dots, T'_{m+q}$ сразу же после фиксации $T_{n+1}, T_{n+2}, \dots, T_{n+k}$ (посредством соответствующего отображения). Операция A не будет отражена полностью до тех пор, пока есть хотя бы одна не зафиксированная транзакция в какой-либо из систем. В свою очередь, общее интеграционное решение будет находиться в согласованном состоянии, только если все поступившие в него операции отражены полностью.

Реализация интеграционного решения на базе такой модели сводится к описанию обмена данными любым из возможных способов, рассматриваемых системах учета. В рассмотренном примере решения задачи на платформе 1С это был обмен xml-файлами по правилам, разработанным в традиционном для 1С инструментарии. Также это может быть использование общей БД (Shared DB), удаленный вызов процедур (Remote Procedure Call) или использование систем обмена сообщениями (Messaging) [9].

Заключение

В работе рассмотрены проблемы проектирования информационных систем для автоматизации хозяйственной деятельности предприятий. Ввиду необходимости вести несколько видов учета (бухгалтерский, налоговый, управленческий) предложено использование интеграционного решения на базе подсистем для каждого из блоков учета.

Построена модель описания отображений данных между информационными системами, позволяющая определить в любой момент времени степень завершенности транзакций бизнес-логики в системе. Для этого введена классификация состояний транзакций и видов отображений транзакционных данных между системами.

Модель была использована при решении задачи автоматизации работы нескольких организаций с давальческим сырьем для конфигураций «ERP 2.4» и «Бухгалтерия 3». Благодаря такому решению пользователям доступна информация о статусе всех операций, что позволяет не упустить из виду важные для отражения в учете данные. Ранее использовались эпизодические выгрузки данных, которые требовали ручного контроля со стороны пользователей. Несмотря на выбор платформы 1С в качестве решения приведенной задачи, предлагаемая модель также может использоваться для любой другой среды разработки.

Для полной модели более чем с двумя информационными системами отображение транзакционных данных значительно сложнее. Рассмотрение этого варианта в совокупности с разработкой инструмента для контроля завершенности транзакционных данных в целом – это предмет дальнейшего исследования.

Список литературы

1. **Друри К.** Управленческий и производственный учет. Вводный курс: Учебник. 5-е изд., перераб. и доп. М.: Юнити-Дана, 2014.
2. **Фаулер М.** Архитектура корпоративных программных приложений. М.: Вильямс, 2008. 544 с.
3. **Андрюшкевич С. К., Гуськов А. Е.** Практика решения задач интеграции информационных систем на основе управления мастер-данными // Вычислительные технологии. 2013. Т. 18, № 6. С. 3–15.
4. **Киркеева Л. И., Кушнер Е. Н.** Переработка дачальческого сырья в АПК: бухгалтерский учет, налоговая оптимизация // Вестник Алтайского гос. аграр. ун-та. 2007. Вып. 2. С. 65–68.
5. **Демиш В. О., Пищик Б. Н., Козьменко Г. Г.** Проблемы автоматизации управления образовательным учреждением // Вестник НГУ. Серия: Информационные технологии. 2009. Т. 7, вып. 3. С. 94–102.

6. **Фаулер М.** Предметно-ориентированные языки программирования. М.: Вильямс, 2011. 576 с.
7. **Ершов Ю. Л., Палютин Е. А.** Математическая логика: Учеб. пособие. 3-е изд., стереотип. СПб.: Лань, 2004.
8. **Грабер М.** Введение в SQL. М.: Лори, 1992.
9. **Hohpe G., Woolf B.** Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. 1st ed. Addison-Wesley Professional, 2003.

References

1. **Drury C.** Management and production accounting. Introductory course. 5th ed. Moscow, Unity-Dana Publ., 2014. (in Russ.)
2. **Fowler M.** Architecture of corporate software applications. Moscow, Williams, 2008, 544 p. (in Russ.)
3. **Demish V. O., Pischik B. N., Kozmenko G. G.** Problems of automation of management of educational institution. *Vestnik NSU. Series: Information Technologies*, 2009, vol. 7, no. 3, p. 94–102. (in Russ.)
4. **Andryushkevich S. K., Guskov A. E.** Practice of solving problems of integration of information systems based on master data management. *Computing Technologies*, 2013, vol. 18, no. 6, p. 3–15. (in Russ.)
5. **Kirkeeva L. I., Kushner E. N.** Processing of tolling raw materials in agriculture: accounting, tax optimization. *Bulletin of the Altai State Agrarian University*, 2007, iss. 2, p. 65–68. (in Russ.)
6. **Fowler M.** Subject-oriented programming languages. Moscow, Williams Publ., 2011, 576 p. (in Russ.)
7. **Ershov Yu. L., Palyutin E. A.** Mathematical logic: A textbook. 3rd ed. Yeah, LAN, 2004, 336 p. (in Russ.)
8. **Graber M.** SQL. Moscow, Lori Publ., 1992. (in Russ.)

*Материал поступил в редколлегию
Received
05.12.2019*

Сведения об авторах

Демиш Всеволод Олегович, старший преподаватель, Новосибирский государственный университет (ул. Пирогова 2, Новосибирск, 630090, Россия); ведущий инженер-программист, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия)
v.demish@gmail.com

Пищик Борис Николаевич, кандидат технических наук, старший научный сотрудник, и.о. заведующего лабораторией, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия); доцент, заведующий кафедрой, Новосибирский государственный университет (ул. Пирогова 2, Новосибирск, 630090, Россия)
Boris.Pishchik@gmail.com

Information about the Authors

Vsevolod O. Demish, Lecturer, Novosibirsk State University (2 Pirogov Str., Novosibirsk, 630090, Russian Federation); Leading Software Engineer, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation)
v.demish@gmail.com

Boris N. Pishchik, Cand. Tech. Sciences, Senior Researcher, Acting Head Laboratory, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation); Associate Professor, Head of Department, Novosibirsk State University (2 Pirogov Str., Novosibirsk, 630090, Russian Federation)
Boris.Pishchik@gmail.com

Оценка точности определения параметров погребенных археологических объектов по данным разновысотной магнитной съемки на основе численного моделирования

Н. Д. Евменов¹, В. В. Воронин², П. Г. Дядьков^{1,2}, О. А. Позднякова³

¹ *Институт нефтегазовой геологии и геофизики им. А. А. Трофимука СО РАН
Новосибирск, Россия*

² *Новосибирский государственный университет
Новосибирск, Россия*

³ *Институт археологии и этнографии СО РАН
Новосибирск, Россия*

Аннотация

Предложен алгоритм решения обратной задачи для определения параметров погребенного археологического объекта в виде эллиптического цилиндра (как модели типичного погребения или ямы). Выполнена оценка точности определения параметров погребенных археологических объектов на основе численного решения обратной задачи магниторазведки для синтетической модели. Ее реализация включала, во-первых, этап моделирования магнитного поля от синтетической модели археологического объекта на трех разных высотах для различных параметров плотности сетки и размеров участка проведения магнитной съемки и, во-вторых, этап решения обратной задачи по синтетическим данным на разных высотах для различных параметров проведения магнитной съемки с оценкой точности решения обратной задачи.

Получено, что на точность определения параметров аномального тела в большей степени влияет размер участка съемки, чем плотность сети измерений. Также отмечается тенденция повышения точности определения параметров при совместном использовании данных разновысотной съемки. Апробация алгоритма решения обратной задачи на реальных данных разновысотной магнитной съемки над погребением на археологическом памятнике Тартас-1 показала достаточно хорошие результаты. Средняя ошибка определения геометрических параметров тела и его залегания составила около 15 %.

Ключевые слова

магнитометрия, археология, численное моделирование, обратная задача, методика магнитной съемки, БПЛА

Благодарности

Работа выполнена при частичной финансовой поддержке РФФИ (грант № 17-29-04314) и проекта ФНИ № 0331-2019-0012

Для цитирования

Евменов Н. Д., Воронин В. В., Дядьков П. Г., Позднякова О. А. Оценка точности определения параметров погребенных археологических объектов по данным разновысотной магнитной съемки на основе численного моделирования // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 16–26. DOI 10.25205/1818-7900-2020-18-1-16-26

Estimation of Accuracy of Determination of Parameters of the Buried Archaeological Objects According to the Data of Different-Altitude Magnetic Survey on the Basis of Numerical Modeling

N. D. Evmenov¹, V. V. Voronin², P. G. Dyadkov^{1,2}, O. A. Pozdnyakova³

¹ *Trofimuk Institute of Petroleum Geology and Geophysics SB RAS
Novosibirsk, Russian Federation*

² *Novosibirsk State University
Novosibirsk, Russian Federation*

³ *Institute of Archaeology and Ethnography SB RAS
Novosibirsk, Russian Federation*

Annotation

An algorithm for solving the inverse problem for determining the parameters of a buried archaeological object in the form of an elliptical cylinder (as a model of a typical burial or pit) is proposed. The estimation of accuracy of determination of parameters of the buried archaeological objects on the basis of the numerical solution of the inverse problem of magnetic prospecting for synthetic model is carried out. Its implementation included, firstly, the stage of modeling the magnetic field from a synthetic model of the archaeological site at three different heights for different parameters of the grid density and the size of the magnetic survey site and, secondly, the stage of solving the inverse problem from synthetic data at different heights, for different parameters of the magnetic survey and assessing the accuracy of the solution of the inverse problem.

It is found that the accuracy of determining the parameters of the anomalous body is more affected by the size of the survey area than the density of the measurement network. There is also a tendency to increase the accuracy of determining parameters when using data of different heights. Approbation of the algorithm for solving the inverse problem on real data of different-altitude magnetic survey over the burial at the archaeological site Tartas-1 showed quite good results. The average error in determining the geometric parameters of the body and its occurrence is ~ 15 %.

Keywords

magnetometry, archaeology, numerical modeling, inverse problem, magnetic survey technique, UAV

Acknowledgments

The work was supported in part by the Russian Foundation for Basic Research (grant no. 17-29-04314) and project BSR № 0331-2019-0012

For citation

Evmenov N. D., Voronin V. V., Dyadkov P. G., Pozdnyakova O. A. Estimation of Accuracy of Determination of Parameters of the Buried Archaeological Objects According to the Data of Different-Altitude Magnetic Survey on the Basis of Numerical Modeling. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 16–26. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-16-26

Введение

Магнитометрия является одним из ведущих геофизических методов поиска и изучения археологических памятников [1; 2]. Однако в археологии, в отличие от геологоразведочных задач, ее использование в абсолютном большинстве случаев ограничивается построением карт магнитного поля и определением планиграфического расположения археологических объектов [3; 4]. При этом не используются потенциальные возможности метода по определению параметров залегания погребенных объектов, в том числе такого важного параметра, как глубина залегания (верхняя и нижняя кромки). Методы определения этих параметров применительно к задачам поиска и разведки полезных ископаемых развивались достаточно давно [5; 6]. Как было показано ранее [7; 8], использование повысотных магнитных съемок может существенно облегчить решение обратных задач магниторазведки.

Несомненно, современный этап развития технологии магнитной съемки при решении задач археологии должен включать разработку специальных методов и алгоритмов для получения максимума возможной информации о погребенных археологических объектах. При этом следует также учитывать возросшие технические возможности данного метода: прецизионную точность используемых квантовых и протонных магнитометров, а также возмож-

ность проведения съемок на нескольких высотах с использованием беспилотных летательных аппаратов (БПЛА) [9; 10].

Данная работа направлена на решение именно этих задач и предполагает оценку погрешностей определения параметров залегания и формы погребенного тела в виде эллиптического цилиндра (как модели реального типичного погребения или ямы) при решении обратной задачи. При этом одной из важнейших задач работы является определение оптимальных параметров проведения магнитной съемки (размеров участка, шага сетки, высоты съемки и комбинаций двух и трех высот) и выработке соответствующих рекомендаций для практического применения.

Таким образом, основной целью данной работы является оценка точности определения параметров погребенных археологических объектов (погребения, ямы) на основе численного решения обратной задачи магниторазведки. Ее реализация включает следующие этапы:

- моделирование магнитного поля от синтетической модели археологического объекта (погребения) на разных высотах для различных параметров плотности сетки и размеров участка проведения магнитной съемки;
- решение обратной задачи по синтетическим данным на разных высотах, для различных параметров проведения магнитной съемки и оценка точности решения обратной задачи;
- апробация алгоритма решения обратной задачи на реальных данных разнорысотной магнитной съемки для погребения на археологическом памятнике Тартас-1.

Алгоритм численного решения прямой и обратной задачи¹

Постановка задачи

Решается задача восстановления параметров намагниченного тела исходя из данных магнитометрии. Известен постоянный вектор нормального поля, т. е. вектор магнитной индукции $\vec{P}_0$. Имеется тело, для которого считается известным коэффициент магнитной восприимчивости χ , т. е. вектор удельной намагниченности тела есть $\vec{m} = \chi \cdot \vec{P}_0$.

Как известно, вектор магнитной индукции, порождаемый точечным диполем, может быть записан в виде

$$\vec{B}(\vec{R}_r, \vec{R}) = 3 \frac{(\vec{m}, \vec{r}) \cdot \vec{r}}{r^5} - \frac{\vec{m}}{r^3}, \quad (1)$$

где m – магнитный момент диполя, $\vec{r} = \vec{R}_r - \vec{R}$ – вектор из источника в точку наблюдения. Из-за наличия аномального тела в каждом приемнике к вектору нормального поля $\vec{P}_0$ прибавляется добавочное поле $\vec{P}(\vec{R}_r)$, и измеряется модуль суммарного поля $|\vec{P}_0 + \vec{P}(\vec{R}_r)| = |\vec{P}_0 + \vec{P}(\vec{R}_r)|$.

Пусть точка наблюдения имеет координаты $\vec{R}_r = (x_r, y_r, z_r)$, а источник с координатами $\vec{R} = (x, y, z)$ пробегает тело V , которое назовем условно «цилиндроидом». А именно: верхнее основание V находится в плоскости $z = z_0$ и представляет собой эллипс с центром x_0, y_0, z_0 , полуосями (A_0, B_0) , причем первая ось повернута относительно координатной оси O_x на угол φ_0 . Нижнее основание – эллипс в плоскости $z = z_0 + H$ с центром $(x_1, y_1, z_1 + H)$, полу-

¹ Алгоритм разработан В. В. Ворониным.

осями (A_1, B_1) и с тем же углом поворота φ_0 . И параметры эллипса линейно меняются при изменении глубины сечения.

Итак, каждый цилиндроида характеризуется набором 11 параметров:

$$\Pi_{11} = (x_0, y_0, z_0, A_0, B_0, x_1, y_1, A_1, B_1, H, \varphi_0). \quad (2)$$

Обратная задача состоит в отыскании этого набора параметров аномального тела, где входной информацией является модуль суммарного поля $|\vec{P}_0 + \vec{P}(\vec{R}_r)|$, измеренный в точках регистрации магнитного поля над телом.

Расчет аномального поля (прямая задача)

Есть тело, для которого имеется свой набор параметров Π_{11} , перечисленных в формуле (2). Весь набор параметров будем обозначать Π_{11} .

Добавочное поле от тела в точке наблюдения представляется интегралом

$$\vec{P}_i(\Pi_{11i}, \vec{R}_r) = \iiint_{V_i} \vec{B}(\vec{R}_r, \vec{R}) dx dy dz, \quad (3)$$

где $\vec{R}_r = (x_r, y_r, z_r)$ – радиус-вектор точки наблюдения, $\vec{R} = (x, y, z)$ – радиус-вектор источника.

Для каждого положения приемника $\vec{R}_r$ относительно тела производится вычисление интеграла в обобщенной полярной системе координат:

$$\vec{P}(\vec{R}_r) = \int_0^{2\pi} dt \int_0^1 \rho d\rho \int_0^1 H((1-\gamma)A_0 + \gamma A_1)((1-\gamma)B_0 + \gamma B_1) \vec{G}(\rho, t, \gamma) d\gamma \quad (4)$$

Поскольку координаты вектора $\vec{r}$ в исходной формуле (1) линейно зависят от γ , то внутренний интеграл по переменной γ может быть взят по аналитическим формулам. А оставшийся от формулы интеграл по переменным ρ, t берется численно, по квадратурным формулам. При этом интеграл по угловой переменной t берется по равномерной на $[0, 2\pi]$ сетке. А для интегрирования по переменной ρ используется кусочно-гауссова квадратурная формула с четырьмя узлами.

Решение обратной задачи

Обратная задача по определению параметров тела решается с помощью оптимизации. Минимизируется квадрат невязки:

$$f_{11}(\Pi_{11}) = \sum_{r=1}^{N_r} \left(\left| \sum_{i=1}^{N_b} \vec{P}_i(\Pi_{11i}, \vec{R}_r) + \vec{P}_0 - \vec{P}_s(\vec{R}_r) \right| \right)^2, \quad (5)$$

где $|\vec{P}_s(\vec{R}_r)|$ – модуль наблюдаемого суммарного поля в приемнике с координатами $\vec{R}_r$.

Минимизация проводится методом Нелдера – Мида [11], который не требует вычисления градиентов. Результат предыдущего этапа берется за начальное приближение для следующего.

За начальное приближение берется прямой круговой цилиндр, то есть начальное $\Pi_{11} = (x_0, y_0, z_0, R, H)$, где x_0, y_0, z_0 – координаты центра верхнего основания, R – радиус цилиндра, H – его высота.

Варируется весь набор из 11 параметров тела, перечисленных в формуле (2), где x_0, y_0, z_0 – координаты центра верхнего основания, H – его высота, $(x_1, y_1, z_0 + H)$ – координаты центра нижнего основания; (A_0, B_0) и (A_1, B_1) – размеры полуосей для верхнего и нижнего основания соответственно, а также угол поворота φ_0 (одинаковый для обоих оснований).

Параметры модели археологического объекта и условий проведения магнитной съемки

Модель археологического объекта (погребения)

Описанный выше метод решает прямые и обратные задачи магниторазведки для тел цилиндрической формы с эллипсом в горизонтальном сечении. В археологии погребения и ямы в плане имеют округлую или овальную форму и, как правило, вытянуты по одной оси. Сверху они перекрыты почвенно-дерновым слоем. Глубина их нижней кромки может изменяться в значительных пределах – вплоть до 2,5 м (памятник Тартас-1). Таким образом, подбираемое в алгоритме тело «цилиндрической» формы подходит для интерпретации подобных археологических объектов. Предложенный алгоритм решения обратной задачи позволяет определять положения и размеры верхних и нижних кромок независимо друг от друга, так как погребения могут быть несимметричными и сужаться или расширяться с глубиной.

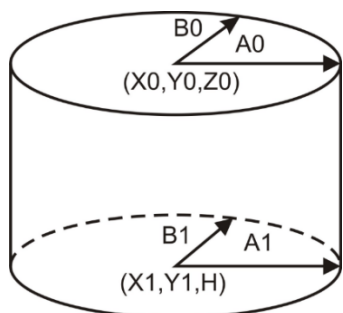


Рис. 1. Модель цилиндра с обозначением подбираемых параметров
Fig. 1. The cylinder model with the designation of the selected parameters

Для оценки точности определения параметров цилиндра были решены прямые задачи от простого тела с параметрами: $x_0, y_0, x_1, y_1 = 0$; $z_0 = 0,5$ м; $H = 1$ м; $a_0, a_1 = 2$ м; $b_0, b_1 = 1$ м; угол поворота тела $\varphi = 0$; магнитная восприимчивость $\chi = 90 \cdot 10^{-6}$ СИ (рис. 1). Значение магнитной восприимчивости соответствует реально измеренным значениям из погребений памятника Тартас-1.

Параметры условий проведения магнитной съемки

Как отмечалось выше, целью работы является определение точности решения обратной задачи в зависимости от таких параметров проведения магнитной съемки как размер участка, плотность сетки измерений и высота проведения магнитной съемки. Размеры участка определялись условием охвата как положительной, так и затухающей отрицательной аномалии от тела. Для данного условия наилучшим размером участка является размер 15×15 м. Худшим вариантом является охват только положительной аномалии (5×5 м). Сетка 10×10 м

взята как промежуточная, в пределах которой не наблюдается полное затухание отрицательной аномалии.

Минимальная плотность сетки для исследования тела с размерами 1×2 м – это профилирование с шагом 1 м (сетка 1×1 м) для того, чтобы хотя бы 3 профиля зафиксировали магнитную аномалию от объекта. Плотность сетки через $0,5 \times 0,5$ и $0,25 \times 0,25$ м взята исходя из необходимости лучшей детализации аномалии.

Высота съемки 0,5 м – это стандартная высота проведения магнитной съемки над археологическими объектами. Высота 1,1 м является характерной высотой расположения верхнего датчика градиентометра, где разнос между датчиками составляет от 0,6 м. Высота 2 м над землей соответствует, как правило, минимальной высоте проведения магнитной съемки с использованием БПЛА.

Таким образом, параметры условий проведения магнитной съемки были приняты следующими:

- размеры участков – 5×5 , 10×10 , 15×15 м;
- плотность измерений (расстояние между точками измерения поля по оси x и по оси y) – 0,25, 0,5, 1 м;
- высота съемки от уровня земли – 0,5, 1,1, 2 м.

Результаты численного моделирования и их анализ

Для оценки точности решения обратной задачи выполнено моделирование магнитного поля от археологического объекта (погребения) на разных высотах для различных параметров плотности сетки и размеров участка проведения магнитной съемки. Были решены прямые задачи для каждой комбинации «размер участка – плотность измерений – высота съемки». Всего получено 27 решений прямых задач.

Обратные задачи с подбором параметров аномального тела (погребения) были решены не только для каждой комбинации «размер участка – плотность измерений – высота съемки», но и для различных комбинаций высот: 0,5–1,1; 0,5–2; 1,1–2; 0,5–1,1–2 м. Количество решенных обратных задач равно 63-м.

Для выявления тенденций улучшения результатов решения обратных задач были составлены таблицы относительных погрешностей определения параметров тела в зависимости от изменения отдельного параметра. Для параметров (X_0 , Y_0 , X_1 , Y_1) и (A_0 , B_0 , A_1 , B_1) в таблице представлено среднее арифметическое значение погрешностей. Погрешность определения всех параметров рассчитывалась как среднее арифметическое по всем 10 параметрам тела за исключением угла поворота (он был неизменен и равен 0°). Отклонения значения параметров X_0 , Y_0 , X_1 , Y_1 брались относительно начального положения эллипсов нижней и верхней кромок соответственно, а именно если положение центров эллипсов оснований отличается на расстояние диаметра изначального тела, то ошибка определения центра тела будет составлять 100 %.

В табл. 1 приведен результат сравнения решения обратной задачи и модельного тела в относительных погрешностях (%) в зависимости от размеров участка. Анализ результатов показал, что наблюдается тенденция к уменьшению погрешности определения параметров тела при увеличении размера участка съемки.

В табл. 2 приведен результат сравнения решения обратной задачи и модельного тела в относительных погрешностях (%) для различной плотности сетки. Анализ свидетельствует о том, что при принятых параметрах плотности сетки не наблюдается явной зависимости погрешности определения параметров тела от плотности сетки.

В табл. 3 приведен результат сравнения решения обратной задачи в относительных погрешностях (%) для трех высот проведения магнитной съемки и их комбинаций. Важным параметром при интерпретации данных магнитометрии является определение глубины залегания нижней кромки. Видно, что наилучший результат достигается при решении обратной

задачи с применением всех трех высот. Также из табл. 3 видно, что геометрия тела подбирается лучше в случае комбинации двух (0,5–2; 1,1–2 м) или трех высот. Анализ результатов показал, что наблюдается тенденция к уменьшению погрешности определения параметров тела при привлечении разновысотных данных.

Таблица 1

Средняя погрешность определения параметра в зависимости от размера участка.

Table 1

The average error in determining the parameter depending on the size of the area.

Размер участка, м	Погрешность определения параметра / параметров, %				
	A0, B0, A1, B1	X0, Y0, X1, Y1	Z0	H	Все параметры
5 × 5	17	25	6	20	19
10 × 10	13	14	3	10	12
15 × 15	12	13	4	6	11

Условные обозначения для табл. 1–3:

- Лучшие результаты
- Промежуточные результаты
- Худшие результаты

Таблица 2

Средняя погрешность определения параметра в зависимости от плотности сетки

Table 2

The average error in determining the parameter depending on the density of the grid

Плотность сетки, м	Погрешность определения параметра / параметров, %				
	A0, B0, A1, B1	X0, Y0, X1, Y1	Z0	H	Все параметры
0,25	13	21	5	13	15
0,5	14	19	3	13	15
1	15	12	5	10	12

Таблица 3

Средняя погрешность определения параметра
в зависимости от уровня высоты съемки

Table 3

The average error in determining the parameter
depending on the altitude of the survey

Высота, м	Погрешность определения параметра / параметров, %				
	H	Z0	X0, Y0, X1, Y1	A0, B0, A1, B1	Все параметры
0,5	21	2	13	17	14
1,1	5	8	25	16	18
2	15	6	26	13	18
0,5–1,1	16	4	18	18	17
0,5–2,0	10	2	11	13	11
1,1–2,0	10	6	15	13	13
0,5–1,1–2,0	5	2	14	9	10

Апробация метода на реальных данных

Для апробации метода были использованы данные разновременной магнитной съемки над погребением № 639 памятника Тартас-1 (Барабинская лесостепь). На момент проведения исследований дерновый слой над погребением был снят. Контраст магнитной восприимчивости (χ) между заполнением погребения и вмещающей породой составлял $90 \cdot 10^{-6}$ СИ. Магнитная съемка над телом была проведена на четырех высотах: 0,25, 0,5, 1 и 1,75 м над землей. Результаты решения обратной задачи для данных разновременной магнитной съемки представлены ниже. На рис. 2 видно, что контуры и амплитуды поля от тела, подобранного при решении обратной задачи на разновременных данных, практически полностью совпадают с реально замеренными данными. Суммарный результат представлен в табл. 4, где приведены значения подобранного тела и реальные значения погребения, которые были замерены при его раскопках. Данный результат показывает, что средняя ошибка определения геометрических параметров тела и его залегания равна $\sim 15\%$. Максимальная ошибка имеет место при определении глубины залегания нижней кромки аномального тела, она равна 30% .

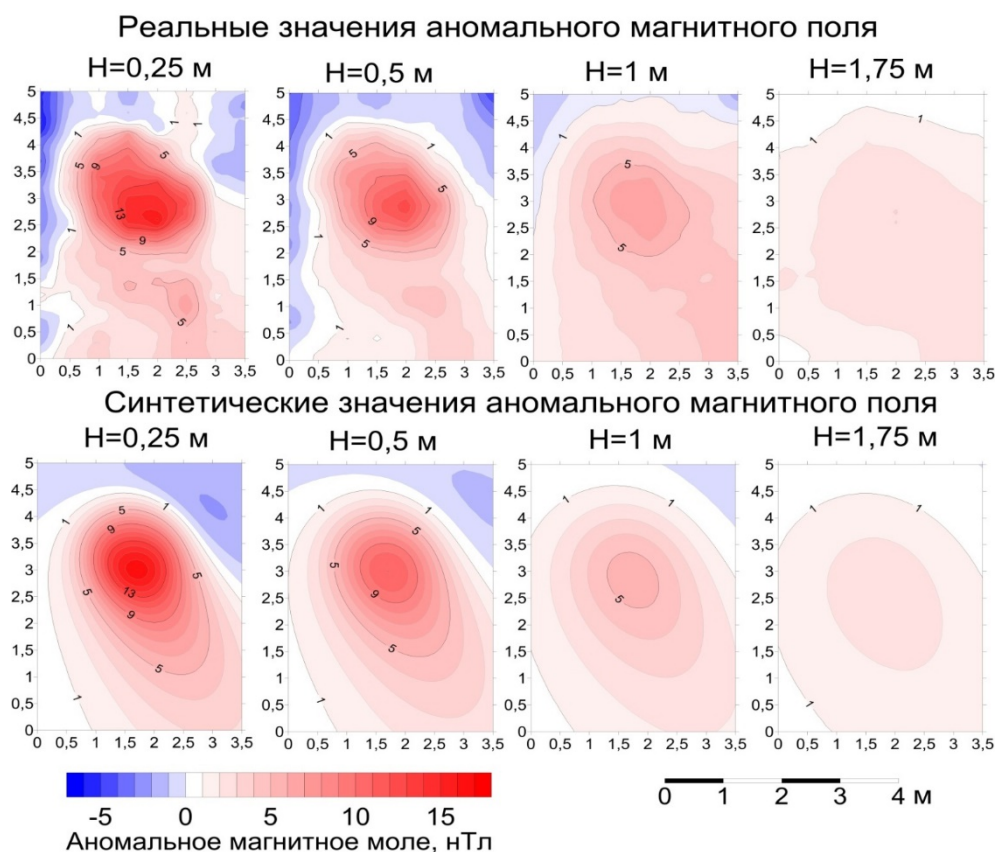


Рис. 2. Замеренное аномальное магнитное поле на высоте 0,25, 0,5, 1 и 1,75 м над погребенным археологическим объектом № 639 памятника Тартас-1 (вверху) и рассчитанное на высоте 0,25, 0,5, 1 и 1,75 м поле от тела, подобранного при решении обратной задачи для разновременных данных (внизу)

Fig. 2. The measured anomalous magnetic field at heights of 0.25, 0.5, 1 and 1.75 meters above the buried archaeological object No. 639 of the Tartas-1 site (above) and the magnetic field from the interpreted body, calculated at heights of 0.25, 0.5, 1 and 1.75 meters, selected from the inversion for different heights data (down)

Таблица 4

Параметры цилиндра после решения обратной задачи
и параметры погребения после раскопок

Table 4

Cylinder parameters after the inversion and burial parameters after excavation

	Параметры цилиндра, м									
	X0	Y0	Z0	A0	B0	X1	Y1	H	A1	B1
Модель	1,65	3,11	0,15	0,81	0,81	1,53	2,41	1,07	1,09	0,81
Раскоп	~ 1,5	~ 3	0	1,1	0,9	~ 1,5	~ 3	0,72	1	0,75

Заключение

В работе предложен алгоритм решения обратной задачи для определения параметров погребенного аномального тела в виде эллиптического цилиндра (как модели типичного погребения или ямы).

Выполнена оценка точности определения параметров погребенных археологических объектов на основе численного решения обратной задачи магниторазведки для синтетической модели. Получено, что на точность определения параметров аномального тела в большей степени влияет размер участка съемки, чем плотность сети измерений. Также отмечается тенденция повышения точности определения параметров при совместном использовании данных разнорельефной съемки.

Апробация алгоритма решения обратной задачи на реальных данных разнорельефной магнитной съемки над погребением на археологическом памятнике Тартас-1 показала достаточно хорошие результаты. Средняя ошибка определения геометрических параметров тела и его залегания равна ~ 15 %. Максимальная ошибка имеет место при определении глубины залегания нижней кромки аномального тела, но она не превышает 30 % (в абсолютном выражении около 30 см при глубине залегания тела около 1 м).

Список литературы

1. **Fassbinder J. W. E.** Magnetometry for Archaeology. In: A. S. Gilbert (ed.) *Encyclopedia of Geoarchaeology. Encyclopedia of Earth Sciences Series*. Dordrecht, Springer, 2017. DOI 10.1007/978-1-4020-4409-0
2. **Дядьков П. Г., Молодин В. И., Чемякина М. А., Михеев О. А.** Магнитометрические исследования археологических памятников Тартас-1 и Преображенка-6 в Барабинской лесостепи // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий: Материалы Годовой сессии Ин-та археологии и этнографии СО РАН. Новосибирск, 2005. Т. 11, ч. 1. С. 304–309.
3. **Becker H., Fassbinder J. W. E.** Magnetometry of a Scythian Settlement in Siberia near Cichah in the Baraba steppe. *Archeological Prospection*, 1999, vol. 31, p. 168–172.
4. **Bedford B.** Magnetic surveying reveals hidden ancient buildings and streets. *Eos*, 11 January 2019. DOI 10.1029/2019EO113625.
5. **Миков Д. С.** Атлас теоретических кривых для интерпретации магнитных и гравитационных аномалий. Томск, 1955. 146 с.
6. **Логачев А. А., Захаров В. П.** Магниторазведка. 4-е изд., перераб. и доп. Л.: Недра, 1973. 351 с.
7. **Ларионов В. А.** Палетки теоретических кривых для интерпретации повысочных измерений магнитного поля // Тр. Ин-та геологии и геофизики. Новосибирск, 1961. Вып. 11: Вопросы разведочной геофизики, сб. 2. С. 99–108.

8. Ларионов В. А. Определение вертикальных размеров возмущающих объектов по функциям $\Delta T(z)$ и $\Delta Z(z)$ // Геология и геофизика. 1964. № 5. С. 96–105.
9. Эпов М. М., Молодин В. М., Манштейн Л. К., Балков Е. В., Дядьков П. Г., Матасова Г. Г., Казанский Л. Ю., Бортникова С. Б., Позднякова О. А., Карин Ю. Г., Кулешов Д. А. Мультидисциплинарные археолого-геофизические исследования в Западной Сибири // Геология и геофизика. 2016. Т. 57, № 3. С. 603–614.
10. Фирсов А. П., Злыгостев И. Н., Дядьков П. Г., Савлук А. В., Вайсман П. А., Шеремет А. С., Вальд А. К., Евменов Н. Д. Применение высокочастотного магнитометра для лёгких БПЛА при геолого-геофизическом изучении трубок взрыва // Интерэкспо ГЕО-Сибирь-2015. XI Междунар. науч. конгр.: Сб. материалов: В 4 т. Новосибирск: СГГА, 2015. Т. 2. С. 299–304.
11. Nelder J. A., Mead R. A Simplex Method for Function Minimization. *Computer Journal*, 1965, vol. 7, p. 308–313.

References

1. Fassbinder J. W. E. Magnetometry for Archaeology. In: A. S. Gilbert (ed.) *Encyclopedia of Geoarchaeology*. Encyclopedia of Earth Sciences Series. Dordrecht, Springer, 2017. DOI 10.1007/978-1-4020-4409-0
2. Dyadkov P. G., Molodin V. I., Chemyakina M. A., Mikheev O. A. Magnetometric studies of archaeological monuments Tartas-1 and Preobrazhenka-6 in the Barabinsk forest-steppe. In: Problems of archaeology, Ethnography, anthropology of Siberia and adjacent territories: Materials of the Annual session of the Institute of archaeology and Ethnography SB RAS 2005. Novosibirsk, 2005, vol. 11, pt. 1, p. 304–309. (in Russ.)
3. Becker H., Fassbinder J. W. E. Magnetometry of a Scythian Settlement in Siberia near Cich in the Baraba steppe. *Archeological Prospection*, 1999, vol. 31, p. 168–172.
4. Bedford B. Magnetic surveying reveals hidden ancient buildings and streets. *Eos*, 11 January 2019. DOI 10.1029/2019EO113625.
5. Mikov D. S. An Atlas teoreticheskikh krivih dly interpretazii magnitnih i gravitazionnih anomalii. Tomsk, 1955, 146 p. (in Russ.)
6. Logachev A. A., Zakharov V. P. Magnitorazvedka. 4th ed. Leningrad, Nedra Publ., 1973, 351 p. (in Russ.)
7. Larionov V. A. Paletki teoreticheskikh krivih dly interpretazii povisotnih izmerenii magnitnogo polya. In: Trudi Instituta geologii i geofiziki. Novosibirsk, 1961, iss. 11, book 2, p. 99–108. (in Russ.)
8. Larionov V. A. Determination of the vertical dimensions of the perturbing object function $\Delta T(z)$ and $\Delta Z(z)$. *Geology and Geophysics*, 1964, no. 5, p. 96–105. (in Russ.)
9. Epov M. M., Molodin V. M., Manstein L. K., Balkov E. V., Dyadkov P. G., Matasova G. G., Kazansky L. Yu., Bortnikova S. B., Pozdnyakova O. A., Karin Yu. G., Kuleshov D. A. Multidisciplinary archaeological and geophysical research in Western Siberia. *Geology and Geophysics*, 2016, vol. 57, no. 3, p. 603–614. (in Russ.)
10. Firsov A. P., Zlygostev I. N., Dyadkov P. G., Savluk A. V., Vaisman P. A., Sheremet A. S., Wald A. K., Evmenov N. D. Application of high-frequency magnetometer for light UAVs in geological and geophysical study of explosion tubes. In: Interexpo GEO-Siberia-2015. The XI Intern. Science. Cong. In 4 vols. Novosibirsk, 2015, vol. 2, p. 299–304. (in Russ.)
11. Nelder J. A., Mead R. A Simplex Method for Function Minimization. *Computer Journal*, 1965, vol. 7, p. 308–313.

Материал поступил в редколлегию

Received
25.11.2019

Сведения об авторах

Евменов Николай Денисович, аспирант лаборатории естественных геофизических полей Института нефтегазовой геологии и геофизики им. А. А. Трофимука СО РАН (Новосибирск, Россия)
evmenovnd@ipgg.sbras.ru
ORCID 0000-0003-0164-8610

Воронин Владислав Владимирович, кандидат физико-математических наук, доцент Новосибирского государственного университета (Новосибирск, Россия)
vladvor1948@bk.ru
ORCID 0000-0002-1727-1873

Дядьков Петр Георгиевич, кандидат геолого-минералогических наук, доцент, заведующий лабораторией естественных геофизических полей Института нефтегазовой геологии и геофизики им. А. А. Трофимука СО РАН (Новосибирск, Россия)
DyadkovPG@ipgg.sbras.ru
ORCID 0000-0001-5345-9331

Позднякова Ольга Анатольевна, научный сотрудник Института археологии и этнографии СО РАН (Новосибирск, Россия)
olka.pozdnyakova@gmail.com
ORCID 0000-0001-8376-0344

Information about the Authors

Nikolay D. Evmenov, Graduate Student at the Laboratory of Natural Geophysical Fields of the Institute of Petroleum Geology and Geophysics named after A. A. Trofimuk SB RAS (Novosibirsk, Russian Federation)
evmenovnd@ipgg.sbras.ru
ORCID 0000-0003-0164-8610

Vladislav V. Voronin, PhD, Associate Professor, Novosibirsk State University (Novosibirsk, Russian Federation)
vladvor1948@bk.ru
ORCID 0000-0002-1727-1873

Petr G. Dyadkov, PhD, Associate Professor, Head of the Laboratory of Natural Geophysical Fields of the Institute of Petroleum Geology and Geophysics named after A. A. Trofimuk SB RAS (Novosibirsk, Russian Federation)
DyadkovPG@ipgg.sbras.ru
ORCID 0000-0001-5345-9331

Olga A. Pozdnyakova, Researcher, Institute of Archeology and Ethnography SB RAS (Novosibirsk, Russian Federation)
olka.pozdnyakova@gmail.com
ORCID 0000-0001-8376-0344

УДК 681.518:622.276
DOI 10.25205/1818-7900-2020-18-1-27-35

Об одном методе классификации нефтяного месторождения с использованием комплекса геолого-промысловых данных и машинного обучения

Д. В. Курганов

*Самарский государственный технический университет
Самара, Россия*

Аннотация

Применение алгоритмов машинного обучения в частности для классификации и зонирования месторождений нефти и газа, является перспективным направлением при анализе разработки месторождений. Существенным условием применения метода является наличие обширной цифровой базы с представительными результатами. В работе рассмотрено применение метода кластеризации k -средних для крупного месторождения в Западной Сибири, а также проведен анализ эффективности применяемых гидроразрывов пласта. Методика позволяет комплексировать такие параметры, как плотность текущих запасов, географическое расположение скважин, параметры проведенных гидроразрывов пласта, пористость, нефтенасыщенность. В целом это позволяет учесть максимум геолого-промысловой информации, создавая предпосылки к автоматизации управления разработкой нефтяного месторождения. Также существенно увеличивается точность прогноза продуктивности коллектора, особенно имеющего сложное строение, в том числе техногенно модифицированного.

Ключевые слова

машинное обучение, классификация, дебит, метод k -средних, нефть, скважина, выборка, месторождение, запасы, гидроразрыв пласта, продуктивность, диаграмма Вороного

Для цитирования

Курганов Д. В. Об одном методе классификации нефтяного месторождения с использованием комплекса геолого-промысловых данных машинного обучения // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 27–35. DOI 10.25205/1818-7900-2020-18-1-27-35

Oil Reservoir Classification by Geological and Production Data Using Unsupervised Machine Learning Algorithm

D. V. Kurganov

*Samara State Technical University
Samara, Russian Federation*

Annotation

In machine learning, k -means unsupervised model is used for classification analysis. In this paper k -means model is applied for productivity prediction of giant Western Siberian oilfield. An essential condition for method's application is availability of digital databases with representative results. Complex method allows combine different reservoir and production parameters: rates, porosity, saturation, frac parameters etc. The method can be particularly useful in complicated reservoirs, e.g. in dual porosity ones, where the relationship between formation parameters (permeability, porosity, saturation) and production rates is unclear and cannot be set by traditional development analysis, particularly in frac environment.

Keywords

machine learning, classification, rate, k -means, oil, well, reservoir, oil in place, sample, frac, productivity, Voronoi diagram

© Д. В. Курганов, 2020

For citation

Kurganov D. V. Oil Reservoir Classification by Geological and Production Data Using Unsupervised Machine Learning Algorithm. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 27–35. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-27-35

Машинное обучение (МО) в настоящее время является перспективным методом, получившим стремительное развитие в нефтегазовой отрасли [1].

Наряду с накоплением фактического и исторического материала по эксплуатации скважин и месторождений появляются новые возможности для установления различных закономерностей и обобщений, связанных с конкретными залежами или типами залежей. Комплекс геолого-промысловых данных, имеющийся на месторождении, подразумевает проведение определенного анализа известными методами МО [2]. В частности, имеется возможность применить к ним известные алгоритмы классификации для выявления зон с совокупностью определенных характеристик, непосредственно влияющих на продуктивность скважин [3]. В отличие от стандартного картирования свойств и дальнейшего анализа набора карт, применяемого в отрасли, данные методы позволяют учесть сразу всю совокупность признаков, влияющих на потенциальную добычу углеводородов, что позволяет значительно упростить процедуру размещения проектных скважин для качественной выработки запасов.

Кластеризация методом k -средних [4] является одним из так называемых неконтролируемых методов МО, в котором исходные данные разбиваются на группы по совокупности определенных признаков.

Процедура кластеризации выглядит следующим образом.

- Первоначально центры кластеров выбираются случайно.
- От каждой точки в пространстве параметров рассчитывается Евклидово расстояние до центра ближайшего кластера.
- Итеративно минимизируется сумма расстояний от каждой точки до центра соответствующего кластера.

Формально задача ставится следующим образом. Пусть имеется m наблюдений (из пространства R^n). Необходимо разбить эти наблюдения на k кластеров, при этом каждое наблюдение относится к тому кластеру, к центру (так называемому центроиду) которого оно ближе всего.

В качестве меры близости используется Евклидово расстояние:

$$\rho(x, y) = \|x - y\| = \sqrt{\sum_{p=1}^n (x_p - y_p)^2},$$

где $x, y \in R^n$.

Таким образом, для ряда наблюдений $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$, $x^j \in R^n$ метод k -средних разделяет m наблюдений на k групп (или кластеров) ($k \leq m$) $S = \{S_1, S_2, \dots, S_k\}$, таким образом, чтобы минимизировать суммарное квадратичное отклонение точек кластеров от центроидов этих кластеров:

$$\min \left[\sum_{i=1}^k \sum_{x^{(j)} \in S_i} \|x^{(j)} - \mu_i\|^2 \right],$$

где $x^{(j)} \in R^n$, $\mu_i \in R^n$, μ_i – центроид для кластера S_i .

Если мера близости до центроида определена, то разбиение объектов на кластеры сводится к определению центроидов этих кластеров. Число кластеров k задается заранее.

Рассмотрим первоначальный набор k -средних (центроидов) $\mu_1, \dots, \mu_k$ в кластерах $S_1, \dots, S_k$. Первоначально центроиды кластеров выбираются случайно или по определенному правилу (например, можно выбрать центроиды, максимизирующие начальные расстояния между кластерами).

Далее следует отнести наблюдения к тем кластерам, чье среднее (центроид) к ним ближе всего. Каждое наблюдение принадлежит только к одному кластеру, даже если его можно отнести к двум и более кластерам.

Затем центроид каждого i -го кластера перевычисляется по следующему правилу:

$$\mu_j = \frac{1}{S_j} \sum_{x^{(j)} \in S_j} x^{(j)}.$$

Таким образом, алгоритм k -средних заключается в перевычислении на каждом шаге центроида для каждого кластера, полученного на предыдущем шаге. Алгоритм останавливается, когда значения μ_i перестают меняться.

Основной проблемой данного алгоритма является то, что оптимальное значение кластеров заранее неизвестно. Одной из методик для его нахождения является определение так называемого *силуэта* выборки [5]. Данный коэффициент не предполагает знания истинных меток объектов и позволяет оценить качество кластеризации, используя только саму неразмеченную выборку и результат кластеризации.

Пусть a – среднее расстояние от данного объекта до объектов из того же кластера, b – среднее расстояние от данного объекта до объектов из ближайшего кластера (отличного от того, в котором лежит сам объект). Тогда силуэтом данного объекта называется величина

$$s = (b - a) / \max(a, b).$$

Силуэтом выборки называется средняя величина силуэта объектов данной выборки. Таким образом, силуэт показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Данная величина лежит в диапазоне $[-1, 1]$. Значения, близкие к -1 , соответствуют плохим (разрозненным) кластеризациям; значения, близкие к нулю, говорят о том, что кластеры пересекаются и накладываются друг на друга; значения, близкие к 1 , соответствуют «плотным» четко выделенным кластерам. Таким образом, чем больше силуэт, тем более четко выделены кластеры, и они представляют собой компактные, плотно сгруппированные облака точек.

Также для дальнейшего изложения необходимо привести определение диаграммы Вороного [6]. Диаграмма Вороного P_i , построенная для точки измерений x_i , характеризуется тем, что содержит те и только те точки, расстояние от которых до точки x_i меньше или равно расстоянию до любой другой точки измерений x_j . При построении диаграмм Вороного используется система соседства, полученная в процессе триангуляции Делоне. Границы диаграмм Вороного P_i состоят из отрезков серединных перпендикуляров, проведенных к сторонам треугольников Делоне.

Применительно к разработке месторождений границы диаграммы Вороного могут рассматриваться как границы зон дренирования скважины.

В качестве примера взято одно из месторождений Западной Сибири, пласт БВ₁₃. Залежь пластово-сводовая, строение сложное. Пласт характеризуется невысокой проницаемостью, значительной изменчивостью параметров по площади. На объекте пробурено 180 скважин (87 действующих добывающих, 50 действующих нагнетательных), реализована очаговая система заводнения, всего было проведено около 90 гидроразрывов пласта (ГРП). Текущая степень выработки 67 %, средняя обводненность продукции – 88 %. Средний дебит скважин по жидкости составляет 60 т/сут. Нефть маловязкая, с высоким газосодержанием. Стандартные методы анализа процесса разработки, в частности фильтрационное моделирование, не

позволяют детально раскрыть процессы, происходящие в пластовой системе. Адаптацию модели нельзя признать удовлетворительной – по-видимому, в процессе проведения многочисленных ГРП в пластовой системе образовалась вторая среда (система трещин), течение флюидов в которой не описывается стандартными законами фильтрации [7].

Для анализа и классификации были выбраны следующие данные: интерпретированный каротаж (пористость, насыщенность, эффективная толщина), добыча первого года промышленной эксплуатации каждой скважины, информация об использованном при ГРП количестве проппанта и разрывающей жидкости. Это важнейшие количественные характеристики процесса ГРП, которые косвенно позволяют судить о его эффективности по данной скважине. В выборку попали скважины, по которым был проведен ГРП после начала закачки на месторождении.

В дополнение к скважинным данным была рассчитана плотность текущих запасов по каждой диаграмме Вороного. Текущие запасы были рассчитаны на основе начальных запасов за вычетом накопленных отборов по каждой скважине [8].

Коллектор был зонально разделен на три типа посредством кластеризации (Класс А, Класс Б и Класс С). С помощью расчета силуэта было установлено, что добавление четвертого класса не оказывает существенного влияния на понимание характеристик пласта и характер разработки месторождения.

В качестве параметров для кластеризации были выбраны следующие: количество проппанта, количество жидкости разрыва, площадное месторасположение скважин, плотность текущих удельных запасов на скважину по диаграмме Вороного, а также величина, характеризующая изначальное качество коллектора в смысле потенциального содержания запасов – произведение начальной нефтенасыщенности, пористости и эффективной толщины ($S_0 \cdot \phi \cdot H_{\text{eff}}$). На рис. 1 приведен результат работы алгоритма с выделением скважин по описанной совокупности признаков. Рисунок 2 показывает примерное эмпирическое зонирование месторождения на основе разбиения по классам скважин, рис. 3 иллюстрирует распределение признаков по классам. Средние значения признаков приведены на рис. 4. Следует отметить, что все исходные данные прошли процедуру нормализации (приведены к интервалу [0, 1] для устранения эффекта масштабирования. Северная зона (розовый цвет) характеризуется худшими показателями, центральная зона – удовлетворительными, южная – лучшими показателями (см. рис. 2). Все распределения параметров, так же как и средние значения, улучшаются от северной части к южной. Особенно важным представляется тот факт, что лучшая зона характеризуется большими трещинами (а точнее, большим количеством проппанта и разрывающей жидкости). Добывающие скважины в коллекторе лучшего качества и с большими трещинами в среднем характеризуются большей добычей.

При этом следует отметить, что в процессе кластеризации начальные дебиты нефти в качестве параметров не использовались, чтобы более точно выявить влияние всех остальных параметров пластовой системы на фактическую добычу.

В среднем добыча нефти скважин класса В превышает добычу скважин класса А в 1,5 раза, а добыча скважин класса С превышает добычу скважин класса В в 1,3 раза (рис. 5).

Чтобы исследовать влияние ГРП на добычу скважин, необходимо выбрать лишь те скважины, которые находятся в качественном коллекторе, исключив из анализа остальные скважины, а именно те, у которых значение ($S_0 \cdot \phi \cdot H_{\text{eff}}$) меньше среднего по классу С. При этом средним значением является 0,7. На рис. 6 цветом выделены скважины, попавшие в качественный коллектор.

Рисунок 7 иллюстрирует влияние параметров проведенных гидроразрывов на отборы нефти. Видно, что наличие качественного коллектора не гарантировало больших дебитов. В то же время большие объемы проппанта и разрывающей жидкости вели в среднем к увеличенным отборам.

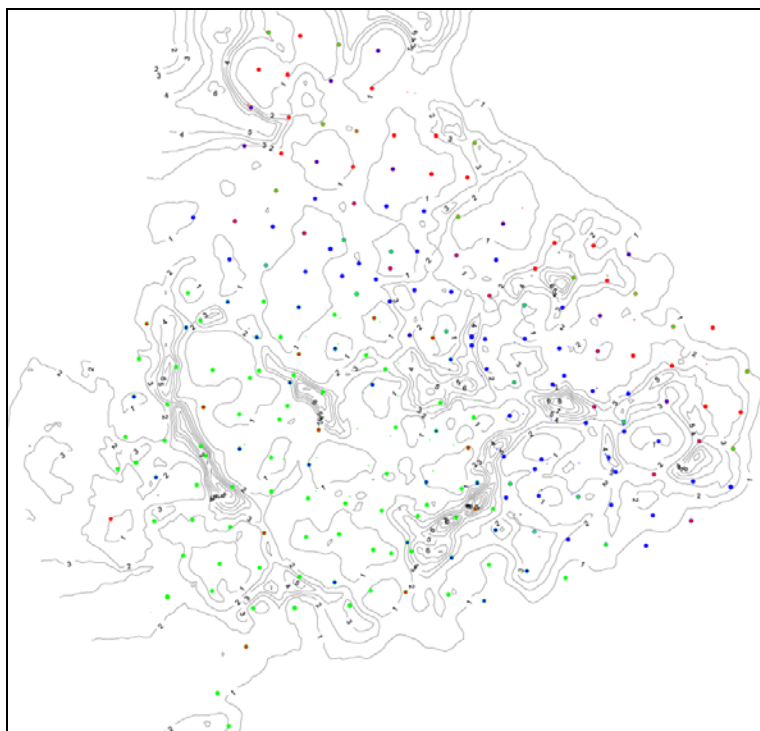


Рис. 1. Фрагмент месторождения с выделенными скважинами по методу k -средних (красные – зона А, синие – зона В, зеленые – зона С)

Fig. 1. Part of the oilfield with colored wells ranked by k -means algorithm (red – zone A, blue – zone B, green – zone C)

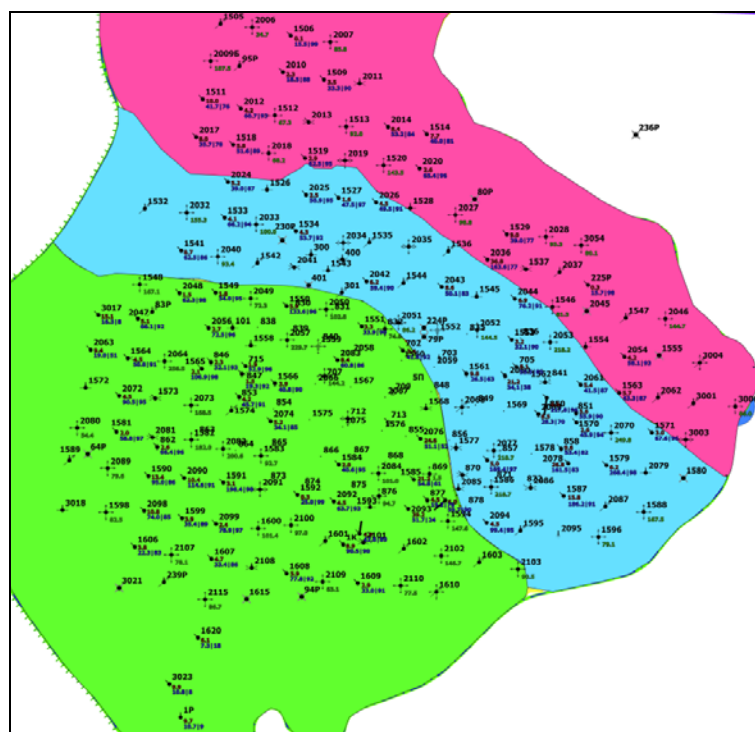


Рис. 2. Фрагмент месторождения с выделенными зонами по методу k -средних

Fig. 2. Part of the oilfield with zones ranked by k -means algorithm

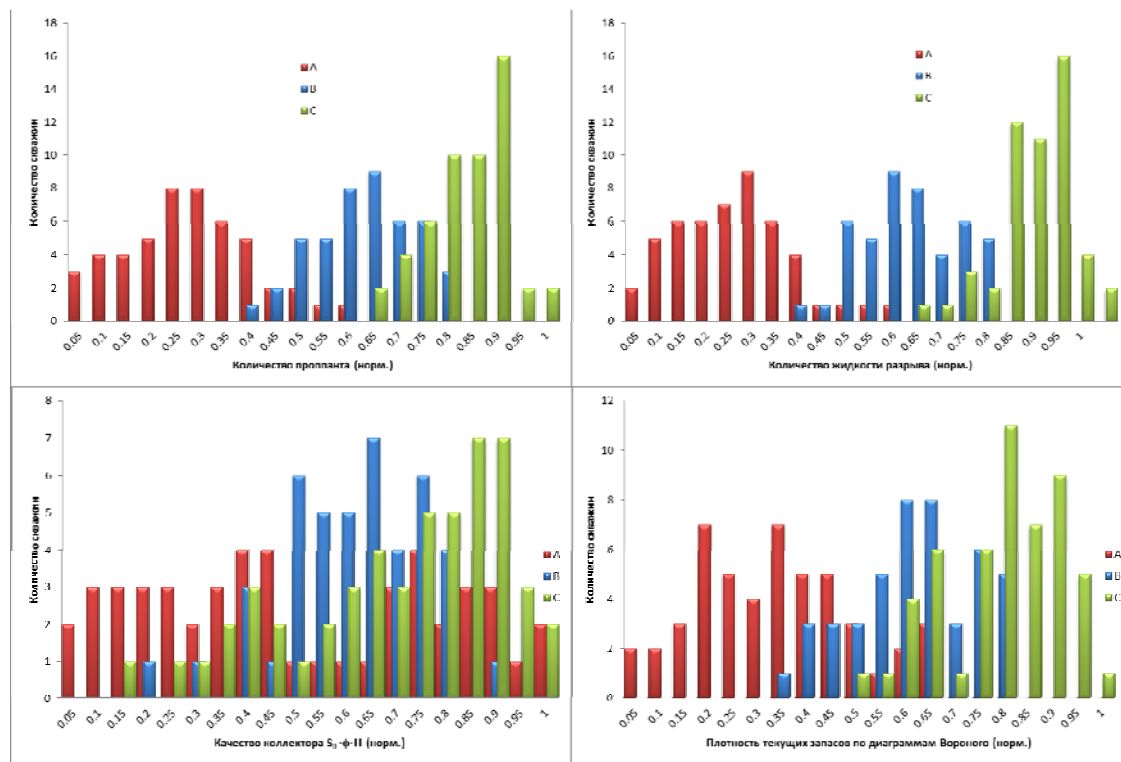


Рис. 3. Распределения нормализованных параметров модели кластеризации

Fig. 3. Distribution of cluster model parameters

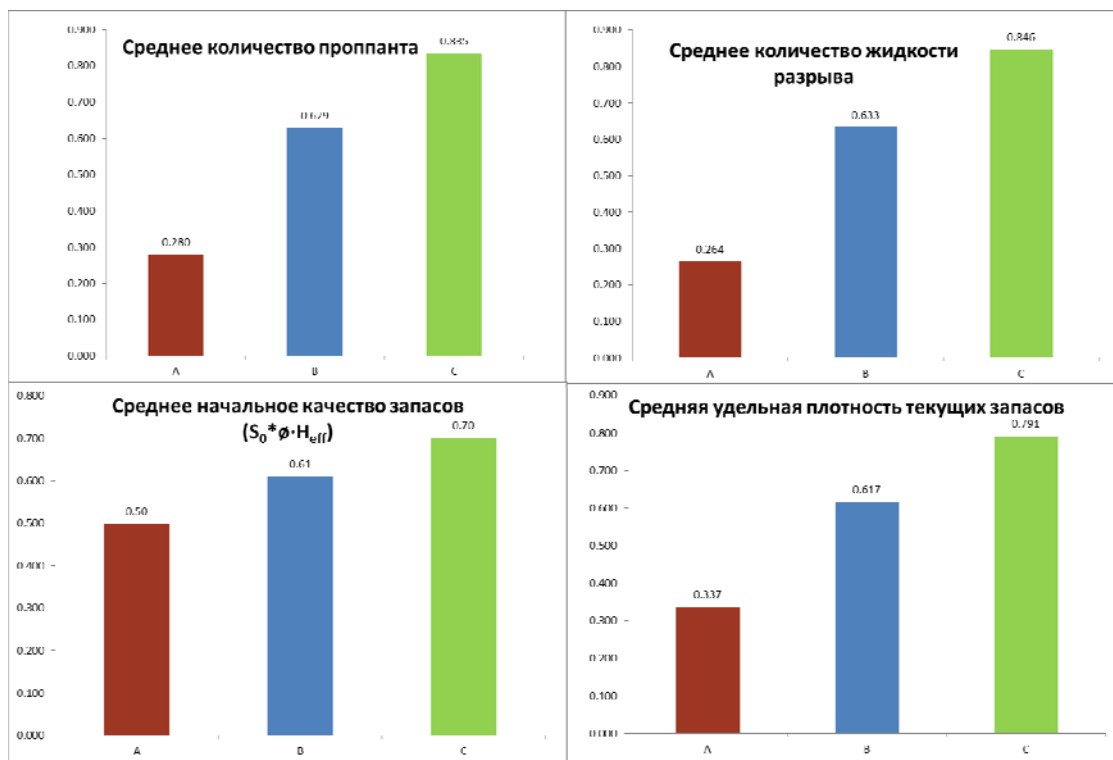


Рис. 4. Средние значения параметров по классам коллектора (нормированные)

Fig. 4. Average parameters distribution classified by reservoir type

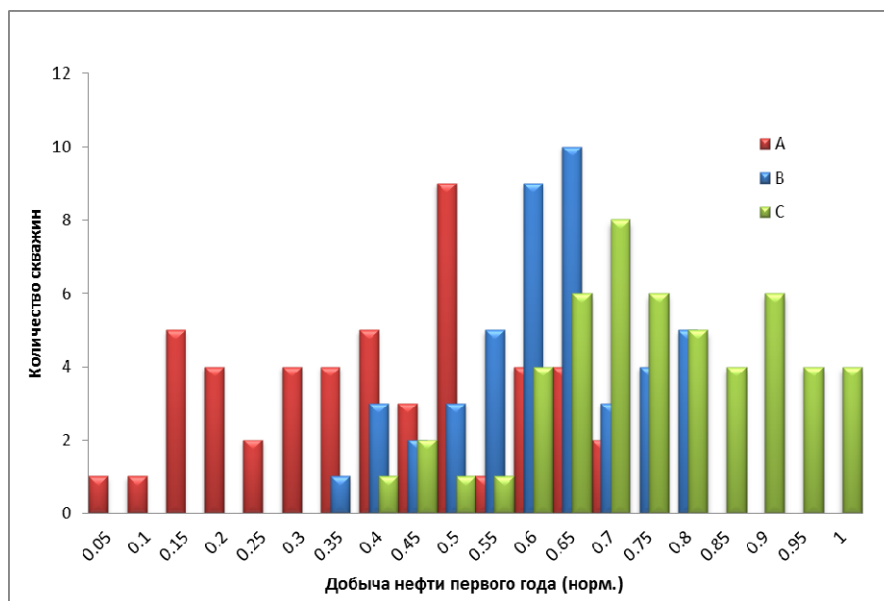


Рис. 5. Распределение дебитов по классам коллектора

Fig. 5. Oil rates classified by reservoir type

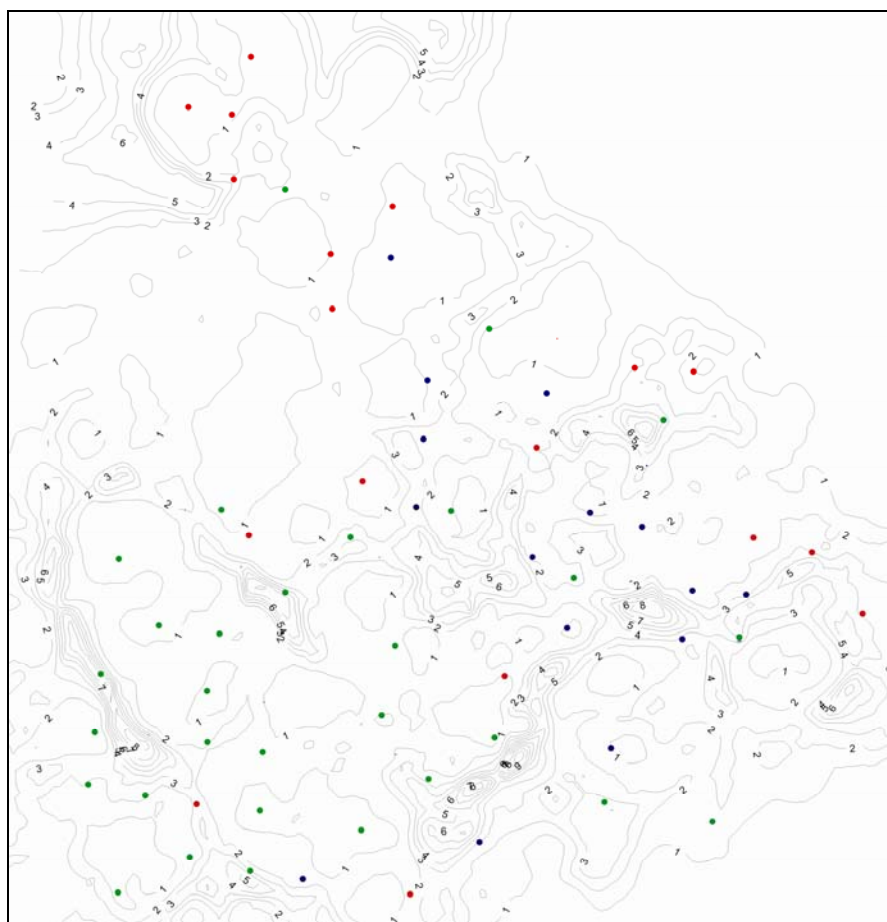


Рис. 6. Распределение скважин, попавших в качественный коллектор

Fig. 6. Well locations of supreme quality reservoir

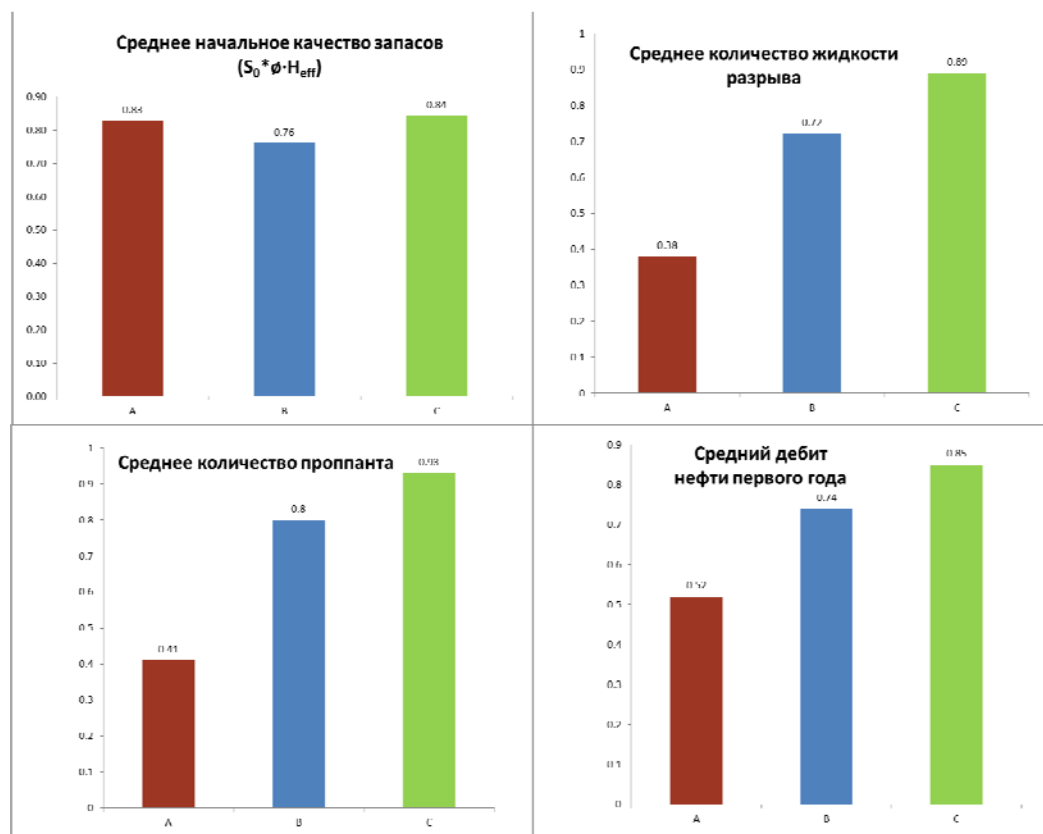


Рис. 7. Распределение параметров модели по скважинам качественного коллектора
 Fig. 7. Average parameters distribution of supreme quality reservoir

Выполненная классификация позволяет оценить различные зоны месторождения на предмет его дальнейшего разбуривания и потенциальной продуктивности с учетом проводимых гидроразрывов пласта, позволяет выявлять значимые параметры и объемы проведения геолого-технических мероприятий, влияющих на продуктивность скважин. Также описанная методика и найденные классы могут использоваться, например, для контролируемого машинного обучения при создании моделей уплотняющего бурения. Обучив модель на найденных образцах, можно прогнозировать продуктивность новых скважин и зон месторождения, в том числе осложненных средой с двойной пористостью [9], принимать управляющие решения для дальнейшей разработки.

Список литературы

1. **Zangl G., Hannerer J.** Data Mining: Applications in the Petroleum Industry. Katy, TX, Round Oak Publ., 2003, 222 p.
2. **Hand D., Manilla H., Smyth P.** Principles of Data Mining. MIT Press, 2001, 546 p.
3. **Курганов Д. В.** О численном решении одной задачи минимизации в моделировании пластовых систем // Сибирский журнал индустриальной математики. 2003. Т. 6, № 1. С. 51–59.
4. **Han J., Kamber M., Pei J.** Data Mining: Concepts and Techniques. USA, Elsevier, 2012, 703 p.
5. **Wu X., Kumar V.** The Top Ten Algorithms in Data Mining. USA, Taylor & Francis Group, 2009, 201 p.

6. Демьянов В. В., Савельева Е. А. Геостатистика: теория и практика. М.: Наука, 2010. 327 с.
7. Курганов Д. В., Попков В. И., Хмелевских Е. И. Оценка влияния низкопроницаемых прослоев коллектора на эффективность выработки запасов нефти с использованием трехмерных гидродинамических моделей // Изв. Самарского Научного центра РАН. 2002. Спец. выпуск: Проблемы нефти и газа. С. 121–126.
8. Колесов В. В., Курганов Д. В. Интегрированное моделирование неоднородности и связности гигантского карбонатного резервуара по геофизическим и промысловым данным на поздней стадии разработки с целью локализации остаточных запасов нефти // Материалы технической конференции SPE «Разработка зрелых месторождений». М., 2017. С. 121–129.
9. Смехов Е. М., Дорофеева Т. В. Вторичная пористость горных пород – коллекторов нефти и газа. Л.: Недра, 1987. 96 с.

References

1. Zangl G., Hannerer J. Data Mining: Applications in the Petroleum Industry. Katy, TX, Round Oak Publ., 2003, 222 p.
2. Hand D., Manilla H., Smyth P. Principles of Data Mining. MIT Press, 2001, 546 p.
3. Kurganov D. V. Numerical solution of one minimization problem in reservoir simulation. *Sibirskiy zhurnal industrial'noy matematiki*, 2003, vol. 6, no. 1, p. 51–59. (in Russ.)
4. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. USA, Elsevier, 2012, 703 p.
5. Wu X., Kumar V. The Top Ten Algorithms in Data Mining. USA, Taylor & Francis Group, 2009, 201 p.
6. Demiyarov V. V., Savelyeva E. A. Geostatistika: teoriya i praktika. Moscow, Nauka, 2010, 327 p. (in Russ.)
7. Kurganov D. V., Popkov V. I., Khmelevskikh E. I. Estimation of low permeability layers on field development process using reservoir simulation. *Izvestiya Samarskogo Nauchnogo tsentra RAN*, 2002, Special issue, p. 121–126. (in Russ.)
8. Kolesov V. V., Kurganov D. V. Integrated modeling of heterogeneity and conductivity of giant mature carbonate reservoir using complex geophysic and production data. In: Proc. of SPE Conference “Razrabotka zrelykh mestorozhdeniy”. Moscow, 2017, p. 121–129. (in Russ.)
9. Smekhov E. M., Dorofeeva T. V. Dual porosity oil and gas reservoirs. Leningrad, Nedra Publ., 1987, 96 p. (in Russ.)

Материал поступил в редколлегию

Received
04.09.2019

Сведения об авторе

Курганов Дмитрий Владимирович, кандидат физико-математических наук, доцент, доцент кафедры разработки и эксплуатации нефтяных и газовых месторождений, Самарский государственный технический университет (Самара, Россия)
Dmitri.Kourganov@inbox.ru

Information about the Author

Dmitri V. Kurganov, PhD, Smara State Technical University (Samara, Russian Federation)
Dmitri.Kourganov@inbox.ru

УДК 519.63, 004.942, 622.245
DOI 10.25205/1818-7900-2020-18-1-36-49

Модель распространения трещины вдоль гидроизоляции скважины

В. Н. Лапин

*Институт вычислительных технологий СО РАН
Новосибирск, Россия*

Аннотация

Предлагается модель инициации и распространения трещины вдоль границы цементной пробки и стальной обсадной колонны в закупоренной скважине, предназначенная для оценки способности пробки противостоять повышению пластового давления. Модель описывает изменение напряженно-деформированного состояния цемента и его разрушение на границе со стальной обсадной колонной под действием давления затекающей в трещину поровой жидкостью, вызванного повышением пластового давления внизу пробки. На основе серии вычислительных экспериментов сделаны выводы о чувствительности скорости распространения трещины, а значит, и стойкости гидроизоляции скважины к значению пластового давления и параметрам преднапряженного состояния цемента. Показано, что качество прилипания цемента к стали и размеры начальных дефектов цементирования слабо влияют на условия инициации и скорость распространения трещины.

Ключевые слова

распространение трещин, течение жидкости, гидроизоляция скважин

Для цитирования

Лапин В. Н. Модель распространения трещины вдоль гидроизоляции скважины // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 36–49. DOI 10.25205/1818-7900-2020-18-1-36-49

Model of Fracture Propagation in Hydraulic Isolation of the Wellbore

V. N. Lapin

*Institute of Computational Technologies SB RAS
Novosibirsk, Russian Federation*

Annotation

A model of fracture initiation and propagation along the boundary of a cement plug and a steel casing in a plugged wellbore is proposed. The model allows predicting the conditions of hydraulic isolation failure of a cement plug induced by reservoir pressure. The model describes the stress state of the cement and its breakaway from the steel casing caused by the pressure of the pore fluid flowing into the fracture. Numerical experiments show that the fracture propagation velocity and the stability of the hydraulic isolation are sensitive to the reservoir pressure and to the pre-stressed state of the cement. It is shown also that the quality of cement adhesion to steel and the size of the initial cementing defects have little effect on the initiation conditions and fracture propagation velocity.

Keywords

fracture propagation, fluid flow, wellbore hydraulic isolation

For citation

Lapin V. N. Model of Fracture Propagation in Hydraulic Isolation of the Wellbore. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 36–49. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-36-49

© В. Н. Лапин, 2020

Введение

При эксплуатации нефтяных скважин часто возникает необходимость изолировать нефтеносные слои от водоносных, а также отделить нефтеносные слои друг от друга. Кроме того, гидроизоляция необходима при консервации скважины или при использовании истощенных месторождений для консервирования углекислого газа, призванного повлиять на его количество в атмосфере и уменьшить парниковый эффект [1]. Для этого применяется технология цементирования скважин, согласно которой внутрь обсадной колонны скважины помещается цементный раствор, образующий после застывания цементную пробку. Развитие добычи сланцевых газа и нефти, требующей большого количества как бурения скважин, так и их последующей консервации, обуславливает необходимость исследования свойств и надежности гидроизоляции. При этом из-за высокой цены специального оборудования и материалов необходимо соблюдать баланс между стоимостью закупоривания скважины и надежностью гидроизоляции [2]. Хотя в настоящее время существуют попытки стандартизировать размеры пробки, параметры цемента и другие параметры цементирования, в них часто не учитываются такие особенности пластов, как температура, пластовое давление и т. д., которые могут существенно влиять на надежность гидроизоляции [3]. Несмотря на то что пробка предназначена для полного изолирования нижней и верхней частей скважины друг от друга, потенциально возможны несколько путей перетока жидкости или газа между ними. Эти пути схематично показаны на рис. 1, заимствованном из работы [1], и включают в себя следующие: через микроканалы на границах стальной обсадной колонны и цемента внутри нее или снаружи (*a, b*), на границе цемента с породой (*f*) или путем фильтрации через материал пробки (*c*).

Настоящая работа посвящена моделированию образования и распространения трещины на стыке обсадной колонны и пробки, которая приводит к образованию микроканалов, соответствующих случаю (*b*) на рис. 1. Предполагается, что пробка находится под действием пластового давления, приложенного к нижнему торцу, и давления в скважине, приложенного к верхнему торцу. Изменения этих давлений, а также других параметров среды, окружающей пробку, могут привести к нарушению целостности пробки или нарушению ее связи с обсадной колонной и прорыву жидкости из одной части скважины в другую. В работах [4–6] приводится количественный анализ риска утечки жидкости в скважину. Авторы работ утверждают, что трещины образуются только внутри обсадной колонны, в канале, образовавшемся в случае неудачной заливки цемента внутри цемента, или, что наиболее распространено, на границе обсадной колонны и цементной пробки либо цементной оболочки и породы. В работе [7] на основе лабораторных и полевых тестов также было получено, что основной причиной утечек жидкости в скважину является отслоение цемента от стальной колонны. Кроме того, модель, разработанная для описания распространения трещины по внутренней границе обсадной колонны, может быть после незначительных модификаций применена для описания трещины, распространяющейся и по ее внешней стороне.

Предполагается, что численное моделирование позволит ответить на следующие вопросы, касающиеся особенностей распространения трещины по границе цементной пробки и стальной обсадной колонны, ответы на которые могут быть использованы при разработке рекомендаций для инженеров, проектирующих гидроизоляцию скважин.

1. Если на нижней стороне пробки есть дефекты цементирования, неплотное прилипание цемента к стальной колонне и т. д., то имеет ли значение размер этого дефекта? Необходимы ли дополнительные меры по укреплению нижней части пробки, или достаточно избегать крупных дефектов цементирования?

2. Если все-таки произошло образование канала по границе цемента и обсадной колонны, то какую форму канала надо использовать при расчете объема протекания жидкости через гидроизоляцию? Происходит ли полный отрыв пробки от обсадной колонны, или канал занимает лишь часть сечения?

3. Какие параметры среди тех, что могут контролироваться при проведении работ по закупориванию скважины, влияют на надежность гидроизоляции?

С точки зрения математического моделирования инициация и эволюция трещины, распространяющейся под действием давления жидкости, является сложной задачей, объединяющей подзадачи из различных разделов механики и вычислительных методов.

Первый и второй разделы работы посвящены описанию постановок задач, уравнений и граничных условий модели, а также ее валидации. В третьем разделе проведены несколько серий вычислительных экспериментов, которые позволили выявить особенности распространения трещины и чувствительность ее скорости распространения к основным параметрам задачи.

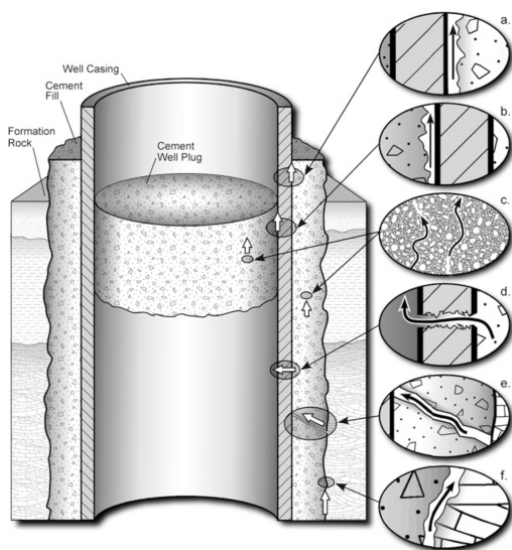


Рис. 1. Схема ствола скважины с гидроизоляцией и возможные причины утечки жидкости (заимствовано из [1])

Fig. 1. Schematics of possible leakage pathways along a well (in accordance with [1])

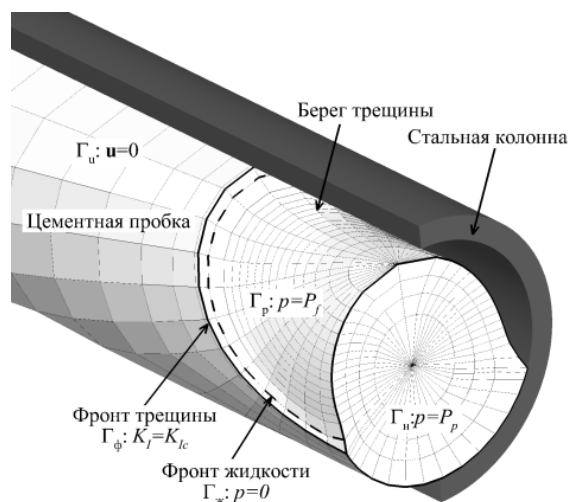


Рис. 2. Постановка задачи о распространении трещины, расчетная сетка и деформация пробки (смещение боковой поверхности увеличено в 1000 раз)

Fig. 2. Statement of the fracture propagation problem, computational mesh and plug deformation (surface displacements are multiplied by 1000)

1. Постановка задачи

Рассматривается распространение трещины по границе цементной пробки и стальной обсадной колонны, которое вызвано проникновением жидкости в трещину под действием по-

вышающего давления в скважине на нижнем торце пробки (см. рис. 1). При моделировании распространения трещины описывается деформация пробки, движение вязкой жидкости внутри нее и разрушение пробки на границе со стальной колонной.

Предполагается, что пробка помещается в скважину путем закачки заданного объема жидкого цемента с помощью давления, которое обеспечивает его продвижение на заданную глубину. Так как поровая жидкость из пласта свободно попадает в ствол скважины ниже пробки, то абсолютное давление на нижней границе пробки в момент ее помещения ($t = t_0$) совпадает с давлением пластовой жидкости $P_{п,0}^{abc}$. Распределение абсолютного давления в остальном цементе $P_{п,0}^{abc}(z)$ в этот момент определяется влиянием силы тяжести и равно

$$P_{п,0}^{abc}(z) = P_{п,0}^{abc}(z) - \rho_c gz.$$

После застывания цемента $t = t_1$ в нем возникает дополнительное давление P_c , и напряженно-деформированное состояние уже твердой пробки характеризуется шаровым тензором напряжений

$$\sigma_0^{abc} = -P_{п,1}^{abc}(z)I, \quad P_{п,1}^{abc}(z) = P_{п,0}^{abc}(z) + P_c = P_{п,0}^{abc}(z) - \rho_c gz + P_c.$$

При решении задачи о распространении трещины вдоль границы пробки это ее состояние будет считаться исходным, и изменение напряжений, деформаций и смещений материала будет рассчитываться относительно этого состояния. В результате фильтрации пластовой жидкости к скважине давление в ее окрестности повышается на некоторую величину P_p . Это вызывает отслаивание цемента от обсадной колонны и заполнение образовавшейся трещины пластовой жидкостью. В этот момент ($t = t_2$) на нижний торец пробки Γ_n и заполненную жидкостью область Γ_p (см. рис. 2), действуют давления, равные

$$\Gamma_n: P_{п,2}^{abc} = P_{п,0}^{abc} + P_p, \Gamma_p: P_{ж,2}^{abc} = P_{п,0}^{abc} - \rho_f gz + p,$$

Здесь давление жидкости в трещине $P_{ж,2}^{abc}$ представлено в виде суммы трех слагаемых: пластового давления на момент помещения цемента $P_{п,0}^{abc}$, гидростатической составляющей $\rho_f gz$ и, в общем случае, переменного, нелинейного распределения давления p , вызванного движением жидкости в трещине и удовлетворяющего условию $p|_{z=0} = P_p$.

Абсолютное поровое давление $P_{п,2}^{abc}$ намного превышает его вариации P_p , а влияние силы тяжести может быть добавлено и движения жидкости и в уравнения упругого равновесия благодаря их линейности. Поэтому для снижения погрешности вычислений при численном моделировании использовались относительные напряжения $\sigma_0^{отн} = \sigma^{abc} - \sigma_0^{abc}$, характеризующие изменение напряженного состояния пробки относительно состояния в момент застывания цемента $t = t_1$, и выше введенное давление p . Давления, действующие на нижний торец пробки и в области, где произошло отслоение, в терминах относительных напряжений $\sigma_0^{отн}$ записываются как

$$\Gamma_n: P_{п,2}^{отн} = P_{п,1}^{abc} - P_{п,1}^{abc}|_{z=0} = P_p - P_c, \\ \Gamma_{ж}: P_{ж,2}^{отн} = P_{ж,2}^{abc} - P_{п,1}^{abc} = p - P_c + (\rho_c - \rho_f)gz.$$

Эти давления используются для постановки задачи упругого равновесия, а задача движения жидкости будет решаться без учета силы тяжести. Символ «отн» будет далее для компактности опускаться.

Деформация материала

Деформация пробки описывается уравнениями упругого равновесия, которые в предположении однородности, изотропности и линейной упругости материала записываются как

$$\text{div} \sigma = 0, \quad \sigma = \frac{E\nu}{(1+\nu)(1-2\nu)} \text{Itr}(\epsilon) + \frac{E}{(1+\nu)} \epsilon, \quad \epsilon = \frac{1}{2} \left(\frac{\partial u}{\partial x} + \left(\frac{\partial u}{\partial x} \right)^* \right), \quad (1)$$

где σ и ϵ – тензоры напряжений и деформаций, u – вектор смещений, E, ν – модуль Юнга и коэффициент Пуассона соответственно.

Расчетная область представляет собой цилиндр (см. рис. 2), на нижнем Γ_n и верхнем Γ_e торцах которого задана нормальная к поверхности компонента тензора напряжений, равная поровому давлению жидкости в резервуаре (пластовому давлению) снизу P_p и сверху P_0 от пробки и давлению P_f движущейся жидкости в области, где произошел отрыв,

$$\Gamma_n: p = P_p, \Gamma_e: p = P_0, \Gamma_p: p = P_f. \quad (2)$$

Во всех случаях силами трения жидкости и цемента пренебрегается, и нормальная к поверхности компонента тензора напряжений задается равной нулю. В области Γ_u , где цемент остается в контакте с обсадной колонной, заданы нулевые смещения

$$\Gamma_u: u = 0. \quad (3)$$

Краевая задача (5)–(7) решается методом граничных элементов, реализация которого описана в [8].

В результате решения рассчитывается вектор смещений u , по которому вычисляется раскрытие трещины W в области Γ_p . Деформация стальной обсадной колонны считается несущественной, так как модуль Юнга стали значительно (до 10 раз) превосходит модуль Юнга цемента, что обуславливает пропорционально меньшие деформации.

Следует отметить, что в момент затвердевания цемента пробка находится в преднапряженном состоянии. Предполагается, что это напряженное состояние обуславливается гидростатическим давлением $\rho_c g z$ цемента с плотностью ρ_c и дополнительным давлением, образующимся в результате гидратации цемента P_c . Поэтому избыточное давление жидкости в граничном условии (2) с учетом преднапряженного состояния рассчитывается как

$$P_f(z) = p - P_c + (\rho_c - \rho_f)gz, \quad (4)$$

где p – давление жидкости, полученное из решения уравнений ее движения без учета влияния силы тяжести, ρ_f – плотность жидкости, а уровень $z = 0$ соответствует нижнему торцу пробки.

Движение жидкости

Жидкость считается ньютоновской и несжимаемой. Предполагается, что высота канала, соответствующая раскрытию трещины, намного меньше ее длины и ширины, поэтому движение жидкости описывается уравнениями смазки. Они состоят из уравнения для давления p , являющегося следствием решения задачи о течении Пуазейля, и уравнения неразрывности

$$\nabla p = -\frac{12\mu}{W^3} \mathbf{q}, \frac{\partial W}{\partial t} + \text{div } \mathbf{q} = 0. \quad (5)$$

Здесь μ – вязкость жидкости, $\mathbf{q} = W\mathbf{v}$ – вектор потока жидкости, $\mathbf{v}$ – скорость движения жидкости. Расчетная область для уравнений движения жидкости соответствует области отрыва цемента от стальной колонны (см. рис. 2). На фронте жидкости $\Gamma_{жс}$ и входной границе ставятся условия для давления

$$\Gamma_{жс}: p = 0, \Gamma_n: p = P_p \quad (6)$$

Уравнения (5) с граничными условиями (6) решаются методом конечных элементов, описанном в [9]. Область течения жидкости разбивается на элементы, каждый из которых считается локально плоским. Затем внутри каждого из них вводится локальная система координат, и функции давления и ширины выражаются в виде суммы базисных функций элемента с неизвестными коэффициентами. Из уравнения (5), переписанного в слабой формулировке, получается система линейных уравнений для коэффициентов разложения. Особенностью применяемого метода является то, что его элементы и базисные функции совпадают с используемыми в методе граничных элементов, который применяется для решения уравне-

ний упругого равновесия [9]. Это позволяет избежать ошибок аппроксимации при передаче значений функций между задачами деформации и течения жидкости.

Положение фронта жидкости $\Gamma_{ж}$ в каждый момент времени определяется из условия Стефана, согласно которому скорость жидкости на фронте равна скорости движения фронта. Настоящая модель трещины, как и модель трещины гидроразрыва, описанная в [9], предполагает наличие отставания фронта жидкости $\Gamma_{ж}$ от фронта трещины Γ_{ϕ} (см. рис. 2). В этой области жидкость отсутствует, поэтому ее абсолютное давление полагается равным нулю.

Критерий разрушения

В качестве критерия разрушения используется классический силовой критерий хрупкого разрушения, предложенный Ирвином [10], согласно которому трещина распространяется, когда коэффициент интенсивности напряжений (КИН) на ее фронте достигает критического значения для рассматриваемого материала. Поэтому на фронте трещины Γ_{ϕ} ставится условие

$$K_I = K_{Ic}, \quad (7)$$

которое и является условием для определения формы фронта трещины на каждом шаге ее распространения. Поскольку модули Юнга стали и цемента отличаются, как минимум, на порядок, предполагается, что обсадная колонна является абсолютно жестким телом.

Это исключает наложение берегов трещины друг на друга и позволяет при расчете КИН учитывать только смещения цемента. КИН вычисляются по раскрытию трещины в окрестности фронта на основе интерполяционных формул, описанных, например, в [10], которые дополнены условием неподвижности одного из берегов трещины. При определении критического значения КИН для трещины, распространяющейся по границе материалов, следует учитывать свойства каждого из материалов и величину сил их сцепления. Однако так как в распоряжении автора нет достаточной информации об этих свойствах, то предполагалось, что значение K_I может принимать любое значение от 0 до критического значения КИН для цемента. Ниже будет показано, что этот параметр слабо влияет на способность пробки противостоять давлению жидкости, поэтому можно ожидать, что приведенные выше упрощения незначительно сказываются на результатах моделирования.

2. Валидация

Ввиду сложности получения полной информации о поведении закупоренной скважины, находящейся в реальном пласте, для валидации построенной модели был выбран лабораторный эксперимент, описанный в [11], в котором трещина распространялась по внешней стороне обсадной колонны, что позволяло визуально наблюдать за процессом. Лабораторные образцы были изготовлены путем «цементирования» алюминиевой ($E=69\text{ ГПа}$, $\nu=0.33$) трубы с внутренним и внешним диаметрами $D_{in}=28\text{ мм}$ и $D_s=40\text{ мм}$, соответственно, имитирующей обсадную колонну ствола скважины, в блоке полиметилметакрилата ($E=3.3\text{ ГПа}$, $\nu=0.35$). «Цемент» представлял собой эпоксидную систему ($E=2.2\text{ ГПа}$, $\nu=0.35$). Размер блока составлял $175 \times 155 \times 140\text{ мм}$, что соответствует удалению границ от скважины примерно на четыре диаметра. Вязкость закачиваемой жидкости равнялась $\mu=11.9\text{ Па}\cdot\text{с}$, а давление на нижнем торце пробки было зафиксировано равным $P_p=10\text{ МПа}$. Трещина распространялась вдоль границы цемент-порода от нижнего конца пробки к ее верхнему концу. В [11] на основе безразмерного анализа утверждается, что параметры эксперимента соответствуют характерным параметрам реальных скважин, а размер блока достаточен, чтобы результаты эксперимента можно было распространять на случай скважины в бесконечной области.

Формы трещин в различные моменты времени в ходе эксперимента показаны на рис. 3. Видно, что в эксперименте наблюдалось распространение трещины как в продольном, так и в окружном направлениях.

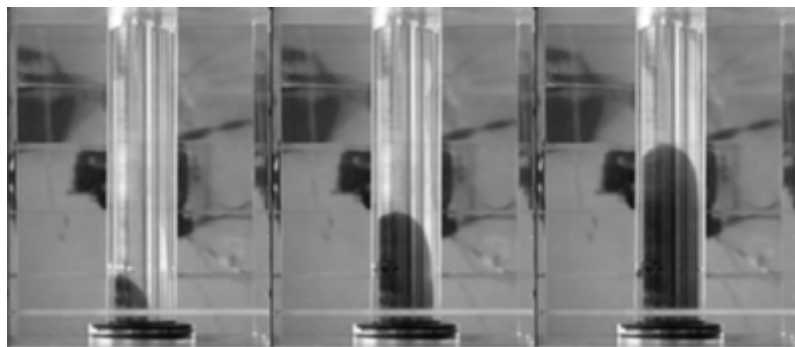


Рис. 3. Формы трещины, наблюдавшиеся в эксперименте [11]

Fig. 3. Fracture forms observed in the experiment [11]

Разработанная численная модель способна описывать деформацию упругой пробки внутри обсадной колонны. Поэтому при выполнении численного моделирования материал со свойствами, соответствующими полиметилметакрилату, используемому в эксперименте, был помещен внутрь обсадной колонны с внутренним радиусом, равным внешнему радиусу «экспериментальной» обсадной колонны. Как и для реальных пробок, жесткость материала обсадной колонны в эксперименте много больше жесткости пробки, поэтому деформациями обсадной колонны пренебрегалось. Из-за неполного соответствия геометрии удалось провести лишь качественное сравнение с экспериментом.

Для оценки величины ошибки, вносимой численным решением уравнений, была проведена серия расчетов на последовательности сгущающихся сеток, содержащих 160, 300 и 600 узлов для покрытия поверхности начальной трещины, с размерами элемента у фронта трещины 2, 1, и 0,5 мм соответственно, что соответствует $1/20 - 1/80$ диаметра скважины. Также изменялся другой параметр численного алгоритма – величина приращения трещины за один шаг, который определяет максимальный размер элементов, добавляемых в сетку при приращении трещины. Его величина варьировалась от 1 до 4 мм ($1/10 - 1/40$ диаметра скважины). В обоих случаях была получена сеточная сходимость зависимости длины трещины от времени.

После выбора параметров численного алгоритма было проведено сравнение с результатами лабораторного эксперимента. На рис. 4 представлены формы фронта трещины, наблюдаемые в лабораторном эксперименте и рассчитанные численно в различные моменты времени. Формы фронта, полученные на первых шагах моделирования, совпадают с формами, полученными в эксперименте. Там же приведены зависимости угла раствора трещины вычисляемого по ее поперечному размеру и диаметру скважины, от длины трещины вдоль скважины. Как видно, рассчитанная зависимость близка к наблюдаемой в лабораторном эксперименте. При дальнейшем распространении трещины асимметричность формы фронта возрастала, что в силу симметричности постановки задачи при проведении вычислительного эксперимента не может быть воссоздано.

Проведенное сравнение показывает, что предложенная модель позволяет описывать распространение трещины по границе обсадной колонны и цементной пробки, вызванной закачкой в нее вязкой жидкости.

3. Оценка стойкости гидроизоляции скважины

Для проведения вычислительных экспериментов были выбраны следующие параметры: характерные для легких цементов модуль Юнга $=9\text{ ГПа}$, коэффициент Пуассона $=0.15$ и плотность $=2000\text{ кг/м}^3$. Коэффициент трещиностойкости варьировался в интервале

от 0.025 до 0.2 МПа, внутренний диаметр обсадной колонны был равен $\varnothing=0.3$ м. Плотность $\rho=800$ кг/м³ и вязкость $\eta=0.2$ Па·с поровой жидкости были выбраны близкими к параметрам легкой нефти. Пластовое давление и давление застывания цемента, если не указано иное, равны $p_0=10$ МПа и $p_{cr}=0$.

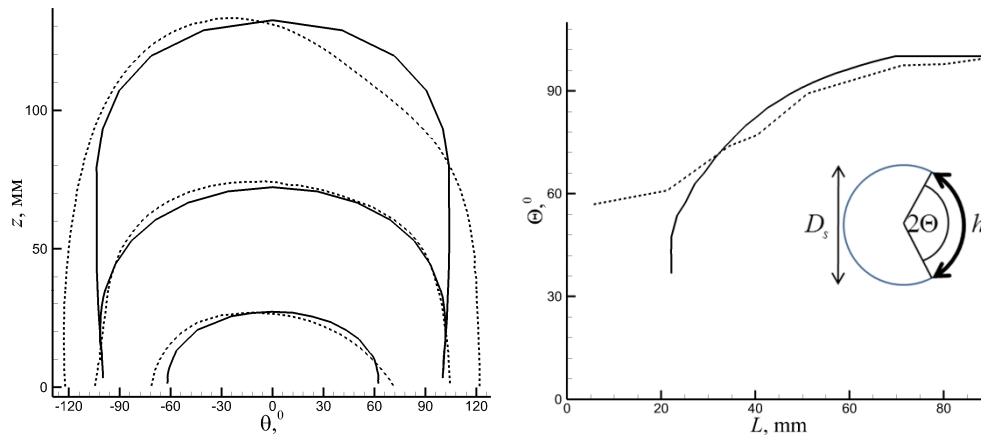


Рис. 4. Формы фронта трещины, полученные в эксперименте (сплошные линии) и в численном моделировании (пунктирные линии), в различные моменты распространения трещины (слева). Зависимость угла раствора θ_0 от длины трещины вдоль оси скважины (справа)

Fig. 4. The fracture front forms obtained in the experiment (solid lines) and in the numerical simulation (dotted lines), at different time moments of crack propagation (left). Fracture angle θ_0 versus fracture length along the well axis (right)

Влияние размера начального дефекта

Одной из целей работы было исследование чувствительности давления, необходимого для начала распространения трещины, к размеру начального дефекта, обусловленному несовершенством цементирования. Для этого были рассмотрены дефекты различного размера и для каждого найдено давление, необходимое для начала распространения. Это давление предполагается постоянным по всей поверхности трещины, так как до начала распространения трещины движение жидкости в ней отсутствует. Поэтому вместо задачи (5)–(6) для определения давления используется условие удовлетворения критерия распространения (7). В качестве параметра, определяющего размер дефекта или начальной трещины, выбран угол раствора трещины (рис. 4). Для упрощения анализа предполагается, что критерий удовлетворяется во всех точках фронта, что позволяет для каждого угла раствора определить длину трещины и форму фронта.

В самом деле, если для распространения трещины в каких-либо точках фронта требуется большее давление, чем для распространения в остальной его части, то это давление приведет к распространению всего фронта до тех пор, пока условие (7) не будет выполнено для всего фронта.

На рис. 5 приведены формы фронта, полученные для различных углов раствора, и зависимость давления, необходимого для начала распространения трещин, имеющих такую форму. Видно, что с ростом размера дефекта давление ведет себя не монотонно. Как и в случае круговых трещин в пространстве, давление, необходимое для начала распространения трещины, падает с ростом ее размера. Затем, из-за того, что трещина находится в ограниченном пространстве, зависимость давления от размера становится обратной, в силу того, что тре-

щина охватывает пробку с разных сторон и сжимающие напряжения в пробке увеличиваются. Этот эффект говорит о том что, если произошла инициация трещины, то угол ее раствора достигнет примерно $\Theta = 90^\circ$, так как для роста трещины требуется меньшее давление, чем для инициации. Таким образом показано, что размер начального дефекта не важен и в практических приложениях следует либо обеспечивать совершенное прилипание цемента к обсадной колонне, либо искать механизмы, предотвращающие распространение трещины, а не ее инициацию. Сходный вывод был сделан в [12], где на основе экспериментального исследования прочности цементных пробок утверждалось о существенном увеличении их стойкости при использовании дополнительного укрепления именно их нижней части (металлические пробки, высоковязкие жидкости и т. д.).

Второе наблюдение из рис. 5, заключается в том, при увеличении размера трещины ее форма становится более вытянутой вдоль скважины. При увеличении угла раствора Θ наряду с ростом давления P_c , необходимого для распространения трещины наблюдается рост и производной $\partial L / \partial \Theta$, что видно по углу наклона графика. Это говорит о том, что с ростом размера трещины распространение вдоль скважины становится предпочтительным, то есть требует меньшего давления в трещине, чем распространение в окружном направлении. Последнее исключает возможность распространения трещины по всей окружности скважины и полный отрыв пробки от обсадной колонны. Тот же вывод был сделан в [11] на основе лабораторных экспериментов и анализа напряженного состояния пробки в двумерной постановке.

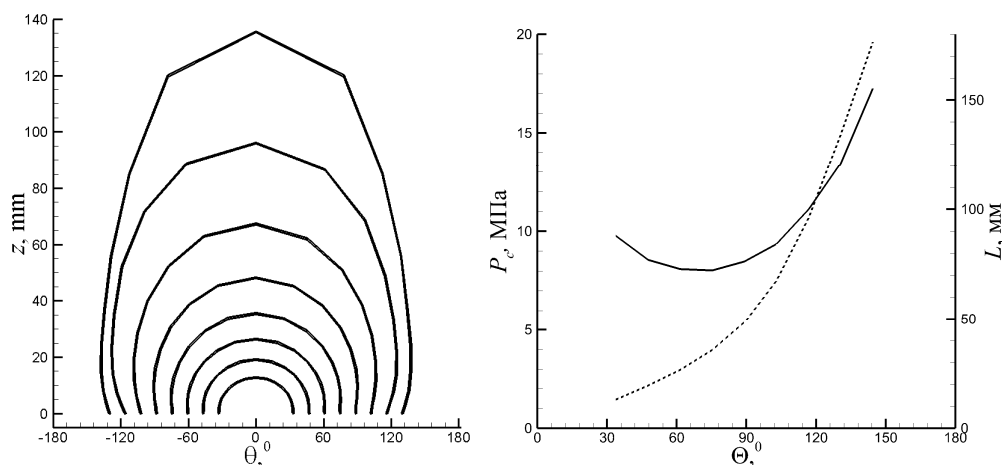


Рис. 5. Форма трещины с постоянным давлением (слева). Зависимости давления, необходимого для распространения трещины, (сплошная линия) и длины трещины (пунктир), от угла раствора трещины Θ (справа)

Fig. 5. Fracture form obtained with constant pressure (left). Pressure that is necessary for fracture propagation (solid line) and fracture length (dotted line) as functions of fracture angle Θ (right)

Влияние пластового давления на скорость распространения трещины

Очевидно, что, так как для инициации трещины требуется увеличение давления, то повышение давления приводит к большей вероятности инициации трещины. Однако важным параметром является скорость, с которой будет распространяться трещина после инициации. Для определения этой величины необходимо описывать движение жидкости в трещине, решая задачу (5)–(6), вместо задания в ней постоянного давления из условия (7), как было сделано в предыдущем разделе. На рис. 6, слева приведены зависимости длины трещины от времени, полученные для различных длин начальной трещины и давлений резервуара. Вид-

но, что трещины различной начальной длины распространяются с одинаковой скоростью, которая зависит от пластового давления P_p . Единственное отличие состоит в том, что трещине, стартующей с меньшего дефекта, требуется некоторое незначительное время для достижения размера большего начального дефекта. Эти зависимости подтверждают вывод, сделанный в предыдущем разделе без учета движения жидкости, об отсутствии влияния размера начального дефекта на распространение трещины. На том же рис. 6, справа в логарифмическом масштабе приведены зависимости скорости распространения трещины v_m от ее длины, полученные при различных значениях пластового давления. Видно, что эта величина является определяющим параметром для скорости распространения трещины, так, например, ее увеличение в 2–3 раза приводит к увеличению скорости распространения трещины в 10–30 раз.

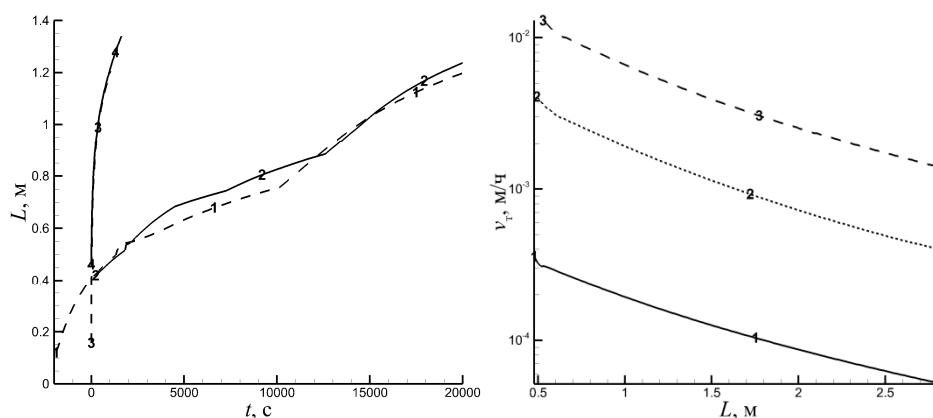


Рис. 6. Зависимости длины трещины от времени при различных длинах начальной трещины L_0 и давлениях в резервуаре P_p (слева): 1, 3 – $P_p = 10$ МПа; 2, 4 – $P_p = 2$ МПа; 1, 2 – $L_0 = 0,1$ м; 3, 4 – $L_0 = 0,4$ м. Зависимости скорости распространения трещины от времени при различных давлениях в резервуаре: 1, 2, 3 – $P_p = 5, 10, 15$ МПа соответственно (справа)

Fig. 6. Fracture length versus time obtained under various initial fracture length L_0 and reservoir pressure P_p (left): 1, 3 – $P_p = 10$ MPa; 2, 4 – $P_p = 2$ MPa; 1, 2 – $L_0 = 0.1$ m; 3, 4 – $L_0 = 0.4$ m. Fracture propagation velocity versus time obtained under various reservoir pressure P_p : 1, 2, 3 – $P_p = 5, 10, 15$ MPa respectively (right)

В отличие от существенного влияния давления P_p , действующего на нижний торец пробки, влияние давления P_0 , действующего на ее верхний торец со стороны столба жидкости в скважине, незначительно. Этот факт был подтвержден серией расчетов, проведенных при различных значениях этого давления, и подкрепляется следующими теоретическими рассуждениями.

После застывания пробка представляет собой цилиндр, большая часть боковой поверхности которого (за исключением области течения жидкости) неразрывно связана с обсадной колонной и потому неподвижна. Это значит, что действие давления P_0 на пробку со стороны верхнего торца компенсируется силами, вызванными условиями прилипания на боковой поверхности в его окрестности. То есть равнодействующая сил, действующих со стороны столба жидкости и обсадной колонны, равна нулю. Так как диаметр пробки не превосходит полуметра, а длина пробки обычно более десяти метров, размер области, где действуют силы, вызванные приложением давления P_0 , много меньше, чем расстояние до нижнего торца пробки – области, где происходит начальное развитие трещины. Это позволяет применить принцип Сен-Венана, согласно которому локальным распределением сил, равнодействующая

которых равна нулю, можно пренебрегать при описании напряженно-деформированного состояния на достаточном удалении от области, где они приложены.

Влияние трещиностойкости цемента на скорость распространения трещины

Важным параметром при изучении механики трещин является трещиностойкость или критическое значение коэффициента интенсивности напряжений K_{Ic} . Значение этого параметра для пары материалов сталь–цемент в общем случае трудно определима, так как зависит от параметров самого цемента и условий его затвердевания. Тем не менее, при проведении работ по закупориванию скважины эта величина поддается влиянию путем выбора параметров цемента, времени его затвердевания до снятия давления с верхнего торца пробки и т.д., поэтому исследование чувствительности стойкости пробки к изменению этой величины представляет интерес для инженеров, проектирующих закупоривание скважины. Достоверно известно, что трещиностойкость соединения ограничена снизу и сверху нулем и трещиностойкостью самого цемента. Поэтому для изучения влияния этого неизвестного параметра его значение варьировалось в интервале $0.025 \text{ МПа} \cdot \text{с} \leq K_{Ic} \leq 0.2 \text{ МПа} \cdot \text{с}$ и оценивалась изменение скорости распространения трещины, как показано на рис. 7. Хотя при малых значениях трещиностойкости из-за численных эффектов скорость распространения в немонотонна, ее усредненное поведение совпадает наблюдаемым в остальных случаях – скорость распространения уменьшается с ростом длины трещины.

Видно, что трещиностойкость материала (или границы между материалами) влияет на скорость распространения трещины, но это влияние снижается с ростом ее длины. Например, когда длина трещины равна двум диаметрам ствола скважины ($\approx 0.6 \text{ м}$), скорость распространения, полученная при $K_{Ic} = 0.025 \text{ МПа} \cdot \text{с}$ в 4 раза больше скорости, полученной при $K_{Ic} = 0.2 \text{ МПа} \cdot \text{с}$ (в 8 раз большей трещиностойкости). Но это соотношение уменьшается примерно до 1.5, когда длина трещины достигает десяти диаметров ствола скважины ($\approx 0.3 \text{ м}$). Таким образом, можно сделать вывод, что влияние трещиностойкости на скорость распространения трещины ограничено начальным этапом распространения трещины, когда ее длина не превышает примерно десяти диаметров ствола скважины.

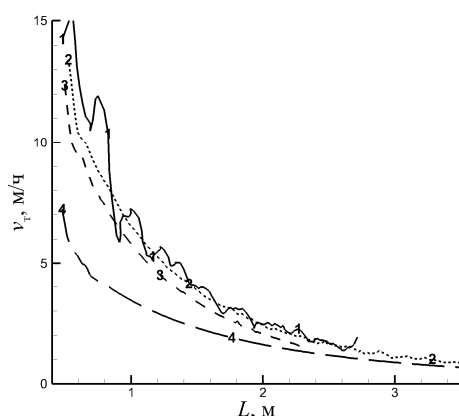


Рис. 7. Зависимости скорости распространения трещины от ее длины при различных значениях трещиностойкости цемента: 1 – $K_{Ic} = 0.025 \text{ МПа} \cdot \text{с}$; 2 – $K_{Ic} = 0.05 \text{ МПа} \cdot \text{с}$; 3 – $K_{Ic} = 0.1 \text{ МПа} \cdot \text{с}$; 4 – $K_{Ic} = 0.2 \text{ МПа} \cdot \text{с}$

Fig. 7. Fracture propagation velocity versus fracture length obtained with various cement fracture toughness: 1 – $K_{Ic} = 0.025 \text{ МПа} \cdot \text{с}$; 2 – $K_{Ic} = 0.05 \text{ МПа} \cdot \text{с}$; 3 – $K_{Ic} = 0.1 \text{ МПа} \cdot \text{с}$; 4 – $K_{Ic} = 0.2 \text{ МПа} \cdot \text{с}$

Влияние преднапряженного состояния цемента на скорость распространения трещины

Еще один параметр, который доступен для влияния со стороны инженеров, проектирующих закупоривание скважины, это давление P_c . Оно характеризует преднапряженное состоя-

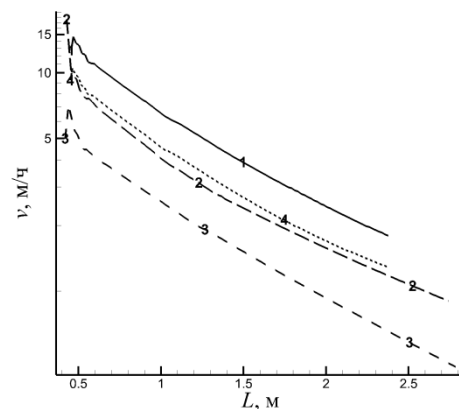
ние цемента, зависящее как от давления, при котором происходит его застывание, так и от добавок, влияющих на степень его сжатия или расширения при застывании. На основе уравнения (4), в котором давление в цементе вычитается из давления жидкости, можно ожидать, что его влияние будет по величине эквивалентно влиянию пластового давления, но противоположным по знаку. Для проверки этого предположения были рассчитаны скорости трещины при двух значениях пластового давления $P_p = 10$ и 9 МПа и трех значений давления цемента $P_c = 0, 0.5$ и 1 МПа. Как видно из рис. 8, скорости, полученные для случаев $P_p=10$ МПа, $P_c=0.5$ МПа и $P_p=9$ МПа, $P_c=0$ МПа близки, то есть увеличение давления цемента на 0.5 МПа оказалось практически эквивалентным снижению пластового давления на 1 МПа. Это объясняется тем, что приведенное выше соображение не учитывает эффект отставания фронта жидкости от фронта трещины, в результате которого образуется область не заполненная жидкостью. В этой области проявляется влияние давления застывания цемента, но отсутствует влияние пластового давления.

Расстояние между фронтами трещины и жидкости зависит, как и в случае классической трещины гидроразрыва, от сжимающих напряжений в материале, в роли которых в данном случае выступает напряжения в цементе. Они складываются из гидростатических напряжений в жидком цементе перед его застыванием и добавочного давления P_c , появившегося при его застывании. При рассматриваемых параметрах максимальный по всему фронту размер расстояния между фронтами трещины и жидкости составил 5.5 см при $P_c = 0$ и 2 см при $P_c = 1$ МПа. Несмотря на то что размер области, не заполненной жидкостью, мал по сравнению с размерами трещины, эта область расположена в окрестности фронта трещины, поэтому напряжения в ней оказывают влияние на ширину всей трещины.

Таким образом, одним из способов повышения стойкости гидроизоляции скважины может являться повышение давления, характеризующего преднапряженное состояние цемента, причем увеличение этого параметра имеет примерно вдвое больший эффект, чем повышение пластового давления.

Рис. 8. Зависимости скорости распространения трещины от ее длины при различных пластовых давлениях и давлениях застывания цемента: 1 – $P_p=10$ МПа, $P_c=0$ МПа; 2 – $P_p=10$ МПа, $P_c=0.5$ МПа; 3 – $P_p=10$ МПа, $P_c=1$ МПа; 4 – $P_p=9$ МПа, $P_c=0$ МПа

Fig. 8. Fracture propagation velocity versus fracture length obtained with various reservoir pressure and cement prestress: 1 – $P_p=10$ МПа, $P_c=0$ МПа; 2 – $P_p=10$ МПа, $P_c=0.5$ МПа; 3 – $P_p=10$ МПа, $P_c=1$ МПа; 4 – $P_p=9$ МПа, $P_c=0$ МПа



Заключение

Разработана модель инициирования и распространения трещины вдоль границы цементной пробки и стальной обсадной колонны. Модель описывает изменение напряженно-деформированного состояния цемента и его разрушение на границе со стальной обсадной колонной под действием давления затекающей в трещину поровой жидкостью, вызванного повышением пластового давления внизу пробки. Уравнения упругого равновесия решаются ранее разработанным методом граничных элементов [8], уравнения движения жидкости –

методом конечных элементов [9], для определения условий разрушения цемента используется силовой критерий Ирвина [10].

Проведена валидация модели путем сравнения с лабораторным экспериментом [11], в котором получено качественное согласование форм фронта трещины в различные моменты времени. На основе серии вычислительных экспериментов сделаны выводы о чувствительности скорости распространения трещины, а значит, и стойкости гидроизоляции скважины к следующим основным параметрам.

1. Повышение пластового давления может вызвать инициацию трещины из дефектов цементирования на нижнем торце пробки. При этом размер дефекта не имеет значения, так как после инициации давление, требуемое для продолжения распространения трещины, падает с ростом ее размера. Поэтому использование подходов, устраняющих дефекты цементирования, может существенно увеличить стойкость гидроизоляции.

2. Определяющим параметром, влияющим на скорость распространения трещины, является давление в пласте.

3. Трещиностойкость цемента или пары цемент-сталь влияет на скорость распространения трещины только на начальном этапе распространения, когда ее длина не превосходит десяти диаметров скважины. Увеличение адгезии цемента не приводит к заметному повышению стойкости гидроизоляции.

4. Преднапряженное состояние цемента оказывает существенное (вдвое большее, чем пластовое давление) влияние на скорость распространения трещины. Увеличение давления застывания цемента на 1 МПа позволяет гидроизоляции противостоять пластовому давлению, увеличенному на вдвое большую величину.

Список литературы / References

1. **Gasda S. E., Bachu S., Celia M. A.** Spatial characterization of the location of potentially leaky wells penetrating a deep saline aquifer in a mature sedimentary basin. *Environmental Geology*, 2004, vol. 46, no. 6, p. 707–720. DOI 10.1007/s00254-004-1073-5.
2. **Vralstad T., Saasen A., Fjer E., Oia T., Ytrehus J. D., Khalifeh M.** Plug and abandonment of offshore wells: Ensuring long-term well integrity and cost-efficiency. *Journal of Petroleum Science and Engineering*, 2019, vol. 173, p. 478–491. DOI 10.1016/j.petrol.2018.10.049.
3. **Godoy R., Fontan M., Capra B., Kvalsund R., Poupard O.** Well integrity support by extended cement evaluation – numerical modeling of primary cement jobs. In: Abu Dhabi International Petroleum Exhibition and Conference. Society of Petroleum Engineers, 2015. DOI 10.2118/177612-ms.
4. **Watson T., Bachu S.** Evaluation of the potential for gas and CO₂ leakage along wellbores. *SPE Drilling & Completion*, 2009, vol. 24, p. 115–126.
5. **Miyazaki B.** Well integrity: An overlooked source of risk and liability for underground natural gas storage. Lessons learned from incidents in the USA. *Geological Society*, London, Special Publications, 2009, vol. 313, no. 1. p. 163–172. DOI 10.1144/sp313.11.
6. **Loizzo M., Akemu O. A., Jammes L., Desroches J., Lombardi S., Annunziatellis A.** Quantifying the risk of CO₂ leakage through wellbores. *SPE Drilling & Completion*, 2011, vol. 26, no. 03, p. 324–331. DOI 10.2118/139635-pa
7. **Carter L., Evans. G.** A study of cement-pipe bonding. *Journal of Petroleum Technology*. 1964, vol. 16, no. 2. p. 157–160. DOI 10.2118/764-PA.
8. Kuranakov D.S., Esipov D.V., Lapin V.N., Cherny S.G. Modification of the boundary element method for computation of three-dimensional fields of strain-stress state of cavities with cracks. *Engineering Fracture Mechanics*, 2016, vol. 153. p. 302–318. DOI 10.1016/j.engfractmech.2015.09.052.
9. **Черный С. Г., Лапин В. Н., Есипов Д. В., Куранаков Д. С.** Методы моделирования зарождения и распространения трещин. Новосибирск: Изд-во СО РАН, 2016, 312 с.

- Cherny S. G., Lapin V. N., Esipov D. V., Kuranakov D. S.** Metody modelirovaniya zarozhdeniya i rasprostraneniya treshchin [Methods for modeling the fracture initiation and propagation]. Novosibirsk, 2016, 312 p. (in Russ.)
10. **Irwin G.** Analysis of stresses and strains near the end of a crack traversing a plate. *J. Applied Mechanics*, 1957, vol. 24, p. 361–364.
 11. **Lecampion B., Bungier A., Kear J., Quesada D.** Interface debonding driven by fluid injection in a cased and cemented wellbore: Modeling and experiments. *International Journal of Greenhouse Gas Control*, 2013, vol. 18, p. 208–223. DOI 10.1016/j.ijggc.2013.07.012.
 12. **Steffek T.** Cement Plug Testing: Weight vs Pressure Testing to Assess Viability of a Wellbore Seal Between Zones. *CSI Technologies*, 2012, 111 p.

Материал поступил в редколлегию

Received

15.01.2020

Сведения об авторе

Лапин Василий Николаевич, кандидат физико-математических наук, старший научный сотрудник, Институт вычислительных технологий СО РАН (Новосибирск, Россия)
lapin@ict.sbras.ru
ORCID 0000-0002-6259-3265
ResearcherID L-5781-2017
AuthorID 14062

Information about the Author

Vasiliy N. Lapin, PhD, Leading Research Scientist, Institute of Computational Technologies SB RAS (Novosibirsk, Russian Federation)
lapin@ict.sbras.ru
ORCID 0000-0002-6259-3265
ResearcherID L-5781-2017
AuthorID 14062

Технология построения локально детализированных разностных сеток для расчета зон поражения при взрывах газа и пыли в угольных шахтах

О. Ю. Лукашов¹, Д. А. Костромин²

¹ ООО «ШахтЭксперт-Системы»
Кемерово, Россия

² Томский государственный университет
Томск, Россия

Аннотация

Описывается технология дискретизации расчетной области для решения нестационарных газодинамических задач в сфере промышленной безопасности угольных шахт. Важнейшей из них является расчет параметров воздушных ударных волн и зон поражения при взрывах метана в горных выработках. Технология позволяет автоматизировать построение разностной сетки для пространственной сети горных выработок с учетом особенностей топологии шахт и моделируемых процессов. При этом сетка может состоять из ячеек различной размерности – одномерных, двумерных и трехмерных. К ячейкам рекурсивно могут применяться операции субдискретизации до заданного уровня глубины и объединения ячеек ранее субдискретизированных областей. Целью работы является создание основ для развития научных исследований газодинамических процессов при взрывах газа в горных выработках за счет учета новых факторов – средств взрывозащиты, участия угольной пыли и др. В задачи работы входят: разработка системы адресации для идентификации и быстрого доступа к произвольным ячейкам разностной сетки; построение однозначных пространственных связей между соответствующими гранями ячеек при их разбиении и объединении; тестирование технологии на сети горных выработок реальной шахты. Новым подходом в работе является возможность автоматизации разбиения расчетной области с одновременным использованием ячеек всех трех размерностей. Статья содержит детали реализации технологии, показаны результаты разбиения сети горных выработок реальной угольной шахты, результаты расчетов с применением полученной сетки, сходимость результатов к некоему решению при уменьшении шага в результате субдискретизации расчетной области с разной глубиной разбиения.

Ключевые слова

угольная шахта, разностная сетка, моделирование, воздушные ударные волны, дискретизация

Для цитирования

Лукашов О. Ю., Костромин Д. А. Технология построения локально детализированных разностных сеток для расчета зон поражения при взрывах газа и пыли в угольных шахтах // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 50–64. DOI 10.25205/1818-7900-2020-18-1-50-64

Technology of Building Locally Detailed Differential Grids for the Calculation of Affected Areas at Gas and Dust Explosions in Coal Mines

O. Yu. Lukashov¹, D. A. Kostromin²

¹ LLC ShatkExpert-Systems
Kemerovo, Russian Federation

² Tomsk State University
Tomsk, Russian Federation

Abstract

The article describes the technology of discretization of the computational domain for solving unsteady gas-dynamic problems in the field of industrial safety of coal mines. The most important of this problem is the calculation of the parameters of air shock waves and the affected areas during methane explosions in mine workings. The technology allows you to automate the construction of a differential grid for a spatial network of mine workings, taking into account the features of the mines topology and simulated processes. At the same time, the grid can consist of cells of various dimensions – one-dimensional, two-dimensional and three-dimensional. The aim of the work is to create the basics for the development of scientific research of gas-dynamic processes in gas explosions in mining workings by taking into account new factors – means of explosion protection, the participation of coal dust, etc. The tasks of the work include: developing an addressing system for identification and quick access to arbitrary cells of the difference grid; the construction of unambiguous spatial relationships between the corresponding faces of the cells during their splitting and combining; testing of the technology on the mine network of a real mine. A new approach in the work is the possibility of automating the partition of the computational domain with the simultaneous use of cells of all three dimensions. The article contains details of the implementation of technology. The article shows the results of splitting the network of mine workings in a real coal mine, the results of calculations using the obtained grid and the convergence of the results to a solution when reducing the step as a result of subsampling the computational domain with different depth of splitting.

Keywords

coal mine, differential grid, modeling, air shock waves, discretization

For citation

Lukashov O. Yu., Kostromin D. A. Technology of Building Locally Detailed Differential Grids for the Calculation of Affected Areas at Gas and Dust Explosions in Coal Mines. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 50–64. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-50-64

Введение

В настоящее время для расчета распространения ударных волн при взрывах метана в угольных шахтах применяется отраслевая методика газодинамического расчета параметров воздушных ударных волн (ВУВ) при взрывах газа и пыли в шахтах (далее – Методика)¹. В ее основе лежат уравнения газовой динамики в нестационарной постановке [1], что значительно выделяет ее на фоне аналогичных методик [2; 3], используемых ранее. Классический подход позволяет развивать методику за счет учета новых факторов, например взрывного горения угольной пыли [4; 5], решать задачи взрыволокализации в горных выработках [6; 7], а также комбинировать модель, используемую в методике, с нестационарными газодинамическими моделями проветривания шахты [8; 9] и решать комплексную задачу математического моделирования развития [10].

Численное решение уравнений, составляющих основу указанной Методики, подразумевает пространственную дискретизацию расчетной области – одномерное разбиение прямолинейных участков горных выработок (ветвей) и по одной трехмерной ячейке в сопряжениях и точках изменения геометрии выработок [11; 12], (рис. 1). Оптимальный шаг по пространству, определенный в ходе многочисленных экспериментов, составляет 4 м.

¹ Методика газодинамического расчета параметров воздушных ударных волн при взрывах газа и пыли. Утверждена Госгортехнадзором России 02.04.2003.

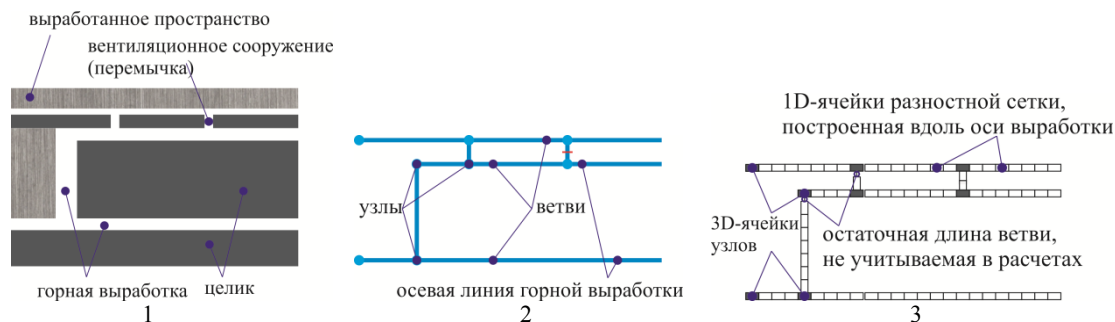


Рис. 1. Существующий принцип построения разностной сетки:

1 – схема добычного участка угольной шахты; 2 – соответствующие осевые линии;
3 – разностная сетка, построенная по осевым линиям

Fig. 1. The existing principle of building a differential grid:

1 – coal mine mining scheme; 2 – respective axis lines; 3 – differential grid which was constructed by the axial lines

Актуальность

Описанная выше схема разбиения простра, обеспечивает высокую скорость и необходимую точность расчетов определения границ опасных зон при взрывах газа в шахте. Но ее обратной стороной является отсутствие детализированной картины происходящих физических процессов в целом ряде характерных мест. Таковыми являются сопряжения, повороты, места изменения геометрии горных выработок, сбойки, ближайшие окрестности перемычек и заслонов (водяных или сланцевых). При экспертизе аварий в шахте изучение процессов, происходящих вблизи перечисленных мест и объектов, часто требуется более детальная картина протекания газодинамических процессов, нежели обеспечивается текущей реализацией Методики.

С другой стороны, простое уменьшение шага по пространству приведет к пропорциональному снижению скорости расчета, в то время как реализация Методики не является научным приложением в полном смысле и должна обеспечивать приемлемое время расчетов на персональных компьютерах.

Такие противоположные друг другу требования делают актуальным разработку и применение в Методике адаптивных сеток. Это позволит динамически менять шаг по пространству, сгущая или укрупняя сетку в характерных местах по мере изменения интенсивности газодинамических процессов (рис. 2). Кроме того, учитывая особенности топологии сетей горных выработок угольных шахт, представляется оправданным и возможным динамическое изменение размерностей ячеек, преимущественно в сторону их снижения, в целях радикального ускорения расчетов.

Постановка задачи

Требуется разработать технологию разбиения ячеек базовой разностной сетки, построенной для пространственной сети горных выработок угольной шахты, до заданного уровня детализации, обеспечив при этом информацию о связях между ячейками и условия для дальнейшей передачи данных в модуль расчета параметров ВУВ. В целях сохранения уже существующей математической модели детализированная сетка должна быть ортогональной.

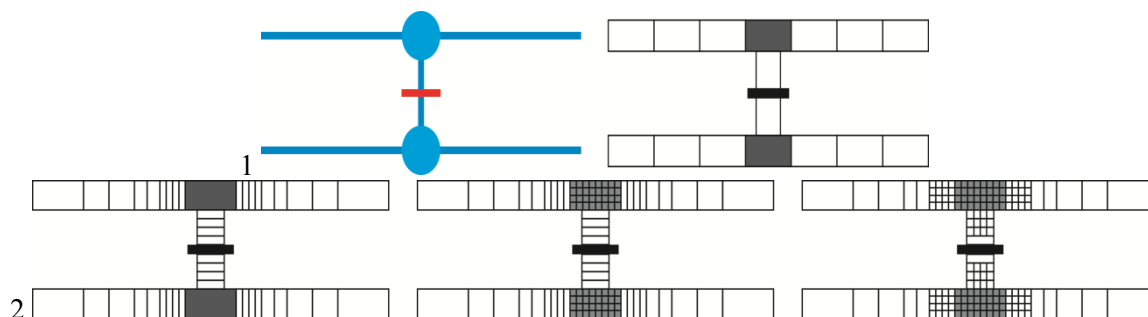


Рис. 2. Варианты построения разностных сеток:

1 – участок горных выработок вблизи сбойки с вентиляционной перемычкой – осевые линии и существующий вариант построения разностных сеток; 2 – возможные варианты детализации разностной сетки вблизи характерных объектов

Fig. 2. Variants of construction of difference grids:

1 – section of mine workings near the bridge with a ventilation jumper; 2 – possible options for detailing the difference grid near characteristic objects

Разбиение должно быть возможным:

- в ручном режиме, за счет явного указания оператора, выполняющего расчеты, отдельных ячеек, подлежащих разбиению;
- в полуавтоматическом режиме, за счет указания лишь характерных точек, с дальнейшим автоматическим разбиением некоторой локальной зоны вокруг них. При этом выбор алгоритма, размерности и детализации разностной сетки могут осуществляться автоматически;
- в полностью автоматическом режиме, когда и выделение характерных объектов, и процесс разбиения выполняются без участия человека.

Ручной режим полностью управляется оператором без применения специальных алгоритмов. Полуавтоматический режим разбиения подразумевает применение некоторой методики, которая позволяла бы определять размеры локально детализируемых зон, глубину их детализации, размерности ячеек, критерии и закономерности разбиения. В качестве критериев могут выступать:

- особенности структуры шахтной сети, например углы поворотов или взаимное расположение горных выработок;
- степень относительного изменения геометрических параметров – величины поперечного сечения выработок, формы крепи и т. д.;
- вид объекта, вокруг которого определяется буферная зона – сопряжение, вентиляционные сооружения, водяной или сланцевый заслон, а также протяженность выработки, в которой они установлены.

Полностью автоматический режим может учитывать дополнительные условия, в частности:

- учитывать расстояние до эпицентра взрыва в целях предварительной оценки интенсивности ожидаемых процессов (например, распространения ВУВ через сопряжение). В данном случае могут быть произведены грубые оценки без учета волновых эффектов на основе экспериментальных данных;
- выполнять динамическую оценку параметров распространения ВУВ на каждом временном слое для адаптивного перестроения разностных сеток.

Особенности топологии сети горных выработок

Для каждого вида точки и объекта (сопряжение, перемычка, ...) могут использоваться и учитываться разные методики и факторы, а также специфические условия, характерные для угольных шахт. Поэтому была проведена систематизация фактов и учтены особенности предметной области при построении разностной сетки:

- разностная сетка для расчета распространения ВУВ строится на основе пространственной сети горных выработок;
- сетка является структурированной с фиксированным шагом h по пространству;
- вдоль ветвей (прямолинейных участков горных выработок) используется одномерное разбиение. Продольный размер (длина) ячейки равен h , поперечный размер соответствует сечению горной выработки;
- в узлах (сопряжениях, на поворотах и в точках изменения геометрии горных выработок) используется по одной трехмерной ячейке. Размеры по трем ортогональным направлениям также равны $h \times h \times h$;
- особенности геологии и проходки горных выработок угольных шахт, для которых и применяется Методика, таковы, что практически отсутствуют гладкие повороты и значительные перепады сечений горных выработок. В сопряжениях выработки примыкают друг к другу, как правило, под значительными углами;
- для правильного расчета распределения энергии и массы воздуха при его движении через сопряжение необходимо знать углы отклонения осевых линий прилегающих выработок к соответствующим граням пространственной расчетной ячейки, моделирующей это сопряжение [1] (рис. 3).

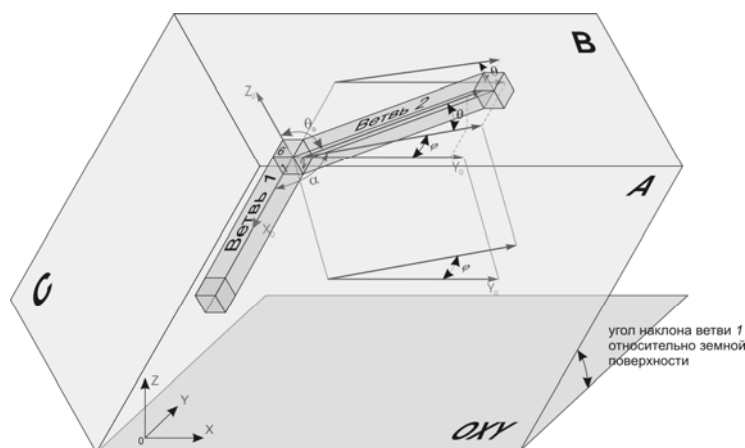


Рис. 3. Схема расчета углов
Fig. 3. Angles calculation scheme

Перечисленные выше особенности шахтных сетей позволяют использовать кубы и прямоугольные параллелепипеды не только в качестве базовых расчетных объемов, но и при любом уровне их дискретизации. Это исключает усложнение математической модели для расчета параметров ВУВ в этих областях.

Реализация технологии

Для реализации поставленной задачи выбран подход, основанный на локальной адаптации декартовых сеток (ЛАД) [13]. Такие сетки позволяют описать геометрию сложных областей; максимально точно передать движущиеся поверхности разрывов, ударных волн, фазовых переходов и другие области больших градиентов функций. Задача построения ЛАД сеток на сегодняшний день является чрезвычайно актуальной и востребованной, о чем свидетельствует большое количество публикаций [14–18]. Основное отличие от известных реализаций заключается в возможности устанавливать связи между смежными ячейками разной размерности и произвольными уровнями дискретизации.

На момент написания статьи адаптация изначально структурированной декартовой сетки, где разбиение производится по какому-то заранее выбранному закону, выполняется в ручном режиме с помощью использования следующих древовидных структур [19]: двоичные, квадродерева и октодеревья. В таких структурах при разбиении исходная (родительская) ячейка сохраняется, являясь своеобразным контейнером для более мелких (дочерних) ячеек, лежащих внутри исходной. При этом бинарные деревья используются для одномерного разбиения ячеек, квадродерева – для двумерного, а октодеревья – для трехмерного (рис. 4). Значимым преимуществом использования древовидных структур является сохранение кратности при дроблении ячеек, что, как будет показано далее, существенно помогает в нахождении соответствий и установлении связей между смежными ячейками. Пример субдетализации расчетной ячейки на основе квадродерева для третьего уровня детализации представлен на рис. 5.

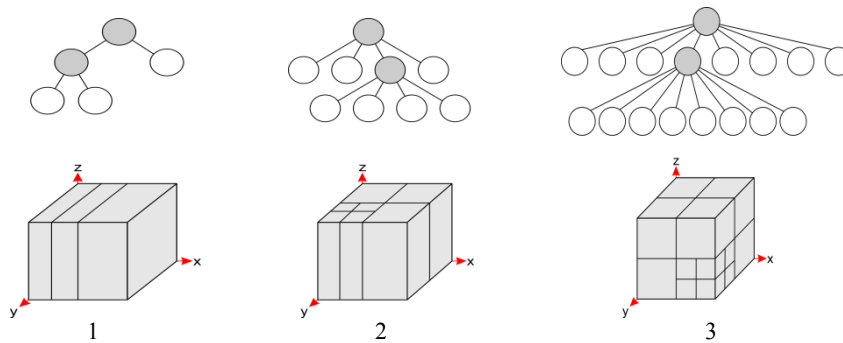


Рис. 4. Структуры данных: 1 – двоичное дерево; 2 – квадродерево; 3 – октодерево

Fig. 4. Data structures: 1 – binary tree; 2 – quadtree; 3 – octotree

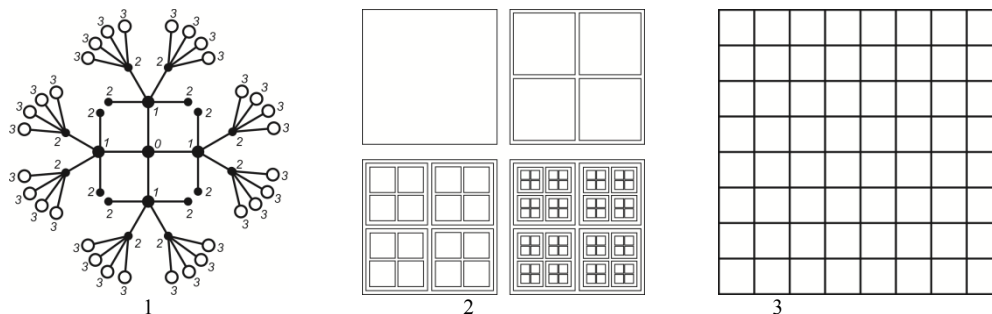


Рис. 5. Схема детализации разностной сетки на основе квадродерева при двумерном разбиении расчетной ячейки: 1 – квадродерево третьего уровня детализации; 2 – схема вложенности расчетных ячеек, соответствующая квадродереву; 3 – итоговая разностная сетка, участвующая в расчете

Fig. 5. Difference grid detailing scheme based on a quadtree with a two-dimensional partition of the computational cell: 1 – quadtree with third level of detail; 2 – nesting scheme of computational cells, corresponding quadtree; 3 – resulting difference grid that involved in the calculation

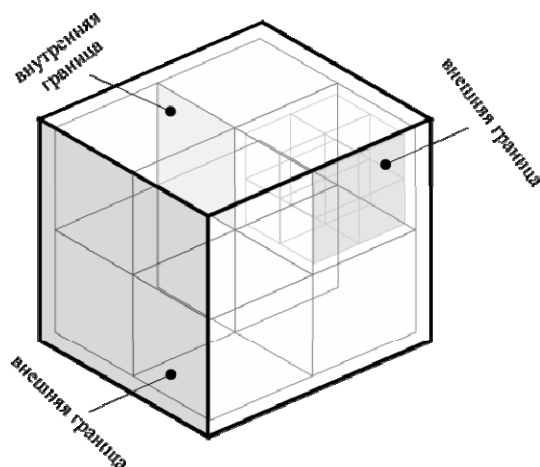


Рис. 6. Элементы технологии
Fig. 6. Elements of technology

Для детализации локальных зон были определены следующие элементы (рис. 6).

Ячейка – элементарный объем пространства, для которого в процессе решения с использованием конечно-объемного метода определяются физические параметры среды. Если ячейка была подвергнута субдетализации, то она содержит список дочерних ячеек, которых в зависимости от размерности может быть две, четыре или восемь. Все дочерние ячейки имеют ссылку на своего «родителя». Размерности дочерних ячеек изначально совпадают с родительской, но в последствии могут быть изменены на более высокие. Так, например, «дочки» двумерной ячейки могут быть также двумерными или трехмерными, но не могут быть одномерными. При этом размерности дочерних ячеек внутри одной родительской могут различаться. Любая дочерняя ячейка может быть впоследствии рекурсивно субдетализирована по тем же правилам, что и родительская. После субдетализации размерность родительской ячейки меняться не может.

Локальная область – это часть расчетной области, которую можно выделить в качестве наиболее крупной логической единицы пространства. Она объединяет в общем случае группу ячеек, которые могут быть разных размерностей. Для Методики в локальные области выделяются узлы и ветви. Каждая локальная область имеет шесть границ, называемых внешними.

Граница – хранит связи, соединяющие две смежные ячейки. Граница называется *внутренней*, если она соединяет две ячейки одного уровня детализации, и *внешней*, если она является границей локальной области. Граница также хранит углы сопряжения смежных ячеек и имеет собственный тип, определяющий граничные условия при решении газодинамической задачи (рис. 7):

- открытая граница – выход из шахты на земную поверхность;
- глухая граница – сплошная поверхность: тупик, грудь подготовительного забоя, борт, почва и кровля горной выработки;
- нормальная – связывает две смежные ячейки.

Создание границ при дискретизации расчетной области происходит следующим образом. При инициализации данных границы локальной области назначаются соответствующим внешним граням ячеек нулевого уровня детализации. При дальнейшем разбиении ячеек выполняется аналогичная операция – каждая родительская ячейка «отдает» свои границы внешней оболочке дочерних ячеек. Таким образом все ячейки всех уровней разбиения имеют одну и ту же внешнюю границу.

Связь определяет контактную границу двух смежных ячеек, через которую происходит движение рассчитываемой физической среды (в нашем частном случае – газа). Связи создаются в процессе субдетализации ячеек и хранятся в границах. Для однозначной идентифика-

ции контактной границы (т. е. граней, через которые осуществляется связка ячеек) была разработана следующая система индексации.

Каждая грань ячейки (контактная граница) имеет адрес, включающий:

- *собственный индекс*, состоящий из:
 - о *размерности* ячейки;
 - о *кода положения* данной ячейки на грани родительской ячейки;
- *индексов* всех родительских ячеек.

Размерность ячейки может принимать значения 1, 2 или 3.

Код положения является комбинацией четырех битов с возможными значениями 0 или 1, объединенных через логическое *ИЛИ*. Значение, равное 1, указывает, что грань дочерней ячейки занимает определенную часть соответствующей границы родительской ячейки по правилу: 0-й бит – верхнюю, 1-й бит – нижнюю, 2-й бит – левую, 3-й бит – правую часть. При этом понятия «лево», «право» применяются для направления вдоль оси *X*; «низ», «верх» – для направления вдоль оси *Z*, а ось *Y* в этом случае является нормалью к данной грани.

Для образования адреса размерность и код положения объединяются в одно трехзначное число. На рис. 8 показаны примеры индексов и полного адреса. Индекс 115 означает, что данная дочерняя ячейка является одномерной (на это указывает первая единица) и занимает всю площадь родительской грани (комбинация бит: 1111 означает, что дочерняя грань занимает все четыре четверти родительской, покрывая ее целиком; в десятичном выражении эта комбинация равна 15). Этот случай на рисунке отражает одномерное разбиение родительской ячейки вдоль оси *Z* (образовано две одномерных ячейки, одну из которых на рисунке не видно). Если продолжить такое одномерное разбиение, то получится еще одна ячейка второго уровня разбиения, имеющая адрес на данной грани *115.115*. Если преобразовать эту дочернюю одномерную ячейку в трехмерную и разбить ее, получится множество дочерних со следующими индексами: *115.309*, *115.305*, *115.306*, *115.310*.

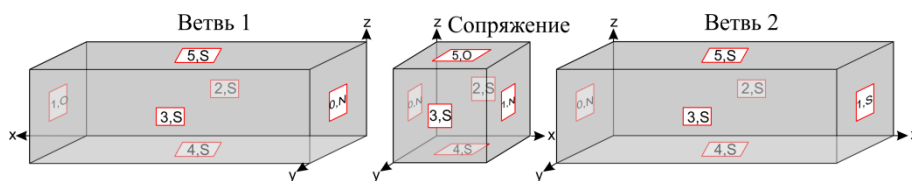


Рис. 7. Правила определения и нумерации границ
(числа – номера границ; О – открытая, S – глухая, N – нормальная границы)

Fig. 7. Rules for defining and numbering borders
(numbers are numbers of borders; O – open, S – solid, N – normal)

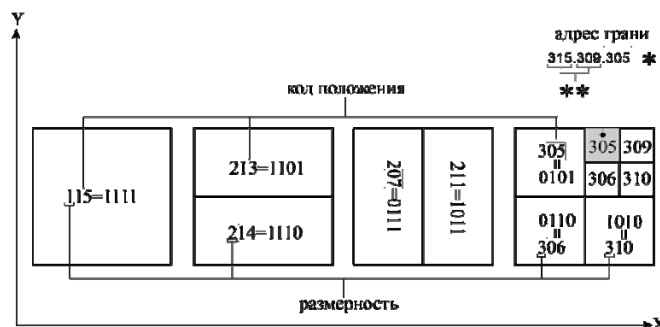


Рис. 8. Схема адресации ячеек
(* – собственный индекс, ** – родительский индекс, длина адрес грани равна трем, что соответствует двум последовательным разбиениям)

Fig. 8. Cell addressing scheme
* – own index, ** – parent index, the length of the address is three, which corresponds to two successive divisions)

На рис. 9 дан пример последовательного разбиения двух ячеек (слева – на трехмерную, справа – на двумерную) и показан результат установления связей между соответствующими гранями.

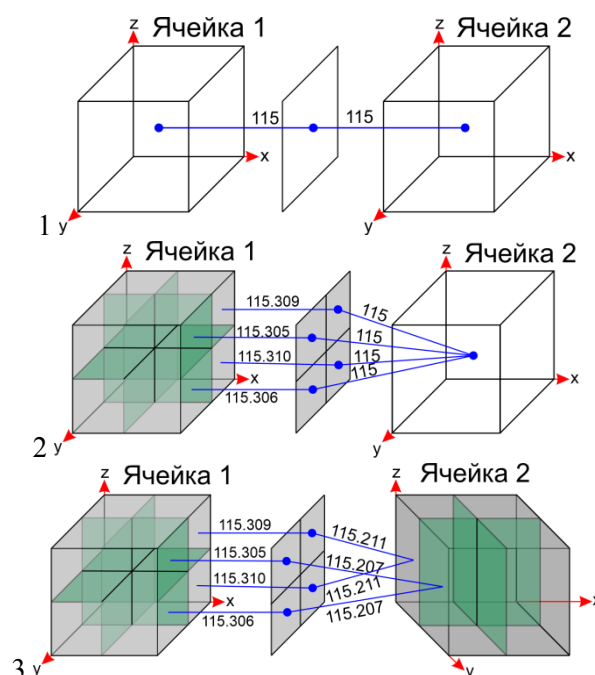


Рис. 9. Пример формирования адресов при разбиении:
 1 – до разбиения; 2 – после разбиения ячейки 1 на трехмерные ячейки;
 3 – после разбиения ячейки 2 на двумерные ячейки
 Fig. 9. An example of the formation of addresses in the process of splitting:
 1 – before splitting; 2 – after splitting cell 1 into three-dimensional cells;
 3 – after splitting cell 2 into two-dimensional cells

На рис. 10 представлены детали проверки существования связи между смежными гранями. Проверка заключается в специальном алгоритме сравнения кодов положения ячеек на определенных уровнях детализации.

Практическая реализация

Описанная технология реализована в виде приложения, разработанного в среде Visual Studio. Приложение позволяет в ручном режиме выполнять детализацию разностной сетки произвольной заданной области, в том числе с возможностью указания произвольной глубины разбиения (уровня детализации) ячеек. Примеры детализации разностных сеток в сопряжениях горных выработок показаны на (рис. 11, 12).

Проверка корректной работы алгоритма проводилась на простой модели, состоящей из двух смежных прямолинейных участков длиной по 200 м каждый (рис. 13). Один край зоны является тупиком, имитируя подготовительный забой в шахте, на расстоянии 90 м от него находится зона взрыва протяженностью 20 м. Давление в зоне составляет 0,6 МПа. Второй край модели имеет открытую границу. Эксперимент заключался в выполнении ряда численных расчетов одного и того же случая с последовательным кратным уменьшением шага по пространству и проверкой давления в контрольной точке, расположенной в 240 м от глухой границы. Зона взрыва расположена таким образом, чтобы исключить влияние фронта ВУВ, отраженной от тупиковой границы, на давление в контрольной точке. Первый расчет прово-

дился с шагом по пространству 4 м, затем шаг последовательно уменьшался в два раза. При проверке отслеживалось значение максимального давления во фронте ударной волны в контрольной точке при различных h (рис. 14).

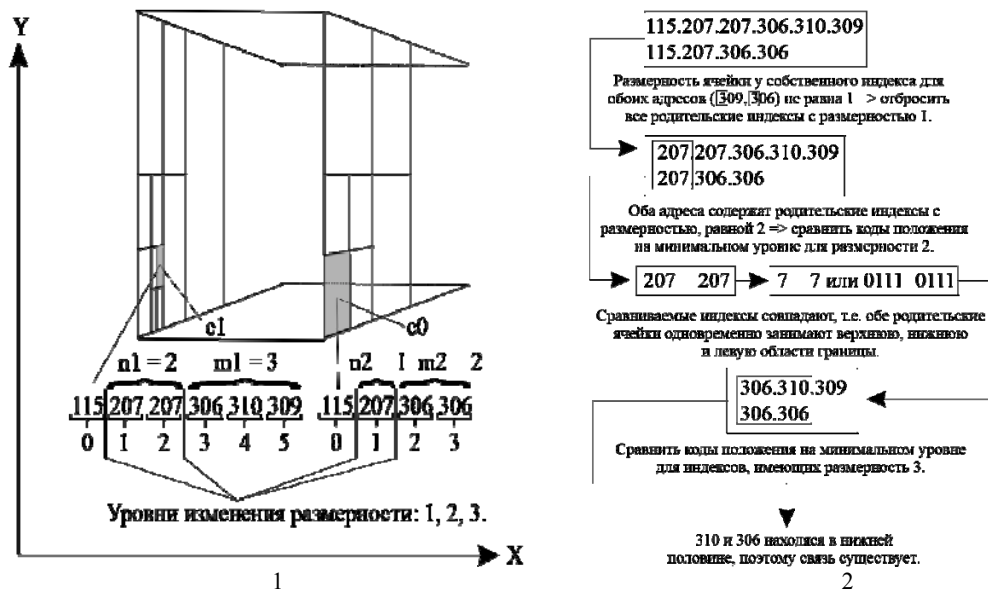


Рис. 10. Проверка существования связи: 1 – пример нахождения индексов изменения размерности в ячейках; $c0$ – ячейка, полученная после дробления; $c1$ – ячейка, составляющая связь с родительской ячейкой $c0$ до дробления; $n1$, $n2$ – количество двумерных разбиений; $m1$, $m2$ – количество трехмерных разбиений; 2 – пример пошаговой проверки существования связи для данных адресов

Fig. 10. Link existence check: 1 – an example of finding dimension change indices in cells; $c0$ is the cell obtained after crushing; $c1$ is the cell making up the connection with the parent cell $c0$ before crushing; $n1$, $n2$ – the number of two-dimensional fragmentations; $m1$, $m2$ – the number of three-dimensional fragmentations; 2 – an example of a step-by-step verification of the existence of communication at given addresses

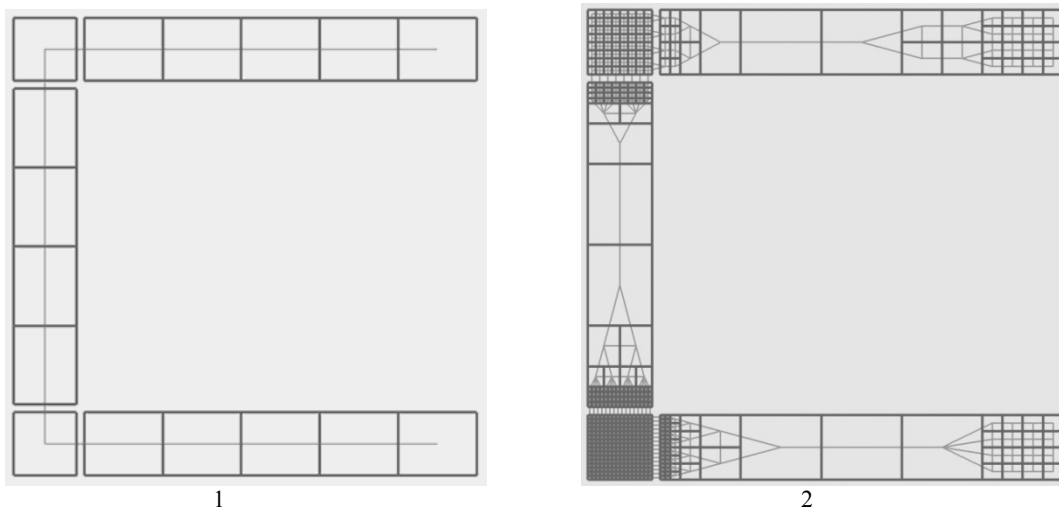


Рис. 11. Пример локальной детализации модели очистного участка угольной шахты:

1 – общий вид; 2 – после детализации

Fig. 11. An example of local detailing model for the coal mine cleanup section:

1 – general view; 2 – after detailing

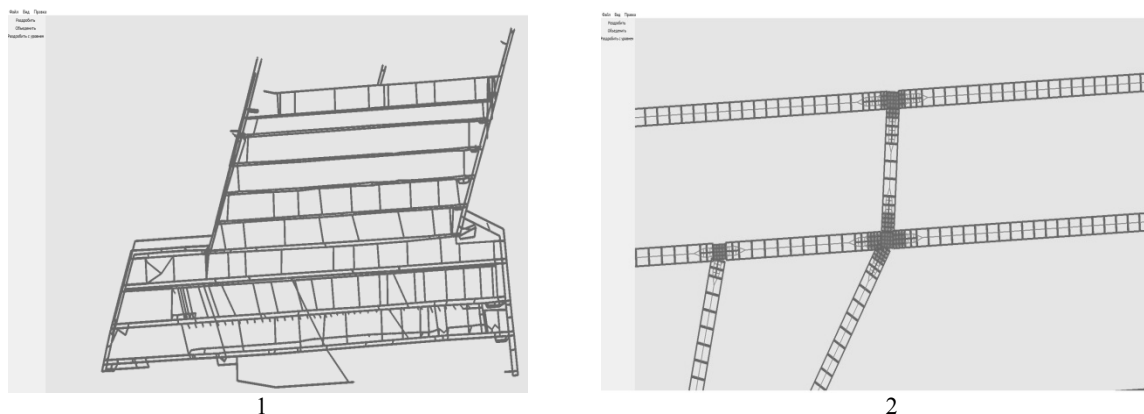


Рис. 12. Внешний вид приложения с загруженной сетью горных выработок шахты «Котинская» АО «СУЭК-Кузбасс»: 1 – общий вид; 2 – участок шахты с детализированной разностной сеткой
 Fig. 12. The view of an application with a loaded network of mine workings for the Kotinsky mine:
 1 – general view; 2 – mine section with a detailed differential grid

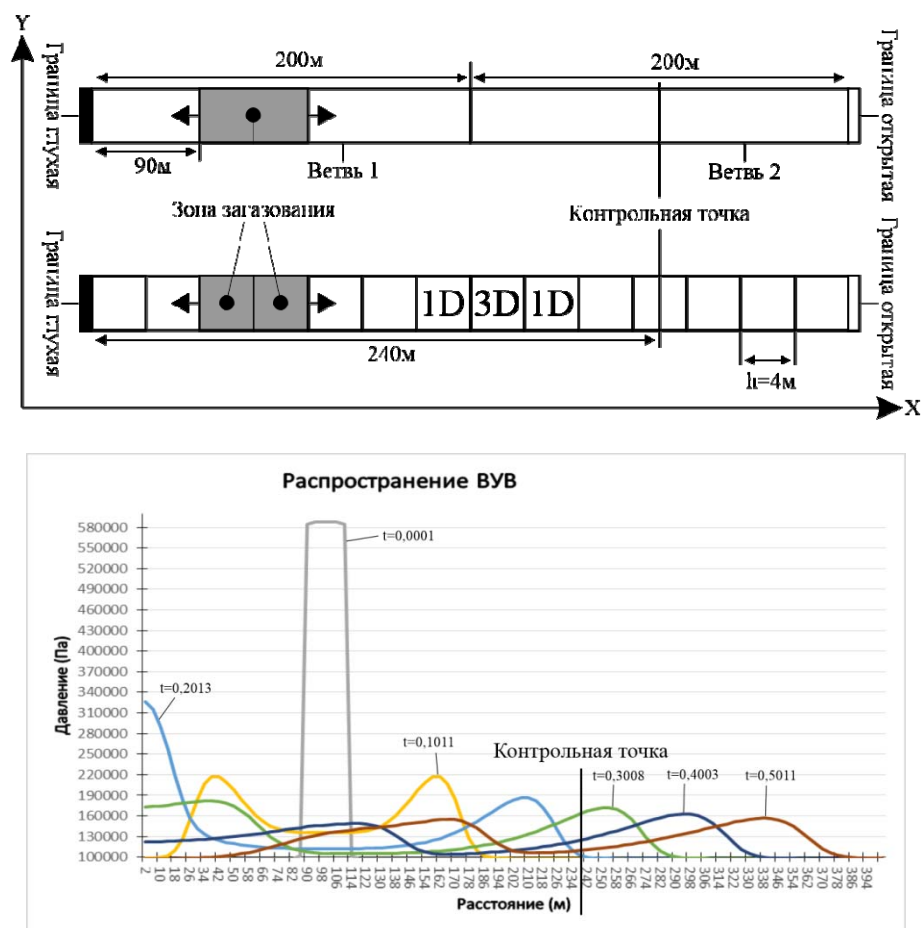


Рис. 13. Эксперимент: верху – модель для проведения численного эксперимента;
 внизу – график распространения ВУВ по модели в различные моменты времени при $h = 4$ м
 Fig. 13. Experiment: upper picture – a model for conducting a numerical experiment;
 lower picture – a graph of the propagation of the air shock wave at various points in time with $h = 4$ m

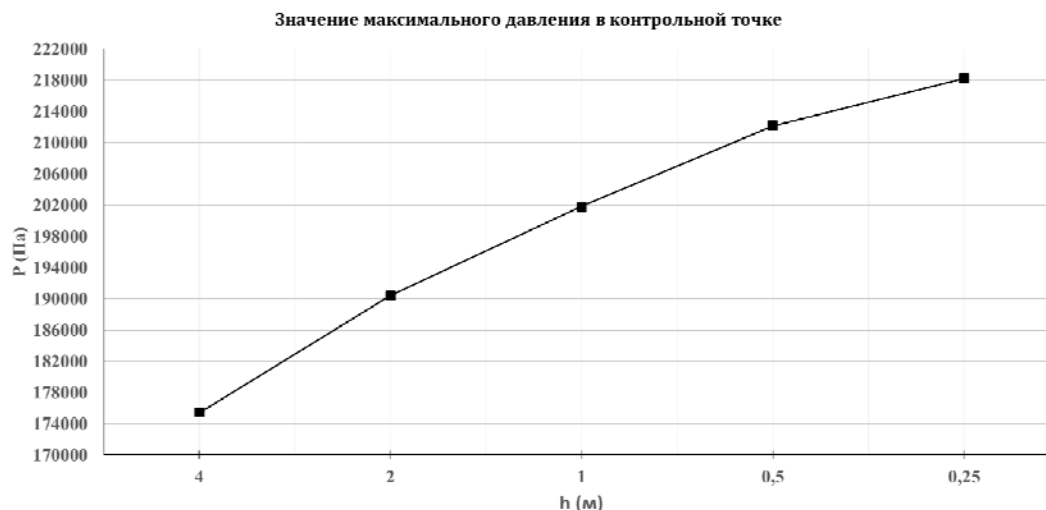


Рис. 14. Зависимость значения максимального давления ВУВ

от размера шага по пространству в контрольной точке: горизонтальная ось (h) – шаг по пространству, вертикальная ось (P) – давление в контрольной точке

Fig. 14. The dependence of the maximum pressure of an air shock wave on the step size in space at the control point: the horizontal axis (h) is the step in space, the vertical axis (P) is the pressure at the control point

Видно, что с уменьшением h давление стремится к некоторому предельному значению – точному решению, что объясняется снижением аппроксимационной вязкости и, как следствие, к более выраженному профилю фронта ВУВ. Этим же подтверждается сеточная сходимость используемого в Методике разностного метода и корректность созданной технологии дискретизации расчетной области.

Заключение

В результате выполненных исследований разработана технология локальной детализации разностных сеток, основанная на использовании двоичных, квадрато- и октодеревьев, и эффективного способа адресации ячеек. Адресация позволяет отказаться от использования методов вычислительной геометрии для связывания смежных ячеек при дроблении расчетной сетки. Важной особенностью технологии является возможность связывать ячейки разной размерности с произвольными уровнями дискретизации. Результаты исследований планируется использовать при исследовании газодинамических процессов при расчетах распространения ВУВ в горных выработках угольных шахт. Технология реализована в виде компьютерной программы на языке программирования C++.

Список литературы

1. Лукашов О. Ю. Исследование волновых эффектов, возникающих при распространении ударных волн по разветвленной сети горных выработок: Дис. ... канд. физ.-мат. наук. Томск, 2003. 141 с.
2. Устав ВГСЧ по организации и ведению горноспасательных работ. М.: Недра. 1986. 254 с.
3. Устав военизированной горноспасательной части (ВГСЧ) по организации и ведению горноспасательных работ на предприятиях угольной и сланцевой промышленности. М., 1997. 201 с.

4. **Палеев Д. Ю., Лукашов О. Ю., Васенин И. М., Шрагер Э. Р., Крайнов А. Ю., Костеренко В. Н.** Моделирование распространения ударных волн от взрыва и горения газопылевой смеси в угольных шахтах // Научные технологии разработки и использования минеральных ресурсов. 2017. № 3. С. 371–377.
5. Математическое моделирование горения и взрыва высокоэнергетических систем / Под ред. И. М. Васенина. Томск, 2006. 322 с.
6. **Палеев Д. Ю., Лукашов О. Ю., Васенин И. М., Шрагер Э. Р., Крайнов А. Ю., Костеренко В. Н.** Гашение энергии ударной волны водяным заслоном при взрыве слоевого скопления метана // Научные технологии разработки и использования минеральных ресурсов. 2017. № 3. С. 385–389.
7. **Палеев Д. Ю., Лукашов О. Ю., Васенин И. М., Шрагер Э. Р., Крайнов А. Ю., Костеренко В. Н.** Взаимодействие ударной волны взрывов метана с облаком порошкового ингибитора // Научные технологии разработки и использования минеральных ресурсов. 2017. № 3. С. 377–381.
8. **Палеев Д. Ю., Аксенов В. В., Лукашов О. Ю., Васенин И. М., Крайнов А. Ю., Шрагер Э. Р.** Моделирование аэродинамических процессов в вентиляционных сетях современных горнодобывающих предприятий // Горный информационно-аналитический бюллетень (научно-технический журнал). 2015. № S7. С. 224–230.
9. **Васенин И. М., Костеренко В. Н., Крайнов А. Ю., Лукашов О. Ю., Палеев Д. Ю., Шрагер Э. Р.** Газодинамический метод расчета нестационарных нормальных и аварийных режимов проветривания шахты // Пожарная безопасность. 2018. № 1. С. 32–41.
10. **Лукашов О. Ю.** О комплексном подходе к моделированию аварийной ситуации при взрыве газа в угольной шахте // Вестник Том. гос. ун-та. Математика и механика. 2014. № 6 (32). С. 86–93.
11. **Палеев Д. Ю., Лукашов О. Ю.** Распространение поражающих факторов взрыва в сети горных выработок // Вопросы безопасности труда на горных предприятиях: Сб. науч. тр. Кемерово: КузГТУ, 2003.
12. **Палеев Д. Ю.** Ударные волны при взрывах в угольных шахтах // Промышленная безопасность. М.: Горное дело; ООО «Киммерийский центр», 2011. 312 с. (Библиотека горного инженера. Т. 6, кн. 3)
13. **Афендинов А. Л., Давыдов А. А., Луцкий А. Е., Меньшов И. С., Меркулов К. Д., Пленкин А. В., Ханхасаева Я. В.** Адаптивные вейвлетные алгоритмы для решения задач гидро- и газовой динамики на декартовых сетках. М.: ИПМ им. М. В. Келдыша, 2016. 234 с.
14. **Сушинов А. А.** Построение декартовых сеток с динамической адаптацией к решению // Математическое моделирование. 2010. Т. 22, № 1. С. 86–98.
15. **Меркулов К. Д., Меньшов И. С.** Динамически перестраиваемые декартовы сетки с локальным измельчением для расчета задач газовой динамики // Физико-химическая кинетика в газовой динамике. 2015. Т. 16 (4). С. 1–11.
16. **Laurmaa V., Picasso M., Steiner G.** An octree-based adaptive semi-Lagrangian VOF approach for simulating the displacement of free surfaces. *Computers and Fluids*, 2016, vol. 131, p. 190–204.
17. **Olshanskii M. A., Terekhov K. M., Vassilevski Y. V.** An octree-based solver for the incompressible Navier – Stokes equations with enhanced stability and low dissipation. *Computers and Fluids*, 2013, vol. 84, p. 231–246.
18. **Aftosmis M. J.** Solution Adaptive Cartesian Grid Methods for Aerodynamic Flows with Complex Geometries. Lecture notes for 28th Computational Fluid Dynamics Lecture Series. Von Karman Institute for Fluid Dynamics. Belgium, 1997.
19. **Ахо А., Хопкрофт Д. Э., Ульман Д. Д.** Структуры данных и алгоритмы: Учеб. пособие: Пер. с англ. М.: ИД Вильямс, 2000. 384 с.

References

1. **Lukashov O. Y.** Investigation of wave effects arising from the propagation of shock waves in a ramified network of mine workings. Diss. candidate of physico-mathematical sciences. Tomsk, 2003, 141 p. (in Russ.)
2. The charter VGSCH on the organization and conduct of mountain rescue. Moscow, Nedra, 1986, 254 p. (in Russ.)
3. The Charter of the militarized mine-rescue part (VGSCH) on the organization and conduct of mine-rescue work at the enterprises of the coal and shale industry. Moscow, 1997, 201 p. (in Russ.)
4. **Paleev D. Y., Lukashov O. Y., Vasenin I. M., Shrager E. R., Kraynov A. Y., Kosterenko V. N.** Simulation of the propagation of shock waves from an explosion and combustion of a gas-dust mixture in coal mines. *High technology development and use of mineral resources*, 2017, vol. 3, p. 371–377. (in Russ.)
5. Mathematical modeling of combustion and explosion of high-energy systems. Ed. by I. M. Vasenin. Tomsk, 2006, 322 p. (in Russ.)
6. **Paleev D. Y., Lukashov O. Y., Vasenin I. M., Shrager E. R., Kraynov A. Y., Kosterenko V. N.** The damping of the energy of the shock wave by a water barrier in the explosion of a layered accumulation of methane. *High technology development and use of mineral resources*, 2017, vol. 3, p. 385–389. (in Russ.)
7. **Paleev D. Y., Lukashov O. Y., Vasenin I. M., Shrager E. R., Kraynov A. Y., Kosterenko V. N.** The interaction of the shock wave of methane explosions with a cloud of powder inhibitor. *High technology development and use of mineral resources*, 2017, vol. 3, p. 377–381. (in Russ.)
8. **Paleev D. Y., Lukashov O. Y., Vasenin I. M., Shrager E. R., Kraynov A. Y., Kosterenko V. N.** Simulation of aerodynamic processes in the ventilation networks of modern mining enterprises. *Mining Information and Analytical Bulletin (scientific and technical journal)*, 2015, vol. S7, p. 224–230. (in Russ.)
9. **Vasenin I. M., Kosterenko V. N., Kraynov A. Y., Lukashov O. Y., Paleev D. Y., Shrager E. R.** Gas-dynamic method of calculating non-stationary normal and emergency modes of mine ventilation. *Fire Safety*, 2018, vol. 1, p. 32–41. (in Russ.)
10. **Lukashov O. Y.** On an integrated approach to modeling an emergency in a gas explosion in a coal mine. *Vestnik of Tomsk State University. Mathematics and mechanics*, 2014, vol. 6 (32), p. 86–93. (in Russ.)
11. **Paleev D. Y., Lukashov O. Y.** The spread of damaging factors of the explosion in the network of mine workings. In: Occupational safety issues in mining enterprises: Collection of scientific papers. Kemerovo, KuzSTU, 2003. (in Russ.)
12. **Paleev D. Y.** Shock waves during explosions in coal mines. In: Industrial Safety. “Mining Engineering” OOO “Cimmerian Center” Publ., 2011, 312 p. (Mining Engineer Library, vol. 6, book 3) (in Russ.)
13. **Afendikov A. L., Davydov A. A., Lutsky A. E., Menshov I. S., Merkulov K. D., Plenkin A. V., Hanhasaeva Ya. V.** Adaptive wavelet algorithms for solving problems of hydro gas dynamics on Cartesian grids. Moscow, IPM named after M.V. Keldysh. 2016. p. 234.
14. **Sukhinov A. A.** Construction of Cartesian grids with dynamic adaptation to a solution. *Mathematical Modeling*, 2010, vol. 22, no. 1, p. 86–98 (in Russ.).
15. **Merkulov K. D., Menchov I. S.** Dynamically tunable Cartesian grids with local grinding for calculating gas dynamics problems. *Journal of Physico-Chemical Kinetics in Gas Dynamics*, 2015, vol. 16 (4), p. 1–11. (in Russ.)
16. **Laurmaa V., Picasso M., Steiner G.** An octree-based adaptive semi-Lagrangian VOF approach for simulating the displacement of free surfaces. *Computers and Fluids*, 2016, vol. 131, p. 190–204.

17. **Olshanskii M. A., Terekhov K. M., Vassilevski Y. V.** An octree-based solver for the incompressible Navier – Stokes equations with enhanced stability and low dissipation. *Computers and Fluids*, 2013, vol. 84, p. 231–246.
18. **Aftosmis M. J.** Solution Adaptive Cartesian Grid Methods for Aerodynamic Flows with Complex Geometries. Lecture notes for 28th Computational Fluid Dynamics Lecture Series. Von Karman Institute for Fluid Dynamics. Belgium, 1997.
19. **Aho A., Hopcroft D. E., Ulman D. D.** Data structures and algorithms: A Stud. Manual. Moscow, Williams Publ., 2000, 384 p. (in Russ.)

Материал поступил в редколлегию
Received
25.04.2019

Сведения об авторах

Лукашов Олег Юрьевич, кандидат физико-математических наук, ООО «ШатхЭксперт-Системы» (пр. Комсомольский 13, офис 4, Кемерово, 650065, Россия)
olukashov@gmail.com

Костромин Дмитрий Александрович, аспирант, Томский государственный университет (пр. Ленина, 36, Томск, 634050, Россия)
kd2206@yandex.ru

Information about the Authors

Oleg Yu. Lukashov, Candidate of Physico-Mathematical Science, LLC ShatkExpert-Systems (13 Komsomolsky Ave., Kemerovo, 650065, Russian Federation)
olukashov@gmail.com

Dmitriy A. Kostromin, Graduate Student, Tomsk State University (36 Lenin Ave., Tomsk, 634050, Russian Federation)
kd2206@yandex.ru

Построение хрупкого цифрового водяного знака, применяемого коллективом авторов в исполняемых файлах

И. В. Нечта

*Сибирский государственный университет телекоммуникаций и информатики
Новосибирск, Россия*

Аннотация

Согласно данным статистических исследований, нарушение лицензионных соглашений ежегодно приносит огромные убытки компаниям, занимающимся разработкой программного обеспечения. С одной стороны, создаются нелегальные копии программного продукта, с другой стороны, несанкционированно используются какие-либо фрагменты программ третьими лицами. Еще одной важной проблемой является нарушение целостности программы, например в части блокировки функций проверки лицензионного ключа. В этой связи задача разработки методов защиты интеллектуальной собственности в программных приложениях является весьма актуальной. Известные ранее методы решают данную задачу при помощи хрупких водяных знаков. Большое внимание исследователи уделяют способам внедрения в программу цифровых водяных знаков (ЦВЗ) и методам их построения. Ниже предлагается метод построения хрупкого цифрового водяного знака, применяемого в исполняемых файлах. Рассматривается модель коллектива разработчиков, создающих защищенный ЦВЗ программный продукт. Применение данного метода позволит выявить факт нарушения целостности контейнера и позволит автору при необходимости подтвердить свое участие в создании и внедрении ЦВЗ. В данном методе используются математические свойства систем линейных алгебраических уравнений, электронно-цифровая подпись и криптографические хэш-функции. Схема базируется на теореме Кронекера – Капелли. Для нахождения группового пароля соавтор, входящий в группу, находит один из корней совместной системы линейных алгебраических уравнений. Указанная система состоит из n уравнений и содержит $n + 1$ неизвестную. Для постороннего, не входящего в группу, система уравнений не имеет единственного решения. Соавтор, входящий в группу, способен вычислить одну неизвестную системы на основе своих паспортных данных. Следовательно, система уравнений для такого соавтора имеет единственное решение. Проанализированы различные атаки на защищенную таким образом программу, показана эффективность предлагаемого метода. Разработанный метод может применяться в компаниях с большим коллективом разработчиков.

Ключевые слова

хрупкий цифровой водяной знак, стеганография, защита авторских прав, контроль целостности программ, коллективное авторское право, нарушение авторских прав, теорема Кронекера – Капелли, электронная цифровая подпись, защищенный программный продукт

Для цитирования

Нечта И. В. Построение хрупкого цифрового водяного знака, применяемого коллективом авторов в исполняемых файлах // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 65–73. DOI 10.25205/1818-7900-2020-18-1-65-73

Construction Fragile Digital Watermark Used by a Team of Authors in Executable Files

I. V. Nechta

*Siberian State University of Telecommunications and Information Sciences
Novosibirsk, Russian Federation*

Annotation

According to statistical research, a violation of license agreements annually causes huge losses to software companies. On the one hand, illegal copies of the software product are created, on the other hand, some fragments of the programs are used by third parties unauthorized. Another important problem is the violation of the program integrity, for example, in terms of blocking functions of the license key checking. In this regard, the task of construction methods for

protecting intellectual property in software applications is highly relevant. Previously known methods solve this problem by means of fragile digital watermarks. Below is presented a method for constructing a fragile digital watermark used in executable files. A model of a developers team creating software product protected by DWM is considered. The application of this method will allow to reveal the fact of the container integrity violation, on the one hand, and, on the other hand, will allow the author, if it is necessary, to confirm his participation in the development and embedding of the DWM. In this method we use mathematical properties of systems of linear algebraic equations, digital signature and cryptographic hash functions. The scheme is based on the Kronecker – Capelli theorem. To find the group password the co-author who is in the group finds one root of the consistent system of linear algebraic equations. The indicated system consists of n equations and contains $n + 1$ variables. For an outsider who is not in the group, the system of equations does not have a unique solution. The co-author of the group is able to calculate one variable in system based on their passport data. Therefore, the system of equations for such co-author has a unique solution. Various attacks on a protected by the new method are analyzed, and it is shown its efficiency. The constructed method can be applied in companies with a large team of developers.

Keywords

fragile digital watermark, steganography, copyright protection, program integrity control, collective copyright, violation of copyright rights, Kronecker – Capelli theorem, electronic digital signature, secure software product

For citation

Nechta I. V. Construction Fragile Digital Watermark Used by a Team of Authors in Executable Files. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 65–73. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-65-73

Введение

При создании и сопровождении цифрового контента одной из важных проблем является защита авторского права. Известно, что коммерческие организации несут значительные убытки вследствие несанкционированного копирования, распространения и неправомерного заимствования защищаемого цифрового объекта. Так, например, изображение, сделанное одним фотографом, может использоваться другим фотографом как собственное. Другим примером является незаконное использование фрагментов чужих программных приложений, например, библиотек.

Развитие стеганографии – науки, изучающей проблемы сокрытия факта передачи секретного сообщения, привело к созданию методов защиты объектов авторского права. Так, в файл может быть внедрено скрытое секретное сообщение, которое предназначено для идентификации автора. Если данный файл будет незаконно использоваться посторонним лицом, то настоящий автор может извлечь сообщение и продемонстрировать его в ходе судебных разбирательств. Таким образом будет установлено истинное авторство.

Встраиваемые в контейнер данные могут быть различными и зависят от решаемой задачи. В научной литературе существует два термина: *цифровой отпечаток пальца* (ЦОП) и *цифровой водяной знак* (ЦВЗ). ЦОП используется для идентификации владельца файла. Например, в каждую продаваемую копию программы автор встраивает паспортные данные лицензионного пользователя. При обнаружении пиратской копии программы, сообщение извлекается, и выявляется нарушитель лицензионного соглашения. С точки зрения правоприменительной практики (западных стран) данный факт является основанием для возбуждения следственных действий в отношении нарушителя.

Понятие ЦВЗ является более общим и включает в себя решение задач идентификации автора [1] и контроля целостности объекта данных или его фрагментов [2]. В случае с программами контроль целостности позволяет конечному пользователю удостовериться, что он использует неповрежденные вредоносными программами приложения.

Одна из задач стеганографии – это скрыть факт передачи секретного сообщения. В случае с защитой авторских прав это уменьшает вероятность успешной атаки злоумышленником, который явно не знает места размещения атакуемых данных. Поэтому сообщение внедряют в безобидный на вид объект данных, так называемый контейнер. В качестве контейнера могут быть текстовые, аудио-, видео- и исполняемые файлы. Каждый тип контейнера имеет свои свойства: вероятность обнаружения секретного сообщения, объем внедрения, устойчивость

внедрения к искажению контейнера и т. д. В силу большого внимания со стороны исследователей к стеганографии указанные свойства часто меняются при разработке новых методов стегоанализа (выявление факта внедрения).

На сегодняшний день существует большое число методов внедрения скрытых сообщений. Группа методов LSB внедрения, например, опубликованные в работах [3–6], применяются для встраивания сообщения в изображение и аудиоконтейнер. Используемый подход базируется на незначительном (с точки зрения восприятия человеком) изменении цвета изображения или звука. Например, запись сообщения производится в последний значащий бит цвета пикселя. Внедрение в текстовые контейнеры может осуществляться методом замены синонимов [7; 8]. Выбор синонима в предложении производится в соответствии со скрываемым сообщением. В данном случае меняется только форма записи смыслового содержания текста.

Для встраивания сообщений в исполняемые файлы используются статические [9; 10] и динамические цифровые водяные знаки [11; 12]. Статические ЦВЗ используют константы, различные формы записи команд (или групп команд) для хранения водяного знака. Однако такие методы не считаются устойчивыми к искажению. Например, исполняемый файл может быть упакован протектором [13], что сделает нечитаемым его водяной знак. Динамические водяные знаки создают некоторые динамические структуры данных в памяти, которые могут быть обнаружены по сигнатуре, известной только автору. Такие методы считаются устойчивыми не только к искажению, но и к стегоанализу.

Постановка задачи

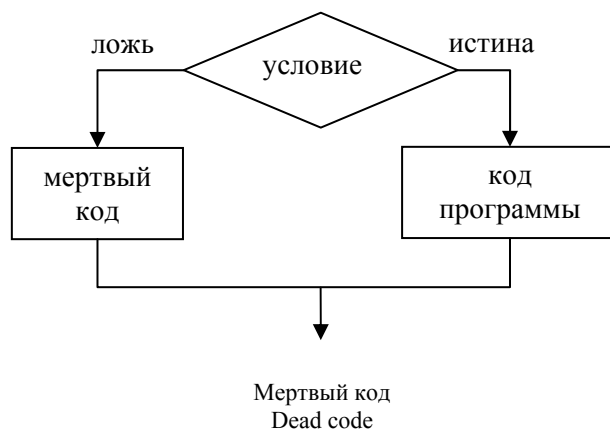
Для повышения устойчивости внедрения в исполняемых файлах используются технологии запутывания кода программы (obfuscation) [14] и технологии контроля целостности (tamper proofing) [15]. Первая предполагает изменение программы так, чтобы устранить возможность анализа программы с целью выяснения логики ее работы. Вторая технология использует эквиваленты цифровой подписи защищаемых участков кода. Согласно определению, сформулированному в работе [2], ЦВЗ считается *хрупким*, если он разрушается при любом редактировании программы. Свойство хрупкости реализуется благодаря использованию хэш-функций, которые меняют свое значение при искажении кода программы даже на один бит. Наличие этих технологий в исполняемых файлах допускает снижение требований к ЦВЗ по устойчивости внедрения, так как устойчивость обеспечивается другими средствами.

Существуют ситуации, когда программное средство разрабатывается коллективом разработчиков. При этом возникает необходимость встроить ЦВЗ, содержащий информацию обо всех авторах. Для решения данной задачи существуют следующие подходы.

Метод Easter egg [16] заключается в создании дополнительного функционала программе. Например, при выборе пункта меню «Об авторах» отображается окно, содержащее список коллектива разработчиков. Такой подход прост в реализации, но является уязвимым. Злоумышленнику легче контролировать процесс искажения программы, так как в явном виде демонстрируются сведения, которые нужно исказить.

Другой подход предполагает использование «мертвого кода», определяющего водяной знак. Такой код встраивается в программу, но никогда не выполняется. Пример представлен на (см. рисунок).

Условие формируется из «непрозрачных предикатов» – логических выражений, которые всегда либо истинны, либо ложны и не могут быть легко проанализированы при дизассемблировании программы [17]. Таким образом, трудно отличить не выполнявшиеся полезные участки кода от мертвого. Недостатком такого метода является увеличение размера программы.



Как уже было отмечено, одним из важных моментов при построении ЦВЗ является вопрос о наполнении водяного знака. Обычно встраивают паспортные данные авторов или их фотографии. Известны методы, например [18], встраивающие значения хэш-функций участков кода.

Целью данной работы является создание метода построения ЦВЗ, применяемого коллективом авторов. Особое внимание уделяется содержимому ЦВЗ. Способ представления этого содержимого (в виде констант, графа, мертвого кода и т. д.) выходит за рамки данного исследования. Предполагается, что данный метод будет использоваться для исполняемого файла.

Рассматривается следующая модель. Пусть имеется организация, создающая программный продукт. Коллектив разработчиков формирует и публикует итоговый исполняемый файл, содержащий ЦВЗ, который включает в себя некоторые параметры и электронную цифровую подпись (ЭЦП) организации (нотариуса).

При необходимости любой автор может подтвердить свое участие в разработке программы при помощи извлечения и анализа ЦВЗ. Целью злоумышленника является либо устранение данных о каком-либо авторе из ЦВЗ, либо замена / добавление новых данных.

Сформулируем предъявляемые требования к ЦВЗ (далее – Требования). Часть Требований к цифровым водяным знакам была уже ранее сформулирована, например, в работе [9].

1. Возможность обеспечить проверку целостности контейнера.
2. Посторонний не может выдать себя за автора.
3. Невозможность произвольно исключать данные автора из ЦВЗ.

Другая часть Требований сформулирована в рамках данной статьи в соответствии с рассмотренной выше моделью.

4. Автор способен подтвердить свои права без раскрытия конфиденциальных сведений о других авторах. Раскрытие конфиденциальных сведений об авторах (например, паспортных данных) возможно только с их согласия, что не считается безопасным действием.

5. Возможность объединять авторов в группы. Данное требование адаптировано к современной практике при создании программ, когда авторы объединены в группы, ответственные за определенный круг задач. Следовательно, появляется возможность определить долю участия автора.

6. Каждая группа имеет возможность реализовать механизм проверки ЦВЗ внутри программы. Здесь предусматривается ситуация, когда злоумышленниками являются члены какой-либо группы и нотариус. При выполнении данного требования достаточно одной группы, не содержащей злоумышленника, для корректной проверки ЦВЗ.

Последний пункт Требований накладывает ограничение на используемый метод встраивания скрытых данных, который должен позволять беспрепятственное считывание ЦВЗ.

Описание предлагаемого метода

Проанализируем требования к ЦВЗ. Достижение поставленной ранее цели может быть осуществлено классической ЭЦП, однако подпись всегда поставляется отдельно от контейнера и требует хранения дополнительных сведений. В отличие от ЭЦП предлагаемый водяной знак позволяет объединять авторов в группы. Указанное свойство требуется для определения доли участия. Например, программа состоит из нескольких модулей, реализованных разными подразделениями организации. ЦВЗ позволяет сгруппировать пользователей по подразделениям.

Рассмотрим предлагаемый метод более подробно. Пусть имеется коллектив авторов из N человек и известно их разбиение на M групп. Обозначим i – номер группы ($i \in [1, M]$) и k_i – размер группы i . Каждому автору ставится в соответствие некоторый элемент $a_{i,j}$, где j – номер автора в i -й группе ($j \in [1, k_i]$). Значение $a_{i,j}$ вычисляется как хэш-функция от паспортных данных автора. Хэш-функция должна обладать свойствами криптографических хэш-функций.

Введем элемент f_i , необходимый для разделения групп. Знание значения f_i есть подтверждение факта вхождения в группу (в таком смысле это групповой пароль). Вычисление происходит следующим образом¹:

$$f_i = \sum_j a_{i,j}.$$

Для каждой группы, состоящей из нескольких авторов, формируется система линейных алгебраических уравнений:

$$\begin{cases} f_i \oplus a_{i,j} = B_{i,j} \\ a_{i,j} \oplus a_{i,j+1} = B_{i,j+1} \\ \dots \\ a_{i,k-1} \oplus a_{i,k} = B_{i,k} \end{cases}$$

Для каждой группы, состоящей из одного автора, формируется уравнение, содержащее a_{null} , рассчитанное для заранее оговоренных паспортных данных (например, нулевые значения серии и номера):

$$a_{i,j} \oplus a_{null} = B_{i,j}.$$

Рассмотрим для примера разбиение на группы:

$$A = \{[a_{1,1}, a_{1,2}], [a_{2,1}, a_{2,2}, a_{2,3}], [a_{3,1}, a_{3,2}, a_{3,3}, a_{3,4}]\}.$$

Получаются следующие системы уравнений:

$$\begin{cases} f_1 \oplus a_{1,1} = B_{1,1} \\ a_{1,1} \oplus a_{1,2} = B_{1,2} \end{cases}$$

¹ Здесь и далее используется сложение по модулю 2.

$$\begin{cases} f_2 \oplus a_{2,1} = B_{2,1} \\ a_{2,1} \oplus a_{2,2} = B_{2,2} \\ a_{2,2} \oplus a_{2,3} = B_{2,3} \end{cases}$$

$$\begin{cases} f_3 \oplus a_{3,1} = B_{3,1} \\ a_{3,1} \oplus a_{3,2} = B_{3,2} \\ a_{3,2} \oplus a_{3,3} = B_{3,3} \\ a_{3,3} \oplus a_{3,4} = B_{3,4} \end{cases}$$

Обозначим H – значение хэш-функции от пустого контейнера, введем $Sign(x)$ – электронную цифровую подпись сообщения x . Предполагается, что ЭЦП использует ключ нотариуса. Конкретизация ЭЦП, которая должна удовлетворять требованиям ГОСТ, выходит за рамки исследования.

Опишем процесс формирования ЦВЗ.

Шаг 1. Нотариус формирует ЭЦП $Sign(H \oplus T)$, где $T = \sum_i B_{i,j}$.

Шаг 2. Последовательно составляются значения: ЭЦП, необходимых данных для ее проверки (например, для RSA – открытая часть ключей) и множество значений $B_{i,j}$ из систем уравнений.

Шаг 3. Полученные значения внедряют в исполняемый файл.

На этапе составления систем уравнений все расчеты доступны внутри группы, следовательно, автор должен запомнить позицию начала собственной группы и позицию собственных данных внутри группы. Для предотвращения добавления ложного автора нотариусом необходимо контролировать число элементов B с помощью средств, реализованных внутри программы.

Рассмотрим действия, необходимые для подтверждения авторства.

Шаг 1. Извлечение ЦВЗ.

Шаг 2. Расчет значения хэш-функции от пустого контейнера. Ранее отмечалось ограничение в выбираемых методах внедрения. Для статических и динамических ЦВЗ пустой контейнер получается путем перезаписи ЦВЗ нулевыми значениями, при этом нужно только знать места его хранения в программе. Это требование не является проблемой для разработчиков программы.

Шаг 3. Вычисление значения $a_{i,j}^*$ на основе собственных паспортных данных и остальных неизвестных $(a_{i,j} : a_{i,j} \neq a_{i,j}^*)$ в системе уравнений соответствующей группы.

Шаг 4. Проверяется корректность подписи $Sign(H \oplus T)$ и числа значений B , если ошибки нет, то переходим к Шагу 5. В противном случае считаем, что контейнер был искажен и авторство не подтверждается.

Шаг 5. Если для группы выполняется условие $\sum a_{i,j} = f_i$, то авторство признается подлинным, иначе ложным.

Анализ выполнения Требований

Проанализируем выполнение предложенным ЦВЗ Требований (в том же порядке).

1. Проверка целостности контейнера осуществляется при помощи вычисления хэш-функции от пустого контейнера и сравнения полученного значения с эталоном. Известно, что хэш-функция изменит свое значение даже при незначительных изменениях в контейнере, в том числе одного бита, что обеспечивает контроль целостности.

2. Посторонний злоумышленник не сможет выдать себя за автора, так как значение хэш-функции от его паспортных данных будет отличаться от используемых в ЦВЗ. Рассмотрим подробнее действия злоумышленника. Предположим, что злоумышленник не входит в состав авторов и хочет предъявить свои права на содержимое контейнера. Тогда ему необходимо указать группу, в которой находятся его данные. На основе паспортных данных злоумышленника вычисляется значение a_{calc} и остальные переменные соответствующей системы уравнения. Использование в методе криптографической хэш-функций делает маловероятным коллизии (совпадений значений хэшей) при различных паспортных данных.

Если группа состоит из 1 человека, то рассчитанное значение B не будет совпадать с имеющимся в ЦВЗ, так как $(a_{i,1} \oplus a_{null}) \neq (a_{злой} \oplus a_{null})$. Если группа состоит из нескольких человек, то злоумышленник не сможет назвать значение f_i , доказывающее принадлежность к группе. Действительно, для нахождения f_i необходимо решить систему уравнений. Уравнения составляются так, что число неизвестных больше на единицу, чем количество уравнений. В таком случае согласно теореме Кронекера – Капелли система не имеет единственного решения. Таким образом, атака постороннего не будет успешной.

3. Произвольное исключение или изменение данных об авторе из ЦВЗ невозможно, так как наличие и корректность этих значений контролируются программой. Рассмотрим случай, когда злоумышленником является один из авторов. Его целью является исключение данных кого-либо из авторов («жертвы») из водяного знака. Допустим, злоумышленник решает обнулить сведения жертвы. Тогда меняются соответствующие значения $B_{i,j}$ и T , следовательно, подпись $Sign(H \oplus T)$ становится неверной. Подделать подпись злоумышленник не может. При выявлении факта разрушения ЦВЗ программа может становиться неработоспособной (например, генерируется программное исключение), что недопустимо для злоумышленника.

Предположим, что нотариус является злоумышленником, тогда он может попытаться либо добавить, либо заменить, либо удалить кого-либо из авторов. Добавление или удаление данных как одного нового автора, так и группы, в ЦВЗ вызовет изменение числа элементов $B_{i,j}$, что отслеживается. Замена данных о каком-либо авторе, во-первых, не дает знания группового пароля f_i , а во-вторых, приведет к изменению значений $B_{i,j}$ в соответствующей системе уравнений. И даже при правильной ЭЦП будет отслежено участниками группы.

Предположим, что злоумышленниками являются один член группы авторов и нотариус. Тогда злоумышленники могут заменить данные какого-либо автора в группе без нарушения ЭЦП. Однако в таком случае также меняется значение группового пароля, что отслеживается средствами контроля целостности ЦВЗ.

4. При доказательстве авторства необходимы только конфиденциальные сведения заявителя. Остальные элементы уравнения (из которого строится ЦВЗ) представляют собой хэш-значения и не раскрывают данных других авторов.

5. Метод позволяет добавлять сведения об авторах, объединенных в группы, как из одного человека, так и из нескольких.

6. Каждая группа имеет возможность реализовать механизм проверки ЦВЗ внутри программы. Действительно, разрабатываемый группой авторов код может содержать в себе функции проверки. Расчет используемых хэш-значений и составление уравнений не является трудной задачей для авторов, чьи данные встраиваются в ЦВЗ.

Таким образом, мы видим, что все сформулированные пункты Требований выполняются.

Заключение

Предложенный в данной работе метод позволяет создавать ЦВЗ, применяемый коллективом авторов, с возможностью подтверждения авторства над исполняемым файлом. Предложенная схема базируется на математических свойствах систем линейных алгебраических уравнений. Применение криптографических хэш-функций и ЭЦП реализует свойство хрупкости водяного знака, что позволяет контролировать целостность программы.

Предложенный метод обладает следующими особенностями. Во-первых, требуется хранение меньшего объема информации в отличие от схемы с ЭЦП. Во-вторых, не требуется регистрация и обслуживание ЭЦП для каждого автора в установленном порядке. В-третьих, метод позволяет объединять авторов в группы, что может быть полезным, например, для установления доли участия автора при разработке программы.

В работе были рассмотрены различные атаки злоумышленника. Показано, что, несмотря на активные действия со стороны злоумышленника, Требования, предъявляемые к ЦВЗ, выполняются. Предложенный подход может применяться при создании систем защиты авторского права.

Список литературы / References

1. **Dey A., Bhattacharya S., Chaki N.** Software watermarking: Progress and challenges. *INAE Letters*, 2019, vol. 4, no. 1, p. 65–75.
2. **Chen Z., Wang Z., Jia C.** Semantic-integrated software watermarking with tamper-proofing. *Multimedia Tools and Applications*, 2018, vol. 77, no. 9, p. 11159–11178.
3. **Zhou X., Gong W., Fu W., Jin L.** An improved method for LSB based color image steganography combined with cryptography. In: *IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 2016, p. 1–4.
4. **Muhammad K., Ahmad J., Rehman N., Jan Z., Sajjad M.** CISSKA-LSB: color image steganography using stego key-directed adaptive LSB substitution method. *Multimedia Tools and Applications*, 2017, vol. 76, no. 6, p. 8597–8626.
5. **Liu J., Tian Y., Han T., Yang C., Liu W.** LSB steganographic payload location for JPEG-decompressed images. *Digital Signal Processing*, 2015, vol. 38, p. 66–76.
6. **Atoum M. S.** A Comparative Study of Combination with Different LSB Techniques in MP3 Steganography. *Information Science and Applications*, 2015, vol. 339, p. 551–560.
7. **Xiang L., Li Y., Hao W., Yang P., Shen X.** Reversible natural language watermarking using synonym substitution and arithmetic coding. *Computers, Materials and Continua*, 2018, vol. 55, no. 3, p. 541–559.
8. **Xiang L., Wang X., Yang C., Liu P.** A novel linguistic steganography based on synonym run-length encoding. *IEICE transactions on Information and Systems*, 2017, vol. 100, no. 2, p. 313–322.
9. **Collberg C., Sahoo T. R.** Software watermarking in the frequency domain: implementation, analysis, and attacks. *Journal of Computer Security*, 2005, vol. 13, no. 5, p. 721–755.
10. **Gong D., Liu F., Lu B., Wang P., Ding L.** Hiding information in java class file. *International Symposium on Computer Science and Computational Technology*, 2008, vol. 2, p. 160–164.
11. **Ma H., Jia C., Li S., Zheng W., Wu D.** Xmark: Dynamic Software Watermarking using Collatz Conjecture. *IEEE Transactions on Information Forensics and Security*, 2019, vol. 14, no. 11, p. 2859–2874.
12. **Wang Y., Gong D., Lu B., Xiang F., Liu F.** Exception handling-based dynamic software watermarking. *IEEE Access*, 2018, vol. 6, p. 8882–8889.

13. **Hai N. M., Ogawa M., Tho Q. T.** Packer identification based on metadata signature. In: Proceedings of the 7th Software Security, Protection, and Reverse Engineering / Software Security and Protection Workshop, 2017, p. 1–11. DOI 10.1145/3151137.3160687
14. **Behera C. K., Bhaskari D. L.** Different obfuscation techniques for code protection. *Procedia Computer Science*, 2015, vol. 70, p. 757–763.
15. **Ahmadvand M., Pretschner A., Kelbert F.** A taxonomy of software integrity protection techniques. *Advances in Computers*, 2019, vol. 112, p. 413–486.
16. **Nematollahi M. A., Vorakulpipat C., Rosales H. G.** Software Watermarking. *Digital Watermarking*, 2017, vol. 11, p. 131–144.
17. **Xu D., Ming J., Wu D.** Generalized dynamic opaque predicates: A new control flow obfuscation method. *International Conference on Information Security*, 2016, vol. 9866, p. 323–342.
18. **Ahmadvand M., Hayrapetyan A., Banescu S., Pretschner A.** Practical Integrity Protection with Oblivious Hashing. In: Proceedings of the 34th Annual Computer Security Applications Conference, 2018, p. 40–52.

Материал поступил в редакцию
Received
08.11.2019

Сведения об авторе

Нечта Иван Васильевич, кандидат технических наук, доцент, начальник отдела подготовки кадров высшей квалификации, Сибирский государственный университет телекоммуникаций и информатики (Новосибирск, Россия)
ivannechta@gmail.com
ORCID 0000-0003-0361-2742
Researcher ID B-1604-2014

Information about Author

Ivan V. Nechta, PhD (Eng), Head of the department of training highly qualified personal Siberian State University of Telecommunications and Information Sciences (Novosibirsk, Russian Federation)
ivannechta@gmail.com
ORCID 0000-0003-0361-2742
ResearcherID B-1604-2014

УДК 81'33: 519.76
DOI 10.25205/1818-7900-2020-18-1-74-82

Автоматическая обработка текстов на основе платформы ТХМ с учетом анализа структурных единиц текста

Ф. Н. Соловьев

*Федеральный исследовательский центр «Информатика и управление» РАН
Москва, Россия*

Аннотация

В настоящей работе мы приводим описание интеграции средств автоматической обработки текста (выделение псевдооснов, именных групп, анализ глагольного управления) для расширения аналитических возможностей платформы корпусного анализа ТХМ. Представленные средства объединены в единый программный комплекс, что позволяет эффективно осуществлять подготовку специальных корпусов для последующего анализа средствами платформы ТХМ.

Ключевые слова

автоматический анализ текстов, платформа ТХМ, псевдоосновы, именные группы, глагольное управление

Благодарности

Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 19-07-00806

Для цитирования

Соловьев Ф. Н. Автоматическая обработка текстов на основе платформы ТХМ с учетом анализа структурных единиц текста // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 74–82. DOI 10.25205/1818-7900-2020-18-1-74-82

Embedding Additional Natural Language Processing Tools into the TXM Platform

F. N. Soloviev

*Federal Research Center "Informatics and Management"
Moscow, Russian Federation*

Abstract

In our work we present a description of integration of natural language processing tools (pseudostem extraction, noun phrase extraction, verb government analysis) in order to extend analytic facilities of the TXM corpora analysis platform. The tools introduced in the paper are combined into a single software package providing TXM platform with an effective specialized corpora preparation tool for further analysis.

Keywords

natural language processing, TXM platform, pseudostems, noun phrases, verb government

Acknowledgements

The research was supported by the RFBR, grant no. 19-07-00806

For citation

Soloviev F. N. Embedding Additional Natural Language Processing Tools into the TXM Platform. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 74–82. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-74-82

© Ф. Н. Соловьев, 2020

Введение

Задачи автоматической обработки текстов, например классификация и выделение нетривиальных языковых конструкций, в настоящее время представляют серьезный интерес как для научного сообщества, так и для решения практических задач, и были, в частности, исследованы нами в [1; 2].

В настоящей работе в контексте решения указанных задач мы опираемся на программный комплекс – платформу ТХМ (<http://textometrie.org>). Платформа ТХМ является эффективным средством корпусного анализа, позволяющим проводить комплексный анализ корпусов (анализ соответствий, кластеризация, построение лексических таблиц, поиск сложных лексических конструкций, выделение подкорпусов по различным параметрам). Платформа ТХМ интегрирована с расширением TreeTagger [3], позволяющим проводить лишь морфологический анализ и лемматизацию словоупотреблений. Она использует словоупотребления в качестве структурных единиц анализа.

Для повышения эффективности таких используемых ТХМ методов, как анализ специфичности и анализ соответствий, целесообразно ввести в рассмотрение новые единицы анализа, опирающиеся на процедуры автоматизированной обработки текстов на естественных языках, описанные в [4].

Мы предлагаем ряд расширений, позволяющих дополнить и усложнить анализ корпусов, включающий автоматический морфологический анализ словоформ и приведение их к канонической форме, выделение псевдооснов, выделение именных и глагольных групп и комбинирование результатов работы предлагаемых расширений. Конечной целью дополнений к платформе ТХМ является создание механизмов для исследования применимости различных дифференцирующих признаков при решении задачи классификации текстов и создания тематических корпусов текстов.

В [5–8] мы провели эксперименты по использованию псевдооснов и именных групп для выявления экстремистской направленности текстов. В данной работе к этим характеристикам добавлены возможности учета глагольного управления.

Псевдоосновы

Для определения дифференцирующих признаков коротких текстов сети Интернет, характеризующихся особыми тематическими и психолингвистическими свойствами, содержащих неологизмы и жаргонизмы, большой интерес представляет использование аналитического метода выделения псевдооснов, так как он позволяет обрабатывать отсутствующие в стандартных словарях формы.

Используемый способ выделения псевдооснов представляет собой метод структурных схем, описанный подробно в [9]. Суть метода состоит в получении псевдоосновы словоформы путем рассмотрения ее словоизменительных аффиксов. Словообразовательные аффиксы считаются в рамках этого метода элементом корневой части и не отбрасываются. Далее под аффиксами мы будем понимать исключительно словоизменительные аффиксы. С каждым словом можно сопоставить отвечающую ему последовательность аффиксов. Такие последовательности называются структурами некорневой части слова. Отсюда происходит название метода. Как и в традиционном морфологическом анализе, аффиксы подразделяются на префиксы и суффиксы в соответствии с их позицией относительно корня слова. Псевдоосновой называется часть слова, не содержащая суффиксов и префиксов. Способ автоматического выделения псевдооснов состоит в сопоставлении рассматриваемой словоформы с множеством допустимых в языке структур некорневой части слова. Псевдооснова слова выделяется отбрасыванием всех соответствующих определенной структурной схеме аффиксов (т. е. допустимой в данном языке максимальной комбинации префиксов и суффиксов). У глаголов, в частности, отбрасываются показатели лица, числа, рода, времени, причастной формы. Видовые префиксы не отбрасываются, так как они могут влиять на лексическое значение слова.

Псевдооснова не всегда совпадает с основой слова в традиционном понимании. Например, в словоформе *людьми* единственным аффиксом, который можно отбросить согласно продуктивной структурной схеме, является *-и*, поэтому выделяется псевдооснова *людьм*.

Данный подход позволяет анализировать текстовые конструкции, опираясь не только на точные словоформы, и тем самым повышает полноту и гибкость корпусного анализа.

Морфологические характеристики

Возможность привести словоформу к канонической форме позволяет анализировать различные элементы словоизменительной парадигмы как одну и ту же структурную единицу текста. Это, в свою очередь, позволяет более корректно проводить содержательный статистический анализ текста, например, путем рассмотрения частот лексем вместо частот отдельных словоформ.

При предобработке всех русскоязычных текстов мы осуществляем автоматический морфологический анализ словоформ на основе словарной компьютерной морфологии, описанной в [4]. Используемая стандартная в отечественной компьютерной лингвистике морфологическая модель относит каждое слово к одному из 24 морфологических классов, включающих, помимо частей речи в традиционном понимании, такие разряды, как «неизменяемое слово», «аббревиатура», «топоним». Каждый из этих морфологических классов характеризуется набором грамматических характеристик: род, падеж, число, наклонение и др. В программной реализации словарной морфологии русского языка применяется специализированная структура данных, позволяющая осуществлять поиск словоформ за линейное по числу букв словоформы время. Каждая словоформа содержит свои грамматические характеристики и ее каноническую (начальную) форму.

В настоящей работе мы также использовали интегрированный в ТХМ программный пакет TreeTagger [3], предоставляющий возможность совместного морфологического анализа слов предложения на основе статистической модели путем сопоставления словоупотреблений, снабженных специальными метками, кодирующими морфологические характеристики. Преимуществом данной процедуры разметки является однозначность морфологического анализа, но при таком анализе существует риск ошибок, который возрастает, если текст содержит большое количество неологизмов и нестандартных написаний слов. Все виды морфологической разметки использовались в дальнейшем для сопоставительного анализа текстов корпуса.

Дополнительную информацию о специфическом содержании текста можно почерпнуть, анализируя не только словоформы, но и именные группы и глагольные группы целиком. В отличие от отдельных слов, выделенные именные и глагольные группы несут информацию о конкретных отдельных аспектах содержания текста.

Выделение именных групп

Именная группа определяется нами как группа слов, у которой главное слово существительное, а другие слова связаны с ним подчинительными синтаксическими связями. Рассмотрение частотных именных групп и их сочетаний в совокупности с анализом отдельных словоупотреблений позволяет получить более полную картину семантических и стилистических характеристик текста, релевантных его содержанию.

Определенную трудность при выделении именных групп представляет разрешение омонимической неопределенности, проистекающей из множественности морфологических разборов отдельных словоупотреблений, которая, как правило, имеет место. Наш метод выделения именных групп предполагает рассмотрение всего множества возможных морфологических разборов каждого слова.

Используемый нами алгоритм подробно описан в [4]. Алгоритм состоит из трех этапов: установление подчинительных синтаксических связей в предложении между парами слов; установление синтаксических связей внутри конструкций с однородными членами; выделение именных групп как цепочки последовательно связанных подчинительными связями слов.

Приведем вкратце основные моменты и определения, относящиеся к алгоритму выделения именных групп. В нашем изложении будем двигаться от простых структурных единиц к сложным: от отдельных словоупотреблений к именным группам.

Словоупотребление w можно рассмотреть как список результатов морфологического анализа, объединенных в группы по частям речи. Введем отображение $MA(w)$ из множества словоупотреблений во множество вариантов морфологического разбора:

$$MA: w \rightarrow (G_1, \dots, G_n),$$

где конкретный морфологический разбор словоупотребления w :

$$G_i = (GI_1, \dots, GI_k), \quad GI_j = (pos, gc, gender, number),$$

$$pos \in PoS, \quad gc \in GC, \quad gender \in Genders, \quad number \in Numbers,$$

где

PoS – множество частей речи русского языка;

GC – множество падежей русского языка;

$Genders$ – множество родов русского языка;

$Numbers$ – множество показателей числа;

GI_k – вариант морфологического разбора (морфологический разбор);

G_i – группа морфологических разборов.

Групп морфологических разборов G_i , вообще говоря, может оказаться несколько из-за явления омонимии в русском языке. Результаты морфологического анализа словоупотребления w могут быть представлены в разгруппированном виде, в качестве плоского списка вариантов морфологического разбора: $MA(w) = (GI_1, \dots, GI_p)$.

Введем подчинительную синтаксическую связь как пару $(w, RL(w))$, где w – словоупотребление, а $RL(w)$ – список подчиненных ему словоупотреблений. Обозначим за R множество подчинительных синтаксических связей.

Входными данными для алгоритма выделения именных групп являются предложения $Sent$, состоящие из словоупотреблений: $Sent = w_1, \dots, w_n$, где $w_i \in \Sigma^+$, при этом множество Σ – алфавит русского языка. Алгоритм состоит из трех этапов.

1. Определение подчинительных синтаксических связей в паре (lw, rw) , lw – левое слово, rw – правое слово в паре слов. Например, если есть предложение $Sent$, состоящее из слов w_1, w_2, w_3 , то пара (w_1, w_2) может образовать синтаксическую связь, в которой $lw = w_1$, $rw = w_2$.

2. Установление синтаксических связей внутри конструкций с однородными членами.

3. Выделение именных групп.

На множестве подчинительных синтаксических связей R как на ребрах (дугах) может быть построен ненаправленный граф $G = (V, E)$, $E = E_q \cup E_u$. Вершины этого графа – слова предложения, дуги – подчинительные синтаксические связи. Различают два типа ребер: q -ребра и u -ребра.

Если

$$e_{q_i} = (v_i, v_j),$$

то

$$\exists r_p \in R : v_j \in RL(v_i),$$

если

$$e_{u_i} = (v_i, v_j),$$

то

$$\exists r_p \in R : v_j \in RL(v_i), pos_{v_j} = ADJ.$$

Построенный граф служит выделению именных групп: в таком графе именная группа – это путь $p = (v_1, \dots, v_k)$, удовлетворяющий одному из следующих условий:

- 1) если $k = 2$, то $e = (v_1, v_2)$, $e \in E_q$;
- 2) если $k > 2$, то $\exists e_i \in E_u$ и $\exists e_j \in E_q$.

Множество $NG = \{p_1, \dots, p_f\}$ путей, полученных в результате работы алгоритма, будет множеством именных групп предложения.

Выделение глагольных групп

Выделение глагольных групп (словосочетаний, главным словом которых является глагол), т. е. установление связей выделенных именных групп с глаголами, представляет важную, необходимую составляющую синтаксического анализа предложения. Данная задача решается нами при помощи анализа глагольного управления.

Глагольным управлением называется разновидность синтаксической подчинительной связи типа управления, в которой главным словом является глагол. Как и в связи управления вообще, при глагольном управлении главным словом (глаголом) накладываются ограничения на употребление зависимого словосочетания в виде набора вариантов допустимых комбинаций грамматических характеристик зависимого словосочетания и, возможно, необходимых служебных частей речи. Такие комбинации мы в соответствующем контексте называем для удобства просто *ограничениями*. Указанные ограничения напрямую зависят от семантических свойств главного слова.

Пусть имеется множество глаголов A и a , типичный представитель этого множества – глагол.

С каждым глаголом может быть сопоставлен набор ограничений, накладываемых им на зависимые словосочетания. Такой набор мы называем *парадигмой управления* данного глагола или *парадигмой глагольного управления*, если мы отвлекаемся от самого глагола и рассматриваем набор его ограничений сам по себе. Обозначим как C_a парадигму управления глагола a .

Таким образом, каждую конкретную глагольную группу vg_a можно представить в виде тройки $vg_a = (a, p, c_a)$ – глагола a , выступающего в качестве главного слова, зависимого словосочетания p и представителя c_a парадигмы управления C_a глагола a , которому удовлетворяет рассматриваемое зависимое словосочетание p .

В нашей работе мы занимаемся выделением только таких глагольных групп, в которых зависимым словосочетанием является именная группа.

В русском языке ограничения на зависимые словосочетания указанного типа имеют простое общее представление в виде требования определенного предлога и падежа следующего за предлогом словосочетания либо требования определенного падежа и отсутствия предлога.

Обозначим множество предлогов как S , а его элементы – s , в которое для удобства включим специальный элемент s_0 , обозначающий «пустой» предлог, или отсутствие предлога. Множество падежей мы ввели выше как GC . Тогда рассматриваемые ограничения – это пары (s, gc) .

Так, к примеру, словосочетание *сделал в последний момент* может быть представлено как глагол *сделал*, зависимое словосочетание *в последний момент* и ограничение употребления словосочетания, состоящее из служебной части речи – предлога *в*, и ограничения на грамматическую характеристику падежа – винительный падеж.

Другой пример: *написал статью*. Здесь главное слово – *написал*, зависимое слово – *статью*, ограничение требует отсутствия предлога и винительного падежа.

Отдельно заметим, что поскольку природа ограничений носит характер, связанный со значением глагола, то сами ограничения не зависят от морфологических характеристик употребления глагола. Из указанного следует, что решение задачи выделения глагольных групп посредством анализа ограничений глагольного управления требует наличия словарей глагольного управления: особых словарей, приводящих глаголы вместе с их парадигмами глагольного управления.

Формально такой словарь можно представить в виде отображения

$$F_c : a \mapsto C_a.$$

Результаты морфологического анализа и процедуры выделения именных групп позволяют, используя словарь глагольного управления, выявить синтаксические связи для определения глагольных групп. Выделение глагольных групп в предложении осуществляется путем анализа всех возможных пар (глагол, именная группа) предложения на предмет соответствия именной группы парадигме управления соответствующего глагола, а именно поиска в парадигме ограничения, которому удовлетворяет рассматриваемая именная группа. Если такая группа находится, принимается решение о наличии связи управления между глаголом и именной группой. Если же в парадигме нет ни одного ограничения, которому бы удовлетворяла рассматриваемая именная группа, то принимается решение об отсутствии связи управления между глаголом и именной группой.

Более формально алгоритм выделения глагольных групп можно описать следующим образом. Пусть в предложении $Sent$ содержатся глаголы $A_{Sent} = \{a \in A | a \in Sent\}$, и именные группы $P_{Sent} = \{p | p \in Sent\}$. Пусть имеется словарь (отображение) глагольного управления F_c . Пусть F_{gc} – отображение, определяющее множество возможных падежей именной группы:

$$F_{gc} : p \mapsto \{gc_1, \dots, gc_k\}.$$

Тогда множество VG_{Sent} глагольных групп может быть определено как

$$VG_{Sent} = \{(a, p, c_a) | c_a \in F_c(a), c_a = (s, gc), gc \in F_{gc}(p), s \wedge p \in Sent\},$$

где выражение $s \wedge p$ обозначает предлог s с последующей именной группой p .

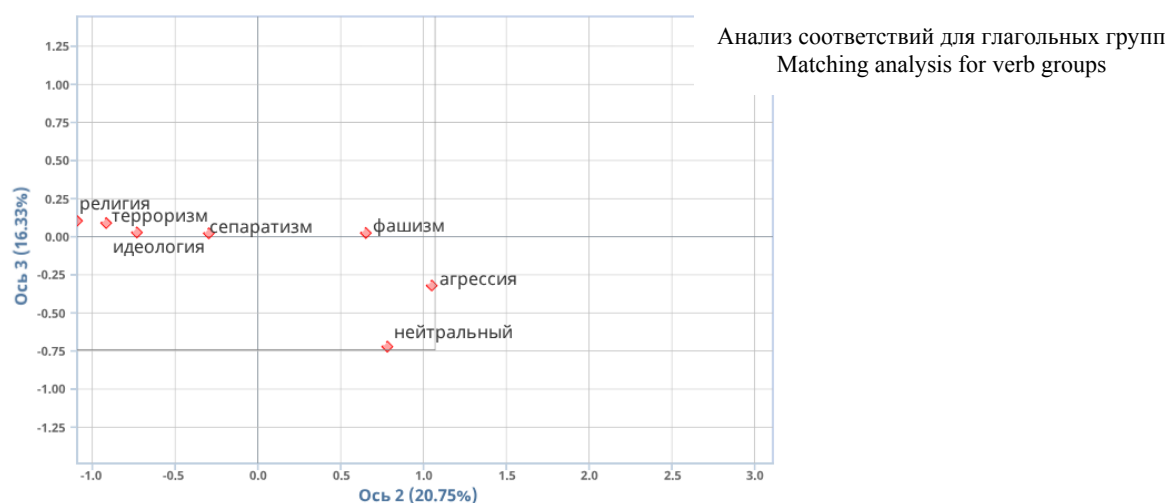
Для анализа глагольного управления был использован электронный словарь глагольного управления, в который вошли первые две тысячи наиболее частотных глаголов русского языка по материалам Национального корпуса русского языка (ruscorpora.ru). Словарь глагольного управления содержит парадигмы глагольного управления, состоящие из ограничений употреблений именных групп вида *предлог + падеж* или *отсутствие предлога + падеж*.

Анализ подкорпусов

Удобным инструментом количественной оценки «необычности» специального подкорпуса относительно всего корпуса является показатель специфичности [10]. Анализ специфичности позволяет составить своего рода «профиль» подкорпуса, выделенного на каких-либо внешних основаниях (например, автор, жанр, тематика или идеологическая направленность текста), путем выявления наиболее характерных или нехарактерных для него словоформ (лексем, псевдооснов, именных и глагольных групп и т. п.). Этот «профиль» может быть использован для диагностики нового текста.

Другим подходом к анализу разделенного на части (подкорпуса) по определенному критерию корпуса является анализ соответствий. Методика анализа соответствий, используемая ТХМ, была предложена Ж.-П. Бензекри [11] и имплементирована в пакете FactoMineR для платформы R [12]. Анализ соответствий демонстрирует взаимную «близость» или «удаленность» подкорпусов на основе анализа частот совместного появления значений переменных (словоформ, начальных форм, псевдооснов, именных групп, морфологических тегов и т. д.).

Экспериментальный корпус был проанализирован с использованием двух обозначенных выше функций ТХМ – специфичность и анализ соответствий. Детально были рассмотрены следующие лексические объекты: словоформы; начальные формы слов, полученные по словарной морфологии; начальные формы слов с морфологическими характеристиками, полученные с помощью TreeTagger; псевдоосновы слов; именные группы, составленные из словоформ; именные группы, составленные из начальных форм; именные группы, составленные из псевдооснов вместо отдельных словоупотреблений; глагольные группы.



На рисунке приведен пример применения анализа соответствий для глагольных групп, в рамках которых слова заменены на их псевдоосновы. Анализируемый корпус состоит из подкорпусов агрессивной, идеологической, националистической, фашистской, религиозной, сепаратистской, террористической и нейтральной направленностей. При делении текстов на подкорпуса есть возможность интерпретировать близость или разделенность значений рассматриваемых характеристик подкорпусов относительно друг друга как оценку, указывающую на сходство или различие маркированных подкорпусов между собой и по отношению к «нейтральному» подкорпусу.

Различные выделяемые дифференцирующие признаки, описанные выше, демонстрируют схожую разделяющую способность.

Заключение

Проведенная работа по интеграции инструментов автоматической обработки текста и платформы корпусного анализа ТХМ показала, что такая интеграция позволяет расширить возможности статистического анализа текстов.

Детально рассмотрены такие лексические объекты, как леммы, псевдоосновы, именные и глагольные группы различной структуры. Упомянутые средства были объединены в набор утилит, позволяющих вычислять для текстовых корпусов ряд характеристик языковых единиц, входящих в их состав. Корпуса с вычисленными характеристиками преобразуются нами в формат для импорта пакетом ТХМ.

В силу выявленных особенностей и противопоставленности нейтрального подкорпуса остальным сформированным корпусам может быть использован для машинного обучения в задачах классификации текстов на предмет выявления заданного содержания с целью их углубленного экспертного анализа.

Список литературы

1. Поляков И. В., Соловьев Ф. Н., Чеповский А. А., Чеповский А. М. Задача распознавания для текстов на естественных языках. М.: Национальный открытый университет «ИНТУИТ», 2017.
2. Поляков И. В., Соколова Т. В., Чеповский А. А., Чеповский А. М. Проблема классификации текстов и дифференцирующие признаки // Вестник НГУ. Серия: Информационные технологии. 2015. Т. 13, № 2. С. 55–63.
3. Schmid H. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of International Conference on New Methods in Language Processing. Manchester, UK, 1994.
4. Чеповский А. М. Информационные модели в задачах обработки текстов на естественных языках. Второе издание, переработанное. М.: Национальный открытый университет «ИНТУИТ», 2015.
5. Ананьева М. И., Девяткин Д. А., Кобозева М. В., Смирнов И. В., Соловьев Ф. Н., Чеповский А. М. Исследование характеристик текстов противоправного содержания // Тр. Ин-та системного анализа РАН. 2017. Т. 67, № 3. С. 86–97.
6. Лаврентьев А. М., Смирнов И. В., Соловьев Ф. Н., Суворова М. И., Фокина А. И., Чеповский А. М. Создание специальных корпусов текстов на основе расширенной платформы ТХМ // Системы высокой доступности. 2018. Т. 14, № 3. С. 76–81.
7. Лаврентьев А. М., Соловьев Ф. Н., Суворова М. И., Фокина А. И., Чеповский А. М. Новый комплекс инструментов автоматической обработки текста для платформы ТХМ и его апробация на корпусе для анализа экстремистских текстов // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2018. Т. 16, № 3. С. 19–31.
8. Chepovskiy A., Devyatkin D., Smirnov I., Ananyeva M., Kobozeva M., Solovyev F. Exploring linguistic features for extremist texts detection (on the material of Russian-speaking illegal texts). In: 2017 IEEE International Conference on Intelligence and Security Informatics: Security and Big Data, ISI 2017. Institute of Electrical and Electronics Engineers Inc., 2017, p. 188–190.
9. Egorova E., Chepovskiy A., Lavrentiev A. A structural pattern based method for automated morphological analysis of word forms in a natural language. *Journal of Mathematical Sciences*, 2016, vol. 214, no. 6, p. 802–813.
10. Lafon P. Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, 1980, no. 1, p. 127–165.
11. Benzécri J.-P. L'analyse des données: l'analyse des Correspondances. 2nd ed. Paris, Dunod, 1979, vol. 2.
12. Lê S., Josse J., Husson F. FactoMineR: An R package for multivariate analysis. *Journal of statistical software*, 2008, no. 25 (1), p. 1–18.

References

1. Polyakov I. V., Soloviev F. N., Chepovskiy A. A., Chepovskiy A. M. Zadacha raspoznavaniya dlya tekstov na yestestvennykh yazykah. Moscow, Natsional'nyy otkrytyy universitet "INTUIT", 2017. (in Russ.)
2. Polyakov I. V., Sokolova T. V., Chepovskiy A. A., Chepovskiy A. M. Problema klassifikatsii tekstov i differentsiruyushchiye priznaki. *Vestnik NSU. Series: Information Technologies*, 2015, vol. 13, no. 2, p. 55–63. (in Russ.)
3. Schmid H. Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of International Conference on New Methods in Language Processing. Manchester, UK, 1994.
4. Chepovskiy A. M. Informatsionnyye modeli v zadachah obrabotki tekstov na yestestvennykh yazykah. 2nd ed. Moscow, Natsional'nyy otkrytyy universitet "INTUIT", 2015. (in Russ.)
5. Ananieva M. I., Devyatkin D. A., Kobozeva M. V., Smirnov I. V., Soloviev F. N., Chepovskiy A. M. Issledovaniye harakteristik tekstov protivopravnogo sodержaniya. *Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk*, 2017, vol. 67, no. 3, p. 86–97. (in Russ.)
6. Lavrentiev A. M., Smirnov I. V., Soloviev F. N., Suvorova M. I., Fokina A. I., Chepovskiy A. M. Sozdaniye spetsial'nykh korpusov tekstov na osnove rasshirennoy platformy TXM. *Sistemy vysokoy dostupnosti*, 2018, vol. 14, no. 3, p. 76–81. (in Russ.)
7. Lavrentiev A. M., Soloviev F. N., Suvorova M. I., Fokina A. I., Chepovskiy A. M. Novyi kompleks instrumentov avtomaticheskoy obrabotki teksta dlya platformy TXM i yego aprobatsiya na korpuse dlya analiza ekstremistskikh tekstov. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2018, vol. 16, no. 3, p. 19–31. (in Russ.)
8. Chepovskiy A., Devyatkin D., Smirnov I., Ananyeva M., Kobozeva M., Solovyev F. Exploring linguistic features for extremist texts detection (on the material of Russian-speaking illegal texts). In: 2017 IEEE International Conference on Intelligence and Security Informatics: Security and Big Data, ISI 2017. Institute of Electrical and Electronics Engineers Inc., 2017, p. 188–190.
9. Egorova E., Chepovskiy A., Lavrentiev A. A structural pattern based method for automated morphological analysis of word forms in a natural language. *Journal of Mathematical Sciences*, 2016, vol. 214, no. 6, p. 802–813.
10. Lafon P. Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, 1980, no. 1, p. 127–165.
11. Benzécri J.-P. L'analyse des données: l'analyse des Correspondances. 2nd ed. Paris, Dunod, 1979, vol. 2.
12. Lê S., Josse J., Husson F. FactoMineR: An R package for multivariate analysis. *Journal of statistical software*, 2008, no. 25 (1), p. 1–18.

Материал поступил в редколлегию

Received

07.06.2019

Сведения об авторе

Соловьев Федор Николаевич, младший научный сотрудник, Федеральный исследовательский центр «Информатика и управление» РАН (Вавилова, 44/2, Москва, 119333, Россия) the0@yandex.ru

Information about the Author

Fyodor N. Soloviev, Junior Researcher, Federal Research Center "Informatics and Management" of the Russian Academy of Sciences (44/2 Vavilov Str., Moscow, 119333, Russian Federation) the0@yandex.ru

Виртуальный атлас персонифицированной анатомии человека «SkiaAtlas» и возможности его применения

Б. О. Щеглов¹, Н. И. Безуленко¹, С. А. Атащиков¹, С. Н. Щеглова²

¹ *Дальневосточный федеральный университет
Владивосток, Россия*

² *Северо-Восточный государственный университет
Магадан, Россия*

Аннотация

Работа посвящена описанию структуры разработанного программного обеспечения «SkiaAtlas», которое ориентировано на работу с индивидуальными анатомическими моделями человеческого тела и физиологических показателей пациента. Показаны проблемы использования макетного и постсекционного материала в обучении студентов-медиков, а также преимущества разработанной информационной системы по сравнению с данными моделями. Виртуальные анатомические модели были получены из анонимизированных DICOM-изображений магнитно-резонансной томографии (МРТ) и компьютерной томографии (КТ). Описаны подсистемы информационной системы: PACS-сервер, где хранятся все данные (серверная часть), и веб-приложение, где пользователь ведет работу с данными (клиентская часть). Подробно описаны модули информационной системы, реализованные в виде различных программных продуктов: модуль импорта данных, модуль анонимизации, модуль СУБД, модуль визуализации и др. Порядок работы этих модулей проиллюстрирован схематично. Показано, на каких языках программирования и фреймворках реализовано данное программное обеспечение, представлены преимущества выбора данных средств реализации относительно других программных продуктов. Детально описан процесс удаления персональных данных из DICOM-файлов; процесс получения «маски» объекта на снимке, которая затем используется для получения объемных моделей внутренних органов пациента. Наглядно описан процесс пользовательской работы с базой данных и поиск патологий с помощью инструментов интерфейса системы. Показаны возможности применения данной информационной системы в образовательной сфере (иллюстрация конкретных клинических случаев с целью поиска причинно-следственных связей в патогенезе различных заболеваний) и для развития клинического мышления у студента. На конкретном клиническом случае приведен пример того, как программа «SkiaAtlas» использовалась для поиска патологии – объемного образования левого полушария головного мозга.

Ключевые слова

анатомия человека, программное обеспечение, PACS-структура сервера, база данных, персонализированная медицина

Благодарности

Благодарим Центр нейрохирургии и неврологии Медицинского центра ДВФУ и КГАУЗ «ВКБ № 2» г. Владивостока за предоставленные для исследования данные

Для цитирования

Щеглов Б. О., Безуленко Н. И., Атащиков С. А., Щеглова С. Н. Виртуальный атлас персонифицированной анатомии человека «SkiaAtlas» и возможности его применения // Вестник НГУ. Серия: Информационные технологии. 2020. Т. 18, № 1. С. 83–93. DOI 10.25205/1818-7900-2020-18-1-83-93

Virtual Atlas of Personified Human Anatomy “SkiaAtlas” and the Possibility of Its Application

B. O. Shcheglov¹, N. I. Bezulenko¹, S. A. Atashchikov¹, S. N. Scheglova²

¹ Far Eastern Federal University
Vladivostok, Russian Federation

² Northeast State University
Magadan, Russian Federation

Abstract

The work is devoted to the description of the structure of the developed SkiaAtlas software, which is focused on working with individual anatomical models of the human body and physiological parameters of the patient. The problem of using mock-up and post-sectional material in teaching medical students, and why the developed information system has advantages over these models, is shown. Virtual anatomical models were obtained from anonymous DICOM images of magnetic resonance imaging (MRI) and computed tomography (CT). The subsystems of the information system are described: a PACS server where all data is stored (server part) and a web application where the user works with data (client part). The information system modules implemented in the form of various software products are described in detail: data import module, anonymization module, DBMS module, visualization module, etc. The operation of these modules is illustrated schematically. It is shown in what programming languages and frameworks this software is implemented, and advantages of choosing these implementation tools relative to software are shown. The process of deleting personal data from DICOM files is described in detail; the process of obtaining the “mask” of the object in the picture, which is then used to obtain three-dimensional models of the patient’s internal organs. The process of user work with the database and the search for pathologies using the system interface tools are clearly described. The possibilities of using this information system in the educational field are shown – an illustration of specific clinical cases in order to search for cause-effect relationships in the pathogenesis of various diseases and the development of clinical thinking in a student. In a specific clinical case, an example is given of how the SkiaAtlas program was used to search for a pathology – a volumetric formation of the left hemisphere of the brain.

Keywords

human anatomy, software, PACS server structure, database, personalized medicine

Acknowledgments

Thank you for providing the research data. Center for Neurosurgery and Neurology, Medical Center, Far Eastern Federal University and Regional state autonomous healthcare institution “Vladivostok Clinical Hospital No. 2”, Vladivostok

For citation

Shcheglov B. O., Bezulenko N. I., Atashchikov S. A., Scheglova S. N. Virtual Atlas of Personified Human Anatomy “SkiaAtlas” and the Possibility of Its Application. *Vestnik NSU. Series: Information Technologies*, 2020, vol. 18, no. 1, p. 83–93. (in Russ.) DOI 10.25205/1818-7900-2020-18-1-83-93

Введение

Основной дисциплиной на начальных этапах медицинского образования являются нормальная человеческая анатомия, патологическая анатомия и топографическая анатомия. Будущие клинические дисциплины, как правило, требуют от студента достаточно широких знаний по этим базовым предметам – знание анатомической топографии является ключевым в изучении внутренних болезней, их инструментальной и лабораторной диагностики, их хирургического лечения. То, как развиваются патологические процессы и каковы изменения в человеческом теле при их развитии, рассказывает патологическая анатомия. Три эти дисциплины развивают главный навык лечащего врача – клиническое мышление [1].

Проблемой современных образовательных медицинских учреждений стала нехватка анатомического постсекционного (после патологоанатомического вскрытия) или макетного материала, на котором студенты отрабатывали бы свои знания¹. Анатомические модели, кото-

¹ Постановление от 21 июля 2012 г. № 750 «Об утверждении правил передачи невостробованного тела, органов и тканей умершего человека для использования в медицинских, научных и учебных целях, а также использования невостробованного тела, органов и тканей умершего человека в указанных целях».

рые выполнены из искусственных материалов, имитирующих настоящие ткани, являются дорогостоящими. Хранение постсекционного материала затруднительно, а также затратно.

Авторами предлагается система виртуального анатомического атласа «SkiaAtlas», которая состоит из двух программных подсистем: PACS-сервера (Picture Archiving and Communication System) – электронной базы файлов DICOM-данных (Digital Imaging and Communications in Medicine) и настольного приложения [1–6].

На базе PACS-сервера реализуется программная обработка DICOM-снимков компьютерной / магнитно-резонансной томографии (КТ / МРТ) – их анонимизация и получение 3D-модели внутренних органов. Таким образом, получают индивидуальные анатомические модели, которые могут быть использованы в различных областях медицины. Изучение этих моделей производится с помощью инструментов веб-приложения «SkiaAtlas» [2; 5; 6].

После обработки снимков КТ / МРТ различных пациентов получают клинические кейсы, при наличии дополнительных данных (результаты биохимического скрининга, клинического анализа крови / мочи, анамнеза пациента) к которым можно проследить патогенез заболевания и все патологические изменения [5].

Постановка задачи

Для использования результативных критериев оценки индивидуального состояния здоровья на каждом этапе жизни человека нужно наблюдение в динамике развития организма в норме и отслеживание трансформации его систем и органов при патологических изменениях биологических макрообъектов в виде снимков лучевых исследований с применением методов математико-статистического аппарата при инактивации гомеостатических констант в целом, в том числе при заболеваниях различного генеза [4; 6]. В связи с этим может использоваться аналитический метод, представленный в виде комплексной информационной системы постоянного наблюдения над морфологическими и патологическими изменениями макроструктур организма [1; 2; 6].

Визуализация с одновременным подключением изменения гомеостатических констант организма позволяет определять и анализировать этапы и процессы развития патологических изменений и дифференцировать определенные виды заболеваний среди сотен и тысяч различных нозологий [5; 6].

Процедура анализа имеет в своей основе проведение лучевого исследования (КТ и МРТ, ультразвуковое исследование) с получением DICOM-файла, выводом из него информации в виде снимков и дополнительных метаданных исследования (пол, возраст пациента, область исследования и др.) [2; 7].

Главным шагом к созданию данной системы является сбор DICOM-данных из открытых интернет-источников или больниц (Медицинский центр Дальневосточного федерального университета, Тысячечокская больница г. Владивостока), изучение заболеваемости в России и мире, анализ возможностей сегментации посредством алгоритмов фильтрации, изучения биофизики поглощения излучения различными тканями (шкала Хаунсфилда для КТ, константы скорости распространения ультразвука для УЗИ, зависимость частоты от магнитного поля Лармора для МРТ) и интеграции обработанных данных в 3D-модель [1; 2; 4–6].

Для практической реализации полученных данных разрабатывается современная комплексная информационная система (ИС) «SkiaAtlas» [2; 5; 6]. Разработка должна дать возможность получения, хранения и обработки данных лучевых исследований, проведения медико-статистических зависимостей между определенными константами организма в норме и при патологии, возможности подключения геоинформационных наблюдений за региональными уровнями заболеваемости и наличием превалирующих нозологий [1; 2; 4]. В дополнение к этому были инициализированы алгоритмы получения, анонимизации и вывода необходимых данных из DICOM-файла, использования методов фильтрации для определения (выделения «маски» объекта) необходимых для данного исследования тканей на основе их

физических свойств и методов машинного обучения (глубокого обучения с учителем, нелинейных полиномиальных регрессионных моделей), объединения последовательных изображений в единую 3D-модель [5–7]. Разрабатываемая система актуальна для медицинских, научных специалистов и студентов, так как она позволяет выполнять просмотр различных органов и тканей человека в норме и при патологии в динамике, проводить исследования корреляционных зависимостей от географии и времени выполненной лучевой диагностики, по полу пациента, а также встречаемости конкретных заболеваний.

Архитектура информационной системы

Механизмы, их функциональность и получаемый на выходе уровень эмерджентности от их взаимодействия друг с другом являются основными факторами для определения архитектуры ИС [3]. Порядок ее функционирования построен на базе технологии «клиент-сервер», в которую входят клиентская и серверная части, а также встраиваемая СУБД SQLite3 [8; 9]. В данной технологии архитектура ИС построена следующим образом: запросы или сетевые нагрузки от клиента (пользовательские команды) поделены между «подрядчиками» (серверами) [3; 8]. В физическом плане эти части являются программным обеспечением, которые согласуются между собой посредством сетевых наборов правил, позволяющих выполнять соединение и обмен данными между включенными в сеть устройствами. Программы сервера ждут от клиентских программ запросы и в ответ передают им свои ресурсы в виде данных или сервисных функций. В рассматриваемой системе клиентом является пользователь браузера, а сервера – веб-сервер [8; 9]. Правила функционирования поделены между клиентом и сервером, компоновка программы выполняется в СУБД SQLite3, сеть в функциональном плане выполняет роль обменного пункта информацией. Данное решение системы обладает важным неоспоримым преимуществом для обычного пользователя в виде кроссплатформенности, т. е. отпадает зависимость использования какой-либо определенной операционной системы [8].

Клиентская часть приложения является самым главным механизмом в функционировании всей системы. В качестве базового компонента применяется веб-фреймворк Django с использованием библиотеки графического интерфейса пользователя (GUI) PyQt5, полностью направленного на применение языка Python, который является высокоуровневым языком программирования при выполнении разнообразных задач. Основной плюс использования – синхронное взаимодействие с веб-сервером при изменении кода программы. В пользу языка Python говорят его быстродействие, «читаемость», развитое комьюнити, невысокие требования к серверному программному компоненту.

Пользовательский интерфейс описывается в клиентской части, также от нее формируются запросы к серверу и обработка ответов от него. Веб-интерфейс пользователя имеет в своем составе методы, посредством которых пользователь взаимодействует с веб-сайтом или каким-либо приложением через веб-обозреватель. Сетевые интерфейсы имеют повсеместное распространение из-за роста популярности Интернета и, соответственно, обширного расширения веб-обозревателей [4; 8].

Главными требованиями к веб-интерфейсам являются приятный человеческому глазу набор визуальных элементов управления (фронтенд) и оптимальная функциональность в рабочем процессе в разных веб-обозревателях. Отсутствие требования в виде использования разнообразных программных продуктов – немаловажное преимущество веб-интерфейсов, так как широко распространенные операционные системы уже имеют в своем составе установленный веб-браузер [8; 9].

Основные модули информационной системы

Работоспособность ИС обеспечивается суммарным взаимодействием программных продуктов, исполненных в виде отдельных модулей. Особенно выделим 9 из них (рис. 1).

1. Модуль импорта данных принимает на вход данные из файла формата *.dcm в базу данных путем выбора определенных тегов:

- a. (0008, 0020) «Study Date» – дата проведенного исследования;
- b. (0008, 0060) «Modality» – вид проведенного исследования (КТ, МРТ, ПЭТ, УЗИ и др.);
- c. (0008, 0070) «Manufacturer» – компания – производитель исследовательского оборудования;
- d. (0008, 0080) «Institution Name» – учреждение, где было произведено исследование;
- e. (0010, 0040) «Patient's Sex» – пол пациента;
- f. (0010, 1010) «Patient's Age» – возраст пациента;
- g. (0018, 0015) «Body Part Examined» – исследованная часть тела;
- h. (0020, 0000) «Group Length» – количество снимков в серии исследования;
- i. (0018, 0050) «Slice Thickness» – толщина слоя;
- j. (0018, 9323) «Exposure Modulation Type» – вид отображения (2D, когда сделано несколько снимков, или 3D, когда получены все снимки, исходя из размеров исследуемой области).

2. Модуль анонимизации – присваивание тегу (0010, 0010) «Patient's Name» сгенерированного ключа-идентификатора с помощью библиотеки uuid.

3. SQLite3 – встраиваемая компактная СУБД, в которой хранится информация о рассматриваемых случаях.

4. Модуль обработки данных на сервере отвечает за обработку поступающих необработанных данных от пользователя и, исходя из уровня доступа и запросов пользователя, выгрузку данных из базы данных и при модификации данных в БД.

5. Модуль наблюдения – принятие мер на действия пользователя в зависимости от поступающего запроса выбор нужного действия.

6. Модуль визуализации – интерфейс и воспроизведение информации на экране пользователя.

7. Модуль модификации – компонент, ответственный за изменение информации.

8. Модуль отчетов – часть системы, функция которой заключается в записи в лог-файл и обработку отчетов о работе системы.

9. Модуль запросов – элемент, отвечающий за отображение информации по вариативности анатомии и констант организма каждого организма.

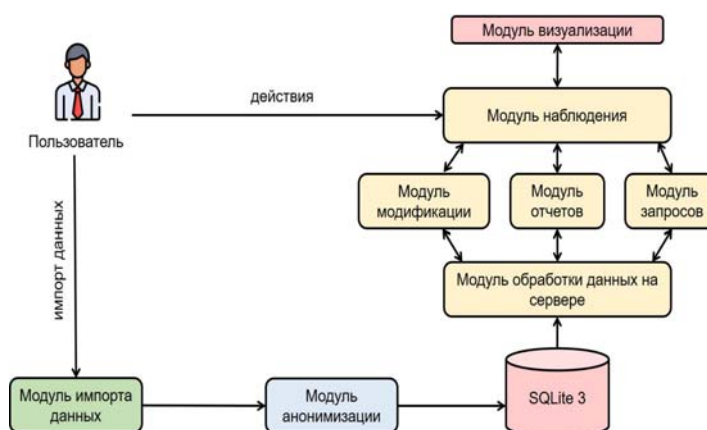


Рис. 1. Схема модулей базы данных SkiaAtlas

Fig. 1. Scheme of SkiaAtlas database modules

Любой модуль имеет строго определенную для него функциональность. При потребности внесения модификаций в тот или иной модуль надо изменить определенную функцию. Функционал каждого модуля можно увеличивать и совершенствовать, добавлять новые процедуры и функции.

Отображение данных и результатов анализа

Запустив браузер, пользователь вводит в строку поиска адреса URL веб-сайта информационной системы, что приводит к загрузке потока HTML-страниц, графики и клиентских сценариев на его компьютер. На стороне клиента веб-обозреватель реализует рендеринг полученного кода с представлением графических частей фронтенда: отображение табличной структуры базы данных по каждому рассматриваемому случаю, средства масштабирования и навигации с инструментами переключения различных типов исследования (рис. 2).

Табличная структура базы данных – основная составляющая интерфейса системы. Встроенное окно 3D-просмотра позволяет без проблем перемещаться по 3D-сцене с рассматриваемым анатомическим органом и изменять масштаб. На сцене можно просмотреть маленькие части, полученные в ходе обработки файлов лучевой диагностики, например сколы, разрывы в ходе механического воздействия на орган в целом, объемные патологические процессы. Элементы навигации и масштабирования состоят из различных кнопок, обеспечивающих навигацию и масштабирование веб-сцены.

Объекты, расположенные в таблице (см. рис. 2), обладают разными свойствами и признаками, полученными при работе модулей импорта, анонимизации и обработки поступающих данных. Они включают в себя: ключ-идентификатор; дата проведенного исследования; вид проведенного исследования; компанию-производителя исследовательского оборудования; учреждение, где было произведено исследование; пол обследованного и его возраст; часть тела, над которой проводилось лучевое исследование; количество снимков в серии исследования; толщина слоя одного снимка; вид отображения (2D, когда сделано несколько снимков, или 3D, когда получены все снимки, исходя из размеров исследуемой области); дополнительные данные (наличие биохимического и общего анализа крови, мочи, сопутствующих заболеваний, особенностей образа жизни). Данную систему можно использовать с целью демонстрации конкретных клинических случаев.

В качестве примера рассмотрим клинический кейс пациентки В (рис. 3).

Пациентка В., 1959 г. р., обратилась в стационар МЦ ДВФУ по направлению с жалобами на головные боли, припухлость левой височной области, судорожные приступы, снижение памяти. С лета 2018 г. появилась шаткость походки, приступы головокружения и дезориентации, по поводу чего пациентка обратилась к неврологу – было рекомендовано ЭЭГ, МРТ головного мозга. По результатам ЭЭГ были выявлены изменения: эпи-активность. По результатам МРТ головного мозга патологии не обнаружено. В стационаре МЦ ДВФУ после осмотра невролога была назначена КТ головного мозга с контрастным усилением. Результат этого исследования – объемное образование левой височной доли. Показано оперативное лечение. Диагноз симптоматической фокальной эпилепсии поставлен с учетом обнаружения объемного образования левой височной доли.

Таким образом, эпилептические приступы пациентки обусловлены объемным образованием. Причинно-следственные связи в патогенезе эпилепсии можно установить с помощью разработанной системы.

Информационную систему «SkiaAtlas» можно использовать для аналогичной демонстрации различных клинических кейсов совместно с анамнезом заболевания и данными сопутствующих исследований, что соответствует трем ключевым аспектам клинического мышления [10]:

- 1) усвоение информации о клинических симптомах, синдромах и нозологических формах;
- 2) накопление персонального клинического опыта;
- 3) развитие навыков решения проблемных клинических задач, особенно диагностических.

Изменить Данные пациента

Тип данных Открытый с номером id 1

Тип of patient:

dfdh3525u5upntr

Data of patient cipher:

Анамнез: Хроническая Почечная Недостаточность.
КТ головы Выполнялось для подтверждения/исключения опухоли.
Анализы на момент поступления в стационар:
Креатинин - 150
Тромбиновое время - 18,5
RV/AsV (Анализ крови RV – это серологический анализ на выявление антител к бледной трепонеме, AsV – анализ на наличие антицеллюлиновых антител) - 3

123

Анамнез: Хроническая Почечная Недостаточность.
КТ головы Выполнялось для подтверждения/исключения опухоли.
Анализы на момент поступления в стационар:
Креатинин - 150
Тромбиновое время - 18,5
RV/AsV (Анализ крови RV – это серологический анализ на выявление антител к бледной трепонеме, AsV – анализ на наличие антицеллюлиновых антител) - 3

Женщина

Data of patient sex:

30

Data of patient age:

CT

Type of study conducted:

-

Place of study conducted:

TOSHIBA

Apparatus on which the study was conducted:

HEAD

Part of body:

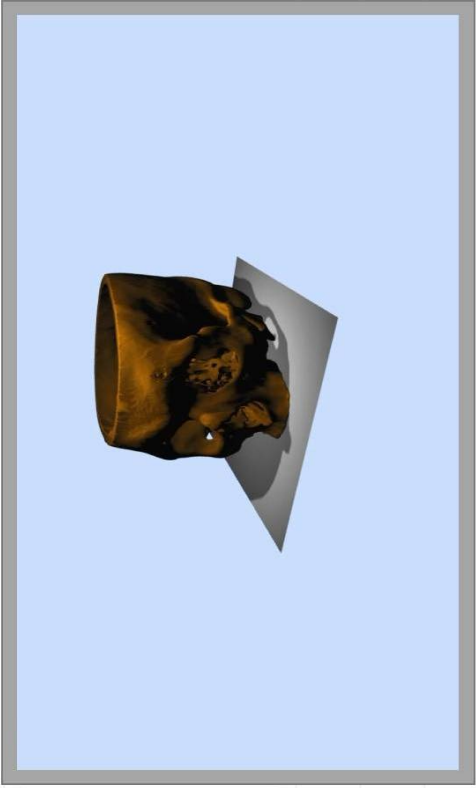


Рис. 2. Окно с информацией по каждому случаю
Fig. 2. Information window for each case

Type of patient:	Тип данных Открытый с номером id 1				
Date of patient entry:	metszhysgvrsh				
Data of patient description:	(МКБ D43.0) Объемное образование левой височной доли. Церебальный синдром Симптоматическая фокальная эпилепсия				
Date of patient sex:	Женщина				
Date of patient age:	60				
Type of study conducted:	CT				
Place of study conducted:	Medical Centre FEU				
Apparatus on which the study was conducted:	TOSHIBA				
Part of body:	HEAD				
Date of the study:	27.06.19				
☒ is active					



[Сохранить и продолжить редактирование](#)
[Сохранить и добавить другой объект](#)
[СОХРАНИТЬ](#)

Рис. 3. Демонстрация клинического случая пациентки В (объемное образование подсвечено красным)
Fig. 3. Demonstration of the clinical case of patient B (The volumetric formation is highlighted in red)

Данную информационную систему можно использовать для моделирования настоящих клинических ситуаций, когда необходимо из исходных данных (анамнеза пациента, данных лабораторных и инструментальных исследований) поставить клинический диагноз. Но возможно создание медицинских задач по моделированию, где необходимо будет не только поставить диагноз, но и назначить лечение согласно клиническим рекомендациям, провести различные прогностические, профилактические действия для предотвращения возникновения данного заболевания [11].

Заключение

Разработанная информационная система определена для работы с результатами проводимых медицинских методов исследования, включая данные по оборудованию, на котором были проведены исследования и места расположения медицинского учреждения, части тела, над которой проводилась диагностика, и дополнительные данные по клинко-диагностической лаборатории.

Информационная система имеет базу данных с таблицами, содержащими информацию по данным проведенного лучевого и клинко-диагностических исследований. Составлены серверная и клиентская части ИС, реализована возможность просмотра 3D анатомических моделей органов и информации о них, увеличения и изменения данных о проведенных методах медицинской диагностики с течением времени, что позволяет прослеживать трансформацию нормального органа в патологический в макромасштабе и при изменении физиологических констант организма.

Список литературы

1. **Безуленко Н. И., Щеглов Б. О.** Разработка виртуального атласа персонифицированной человеческой анатомии «SkiaAtlas» и изучение возможностей его применения в обучении студентов медицинских вузов // Информационные технологии: Материалы 57-й Междунар. науч. студ. конф. Новосибирск: ИПЦ НГУ, 2019. С. 41. ISBN 978-5-4437-0871-3.
2. **Безуленко Н. И., Щеглов Б. О.** Разработка системы SkiaAtlas для автоматизированного определения плотности костей черепа по КТ снимкам // Инновации и технологии в биомедицине: Сб. материалов науч.-практ. конф. Владивосток: Изд-во ДВФУ, 2019. С. 12–15. ISBN 9-78-5-7444-4583-6.
3. **Ружников Г. М. и др.** Современные технологии информационно-аналитической оценки // Бюл. СО РАМН. 2012. Т. 32, № 6 (57). С. 55–59.
4. **Щеглов Б. О., Багрянцев В. Н., Атарщиков С. А. и др.** Оценка качества диагностики спондилолистеза с помощью программ-слайсеров // Современные проблемы физики и технологий: Тез. докл. VII Междунар. молодежной науч. школы-конференции. М.: НИЯУ МИФИ, 2018. Ч. 2. С. 156–157. ISBN 978-5-7262-2467-1.
5. **Щеглов Б. О., Щеглова С. Н., Щелканов М. Ю. и др.** Эффективность внедрения лаборатории 3D-моделирования в лечебно-диагностических медицинских учреждениях // Якутский медицинский журнал. 2019. № 3 (67). С. 109–111. DOI 10.25789/YMJ.2019.67.30
6. **Щеглов Б. О.** Разработка программного обеспечения SkiaSorter для визуализации поликистоза печени // Современные проблемы физики и технологий: Тез. докл. VIII Междунар. молодежной науч. школы-конференции. М.: НИЯУ МИФИ, 2019. Ч. 1. С. 172–173. ISBN 978-5-7262-2576-0.
7. **Щеглов Б. О.** Восстановление некачественных КТ снимков с помощью нейронной сети // Дни науки: Сб. материалов науч.-практ. конф. студентов, аспирантов и молодых уче-

- ных: В 3 ч. Владивосток: Изд-во ДВФУ, 2018. Ч. 1: Материалы Школы биомедицины и Школы искусств и гуманитарных наук ДВФУ. С. 91–92. DOI 10.24866/7444-4309-2
8. **Загорулько Ю. А.** Технология разработки интеллектуальных научных интернет-ресурсов, ориентированная на экспертов предметной области // Инфраструктура научных информационных ресурсов и систем: Сб. избр. науч. ст. Труды IV Всерос. симп. / Под ред. Е. В. Кудашева, В. А. Серебрякова. М.: ВЦ РАН, 2014. Т. 1. С. 69–86.
 9. **Грегер С. Э.** Проектирование и реализация онтологии навигационной системы сайта // Объектные системы. 2012. № 1 (6). С. 19–24.
 10. **Кузьминов О. М., Крупенькина Л. А.** Формирование клинического мышления и дидактические задачи программных средств учебного назначения // Научные ведомости. Серия: Медицина. 2012. № 18 (161). С. 99–102.
 11. **Турчина Ж. Е., Белобородов А. А., Данилина Е. П.** Некоторые аспекты формирования клинического мышления у студентов младших курсов // Сибирское медицинское обозрение. 2013. № 4. С. 88–90.

References

1. **Bezulenko N. I., Shcheglov B. O.** Development of a virtual atlas of personal human anatomy “SkiaAtlas” and the study of the possibilities of its use in teaching students of medical universities. In: Information technology: Materials of the 57th International scientific stud. conf. Novosibirsk, NSU Press, 2019, p. 41. ISBN 978-5-4437-0871-3. (in Russ.)
2. **Bezulenko N. I., Shcheglov B. O.** Development of the SkiaAtlas system for automated determination of the density of skull bones from CT images. Innovations and technologies in biomedicine. Collection of materials of scientific and practical conf. Vladivostok, FEFU Press, 2019, p. 12–15. ISBN 9-78-5-7444-4583-6. (in Russ.)
3. **Ruzhnikov G. M. et al.** Modern technologies of information-analytical assessment. *Bull. SB RAMS*, 2012, vol. 32, no. 6 (57), p. 55–59. (in Russ.)
4. **Shcheglov B. O., Bagryantsev V. N., Atarshchikov S. A. et al.** Evaluation of the quality of diagnosis of spondylolisthesis using slicer programs. In: Modern problems of physics and technology. VII International Youth Scientific Shchool-Conference. Abstracts. Moscow, NRNU MEPhI, 2018, pt. 2, p. 156–157. ISBN 978-5-7262-2467-1. (in Russ.)
5. **Shcheglov B. O., Shcheglova S. N., Shchelkanov M. Yu. et al.** Efficiency of introducing a 3D-modeling laboratory in medical diagnostic medical institutions. *Yakutsk Medical Journal*, 2019, no. 3 (67), p. 109–111. DOI 10.25789/YMJ.2019.67.30 (in Russ.)
6. **Shcheglov B. O.** SkiaSorter software development for imaging of polycystic liver disease. In: Modern problems of physics and technology. VIII International Youth Scientific Shchool-Conference. Abstracts. Moscow, NRNU MEPhI, 2019, pt. 1, p. 172–173. ISBN 978-5-7262-2576-0. (in Russ.)
7. **Shcheglov B. O.** Recovery of low-quality CT images using a neural network. In: Days of Science. Collection of materials of the scientific and practical conference of students, graduate students and young scientists. Vladivostok, FEFU Press, 2018, part 1. Materials of the Shchool of Biomedicine and the School of Arts and Humanities of the FEFU, p. 91–92. DOI 10.24866/7444-4309-2 (in Russ.)
8. **Zagorulko Yu. A.** The development technology of intelligent scientific Internet resources, focused on domain experts. In: Infrastructure of scientific information resources and systems: Sat. Fav. Scientific Art. Proceedings of the Fourth All-Russian Symposium. Eds. E. V. Kudashева, V. A. Serebryakova. Moscow, 2014, vol. 1, p. 69–86. (in Russ.)
9. **Greger S. E.** Design and implementation of the ontology of the site’s navigation system. *Object systems*, 2012, no. 1 (6), p. 19–24. (in Russ.)

10. **Kuzminov O. M., Krupenkina L. A.** Formation of clinical comprehension and educational software didactic tasks. *Scientific sheets. Medicine series*, 2012, no. 18 (161), p. 99–102. (in Russ.)
11. **Turchina J. E., Beloborodov A. A., Danilina E. P.** Some aspects of the clinical thinking forming among undergraduate students. *Siberian medical review*, 2013, no. 4, p. 88–90. (in Russ.)

*Материал поступил в редколлегию
Received
26.11.2019*

Сведения об авторах

Щеглов Богдан Олегович, студент 4 курса Школы биомедицины, руководитель студенческого объединения «3D-моделирование в биомедицине» Центра проектной деятельности Дальневосточного федерального университета (Владивосток, Россия)
b.shcheglov@mail.ru

Безуленко Никита Иванович, студент 5 курса Школы биомедицины Дальневосточного федерального университета (Владивосток, Россия)
bezulenko.nik@bk.ru

Атарщиков Сергей Анатольевич, старший преподаватель Школы биомедицины Дальневосточного федерального университета (Владивосток, Россия)
kadmag@mail.ru

Щеглова Светлана Николаевна, кандидат педагогических наук, доцент, доцент кафедры математики и информатики, Северо-Восточный государственный университет (Магадан, Россия)
schegloff@mail.ru

Information about the Authors

Bogdan O. Shcheglov, 4th year student of the School of Biomedicine, head of the student association “3D-modeling in biomedicine” of the Center for Project Activities, Far Eastern Federal University (Vladivostok, Russian Federation)
b.shcheglov@mail.ru

Nikita I. Bezulenko, 5th year student of the School of Biomedicine, Far Eastern Federal University (Vladivostok, Russian Federation)
bezulenko.nik@bk.ru

Sergey A. Atarschikov, Senior Lecturer, School of Biomedicine, Far Eastern Federal University (Vladivostok, Russian Federation)
kadmag@mail.ru

Svetlana N. Shcheglova, Candidate of Pedagogical Sciences, Associate Professor, Associate Professor of the Department of Mathematics and Computer Science, North-East State University (Magadan, Russian Federation)
schegloff@mail.ru

Правила оформления текста рукописи

Авторы представляют статьи на русском или английском языке объемом от 0,5 авторского листа (20 тыс. знаков) до 1 авторского листа (40 тыс. знаков), включая иллюстрации (1 иллюстрация форматом 190×270 мм = 1/6 авторского листа, или 6,7 тыс. знаков). Публикации, превышающие указанный объем, допускаются к рассмотрению только после индивидуального согласования с редакцией журнала.

Текст рукописи должен быть представлен в редколлегию в виде файла MS Word (.doc, .docx). Гарнитура Times New Roman, размер шрифта 11, межстрочный интервал 1, размеры полей – стандартные значения текстового редактора. Форматирование – выравнивание по ширине страницы, переносы слов включены, каждый новый абзац начинается с красной строки. Не допускается ручное форматирование абзацев (пробелами, лишними переводами строк, разрывами страниц).

Структура статьи

- Индекс УДК (универсальной десятичной классификации). Выравнивание по левому краю
- Название статьи. Выравнивание по центру, полужирный шрифт
- ФИО авторов (инициалы, фамилия). Выравнивание по центру, полужирный шрифт
- Места работы всех авторов. Выравнивание по центру, курсив
- Аннотация статьи
- Ключевые слова, не более 10
- Благодарности, сведения о финансовой поддержке
- Название статьи **на английском языке**. Выравнивание по центру, полужирный шрифт
- ФИО авторов **на английском языке** (инициалы, фамилия). Выравнивание по центру, полужирный шрифт
- Места работы авторов **на английском языке**. Выравнивание по центру, курсив
- Аннотация статьи **на английском языке (Abstract)**, 200–250 слов
- Ключевые слова **на английском языке (Keywords)**, не более 10
- Благодарности, сведения о финансовой поддержке **на английском языке**, если есть соответствующий раздел на русском языке (**Acknowledgements**)
- Основной текст
- Список литературы / **References**
- Сведения об авторах

Требования к оформлению основного текста и иллюстративных материалов

Основной текст должен быть представлен в структурированном виде, рекомендуется использовать подзаголовки – например: Введение, Методика..., Выводы, Результаты, Заключение.

Подзаголовки отделяются и набираются полужирным шрифтом. В целях выделения частей текста и отдельных слов и словосочетаний допускается использование курсива или полужирного шрифта. Подчеркивание, разрядка, изменение основного кегля и выделение цветом не используются.

Иллюстрации к рукописи статьи должны быть приложены в виде отдельных файлов. При этом в тексте должно содержаться включенное изображение с указанием имени файла. Все

иллюстрации, содержащие схемы, графики, алгоритмы и т. п., должны быть представлены в векторном виде (.ai, .eps, .cdr). Скриншоты и другие растровые изображения должны быть представлены в максимально высоком качестве, без каких-либо потерь и искажений (.jpg, .tif). Все иллюстрации должны иметь подрисуночную подпись – свое название. Надписи к таблицам и подписи к иллюстрациям приводятся **на двух языках (русском и английском).**

Примеры:

Рис. 1. Диаграмма производительности...

Fig. 1. Performance diagram...

Таблица 1

Сравнение алгоритмов...

Table 1

Comparison of algorithms...

Нумерация последовательная и неразрывная от начала статьи. Не допускается использование других наименований, кроме «Рис.» / «Fig.», «Таблица» / «Table», и усложнение нумерации (например, «Рис. 3.2.»). Ссылка на иллюстрацию в тексте должна быть приведена в круглых скобках, например: (рис. 1), (табл. 1).

Формулы должны быть набраны с использованием редактора MathType либо встроенного редактора формул MS Word. Кегль основных символов – 11, греческие символы набираются прямым шрифтом, латинские – курсивом. Нумеруются только те формулы, на которые автор ссылается в тексте.

Abstract

Аннотация статьи на английском языке (Abstract) не должна быть дословным переводом русскоязычной аннотации. Раздел Abstract, как и основной текст, должен быть структурирован, в нем должно содержаться описание цели работы, методов исследования, научной значимости, выводов / результатов. Требуется качественный перевод на английский язык (при необходимости просим авторов обращаться к профессиональным переводчикам). **Объем Abstract 200–250 слов.**

Список литературы / References

Список литературы и список литературы на английском языке (References) размещаются в общем разделе. Рекомендованное количество цитируемых в статье источников – не менее 10, в список желательно включать ссылки на актуальные работы по теме исследования, особенно в иностранных периодических изданиях.

В тексте статьи ссылки на литературу указываются цифрами в квадратных скобках, при необходимости указываются номера страниц, например: [2; 3. С. 15].

Список литературы нумеруется в порядке цитирования и оформляется в соответствии с ГОСТ Р 7.0.5-2008 на библиографическое описание (знаки тире в описании опускаются). Ссылки на неопубликованные работы, а также на Интернет-ресурсы (кроме электронных изданий, поддающихся библиографическому описанию) оформляются в виде сноски.

В Список литературы ссылки на источники следует включать на оригинальном языке опубликования. Каждый источник должен быть также оформлен на английском языке (References) по международному стандарту для публикаций в области информатики IEEE Style со следующими отличиями:

- инициалы авторов указываются после фамилии;
- название статьи не берется в кавычки, отделяется точкой;
- отсутствует союз «and» перед фамилией последнего автора;
- в диапазоне страниц указывается одна «р» (вместо «pp. 2–9» – «р. 2–9»);
- год издания указывается после места издания (для книг) и сразу после названия журнала (для периодики).

Перевод источника на английский язык:

- если источник имеет выходные данные на английском языке, то для формирования References **следует использовать именно эти данные**;

- если оригинальная публикация не содержит выходных данных на английском языке, то допускается транслитерация названия материала на латинский алфавит в сочетании с переводом на английский язык в квадратных скобках. В конце описания указывается, на каком языке написана эта работа, например, (in Russ.). При транслитерации можно воспользоваться Интернет-ресурсом <http://ru.translit.ru/>, рекомендуется выбрать стандарт BSI. Место издания не транслитерируется, указывается полностью на английском языке, например: Moscow. Название издательства / издателя, как правило, транслитерируется. Для журналов, у которых есть официальное название на английском языке, – использовать его (проверить на сайте журнала, или, например, в библиотеке WorldCat), если названия на английском языке нет, использовать транслитерацию по системе BSI. Не следует самостоятельно переводить названия журналов.

Если у цитируемого источника есть **цифровой идентификатор DOI** (<https://search.crossref.org/>), его требуется обязательно указывать в конце библиографической ссылки.

Примеры оформления ссылок. Каждый источник в том же пункте дублируется на английском языке (References).

Источник на русском языке, перевод на английский доступен в метаданных статьи

1. Журавлев С. С., Рудометов С. В., Окольников В. В., Шакиров С. Р. Применение модельно-ориентированного проектирования к созданию АСУ ТП опасных промышленных объектов // Вестник НГУ. Серия: Информационные технологии. 2018. Т. 16, № 4. С. 56–67. DOI 10.25205/1818-7900-2018-16-4-56-67

Zhuravlev S. S., Rudometov S. V., Okolnishnikov V. V., Shakirov S. R. Model-Based Design Approach for Development Process Control Systems of Hazardous Industrial Facilities. *Vestnik NSU. Series: Information Technologies*, 2018, vol. 16, no. 4, p. 56–67. (in Russ.) DOI 10.25205/1818-7900-2018-16-4-56-67

Источник на английском языке. Оформляем согласно требованиям для References. Приводим только 1 раз.

2. Telnov V. I. Optimization of the Beam Crossing Angle at the ILC for E + e- and yy Collisions. *Journal of Instrumentation*, 2018, vol. 13, no. 03, p. P03020–P03020. DOI 10.1088/1748-0221/13/03/p03020

Метаданные источника доступны только на русском языке

3. Жижимов О. Л., Федотов А. М., Шокин Ю. И. Технологическая платформа массовой интеграции гетерогенных данных // Вестник НГУ. Серия: Информационные технологии. 2013. Т. 11, вып. 1. С. 24–41

Zhizhimov O. L., Fedotov A. M., Shokin Yu. I. Tekhnologicheskaya platforma massovoi integratsii geterogennykh dannykh [Technology Platform For the Mass Integration of Heterogeneous Data]. *Vestnik NSU. Series: Information Technologies*, 2013, vol. 11, no. 1, p. 24–41. (in Russ.)

Сведения об авторах

Последний раздел статьи – информация об авторе / авторах **на русском и английском языках**:

- ФИО полностью, ученая степень, ученое звание, должность, место работы, адрес места работы
- e-mail
- идентификаторы автора, такие как ORCID или ResearcherID (всем авторам рекомендуется использовать данные сервисы для ведения актуального списка своих публикаций)
- контактный телефон (не публикуется)

Если статья представляется на английском языке, необходимо приложить перевод на русский язык названия, аннотации, ключевых слов, сведений об авторе.

Доставка материалов

Материалы предоставляются в редакцию по электронной почте inftech@vestnik.nsu.ru.

Порядок рецензирования

Все статьи сначала проходят проверку на заимствование и только после этого отправляются на рецензирование. Редакционный совет не допускает к публикации материал, если имеется достаточно оснований полагать, что он является плагиатом.

Тип рецензирования статей – двухуровневое, одностороннее анонимное («слепое»).

Для каждой статьи редколлегией выбираются рецензенты, научная деятельность которых связана с темой представленного материала. Ответственный секретарь журнала обращается к ним с просьбой дать экспертную оценку статье либо помочь организовать рецензирование.

Рецензии для журнала «Вестник НГУ. Серия: Информационные технологии» составляются по единой схеме и подразумевают оценку по следующим критериям: соответствие тематике журнала, оригинальность и значимость результатов, качество изложения материала.

Заполненный бланк рецензии высылается на электронный адрес редакции. В зависимости от экспертных заключений статья может быть принята редакционным советом к опубликованию, рекомендована автору к доработке (с последующим повторным рецензированием либо без него) или отклонена (с предоставлением автору мотивированного отказа). Автору на электронный адрес высылается текст рецензии без указания ФИО рецензента и его контактных данных.

Все рецензии хранятся в редакции журнала не менее 5 лет. Редколлегия журнала обязуется при поступлении соответствующего запроса направлять копии рецензий в Министерство науки и высшего образования Российской Федерации.