

ОГЛАВЛЕНИЕ

ОТ СОСТАВИТЕЛЕЙ	2
О. М. Атаева, Н. Е. Каленов, В. А. Серебряков ОНТОЛОГИЧЕСКИЙ ПОДХОД К ОПИСАНИЮ ЕДИНОГО ЦИФРОВОГО ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ	3–19
В. И. Балута, В. П. Осипов, Т. В. Сивакова ПРЕДЛОЖЕНИЯ ПО РАЗРАБОТКЕ СРЕДСТВ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ В УСЛОВИЯХ ЭПИДЕМИЙ	20–41
Т. П. Баранова, А. Б. Бугеря, К. Н. Ефимкин МЕТОД БАЛАНСИРОВКИ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ ДЛЯ ГИБРИДНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ	42–56
К. П. Беляев, Г. М. Михайлов, А. Н. Сальников, Н. П. Тучкова СЕЗОННАЯ И МНОГОЛЕТНЯЯ ИЗМЕНЧИВОСТЬ АТМОСФЕРНОГО ДАВЛЕНИЯ В АРКТИКЕ, СТАТИСТИЧЕСКИЙ И ВРЕМЕННОЙ АНАЛИЗ	57–73
М. М. Горбунов-Посадов РОССИЙСКАЯ НАУЧНАЯ ПУБЛИКАЦИЯ ПЕРИОДА COVID-19	74–87
А. В. Ермаков НЕКОТОРЫЕ ИНСТРУМЕНТЫ ДЛЯ РАБОТЫ С БИБЛИОГРАФИЧЕСКИМИ ССЫЛКАМИ	88–100
Н. Е. Каленов ТЕХНОЛОГИЯ НАПОЛНЕНИЯ ПРЕДМЕТНЫХ ОНТОЛОГИЙ ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ	101–115
Ф. О. Каспаринский ВАРИАНТЫ ВИДЕОЛЕКЦИЙ	116–156
Н. А. Катаев СРЕДСТВА ИНТЕРАКТИВНОГО ВЗАИМОДЕЙСТВИЯ С ПОЛЬЗОВАТЕЛЕМ В СИСТЕМЕ SAPFOR	157–183
А. А. Печников, Д. Е. Чебуков О НЕКОТОРЫХ СВОЙСТВАХ ГРАФОВ СОТРУДНИЧЕСТВА УЧЁНЫХ В MATH-NET.RU	184–197

ОТ СОСТАВИТЕЛЕЙ

Настоящий тематический выпуск журнала «Электронные библиотеки» состоит из двух частей и включает статьи, подготовленные их авторами на основе материалов, представленных в 2020 году на XXII Всероссийской научной конференции «Научный сервис в сети Интернет».

Эта конференция была проведена с 21 по 25 сентября 2020 года и традиционно была посвящена направлениям и тенденциям использования интернет-технологий в современных научных исследованиях. Основная цель конференции — предоставить возможность для обсуждения, апробации и обмена мнениями о наиболее значимых результатах, полученных ведущими российскими учеными за последнее время в данной области деятельности. Организатором конференции был Институт прикладной математики им. М.В. Келдыша Российской академии наук. В связи со сложившейся эпидемической обстановкой конференция была проведена в режиме онлайн.

Первая часть тематического выпуска размещена в №1 журнала «Электронные библиотеки» за 2021 год, вторая часть – в №2.

М. М. Горбунов-Посадов, А. М. Елизаров

УДК 004.65 + 004.053 + 005 + 001.5

ОНТОЛОГИЧЕСКИЙ ПОДХОД К ОПИСАНИЮ ЕДИНОГО ЦИФРОВОГО ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ

О. М. Атаева¹ [0000-0003-0367-5575], Н. Е. Каленов² [0000-0001-5269-0988],
В. А. Серебряков³ [0000-0003-1423-621X]

¹Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва, ул. Вавилова, 40;

²Межведомственный суперкомпьютерный центр РАН, г. Москва, Ленинский проспект, 32а; Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва, ул. Вавилова, 40;

³Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва, ул. Вавилова, 40

¹oli@ultimeta.ru, ²nekalenov@mail.ru, ³serebr@ultimeta.ru

Аннотация

Несмотря на развитие технических средств, усложняются процессы, связанные с поиском полной и точной научной информации в огромном количестве источников данных. Для выхода на новый уровень в использовании технологий обработки информации, в первую очередь, необходим переход к семантически значимому представлению научных знаний, извлекаемых из информации в цифровой среде. В современных условиях, характеризующихся мультидисциплинарностью исследований, необходимого эффекта можно достичь, разработав универсальные подходы к хранению и представлению научных знаний. Эти подходы нашли свое отражение в концепции Единого цифрового пространства научных знаний. В работе представлен обзор основных понятий в этой области, используемых как для представления элементов пространства, так и для обеспечения доступа к ним не только для человека, но и для программных агентов. В качестве инструментария для конструирования пространства знаний рассмотрены семантические библиотеки.

Ключевые слова: *пространство знаний, цифровое пространство знаний, онтологии, метаданные, научные знания, уровни метаданных, проектирование онтологии, семантические библиотеки.*

ВВЕДЕНИЕ

Развитие цифровизации многих сторон жизни общества поставило его перед необходимостью накопления и обработки большого объема информации. Происходит интенсивное развитие информационных ресурсов нового типа, возникают новые научные ресурсы, широко использующие цифровое представление. Появилось большое количество источников информации, которые предоставляют данные в разных видах и форматах. Несмотря на развитие технических средств, усложняются процессы, связанные с поиском полной и точной научной информации, резко возрастает время, необходимое на обработку данных. С появлением парадигмы Semantic Web для решения этих проблем делаются попытки формализации знаний в различных областях науки на основе разрабатываемых онтологий. Это дает возможность семантической обработки информации, извлечения новых знаний.

Для выхода на новый уровень в использовании возможностей технологий обработки информации, бурно развивающихся сегодня, в первую очередь, необходим переход к семантически значимому представлению научных знаний, извлекаемых из информации в цифровой среде. Хотя каждая область науки имеет свою специфику, в современных условиях, характеризуемых мультидисциплинарностью исследований и взаимопроникновением научных направлений, необходимого эффекта можно достичь, только разработав универсальные подходы к хранению и представлению научных знаний. Эти подходы нашли свое отражение в концепции создания Единого цифрового пространства научных знаний (ЕЦПНЗ) [1, 2].

Развитие технологий позволяет шире взглянуть на определение ЕЦПНЗ и обобщить накопившийся опыт в реализации различных решений в этой области. Далее представлен обзор основных понятий в этой области, используемых как для представления элементов пространства, так и для обеспечения доступа к ним не только человека, но и программных агентов, что открывает широкие

возможности их обработки и использования в различных областях потребления заинтересованными участниками научной деятельности.

Системность научной информации [4, 5] подразумевает опору на исследование разнообразных зависимостей. Спецификой такой информации является четкая структура организации научных данных в иерархические структуры, пронизанные горизонтальными связями. Как следствие, обеспечивается однозначная интерпретация научного знания различными исследователями.

Определение *научных знаний* тесно связано с понятием *научной информации*, которая определяется по ГОСТ 7.0-99 [6] как логически организованная информация, получаемая в процессе научного познания и отображающая явления и законы природы, общества и мышления. Как видно из определения, эти два понятия (*научная информация* и *научные знания*) часто используются как синонимы. Далее в тексте использован термин *научные знания*, на наш взгляд, наиболее точно отражающий смысл.

1. СОСТАВЛЯЮЩИЕ ЕЦПНЗ

Под *пространством научных знаний* понимается проверенная научным сообществом система знаний из разных областей науки. При этом *цифровое пространство научных знаний* (ЦПНЗ) представляет собой цифровую среду, в которую интегрированы проверенные научным сообществом информационные ресурсы и объекты, отражающие научные знания из некоторой области науки. Подпространство ЦПНЗ является частью пространства, ограниченного рамками определенной предметной области. Фактически ЕЦПНЗ состоит из совокупности подпространств, относящихся к различным направлениям науки и построенных по единым принципам.

Несмотря на то, что существуют отдельные примеры формализации знаний в разных предметных областях [7–15], обобщенный подход к определению цифрового пространства научных знаний отсутствует. Анализ примеров формализации пространства знаний в различных областях свидетельствует о том, что основными составляющими ЕЦПНЗ в целом и каждого его подпространства в частности являются *онтология* и *контент*.

В качестве *контента* рассматриваются совокупность цифровых копий объектов реального мира и описание профилей их метаданных, тогда как *онто-*

логия включает универсальное описание структуры данных ЕЦПНЗ, т. е. онтология ЕЦПНЗ содержит классы объектов, отражаемых в каждом подпространстве, виды связей между этими классами и их объектами как внутри одного подпространства, так и между подпространствами, а также правила отражения объектов в ЕЦПНЗ.

2. ПОДХОДЫ К ПОСТРОЕНИЮ ОНТОЛОГИИ ЕЦПНЗ

Построение онтологии подпространства ЕПНЦЗ может быть представлено с точки зрения двух ортогональных подходов:

1. Вводятся термины, характерные для рассматриваемой научной предметной области, соединенные различными связями (как иерархическими, так и горизонтальными);

2. Вводится набор определений, который на более абстрактном уровне описывает множества объектов научной предметной области, фактически задавая структуру их описания и отношений между ними.

В различных исследованиях [3, 7–10, 16, 17] в обоих случаях говорят о построении или тезауруса, или онтологии предметной области. Но, по сути, это два совершенно разных подхода к описанию предметной области, которые не являются при этом взаимоисключающими, а должны дополнять друг друга. Такой подход позволяет, с одной стороны, отдельно сконцентрироваться только на типах информационных ресурсов, которые являются элементами пространства знаний, и описать основные понятия, характерные для этой предметной области. С другой стороны, говоря о тезаурусе, надо иметь в виду набор понятий и терминов, которые обеспечивают терминологическую поддержку понятий онтологии предметной области. Исходя из вышесказанного, считаем, что тезаурус пространства знаний – это полный систематизированный набор терминов какой-либо области знаний, в значительной мере относящийся к лексике, используемой в конкретной области, тогда как онтология описывает ресурсы предметной области и их взаимосвязи. Для каждой предметной области набор ресурсов может отличаться как по формату, так и по набору самих ресурсов.

Онтология научного пространства знаний [18]– это сложная многоуровневая система понятий, описывающих ресурсы и объекты предметной области, концептов, терминов и связей между ними, характеризующаяся открытой

иерархической и динамичной структуризацией и служащая как для хранения существующих знаний и их структуризации, так и для извлечения новых.

3. ПРОЕКТИРОВАНИЕ ОНТОЛОГИИ ЕЦПНЗ

Проектирование общесистемной части онтологии ЕЦПНЗ предполагает выделение совокупности универсальных и общесистемных классов, а также определение их атрибутов. Среди универсальных классов выделяются общесистемные классы, экземпляры объектов которых могут одновременно относиться к различным подпространствам. К ним относятся персоны (один ученый может работать в различных областях науки), организации (одна организация может заниматься политематическими исследованиями), географические понятия, политематические журналы, сборники, базы данных и др.

Наряду с такими классами в каждом подпространстве могут существовать классы, специфичные именно для данного подпространства. Проектирование контента тематического подпространства включает создание его предметной онтологии, определение классов объектов и их атрибутов, специфичных для данного научного направления. Предметная онтология подпространства включает в себя совокупность индексов классификационных систем, ключевых терминов с их тезаурусными связями, относящихся к данному научному направлению, совокупность метаданных, специфичных для этого подпространства.

4. УРОВНИ МЕТАДАНЫХ ОНТОЛОГИИ ЕЦПНЗ

Фактически, решая задачу проектирования онтологии, приходим к необходимости использования метаданных разного уровня:

1. метаданные как универсальные понятия ЕЦПНЗ;
2. метаданные как часть описания объектов прикладной области или подмножества ЕЦПНЗ;
3. метаданные прикладной области как таковые.

В такой онтологии на верхнем уровне используются понятия, по сути своей относящиеся к высокоуровневым онтологиям и не связанные со спецификой какой-то конкретной предметной области. На втором уровне используются понятия, описывающие предметную область, при этом являющиеся экземплярами классов, определенных на первом уровне, но используемые как определения

классов для описания данных третьего уровня уже в конкретной предметной области. Иными словами, на первом уровне даются определения основных понятий, которые используются при формировании ЕЦПНЗ, в том числе:

- тематическое подпространство;
- контент ЕЦПНЗ (совокупность информационных объектов);
- информационный объект (цифровая копия объекта реального мира или специально созданный цифровой объект, отражающий определенные свойства реального объекта);
- идентификатор информационного объекта – элемент данных, позволяющий однозначно идентифицировать объект в ЕЦПНЗ;
- атрибуты цифрового объекта (набор метаданных (свойств объекта), характеризующих объект с точки зрения задач ЕЦПНЗ);
- источник данных (объект реального мира, содержащий информацию, отраженную в атрибутах цифрового объекта);
- предметная онтология подпространства – совокупность индексов классификационных систем, ключевых терминов с их тезаурусными связями, относящихся к данному научному направлению;
- предметная онтология ЕЦПНЗ – совокупность предметных онтологий отдельных подпространств;
- тезаурусные связи – связи между двумя элементами А и В предметной онтологии, принимающие одно из 4-х значений: «А эквивалентно В», «А входит в В», «А содержит В», «А пересекается с В»;
- локальный класс объектов – объекты, относящиеся к одному тематическому подпространству
- универсальный класс объектов – объекты, связанные с несколькими тематическими подпространствами.

На втором уровне мы описываем понятия конкретной предметной области как экземпляры классов первого уровня, т. е., например, конкретный тезаурус, конкретные типы информационных ресурсов, источников данных и т. д.

Понятия второго уровня используются как определения классов на третьем уровне при заполнении онтологии данными, которые являются экземплярами классов второго уровня. При этом, если новые введенные понятия являются

на втором уровне экземплярами обозначенных ресурсов первого уровня, то при наполнении онтологии ЕЦПНЗ мы используем их в качестве классов для описания данных. Рассмотрение экземпляров в качестве классов называют *метамоделированием*. И хотя даже прямая семантика языка онтологий OWL2, используемого для описания онтологий, не позволяет такого метамоделирования, это ограничение в языке обходится с помощью синтаксического трюка, известного под названием *running*. Это означает, что когда идентификатор экземпляра встречается в аксиоме класса, он рассматривается как класс, а когда этот же идентификатор встречается в отдельном утверждении, то рассматривается как экземпляр.

Итак, выполняя построение онтологии подпространства ЕЦПНЗ или конкретной предметной области, фактически конструируют трехуровневую онтологию, в которой экземпляры первого уровня – это высокоуровневые понятия, на втором уровне описываются понятия конкретной предметной области как экземпляры в терминах первого уровня и используются как определения классов на третьем уровне при заполнении онтологии данными.

5. СЕМАНТИЧЕСКАЯ БИБЛИОТЕКА КАК ИНСТРУМЕНТ КОНСТРУИРОВАНИЯ ЕЦПНЗ

Формирование модели с перечисленными свойствами отвечает требованиям конструирования онтологии семантической научной библиотеки, близкой, по сути, к высокоуровневым онтологиям [20] для предметных областей науки. Фактически понятия делятся на три категории: первая включает определения понятий контента семантической библиотеки; вторая категория относится к определению понятий, необходимых для поддержки терминов в тезаурусе предметной области, и третья включает определения, необходимые для определения процессов интеграции контента этих ресурсов [21–24]. На основе этих определений могут быть описаны такие основные процессы, как, например, интегрирование данных из разных источников, категоризация/классификация, отображение разных моделей данных источников на заданную предметную область, построение классов эквивалентности и т. д. Такой подход хорошо согласуется с вышеописанной трехуровневой онтологией и позволяет говорить о семантических библиотеках как инструменте для построения подпространств ЕЦПНЗ.

Семантическая библиотека должна поддерживать модель данных для описания научных ресурсов и позволяет не ограничиваться при разработке строго очерченным набором ресурсов. Применение описанной модели позволяет понизить сложность (размерность) как самой модели данных, так и систем, разрабатываемых на ее основе. Получаемые модели более абстрактны, состоят из меньшего количества понятий с более простыми связями и не привязаны к определенным предметным областям. Использование этой модели данных делает возможной динамическую трансформацию и интерпретацию модели данных в приложении, позволяет настраивать решения под определенную предметную область. Фактически появляется возможность воспроизвести и поддерживать в процессе развития описание различных структур и процессов, используемых в рассматриваемой предметной области. Такой подход позволяет значительно улучшить качество обработки и поиска поступающих ресурсов и данных в рамках ограниченной предметной области не только за счет использования ее тезауруса, но и за счет гибкости описания представления имеющихся ресурсов. А также позволяет структурировать и связать различные ресурсы, извлечь из них и контекстуализировать разнообразные данные, превращая их в знания.

Приведем основные виды задач, реализуемых в семантической библиотеке, предназначенной для конструирования подпространства ЕЦПНЗ:

- описание контента информационной системы;
- реализация задач интеграции данных из внешних источников;
- поддержка коллекций;
- поиск и навигация по объектам системы;
- поддержка пользователей.

6. ПРЕДМЕТНАЯ ОБЛАСТЬ «МАТЕМАТИКА»

Рассмотрим в качестве примера реализации подпространства ЕЦПНЗ предметную область «Математика» и ее раздел «обыкновенные дифференциальные уравнения» (далее ОДУ). На основе предложенного подхода было выполнено конструирование многоуровневой онтологии. В качестве тезауруса использован тезаурус ОДУ [19]. Особенность этого тезауруса заключается в том, что он содержит не только сами понятия и термины, но и ссылки на публикации, в которых вводятся/определяются эти понятия, их математические записи. Так-

же использованы различные математические классификаторы, такие, как MSC и математическая часть УДК, статьи математической энциклопедии. Структура понятий математической энциклопедии не обладает иерархией как таковой, но благодаря использованию связанных с понятиями кодов MSC удалось выделить тематически связанные термины отдельных разделов математики. Были отдельно выделены формулы и каждому понятию при возможности был сопоставлен набор соответствующих формул.

В качестве информационных ресурсов здесь использовались такие ресурсы, как события, теоремы, персоны, публикации. Отдельно выделяются формулы, поскольку математика подразумевает их наличие. Это семантический объект с разными связями. Формулы могут быть связаны с разными объектами, обладать различными метками и т. д. В качестве источников данных были использованы два крупных источника: это DBpedia и MathNet.

В качестве контента были использованы около 4000 публикаций, формулы, персоны, статьи математической энциклопедии. Формулы были извлечены из описаний математических текстов и на основе этих данных были сформулированы и выведены дополнительные связи: между MSC и УДК, между формулами и MSC, формулами и УДК.

Рассмотрим кратко как для описания тезауруса ОДУ базовая онтология тезауруса расширяется на втором уровне для того, чтобы учесть все особенности модели этого тезауруса. Рассмотрим понятия, необходимые для описания на всех уровнях онтологии и связи между ними:

– На первом уровне используются классы, необходимые для описания общей модели, такие как *информационный ресурс*, *тезаурус*, *концепт*, *атрибут тезауруса* и т. д.

– На втором уровне описываются понятия конкретной предметной области как экземпляры в терминах первого уровня:

- *Математическая запись* – экземпляр класса *атрибут тезауруса*. Используется для хранения строки формулы;
- *Математическое примечание* – также экземпляр класса *атрибут тезауруса*; используется для хранения текста с формулами;

- *Литература* – экземпляр класса *информационный ресурс*, для описания литературы, включённой в тезаурус ОДУ;
- На третьем уровне используем понятия первого и экземпляры второго уровней как определения классов на третьем уровне.

Для поддержки формул в онтологию на втором уровне было введено понятия *Формула*, которое позволяет хранить оригинальную строку формулы из источника и связано отношениями с *информационными объектами* и *понятиями* тезауруса. Таким образом, можно построить сеть связей формулы с различными объектами, составляющими контент рассматриваемого подпространства.

С использованием такого подхода к описанию онтологии для каждой публикации на основе ее названия, аннотации и ключевых слов были выявлены связи с тезаурусом ОДУ. В качестве семантических меток были использованы термины математической энциклопедии. Такое связывание позволило выявить с некоторой долей вероятности статьи, относящиеся к предметной области ОДУ в имеющемся наборе публикаций, выявить межпредметные связи и рубрики и организовать их в коллекции на основе тезауруса и выявленных семантических меток.

ЗАКЛЮЧЕНИЕ

Рассмотрены основные принципы построения онтологии ЕЦПНЗ и набор основных понятий для построения описания произвольной предметной области. Продемонстрирован пример разработки онтологии ЕЦПНЗ для предметной области «Математика». Дальнейшие работы ориентированы на использование математического аппарата, лежащего в основе дескриптивных логик, на которых базируются онтологии, и использование средств логического вывода новых фактов. Такой подход позволяет выявлять скрытые знания и находить противоречия в имеющихся, что повышает достоверность знаний.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты №№20-07-00324, 18-00-00297, 18-00-00372.

СПИСОК ЛИТЕРАТУРЫ

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89, № 7. С. 728–735.
 2. Антопольский А.Б. и др. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.
 3. Муромский А.А., Тучкова Н.П. Представление математических понятий в онтологии научных знаний // Онтология проектирования. 2019. Т. 9, № 1 (31). С. 50–69.
 4. Губанов Н.И., Губанов Н.Н., Волков А.Э. Критерии истинности и научности знания // Философия и общество. 2016. №3 (80). С. 78–95.
URL: <https://cyberleninka.ru/article/n/kriterii-istinnosti-i-nauchnosti-znaniya>
 5. Ильин В.В., Калинин А.Т. Природа науки: Гносеологический анализ. М.: Высшая школа, 1985. 230 с.
 6. ГОСТ 7.0-99. Межгосударственный стандарт ГОСТ 7.0-99 «Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения». Постановление Госстандарта РФ от 7 октября 1999 г. N 334-ст.
 7. Гуревич И.Б., Трусова Ю.О. Тезаурус и онтология предметной области «Анализ изображений» // Всероссийская конф. с междунар. участием «Знания – Онтологии – Теории» (ЗОНТ–09). Новосибирск: Институт математики им. С.Л. Соболева СО РАН. 2009. 10 с.
 8. Hlava M.M.K. The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. 2014. V. 6, No. 3. P. 1–80.
 9. Hlava M.M.K. The taxobook: Principles and practices of building taxonomies, part 2 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. 2014. V. 6, No. 4. P. 1–164.
 10. Hlava M.M.K. The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. 2014. V. 6, No. 4. P. 1–156.
-

11. *Бездушный А.Н., Жижченко А.Б., Кулагин М.В., Серебряков В.А.* Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек // Программирование. 2000. №4. С. 3–14.

12. *Ахлестин А.Ю., Лаврентьев Н.А., Фазлиев А.З.* Систематизация научных графических ресурсов по молекулярной спектроскопии // Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18–23 сентября 2017 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2017. С. 34–42.

URL: <http://keldysh.ru/abrau/2017/39.pdf> doi:10.20948/abrau-2017-39

13. *Сотников А.Н. и др.* Принципы построения и формирования электронной библиотеки «Научное наследие России» // Программные продукты и системы. 2012. № 4. С. 30–40.

14. *Елизаров А.М. и др.* Онтологии математического знания и рекомендательная система для коллекций физико-математических документов // Докл. РАН. 2016. Т. 467, № 4. С. 392–395.

15. *Митрофанова О.А., Константинова Н.С.* Онтологии как системы хранения знаний // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». 2008. 54 с.

16. *Dextre Clarke S.G., Zeng M.L.* From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling // Information Standards Quarterly (ISQ). 2012. V. 24, No. 1. P. 20–26.

17. *Костин В.В.* Обзор семантических моделей, описывающих научные публикации и научно-исследовательскую деятельность // Электронные библиотеки: перспективные методы и технологии, электронные коллекции (RCDL–2014). 2014. С. 131–136.

18. *Gruber T.* Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). 2007. V. 3, No. 1. P. 1–11.

19. *Муромский А.А., Тучкова Н.П.* О тезаурусе для предметной области «Обыкновенные дифференциальные уравнения». Вычисл. центр им. А.А. Дородницына РАН. 2004.

20. *Mascardi V., Cordi V., Rosso P.* A Comparison of Upper Ontologies. 2007. 10 p.

21. Katsis Y., Papakonstantinou Y. View-based data integration // Encyclopedia of Database Systems. 2009. P. 3332–3339.
 22. Xu L., Embley D.W. Combining the Best of Global-as-View and Local-as-View for Data Integration // ISTA. 2004. V. 48. P. 123–136.
 23. Noy N.F. Semantic integration: a survey of ontology-based approaches // ACM Sigmod Record. 2004. V. 33, No. 4. P. 65–70.
 24. Zhao L., Ichise R. Ontology integration for linked data // Journal on Data Semantics. 2014. V. 3, No. 4. P. 237–254.
-

ONTOLOGICAL APPROACH TO THE DESCRIPTION OF A COMMON DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE

O. M. Ataeva¹ [0000-0003-0367-5575], N. E. Kalenov² [0000-0001-5269-0988],
V. A. Serebryakov³ [0000-0003-1423-621X]

¹*Dorodnicyn Computing Centre FRC CSC RAS, Moscow*

²*Joint Supercomputer Center of the RAS, Moscow;* ¹*Dorodnicyn Computing Centre FRC CSC RAS, Moscow*

³*Dorodnicyn Computing Centre FRC CSC RAS, Moscow*

¹oli@ultimeta.ru, ²nekalenov@mail.ru, ³serebr@ultimeta.ru

Abstract

Despite the development of technical means, the processes associated with the search for complete and accurate scientific information in a huge number of data sources are becoming more complicated. To reach a new level in the use of information processing technologies, first of all, a transition to a semantically meaningful representation is necessary for scientific knowledge extracted from information in a digital environment. In modern conditions, characterized by multidisciplinary research, the desired effect can be achieved by developing universal approaches to the storage and presentation of scientific knowledge. These approaches are reflected in the concept of the Common Digital Space of Scientific Knowledge. The paper presents an overview of the basic concepts in this area, which are used both to represent the elements of space and to provide access to them not only for humans, but also for

software agents. Semantic libraries are considered as tools for constructing the knowledge space.

Keywords: *knowledge space, digital knowledge space, ontologies, metadata, scientific knowledge, metadata levels, ontology design, semantic libraries.*

REFERENCES

1. *Antopolskiy A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N.* Common digital space of scientific knowledge // *Vestn. Ros. akad. nauk.* 2019. T. 89, № 7. S. 728–735.
2. *Antopolskiy A.B. i dr.* Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy // *Nauchno tekhnicheskaya informaciya. Ser. 1.* 2020. № 4. S. 9–17.
3. *Muromskij A.A., Tuchkova N.P.* Predstavlenie matematicheskikh ponyatij v ontologii nauchnyh znaniy // *Ontologiya proektirovaniya.* 2019. T. 9, № 1 (31). S. 50–69.
4. *Gubanov N.I., Gubanov N.N., Volkov A.E.* Kriterii istinnosti i nauchnosti znaniya // *Filosofiya i obshchestvo.* 2016. №3 (80). S. 78–95.
URL: <https://cyberleninka.ru/article/n/kriterii-istinnosti-i-nauchnosti-znaniya>
5. *Il'in V.V., Kalinkin A.T.* Priroda nauki: Gnoseologicheskij analiz. M.: Vysshaya shkola, 1985. 230 s.
6. GOST 7.0-99 Mezhdunarodnyy standart GOST 7.0-99 «Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu. Informacionno-bibliotechnaya deyatel'nost', bibliografiya. Terminy i opredeleniya» (Postanovlenie Gosstandarta RF ot 7 oktyabrya 1999 g. N 334-st)
7. *Gurevich I.B., Trusova Yu.O.* Tezaurus i ontologiya predmetnoj oblasti «Analiz izobrazhenij» // *Vserossiyskaya konf. s mezhdunar. uchastiem "Znaniya – Ontologii – Teorii" (ZONT-09).* Novosibirsk: Institut matematiki im. S.L. Soboleva SO RAN, 2009. 10 s.
8. *Hlava M.M.K.* The Taxobook: History, Theories, and Concepts of Knowledge Organization, Part 1 of a 3-Part Series // *Synthesis Lectures on Information Concepts, Retrieval, and Services.* 2014. V. 6, No. 3. P. 1–80.

9. *Hlava M.M.K.* The taxobook: Principles and practices of building taxonomies, part 2 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. 2014. V. 6, No. 4. P. 1–164.

10. *Hlava M.M.K.* The Taxobook: Applications, Implementation, and Integration in Search: Part 3 of a 3-Part Series // Synthesis Lectures on Information Concepts, Retrieval, and Services. 2014. V. 6, No. 4. P. 1–156.

11. *Bezdushnyj A.N., Zhizhenko A.B., Kulagin M.V., Serebryakov V.A.* Integrirovannaya sistema informacionnyh resursov RAN i tekhnologiya razrabotki cifrovyyh bibliotek // Programmirovaniye. 2000. №4. S. 3–14.

12. *Ahlyostin A.Yu., Lavrent'ev N.A., Fazliev A.Z.* Sistematizatsiya nauchnyh graficheskikh resursov po molekulyarnoy spektroskopii // Nauchnyj servis v seti Internet: trudy XIX Vserossiyskoj nauchnoj konferencii (18–23 sentyabrya 2017 g., g. Novorossiysk). M.: IPM im. M.V. Keldysha. 2017. S. 34–42.

URL: <http://keldysh.ru/abrau/2017/39.pdf> doi:10.20948/abrau-2017-39

13. *Sotnikov A.N. i dr.* Principy postroeniya i formirovaniya elektronnoj biblioteki «Nauchnoe nasledie Rossii» // Programmnye produkty i sistemy. 2012. No. 4. S. 30–40.

14. *Elizarov A.M. i dr.* Ontologii matematicheskogo znaniya i rekomendatel'naya sistema dlya kollekcij fiziko-matematicheskikh dokumentov // Dokl. RAN. 2016. T. 467, No. 4. S. 392–395.

15. *Mitrofanova O.A., Konstantinova N.S.* Ontologii kak sistemy hraneniya znaniy // Vserossiyskij konkursnyj otbor obzorno-analiticheskikh statej po prioritetnomu napravleniyu «Informacionnotelekkommunikacionnye sistemy». 2008. S. 54.

16. *Dextre Clarke S.G., Zeng M.L.* From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling // Information Standards Quarterly (ISQ). 2012. V. 24, No. 1. P. 20–26.

17. *Kostin V.V.* Obzor semanticheskikh modelej, opisyvayushchih nauchnye publikatsii i nauchno-issledovatel'skuyu deyatel'nost' // Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollekcii (RCDL–2014). 2014. S. 131–136.

18. *Gruber T.* Ontology of folksonomy: A mash-up of apples and oranges // International Journal on Semantic Web and Information Systems (IJSWIS). 2007. V. 3, No. 1. P. 1–11.
19. *Muromskij A.A., Tuchkova N.P.* O tezauruse dlya predmetnoj oblasti "Obyknovennye differencial'nye uravneniya" // Vychisl. centr im. A.A. Dorodnicyna RAN, 2004.
20. *Mascardi V., Cordi V., Rosso P.* A Comparison of Upper Ontologies. 2007. 10 p.
21. *Katsis Y., Papakonstantinou Y.* View-based data integration // Encyclopedia of Database Systems. 2009. P. 3332–3339.
22. *Xu L., Embley D.W.* Combining the Best of Global-as-View and Local-as-View for Data Integration // ISTA. 2004. V. 48. P. 123–136.
23. *Noy N.F.* Semantic integration: a survey of ontology-based approaches // ACM Sigmod Record. 2004. V. 33, No. 4. P. 65–70.
24. *Zhao L., Ichise R.* Ontology integration for linked data // Journal on Data Semantics. 2014. V. 3, No. 4. P. 237–254.

СВЕДЕНИЯ ОБ АВТОРАХ



АТАЕВА Ольга Муратовна – научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат техн. наук, специалист в области системного программирования и баз данных.

Olga Muratovna ATAeva – researcher of the of Dorodnicyn computing center FRC SCS RAS, PhD, expert in the field of system programming and databases.

email: oli@ultimeta.ru



КАЛЁНОВ Николай Евгеньевич – главный научный сотрудник Межведомственного суперкомпьютерного центр РАН – филиала ФГУ НИЦ НИИСИ РАН, доктор технических наук, профессор.

Nikolay Evgenevich KALENOV – chief Researcher in Joint Supercomputer Center of the Russian Academy of Sciences, doctor of sciences, professor.

email: nekalenov@mail.ru



СЕРЕБРЯКОВ Владимир Алексеевич – специалист в области теории формальных языков и её приложений, доктор физ.-мат. наук, профессор, зав. отделом Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН. Руководитель и участник разработки ряда известных программных проектов, в частности, ИСИР и ИСИР РАН, «Научный портал РАН».

Vladimir Alekseevich SEREBRIAKOV – expert in the field of theory of formal languages and its applications, doctor of sciences, professor, head of Dorodnicyn computing center FRC SCS RAS department. Head and participant in the development of a number of well-known program projects, in particular, ISIR and ISIR RAS, Scientific portal RAS.

email: serebr@ultimeta.ru

Материал поступил в редакцию 25 ноября 2020 года

УДК 519.218.28+ 303.725.36

ПРЕДЛОЖЕНИЯ ПО РАЗРАБОТКЕ СРЕДСТВ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ В УСЛОВИЯХ ЭПИДЕМИЙ

В. И. Балута^{1,2} [0000-0002-4399-0943], В. П. Осипов¹ [0000-0002-8237-1053], Т. В. Сивакова^{1,2} [0000-0001-8026-2198]

¹Институт прикладной математики им. М.В. Келдыша Российской академии наук, г. Москва

²Российский экономический университет им. Г.В. Плеханова

vbaluta@keldysh.ru, osipov@keldysh.ru, sivakova15@mail.ru

Аннотация

Статья посвящена рассмотрению методов моделирования эпидемий применительно к COVID-19 и обоснованию путей повышения эффективности управленческих решений с учетом прогнозируемых последствий. В работе приведён обзор методов моделирования для прогнозирования и оценки последствий эпидемиологической обстановки. Научная новизна работы заключается в использовании средств поддержки принятия решений для оперативной оценки ситуации и прогноза ее развития. Для поставленной задачи предлагается использовать мультиагентный подход имитационного моделирования.

Ключевые слова: поддержка принятия решений, эпидемическая ситуация, суперкомпьютерное моделирование, социально-экономические последствия, математические модели, мультиагентный подход.

ВВЕДЕНИЕ

Охватившая мир пандемия COVID-19 коснулась не только сферы здравоохранения. Её последствия сегодня проявляются в экономической, политической, социальной сферах жизни общества. Это связано как с особенностями самой инфекции, так и с характером и масштабом мер реагирования, принимавшихся в оперативном порядке. Длительный инкубационный период развития вируса в организме человека, наличие случаев бессимптомного протекания болезни, отсутствие на первых порах или недостаточная эффективность средств выявления фактов наличия заболевания у конкретных лиц существенно ослож-

нили создание эффективных барьеров на пути распространения инфекции. В сложившихся условиях органы управления сосредоточились на снижении темпов распространения болезни и создании резерва времени на мобилизацию и перестройку системы здравоохранения для её более эффективного функционирования в условиях сложной эпидемиологической обстановки, а вопросы должного учета сопутствующих социальных и экономических последствий были временно отодвинуты на второй план. Введенные карантинные меры с изоляцией граждан, ограничением функционирования или закрытием ряда предприятий для уменьшения количества прямых контактов между людьми позволили лишь несколько сгладить темпы роста заболеваемости, однако не решили задачу локализации очагов инфекции. Существенными побочными эффектами принятых мер стали потеря доходов у части граждан, рост недовольства среди населения и ущерб экономике, особенно заметный в сегментах малого и среднего бизнеса.

Специалисты заявляют о высоком уровне угроз появления новых видов подобных инфекционных болезней, причем с неизвестными пока характеристиками периодов инкубации, механизмов распространения вирусов и тяжести заболеваний, что актуализирует необходимость целенаправленной превентивной проработки мер повышения эффективности государственного управления в условиях сложной эпидемиологической обстановки, обеспечивающих в динамических условиях развития эпидемии необходимый уровень безопасности населения при минимизации экономического ущерба и социальных издержек.

В качестве одного из важных направлений этой работы можно выделить разработку более совершенных средств поддержки принятия решений для оперативной оценки ситуации и прогноза ее развития, включая выработку соответствующей стратегии на основе оценки уровней эффективности альтернативных вариантов реагирования. Создание таких средств связано со сменой парадигмы подходов к моделированию эпидемий в контексте развития платформенных решений поддержки принятия решений органами государственного управления. Более того, необходимо создание универсальных средств моделирования, ориентированных не только на эпидемии, но и на другие виды чрезвычайных ситуаций.

Целесообразно одновременно моделировать и прогнозировать возможные социальные и экономические последствия не только самой эпидемии (или другой чрезвычайной ситуации), но и предпринимаемых защитных мер. При этом нужно обеспечить возможность находить сбалансированные управленческие решения с оценкой прямых и отдаленных последствий в рамках концепции приемлемого риска, которая предполагает, что предпринимаемые в рамках реагирования меры не повлекут больший ущерб, чем ожидаемый при отсутствии вмешательства.

Напомним, что тема необходимости разработки и практического создания программно-аналитических инструментов поддержки принятия решений органами государственного управления, в которых большие объемы разнообразных мониторинговых данных должны использоваться в качестве исходной информации для оценки и прогностического моделирования обстановки при различных стратегиях управления, обсуждается в научно-практическом сообществе достаточно давно [1, 2]. Тем не менее, проблема до сих пор остается открытой, невзирая на наличие прямой потребности её решения в интересах оснащения системы ситуационных центров органов государственной власти [3, 4]. Между тем, как следует из вышеприведенного обзора, предпосылки для её успешного решения формируются множеством различных научных групп и коллективов в рамках отдельных локальных направлений в виде разработки и совершенствования подходов к моделированию различных социальных, экономических, общественных явлений и процессов. В последние годы интенсивно развиваются методы работы с большими массивами данных, методы обработки зашумленных данных и т. п. Поэтому речь может идти не о недостатке научных разработок, а о способах их комплексного практически значимого применения.

По нашему мнению, в качестве базового подхода при такой постановке вопроса представляет интерес мультиагентный подход имитационного моделирования, реализуемый в концепции агентно-ориентированных моделей (АОМ) [5]. Посредством мультиагентных моделей возможно имитирование поведения больших социально-экономических или других видов систем на основе реконструкции их внутренней структуры, а также структуры и поведения включенных в них более мелких акторов. По сути, динамика характеристик этих систем прояв-

ляется через результат действия совокупности самостоятельных акторов. АОМ позволяют смоделировать любую систему в приближении к реальности. Известно, что применение мультиагентных подходов в разработке имитационных моделей хорошо зарекомендовало себя в задачах, в которых учет индивидуального поведения объектов является существенным, определяющим динамику всей системы. Например, авторы с успехом применяли этот подход для решения задач моделирования динамики транспортной загрузки городской дорожно-уличной сети. На сегодняшний день есть примеры успешного применения этого подхода как для моделирования распространения инфекционных заболеваний, так и для описания экономических процессов в обществе, что позволяет предложить его в качестве базового для решения сформулированной проблемы.

МУЛЬТИАГЕНТНЫЕ ПОДХОДЫ К МОДЕЛИРОВАНИЮ ЭПИДЕМИЙ

Базовые направления развития агентных подходов в области эпидемиологии хорошо отражены в обзоре [6]. Отмечено, что при их реализации в области моделирования инфекционных эпидемий в качестве основных компонентов принято рассматривать в различной степени приближения четыре сферы, которые условно можно обозначить как «болезнь», «общество», «транспорт» и «окружающая среда». При создании агентной модели распространения инфекционных заболеваний необходимо разработать модель описания каждого из четырех компонентов.

При моделировании распространения заболевания определяется, как инфекционное заболевание передается между агентами и как болезнь прогрессирует у отдельного инфицированного агента. Моделирование общества включает в себя моделирование структуры населения и восприимчивости различных страт к инфекции. Моделирование транспорта определяет, как и по каким путям происходит перемещение агентов, в заметной мере сопутствующее распространению инфекции. Моделирование окружающей среды предполагает создание пространства, в котором происходит взаимодействие агентов, а также схемы этого взаимодействия.

Разделение на компоненты весьма условно, поскольку они переплетены между собой. Например, моделирование болезни будет определять, с какой ве-

роятностью заражается восприимчивый агент, вступающий в контакт с инфицированным агентом, однако контакт между агентами предопределяется тем, как моделируется их перемещение, а моделирование пространства формирует места возможных контактов агентов – дом, работа, транспорт, торговые точки и т. п.

В компонент «болезнь» включается рассмотрение ряда характеристических особенностей распространения и развития болезни, в частности, иммунные особенности подверженности заболеванию, формы и вероятности передачи инфекции, тяжесть и временные интервалы протекания болезни. Спектр различий в их значениях имеет достаточно большое разнообразие. Распространение инфекции может осуществляться через воздушную среду, загрязненную воду, зараженную пищу, при контакте с загрязненными поверхностями, либо через промежуточных носителей, таких как кровососущие насекомые или грызуны. Для многих возбудителей вероятны несколько путей одновременно. В этом аспекте рассматривается и опасность скрытого периода развития болезни. Многие виды инфекций имеют скрытый (латентный) период развития, когда инфицированный человек не имеет признаков заболевания, но является переносчиком возбудителя и инфицирует контактирующих с ним людей.

В компоненте «общество» описываются способы представления структуры населения в форме агентов. В большинстве имеющихся моделей, как правило, исходят из примитивных форм гомогенной структуры с рассмотрением простых форм взаимодействия относительно небольшой выборки представителей общества. В упомянутой работе предложено выделять модели, основанные на использовании хоть каких-то реальных данных в качестве «конкретных» моделей в отличие от «общих» моделей, описывающих только некоторые процессы без привязки к реальности. Общие модели общества могут быть созданы путем случайной генерации агентов в моделируемой обстановке. Например, в модели Данхема [7] рассматривается процесс распространения инфекции при взаимодействии пяти десятков одинаковых агентов по дороге на работу и обратно. А в работе [8] описана модель, основанная на сгенерированном населении жилого сектора со случайной разбивкой их на рабочих и студентов, имеющих разные схемы поведения.

Конкретные модели опираются на использование каких-либо источников фактических данных. Так, в модели [9] на основе переписи населения небольшого австралийского городка задана структура моделируемого общества путем усреднения параметров возрастного, гендерного состава и семейного положения граждан. Однако отмечается, что построение конкретных моделей общества чрезвычайно затруднено из-за недостатка данных, поэтому в большинстве своем модели с детализацией видов агентов, как правило, базируются на субъективных представлениях авторов и зачастую имеют характер виртуальных исследовательских экспериментов.

Транспортный компонент при моделировании определяет способ перемещения агентов между различными местоположениями. В зависимости от подробности модели в ней могут использовать различные подходы. Так, Данхем [7] применяет простую транспортную модель, в которой агенты просто перемещаются между точками по прямой линии с постоянной скоростью. Более реалистичные транспортные модели опираются на данные, содержащие информацию о транспортной инфраструктуре территории. В ряде случаев разработчики опираются на представления, уже развитые в научных областях моделирования транспортных потоков в интересах определения динамики загруженности улично-дорожной сети. В этом случае перемещения людей по схемам дорожной сети задаются на основе правил выбора маршрута агентами, например, по кратчайшему пути, как в [10]. В более сложных моделях включается рассмотрение общественного транспорта как мест локальной скученности людей, способствующей передаче инфекции [11].

Принципы определения мотивов выбора маршрутов хорошо развиты в сфере моделирования динамики транспортных потоков на основе суточного цикла деятельности каждого индивида для выделения потребности в перемещениях. Такой подход, с одной стороны, наиболее точно отражает вторичность транспортного спроса по отношению к спросу на деятельность вне дома [12], с другой, позволяет моделировать меры, стимулирующие изменение вида транспорта и/или времени поездки, а также совместное использование транспортных средств.

В основе этих подходов лежит предположение о том, что деятельность каждого человека является результатом последовательно принятых решений, в процессе которых осуществляется дискретный выбор одной из нескольких альтернатив. Выбор каждого из вариантов определяется его привлекательностью для агента, принимающего решение. Для моделирования выбора вводится количественная оценка привлекательности рассматриваемых альтернатив, в качестве которой используется функция привлекательности. Последняя зависит от характеристик альтернатив и свойств принимающего решение. Задача усложняется тем, что помимо количественных характеристик (например, время, стоимость, выгода и т. д.) каждая альтернатива обладает и трудно формализуемыми качественными (комфортность, привычность, соответствие стилю жизни и т. д.).

В работе [13] для адекватного учета качественных характеристик используется модель скрытых переменных, в которой вводится предположение, что помимо наблюдаемых переменных существуют некие скрытые переменные, а индикаторы, характеризующие качественные характеристики альтернативы и важность той или иной характеристики для индивида, а также привлекательность альтернативы зависят как от наблюдаемых, так и от скрытых переменных. При этом в функцию привлекательности вводят случайную составляющую, отражающую:

- случайный характер принятия многих решений (влияние настроения, самочувствия и т. п.);
- неполную информированность о характеристиках альтернатив, которые также случайны, т. к. являются результатом взаимодействия многих факторов, в том числе, выбора остальных участников движения;
- неучтенную информацию о характеристиках альтернатив;
- неучтенную информацию о характеристиках принимающих решение лиц и последствия агрегирования такой информации по группам населения.

Случайная составляющая функции привлекательности необходима для адекватного моделирования принятия решений индивидами. На практике для моделирования реального распределения случайных величин или аппроксимации случайных функций используют априорно задаваемые аналитические распределения.

Описание «окружающей среды» является важным элементом модели, поскольку определяет пространство местоположений, в котором находятся и взаимодействуют агенты. Уровень сложности этого описания определяется задачами исследования и может быть как просто выделенной областью, которую можно сравнить с площадью, на которой находится толпа перемещающихся по каким-то правилам агентов (сеточная схема) [14], так и иметь сложную конфигурацию на основе картографических схем населенных пунктов или транспортных схем [15]. Так, в работе [16] схема расположения школ в городской среде использовалась для оценки сроков закрытия школ на карантин по гриппу. В некоторых работах рассматриваются и другие параметры окружающей среды, в частности, в [17] при моделировании распространения холеры учитывались температурный режим окружающей среды и гидрология территории.

Важным аспектом оценки качества модели при агентном моделировании является ее валидация. Наиболее приемлемой считается проверка на данных о характере распространения вспышек эпидемий в прошлых периодах [18]. Однако такие примеры можно найти только для случаев моделирования тех видов инфекций, по которым имеется достаточная статистика. В других ситуациях необходимо разрабатывать другие способы валидации моделей.

В текущем периоде уже появились работы со ссылкой на применение агентных моделей для моделирования COVID-19. Так, в статье [19] рассмотрен вариант общей (в терминах выше приведенной классификации Hunter et al.) модели в форме графа некоторого сетевого сообщества взаимно связанных агентов (граф Watts-Strogatz с 1000 вершинами), которые попарно взаимодействуют между собой случайным образом. В результате взаимодействия происходит инфицирование восприимчивых от заболевших. В модели учитываются не только процесс передачи инфекции, но и обратный процесс выздоровления, когда через некоторое условное время заболевшие считаются выздоровевшими, приобретая иммунитет, то есть они перестают быть переносчиками инфекции и не заражаются при контакте с инфицированными. Автор проводит аналогию с SIR-моделированием. Время в представленной модели является дискретным и измеряется «шагами» смены состояний системы, когда на каждом шаге происходит какое-то количество заражений или выздоровлений. Посредством модели-

рования проводится исследование вероятной динамики распространения инфекции в подобном сообществе и зависимость этой динамики от различных видов распределения входных параметров. Автор рассматривает также результаты исследования зависимости и масштабов эпидемии при вариациях форм сетевого сообщества, скоростей восстановления инфицированных, динамики и характера взаимодействий между агентами (часто–редко, один к одному, один ко многим, и т. п.). В процессе проведенных исследований получены, в общем-то, ожидаемо тривиальные выводы, что при уменьшении среднего числа контактов между агентами снижается скорость распространения эпидемии, что карантинные мероприятия могут дать положительный эффект, а проведение массовых мероприятий, напротив, негативный. Собственно, к моделированию непосредственно COVID-19 эта модель имеет весьма слабое отношение, поскольку подобным образом можно исследовать, например, распространение горячих новостей в интернет-сообществе или какие-то другие сходные процессы.

В отечественной литературе также появилась публикация с результатами мультиагентного моделирования процессов распространения пандемии COVID-19 [20]. Большой коллектив исполнителей под руководством Г. Рыкованова провел исследование эффективности различных подходов к моделированию пандемии, включая как традиционное решение системы дифференциальных уравнений по модели SEIRD, так и мультиагентное моделирование поведения людей в процессе их жизнедеятельности. Параметры моделей задавались экспертным путем, а результаты сопоставлялись с имеющимися данными по развитию эпидемий в таких городах, как Москва, Нью-Йорк и Ухань.

Помимо подробных результатов исследований, демонстрируемых графиками, и их сопоставления с данными статистики, приводимыми в открытых источниках, в качестве еще одного интересного результата применения мультиагентного подхода в этой работе можно считать отсылки к источникам получения данных для построения самой модели и калибровки параметров поведения агентов.

Из зарубежных работ особый интерес представляет работа [21], в которой рассмотрены процессы распространения заболевания и их влияние на экономические показатели при различных сценариях реагирования. Нужно заметить, что

в этой статье на стадии формулировки задачи и выбора методов её решения приведен достаточно детальный анализ работ других исследователей, имеющих на данном направлении и опубликованных в научных изданиях, что само по себе может представлять интерес в целях ознакомления с проблемой мультиагентного моделирования эпидемий.

В своей работе авторы опираются на традиционную модель распространения эпидемических процессов – модель SEIR. Однако используют парадигму агентного моделирования не только для оценки чисто эпидемиологических, но и экономических последствий эпидемии COVID-19. В качестве агентов в модели представлены следующие объекты: группа объектов A1 – люди, A2 – домовладения (семьи), A3 – производственные или сервисные предприятия, A4 – органы государственного управления, A5 – система здравоохранения с ее медицинскими учреждениями. Функции соответствующих агентов, позволяющие оценить степень приближений, принятых в модели, представлены ниже.

A1: Человек. Группа A1 – это основной тип агента. Его динамическое положение изменяется в зависимости от окружающей среды и может быть связано или нет (бездомный) с объектами группы A2 либо связано или нет (безработный) с объектами группы A3. В качестве атрибутов используются местоположение (динамическое), возраст, дом (A2), работодатель (A3), эпидемиологический статус, инфекционный статус, богатство, доход и положение в социальной структуре. Возможные действия: свободное перемещение (ежедневно), следование домой (ежедневно), следование на работу (ежедневно), личные контакты (ежечасно), деловые контакты (ежечасно), следование в больницу (при наличии инфицирования).

A2: Дом. Группу A2 представляют семьи, у которых одно место проживания и общие финансовые счета. В качестве атрибутов используются местоположение (статическое), социальный слой, соседи по дому (из группы A1), богатство, доходы и расходы. Агенты A2 осуществляют домашнюю регистрацию (ежедневно), ведут бухгалтерию (ежемесячно).

A3: Бизнес. Группа A3 – это экономические агенты, например, промышленность, магазины или рынки. Они взаимодействует с A1, выплачивая зарплату или продавая продукт. В качестве атрибутов используются местоположение (ста-

тическое), социальный слой, работники (группа А1), богатство, доходы и расходы. Действия включают бухгалтерский учет (ежемесячно), деловые контакты (ежечасно).

А4: Правительство. А4 – это единый агент, который получает налоги от А2 и А3, предоставляет средства А5 и страхует бездомных и безработных А1. В качестве атрибутов используются положение (статическое) и богатство. Осуществляемые действия – бухгалтерский учет (ежемесячно).

А5: Система здравоохранения. А5 – это также единственный агент (сингтон), представляющий собой систему здравоохранения в целом, которая в идеале должна быть способна обслуживать всё население (в котором объединяются возможности осуществлять контроль, выявлять больных, принимать их в больницах и лечить). Из атрибутов – только положение и богатство.

В работе рассмотрены семь сценариев возможного государственного вмешательства, которые были условно поименованы следующим образом: (1) – невмешательство (ничего не делать), (2) – блокировка, (3) – условная блокировка, (4) – вертикальная изоляция, (5) – частичная изоляция, (6) – использование масок для лица и (7) – использование масок для лица вместе с 50% социальной изоляцией. Все сценарии рассмотрены применительно к особенностям систем управления и жизнедеятельности населения в Бразилии. В модели представлена достаточно сложная структура взаимоотношений людей, организованных в семье, бизнес и правительство, которые взаимодействуют друг с другом. Для людей рассмотрены три действия: «иди на работу», «иди домой», «гуляй свободно». Для инфицированных при переходе болезни в тяжелую стадию добавляется действие «иди в больницу». Умершие в своих действиях обнуляются. При этом в качестве базового принят сценарий, который имитирует экономическое поведение общества в нормальных условиях – без пандемии. Моделирование проводится в виде почасовых итераций, в ходе которых рассматриваются взаимодействия агентов с определенными вероятностями. В ходе таких взаимодействий производится учет двух процессов. Один процесс отражает процесс распространения инфекции и протекания заболевания. Этот процесс включает:

- передачу заболевания с какой-то вероятностью от инфицированных агентов группы А1 к восприимчивым, к которым отнесены все не переболевшие агенты этой же группы, при их контакте;
- различный характер развития заболевания у отдельных агентов: либо в бессимптомной или слабой форме, либо в тяжелой форме, требующей госпитализации;
- выздоровление заболевших или их смерть с определенной вероятностью, зависящей от возраста.

Второй процесс отражает экономические процессы, влияющие на изменение богатства (доходы и расходы), условно отражающиеся в бухгалтерском учете.

Несмотря на то, что взаимодействие различных категорий агентов рассматривается весьма условно, особенно в экономических ее аспектах, подобное моделирование позволяет получить качественную картину эффективности тех или иных мер реагирования для сопоставления их между собой, что позволяет выработать соответствующие рекомендации для органов управления. Уже на таком основании видна целесообразность применения агентных моделей для выработки и оценки эффективности различных сценариев реагирования в случае эпидемий. Таким образом, на основании результатов подобной работы можно сделать вывод о возможности и осуществимости решения поставленной фундаментальной проблемы средствами мультиагентного моделирования при комплексной постановке задачи.

НОВАЯ ПАРАДИГМА МОДЕЛИРОВАНИЯ

Как показывает приведенная работа, в ситуации анализа эпидемических процессов большим достоинством мультиагентного моделирования является возможность проведения прогностических исследований для оценки влияния на состояние социально-экономической системы и её элементов различных вариантов изменения внешней среды (в рассматриваемом случае – появления угрозы инфицирования) при одновременном рассмотрении изменения условий функционирования агентов (например, ввода каких-то ограничительных мер).

Если подходить к формулированию задачи более детально, то для моделирования распространения инфекций, передающихся воздушно-капельным путем, к которой относится и COVID-19, необходимо прежде всего построить модель, имитирующую все возможные коммуникации людей. Люди встречаются дома, на работе, на улице, в транспорте, других общественных местах, при этом контакт может быть тесным или мимолетным, что сказывается на возможности передачи инфекции. Необходимо учитывать, что сама вероятность таких контактов предопределяется множеством параметров. Люди делятся по гендерным признакам, возрастным категориям, социальному положению, уровням дохода, профессиональной принадлежности, месту и условиям проживания, состоянию здоровья и т. д. Каждый человек одновременно может выступать в разных ролях в различных социумах (в семье, трудовом коллективе, общественном транспорте, торговой точке, дружеском застолье и т. п.). В зависимости от любого из названных факторов человек с какой-то вероятностью может оказаться в конкретное время в том или ином месте.

В качестве базовой модели необходимо выбрать нормальный режим жизнедеятельности социума, для чего имеет смысл воспользоваться известным фактом «типового распорядка». В статистически значимом большинстве каждый человек выстраивает свой день по определенному распорядку, будь то работник предприятия, который утром должен прибыть на работу к определенному времени, используя какой-то маршрут в общественном транспорте, а затем вечером вернуться домой, возможно, по другому маршруту, если есть потребность куда-либо зайти (спортзал, магазин), или это мама-домохозяйка, которая должна собрать и проводить детей в школу или садик и забрать их после смены. Отклонения от типовой схемы также имеют более-менее статистически значимый характер, имеющий сезонные колебания, – проценты заболевших, командированных, отпускников и т. п. Значения характерных величин, определяющих эти соотношения, изначально могут быть заданы на основании экспертных оценок. Однако в процессе отладки модели исходные оценки могут быть откорректированы. Коррекция значений статистически значимых параметров общей модели функционирования социума целесообразно проводить путем установления согласованности со множеством доступных фактических данных: по базам данных пас-

портных столов, ЗАГСов, учета численности школьников и т. п. – в части состава населения на изучаемой территории; по данным систем видеонаблюдения, валидаторов в общественном транспорте, регистрации транзакций на кассах торговых точек, динамике привязки мобильных телефонов к сотовым вышкам – в части активности их перемещений; по регистрации приема в поликлиниках, билетов в транспортных узлах (аэропорт, железнодорожный или автовокзал) – в части процентных отклонений от типового распорядка, и так далее. В современном мире построения цифрового общества такие возможности привлечения дополнительных данных постоянно расширяются. И чем шире будет круг привлекаемых фактических данных прямых и косвенных систем измерений, используемых для калибровки имитационной модели, тем выше будут обеспечиваться уровень её достоверности и, соответственно, прогностический и исследовательский потенциал такой модели. При дополнении этой модели другими акторами жизнедеятельности общества – субъектами экономических, социальных, общественных отношений, тесно связанными с элементарными акторами-людьми, – возможно построение модели общества, коррелированной с реальной картиной жизни. Такая модель при коррекционном уточнении её текущих параметров за счет перманентно накапливаемых разнородных данных может служить целям анализа ситуации в реальном масштабе времени, что позволяет использовать её для исследования возможного изменения обстановки при трансформации внешних условий (инфекции, погодные аномалии, экономические шоки), а также применять для прогностического изучения последствий реализации различных управленческих решений.

Разработать принципиальную архитектуру и ключевые субмодели такой комплексной модели общества, создать и апробировать алгоритмы моделирования и калибровки параметров модели вполне по силам опытному научному коллективу, однако реализовать её полномасштабную версию для практического использования без подключения административных механизмов невозможно [13]. А именно, помимо разработки математического и программно-инструментального аппарата моделирования необходимо обеспечить создание отдельной системы полноценного наполнения созданной комплексной модели и её составных частей реальными данными в постоянном режиме. Для повыше-

ния уровня адекватности комплексной модели целесообразно применение методов когнитивного моделирования и технологий искусственного интеллекта.

Возвращаясь к вопросу о новой парадигме моделирования эпидемий, имеет смысл подчеркнуть уже имеющиеся условия и постоянно наращиваемые возможности для её успешной реализации. В числе этих условий и возможностей:

- обоснованные теоретически и апробированные при решении различных задач методы мультиагентного моделирования сложных систем;
- развитые технологии высокопроизводительных вычислений, технологии облачного хранения данных, технологии дистанционных коммуникаций и удаленного доступа;
- развернутая сетевая инфраструктура взаимодействия органов государственного управления и сформулированная потребность в подобных платформенных решениях;
- интенсивное внедрение цифровых технологий во все сферы жизни общества, естественным образом способствующее производству огромного количества фактологических данных, которые могут быть использованы для калибровки моделей в условиях повседневной жизнедеятельности.

ЗАКЛЮЧЕНИЕ

В качестве одного из возможных вариантов практической реализации предлагаемого подхода можно предложить следующую схему. Собственно комплексная модель социума и расчетный аппарат для его имитационного моделирования на базе высокопроизводительных вычислительных систем может разрабатываться группой специалистов прикладных академических институтов в рамках отдельной комплексной программы или государственного задания. С учетом имеющегося опыта, как нам представляется, в качестве базовых организаций для формирования такого коллектива могут выступить ЦЭМИ РАН и ИПМ им. М.В. Келдыша РАН. Поскольку разрабатываемую модель необходимо адаптировать для применения в качестве одного из платформенных решений Системы распределенных ситуационных центров органов государственного управления, в команду разработчиков целесообразно включить также специалистов ФИЦ ИУ РАН и ИПУ РАН. Внедряемая модель в качестве одного из элементов си-

стемы поддержки управления должна быть доведена до уровня отдельных муниципальных образований. Именно на этом уровне должно производиться наполнение модели фактическими данными, первоначально путем экспертных оценок, с последующей их корректировкой и постоянной калибровкой с помощью механизмов искусственного интеллекта и обработки больших объемов разнородных данных, которые накапливаются в базах данных информационных систем, а также на основе обработки информации из социальных сетей.

СПИСОК ЛИТЕРАТУРЫ

1. *Зацаринный А.А.* О повышении эффективности информационно-аналитической поддержки принятия стратегических решений в органах государственной власти // Межотраслевая информационная служба. 2015. № 1. С. 11–22.
2. *Балута В.И., Осипов В.П., Яковенко О.Ю.* Среда моделирования, прогнозирования и экспертиз как интеллектуальное ядро поддержки управления сложными системами // М.: Препринты ИПМ им. М.В. Келдыша. 2015. №82. 16 с. URL: https://keldysh.ru/papers/2015/prep2015_82.pdf
3. *Зацаринный А.А., Ильин Н.И., Колин К.К., Лепский В.Е., Малинецкий Г.Г., Новиков Д.А., Райков А.Н., Сильвестров С.Н., Славин Б.Б.* Ситуационные центры развития в полисубъектной среде // Проблемы управления. 2017. №5. С. 31–42.
4. *Ильин Н.И.* Интервью Национальному центру цифровой экономики МГУ им. М.В. Ломоносова, 10.12.2018 г. URL: <https://digital.msu.ru>
5. *Макаров В.Л., Бахтизин А.Р.* Современные методы прогнозирования последствий управленческих решений // Управленческое консультирование. 2015. №7. С. 12–24.
6. *Hunter E., Mac Namee B., Kelleher D.* A taxonomy for agent-based models in human infectious disease epidemiology // Journal of Artificial Societies and Social Simulation. 2017. V. 20, No. 3. P. 2. URL: <http://jasss.soc.surrey.ac.uk/20/3/2.html>
7. *Dunham J.B.* An agent-based spatially explicit epidemiological model in MASON // Journal of Artificial Societies and Social Simulation. 2005. V. 9, No. 1. P. 3.

8. *Perez L., Dragicevic S.* An agent-based approach for modeling dynamics of contagious disease spread // *International Journal of Health Geographics*. 2009. V. 8. No. 50. P. 1–17. URL: <https://doi.org/10.1186/1476-072X-8-50>

9. *Skvortsov A.T., Connell R.B., Dawson P.D. and Gailis R.M.* Epidemic modelling: Validation of agentbased simulation by using simple mathematical models // *MODSIM 2007 International Congress Modelling on and Simulation*. Modelling and Simulation Society of Australia and New Zealand. 2007. P. 657–662.

URL:

https://www.mssanz.org.au/MODSIM07/papers/13_s20/EpidemicModeling_s20_Skvortsov_.pdf

10. *Crooks A.T., Hailegiorgis A.B.* An agent-based modeling approach applied to the spread of cholera // *Environmental Modelling & Software*. 2014. V. 62. P. 164–77. URL: <https://doi.org/10.1016/j.envsoft.2014.08.027>

11. *Rakowski F., Gruziel M., Bieniasz-Krzywiec L., Radomski J.P.* Influenza epidemic spread simulation for Poland – a large scale, individual based model study // *Physica A: Statistical Mechanics and its Applications*. 2010. V. 389 (16). P. 3149–3165. URL: <https://doi.org/10.1016/j.physa.2010.04.029>

12. *Armstrong J.S., Green K.C.* Demand Forecasting: Evidence-based Methods // *Strategic Marketing Management: A Business Process Approach*. 2005. V. 24. URL: https://www.researchgate.net/publication/5179920_Demand_Forecasting_Evidence-Based_Methods

13. *McFadden D.L.; Train K.* Mixed MNL Models for Discrete Response // *Journal of Applied Econometrics*. 2000. V. 15. No. 5. P. 447–470.

URL: [https://doi.org/10.1002/1099-1255\(200009/10\)15:5<447::AID-JAE570>3.0.CO;2-1](https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1)

14. *Duan W., Qiu X., Cao Z., Zheng X., Cui K.* Heterogeneous and stochastic agent-based models for analyzing infectious diseases' super spreaders // *IEEE Intelligent Systems*. 2013. V. 13. P. 1541–1672.

15. *Mao L.* Modeling triple-diffusions of infectious diseases, information, and preventive behaviors through a metropolitan social network – an agent-based simulation // *Applied Geography*. 2014. V. 50. P. 31–39.

URL: <https://doi.org/10.1016/j.apgeog.2014.02.005>

16. Lee B.Y., Brown S.T., Cooley P., Potter M.A., Wheaton W.D., Voorhees R.E., Stebbins S., Grefenstette J.J., Zimmer S.M., Zimmerman R.K., Assi T.-M., Bailey R.R., Wagener D.K., Burke D.S. Simulating school closure strategies to mitigate an influenza epidemic // *Journal of Public Health Management and Practice*. 2008. V. 16. No. 3. P. 252–261. URL: <https://doi.org/10.1097/PHH.0b013e3181ce594e>

17. Crooks A.T., Hailegiorgis A.B. An agent-based modeling approach applied to the spread of cholera // *Environmental Modelling & Software*. 2014. V. 62. P. 164–177. URL: <https://doi.org/10.1016/j.envsoft.2014.08.027>

18. Merler S., Ajelli M., Fumanelli L., Gomes M.F.C., Piontti A.P., Rossi L., Chao D.L., Jr I. M.L., Halloran M.E., Vespignani A. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *The Lancet Infectious Diseases*. 2015. V. 15. No. 2. P. 204–211. URL: [https://doi.org/10.1016/S1473-3099\(14\)71074-6](https://doi.org/10.1016/S1473-3099(14)71074-6)

19. Wolfram C. An Agent-Based Model of COVID-19 // *Complex Systems*. 2020. V. 29. No. 1. P. 87–105. URL: <https://doi.org/10.25088/ComplexSystems.29.1.87>

20. Адарченко В.А. и др. Моделирование развития эпидемии коронавируса по дифференциальной и статистической моделям // Снежинск. Изд-во РФЯЦ-ВНИИТФ. 2020. Препринт №264. 29 с. URL: <http://vniitf.ru/data/files/pdf/corona.pdf>

21. Silva P.C.L., Batista P.V.C., Lima H.S., Alves M.A., Guimarães F.G., Silva R.C.P. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions // *Chaos, Solitons & Fractals*. 2020. P. 37. E-print: arXiv:2006.10532 [cs.AI] URL: <https://doi.org/10.1016/j.chaos.2020.110088>

SUGGESTIONS FOR DEVELOPING TOOLS TO IMPROVE THE EFFECTIVENESS OF MANAGEMENT IN THE CONTEXT OF EPIDEMICS

V. I. Baluta^{1,2}, V. P. Osipov¹, T. V. Sivakova^{1,2}

¹*Keldysh Institute of Applied Mathematics, Moscow*

²*Plekhanov Russian University of Economics, Moscow*

vbaluta@keldysh.ru, osipov@keldysh.ru, sivakova15@mail.ru

Abstract

The article is devoted to the consideration of methods for modeling epidemics in relation to COVID-19 and substantiation of ways to improve the efficiency of management decisions, taking into account the predicted consequences. The paper provides an overview of modeling methods for predicting and assessing the consequences of the epidemiological situation. The scientific novelty of the work lies in the use of decision support tools for the operational assessment of the situation and forecast of its development. For the task at hand, it is proposed to use a multi-agent approach to simulation.

Keywords: *decision support, epidemic situation, supercomputer modeling, socio-economic consequences, mathematical models, multi-agent approach*

REFERENCES

1. Zacarinyj A.A. O povyshenii effektivnosti informacionno-analiticheskoy podderzhki prinyatiya strategicheskikh reshenij v organah gosudar-stvennoj vlasti // Mezhotraslevaya informacionnaya sluzhba. 2015. № 1. S. 11–22.
2. Baluta V.I., Osipov V.P., Yakovenko O.YU. Sreda modelirovaniya, prognozirovaniya i ekspertiz kak intellektual'noe yadro podderzhki upravleniya slozhnymi sistemami // M.: Preprinty IPM im. M.V. Keldysha. 2015. №82. 16 s. URL: https://keldysh.ru/papers/2015/prep2015_82.pdf
3. Zacarinyj A.A., Il'in N.I., Kolin K.K., Lepskij V.E., Malinec-kij G.G., Novikov D.A., Rajkov A.N., Sil'vestrov S.N., Slavin B.B. Situacionnye centry razvitiya v polisub'ektnoj srede // Problemy upravleniya. 2017. №5. S. 31–42.
4. Il'in N.I. Interv'yu Nacional'nomu centru cifrovoj ekonomiki MGU im. M.V. Lomonosova, 10.12.2018 g. URL: <https://digital.msu.ru>

5. *Makarov V.L., Bahtizin A.R.* Sovremennye metody prognozirovaniya posledstvij upravlencheskih reshenij // Upravlencheskoe konsul'tirovanie. 2015. №7. S. 12–24.

6. *Hunter E., Mac Namee B., Kelleher D.* A taxonomy for agent-based models in human infectious disease epidemiology // Journal of Artificial Societies and Social Simulation. 2017. 20(3) 2. URL: <http://jasss.soc.surrey.ac.uk/20/3/2.html>

7. *Dunham J.B.* An agent-based spatially explicit epidemiological model in MASON // Journal of Artificial Societies and Social Simulation. 2005. V. 9(1). P. 3.

8. *Perez L., Dragicevic S.* An agent-based approach for modeling dynamics of contagious disease spread // International Journal of Health Geographics. 2009. V. 8. No. 50. P. 1–17. URL: <https://doi.org/10.1186/1476-072X-8-50>

9. *Skvortsov A.T., Connell R.B., Dawson P.D. and Gailis R.M.* Epidemic modelling: Validation of agentbased simulation by using simplemathematical models // MODSIM 2007 International Congress Modelling on and Simulation. Modelling and Simulation Society of Australia and New Zealand. 2007. P. 657–662. URL: https://www.mssanz.org.au/MODSIM07/papers/13_s20/EpidemicModeling_s20_Skvortsov_.pdf

10. *Crooks A.T., Hailegiorgis A.B.* An agent-based modeling approach applied to the spread of cholera // Environmental Modelling&Soware. 2014. V. 62. P. 164–77. URL: <https://doi.org/10.1016/j.envsoft.2014.08.027>

11. *Rakowski F., Gruziel M., Bieniasz-Krzywiec L., Radomski J.P.* Influenza epidemic spread simulation for Poland – a large scale, individual based model study // Physica A: Statistical Mechanics and its Applications. 2010. V. 389 (16). P. 3149–3165. URL: <https://doi.org/10.1016/j.physa.2010.04.029>

12. *Armstrong J.S., Green K.C.* Demand Forecasting: Evidence-based Methods // Strategic Marketing Management: A Business Process Approach. 2005. V. 24. URL: https://www.researchgate.net/publication/5179920_Demand_Forecasting_Evidence-Based_Methods

13. *McFadden D.L.; Train K.* Mixed MNL Models for Discrete Response // Journal of Applied Econometrics. 2000. V. 15. No. 5. P. 447–470. URL: [https://doi.org/10.1002/1099-1255\(200009/10\)15:5<447::AID-JAE570>3.0.CO;2-1](https://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1)

14. *Duan W., Qiu X., Cao Z., Zheng X., Cui K.* Heterogeneous and stochastic agent-based models for analyzing infectious diseases' super spreaders // *IEEE Intelligent Systems*. 2013. V. 13. P. 1541–1672.

15. *Mao L.* Modeling triple-diffusions of infectious diseases, information, and preventive behaviors through a metropolitan social network – an agent-based simulation // *Applied Geography*. 2014. V. 50. P. 31–39.

URL: <https://doi.org/10.1016/j.apgeog.2014.02.005>

16. *Lee B.Y., Brown S.T., Cooley P., Potter M.A., Wheaton W.D., Voorhees R.E., Stebbins S., Grefenstette J.J., Zimmer S.M., Zimmerman R.K., Assi T.-M., Bailey R.R., Wagener D.K., Burke D.S.* Simulating school closure strategies to mitigate an influenza epidemic // *Journal of Public Health Management and Practice*. 2008. V. 16. No. 3. P. 252–261. URL: <https://doi.org/10.1097/PHH.0b013e3181ce594e>

17. *Crooks A.T., Hailegiorgis A.B.* An agent-based modeling approach applied to the spread of cholera // *Environmental Modelling & Software*. 2014. V. 62. P. 164–177. URL: <https://doi.org/10.1016/j.envsoft.2014.08.027>

18. *Merler S., Ajelli M., Fumanelli L., Gomes M.F.C., Y Piontti A.P., Rossi L., Chao D.L., Jr I. M.L., Halloran M.E., Vespignani A.* Spatiotemporal spread of the 2014 outbreak of ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *The Lancet Infectious Diseases*. 2015. V. 15. No. 2. P. 204–211.

URL: [https://doi.org/10.1016/S1473-3099\(14\)71074-6](https://doi.org/10.1016/S1473-3099(14)71074-6)

19. *Wolfram C.* An Agent-Based Model of COVID-19 // *Complex Systems*. 2020. V. 29. No. 1. P. 87–105.

URL: <https://doi.org/10.25088/ComplexSystems.29.1.87>

20. *Adarchenko V.A. i dr.* Modelirovanie razvitiya epidemii koronavirusa po differencial'noj i statisticheskoy modelyam // *Snezhinsk. Izd-vo RYAC-VNIITF*. 2020. Preprint №264. 29 s. URL: <http://vniitf.ru/data/files/pdf/corona.pdf>

21. *Silva P.C.L., Batista P.V.C., Lima H.S., Alves M.A., Guimarães F.G., Silva R.C.P.* COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions // *Chaos, Solitons & Fractals*. 2020. P. 37. E-print: arXiv:2006.10532 [cs.AI]

URL: <https://doi.org/10.1016/j.chaos.2020.110088>

СВЕДЕНИЯ ОБ АВТОРАХ



БАЛУТА Виктор Иванович – старший научный сотрудник ИПМ им. М.В. Келдыша РАН, к. т. н., с.н.с.;

Victor Ivanovich BALUTA – Keldysh Institute of Applied Mathematics, senior research, Candidate of Engineering Sciences, Associate Professor.

vbaluta@keldysh.ru; ORCID: 0000-0002-4399-0943



ОСИПОВ Владимир Петрович – ведущий научный сотрудник ИПМ им. М.В. Келдыша РАН, к. т. н., доцент;

Vladimir Petrovich OSIPOV – Keldysh Institute of Applied Mathematics, lead researcher, Candidate of Engineering Sciences, Associate Professor.

osipov@keldysh.ru; ORCID: 0000-0002-8237-1053



СИВАКОВА Татьяна Владимировна – научный сотрудник ИПМ им. М.В. Келдыша РАН;

Tatiana Vladimirovna SIVAKOVA – Keldysh Institute of Applied Mathematics, research.

email: sivakova15@mail.ru; ORCID: 0000-0001-8026-2198

Материал поступил в редакцию 27 ноября 2020 года

УДК 004.4'242, 004.4'422, 004.432.4

МЕТОД БАЛАНСИРОВКИ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ ДЛЯ ГИБРИДНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Т. П. Баранова¹ [0000-0002-2003-3737], А. Б. Бугеря² [0000-0002-9698-458X], К. Н. Ефимкин³ [0000-0002-1087-7645]

¹⁻³Институт прикладной математики им. М.В. Келдыша
Российской академии наук, г. Москва

¹bart1950@yandex.ru, ²shurabug@yandex.ru, ³bigcrocodile@yandex.ru

Аннотация

Рассмотрены вопросы распределения вычислений внутри одного узла гибридной вычислительной системы для прикладных программ расчётного характера. Предложены метод статического распределения вычислений, а также метод автоматической балансировки вычислительной нагрузки в процессе выполнения программы. Метод автоматической балансировки основан на периодическом анализе величины загрузки центрального процессора выполняемой программой и принятии решения о перераспределении вычислительной нагрузки в случае необходимости. Приведённые методы реализованы в прикладной программе, решающей задачу из области газодинамики с использованием вычислительных ресурсов многоядерного центрального процессора и графических ускорителей. Получены и проанализированы результаты выполнения программы с различными распределениями данных, как с включённым механизмом автоматической балансировки вычислительной нагрузки, так и без него.

Ключевые слова: параллельное программирование, автоматизация программирования, балансировка вычислительной нагрузки, гибридные архитектуры, язык NORMA, автоматическая генерация программ

ВВЕДЕНИЕ

В современном мире существует огромное количество различных задач, решение которых требует наличия значительных вычислительных мощностей. Такие задачи имеются во всех областях науки и промышленности, в бизнесе и даже в сфере индивидуального применения. Типичными примерами таких ре-

сурсоёмких задач могут служить решение задач математической физики численными методами (например, моделирование процессов, происходящих в ядерном реакторе), моделирование физических, химических и биологических процессов. Постоянно появляются новые задачи подобного рода. Современные вычислительные системы, на которых решаются подобные задачи, предоставляют возможность параллельных вычислений. Поэтому современная эффективная программа обязана быть параллельной.

Существуют различные методы автоматизации разработки параллельных программ. Отметим лишь монографии [1, 2], в которых строго сформулированы математические основы совместного изучения параллельных численных методов и параллельных вычислительных систем и исследована задача отображения программы на архитектуру параллельного компьютера. В частности, показано, что автоматическое отображение уже написанной последовательной программы на параллельную вычислительную систему, в общей постановке, является NP-полной задачей. Это и объясняет отсутствие, до настоящего времени, удовлетворительного с практической точки зрения универсального метода разработки параллельных программ.

Теоретические сложности, возникающие в области разработки параллельных программ, усугубляются постоянным развитием и усложнением архитектур вычислительных систем. Эти возможности, с одной стороны, дают новый потенциал ускорения вычислений, с другой стороны, ставят проблему утилизации этого потенциала, разработки методов и средств программирования в условиях этих новых возможностей.

Помимо центральных процессоров общего назначения (CPU) современные вычислительные системы, как правило, содержат дополнительные вычислители, предназначенные для быстрого и энергоэффективного параллельного выполнения массовых вычислительных операций, одинаковых для большого объёма обрабатываемых данных. Примерами таких вычислителей могут служить графические процессоры (GPU) и ускорители Xeon Phi. Для своего эффективного выполнения параллельная программа должна обеспечить все имеющиеся в её распоряжении вычислители непрерывной загрузкой данными для вычислений. Также она должна обеспечить синхронизацию вычислений, где необходимо, при об-

ращении к общим данным, минимизировать простои вычислителей при синхронизации и доступе к прочим ресурсам, как программным, так и аппаратным. В случае, если некоторые вычислители обрабатывают выделенный им объём данных быстрее других и затем простаивают в ожидании синхронизации, возникает потребность перераспределить обрабатываемые данные между вычислителями в процессе выполнения программы.

Решение задачи распределения данных между вычислителями называется балансировкой вычислительной нагрузки (computational load balancing), а в случае периодического решения её в процессе выполнения программы – динамической балансировкой вычислительной нагрузки. От успешного решения этой задачи в значительной степени зависит и эффективность выполнения всей программы в целом.

Исследования в области создания как методов программирования для новых архитектур, так и реализации этих методов в языковых средствах параллельного программирования, ведутся весьма активно и поддерживаются фирмами-производителями вычислительных систем. Достаточно полная классификация архитектур, методов и средств параллельного программирования представлена на сайте [3], посвященного, в частности, технологиям параллельных вычислений. Из уже реализованных подходов перспективными являются, по нашему мнению, подходы, базирующиеся на вполне разумном симбиозе распараллеливающего компилятора и подсказок со стороны программиста, выполненных в виде специальных программных директив, например, [4, 5].

Интерес также представляет непроцедурный подход к построению параллельных программ [6], в котором точно определяются границы того, что и как можно автоматически распараллелить, и который автоматизирует процесс построения эффективных параллельных программ. Этот подход использует в качестве языка программирования непроцедурный язык NORMA [7].

В данной статье предложен метод организации автоматической динамической балансировкой вычислительной нагрузки в пределах одного узла гибридной вычислительной системы, в составе которого есть один или более CPU и один или более дополнительных вычислителей. Вопрос распределения вычислительной нагрузки между узлами вычислительной системы остаётся за рам-

ками данной работы. Описываемый метод был разработан для применения в системе программирования HOPMA [6] при трансляции программ для вычислительных систем с графическими ускорителями (GPU). Следует отметить, что метод универсален, не зависит от типа конкретной гибридной вычислительной системы и применяемых средств программирования и, как кажется авторам, с успехом может быть применён при создании параллельных программ вычислительного характера, как при ручном программировании, так и в случае автоматизированного подхода.

Предложенный метод был успешно опробован на программе вычислительного характера из области газовой динамики, на параллельных системах с гибридной архитектурой (графические процессоры).

СТАТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ

Рассмотрим вопрос распределения вычислительной нагрузки в пределах одного узла гибридной вычислительной системы. В каждом таком узле имеются один или более центральный процессор, CPU. Так как все современные CPU многоядерные и имеют доступ ко всей оперативной памяти узла, то неважно, сколько их в узле – вся их совокупность всегда рассматривается прикладной программой и операционной системой как единый многоядерный процессор. Также в узле гибридной вычислительной системы имеются один или более специальный вычислитель (ускоритель, GPU или Xeon Phi, или, возможно, какие-то другие). Их количество уже имеет важное значение, так как, как правило, каждый такой вычислитель имеет доступ и может обрабатывать данные только из собственной памяти.

Эффективная параллельная программа должна использовать все имеющиеся вычислительные мощности. Соответственно в узле гибридной вычислительной системы весь объём производимых вычислений должен быть как-то распределён между CPU и ускорителями. Вычисления, выполняемые на CPU, должны производиться с использованием технологий для многопоточного программирования, например, OpenMP. Вычисления, выполняемые на ускорителе, должны производиться с использованием доступной для данного типа ускорителей технологии, например, NVIDIA CUDA для GPU фирмы NVIDIA.

Процесс автоматического статического распределения вычислений между CPU и GPU при трансляции программ, написанных на языке HORMA, подробно описан в работах [8, 9]. Методы и идеи, изложенные в них, могут быть применены к любой параллельной программе и любому другому типу ускорителей. Кратко, эти методы заключаются в следующем.

В языке HORMA оператор, описывающий какие-то вычисления, может производиться над областью. Область – аналог математического понятия сетки. Таким образом, оператор описывает набор одинаковых вычислений, производимых в точках области (узлах сетки) произвольного типа. Как правило, вычисления в каждой точке области не зависят от значений в других точках, получаемых в этом же операторе на этом же шаге итерации. Тогда вычисления, производимые одним оператором, в каждой точке области не зависят друг от друга и могут быть выполнены параллельно.

Для распределения таких вычислений между CPU и ускорителями предлагается использовать такой подход, что каждый такой оператор выполняется и на CPU, и на каждом ускорителе, имеющемся в системе. При этом вся область такого оператора (точнее, массив точек, эту область составляющих) распределяется между вычислителями, и каждый вычислитель выполняет оператор только для тех точек области, которые были ему распределены. Для того чтобы вычислитель мог выполнить оператор для своих точек, также необходимо, чтобы переменные-массивы, определённые в этих точках области, как требуемые для вычислений, так и те, которые вычисляются в результате оператора, также были физически распределены между памятью CPU и памятью каждого ускорителя. То есть, фактически, процесс распределения вычислений сводится к процессу распределения данных и затем к синхронному выполнению вычислений каждым вычислителем над предназначенной ему частью общих данных.

Для распределения таких переменных предлагается следующий способ. Вначале весь объём обрабатываемых данных делится на две неравные части: зона, обрабатываемая CPU, и зона, обрабатываемая ускорителем (ускорителями). Размеры зон выбираются в соответствии с предполагаемым соотношением производительности CPU и суммарной производительностью ускорителей. Затем, если ускорителей несколько, их зона делится на соответствующее число подзон и рас-

пределяется между ними поровну. На рис. 1 приведён пример такого распределения.

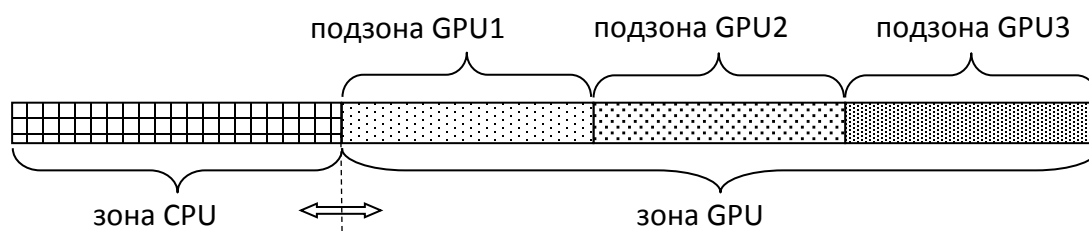


Рисунок 1. Распределение обрабатываемых данных между зонами.

Каждый вычислитель производит вычисления над теми данными, которые попали в его зону (подзону). При этом дополнительно решается задача передачи данных между зонами и подзонами, если в этом есть необходимость.

ДИНАМИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ

Граница распределения вычислений между зоной CPU и зоной ускорителей может быть фиксированной, а может быть изменяемой в процессе выполнения программы. С помощью периодического изменения границы зон в процессе выполнения программы осуществляется динамическая балансировка вычислительной нагрузки, о которой пойдёт речь ниже. При изменении размера зоны GPU соответственно пересчитываются размеры подзон каждого ускорителя.

Чтобы оценить необходимость внести корректировку в положение границы зон, периодически (например, на каждом n -ом шаге итерации) запускается процесс определения общей эффективности вычислений при текущем распределении вычислительной нагрузки. В результате этого процесса принимается решение о сохранении текущего положения границы зон или о сдвиге его на определённую величину в ту или иную сторону. В случае сдвига необходимо организовать перераспределение данных, которые в результате сдвига начинают обрабатываться другим вычислителем.

МЕТОД ОПРЕДЕЛЕНИЯ НЕОБХОДИМОСТИ КОРРЕКТИРОВКИ РАСПРЕДЕЛЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ НАГРУЗКИ

Для определения общей эффективности вычислений предлагается использовать следующий метод, основанный на оценке использования программой ресурса CPU. При разработке этого метода мы исходим из предположения, что программа, решающая задачу вычислительного характера, должна постоянно заниматься вычислениями и полностью потреблять ресурс вычислителей. Конечно, в ней могут быть точки синхронизации, когда отдельные процессы и потоки могут ожидать другие процессы и потоки, но время такого ожидания должно быть достаточно мало по сравнению со временем счёта. Если программа периодически ожидает какие-то внешние события и проводит в этом ожидании значительное время, то данный метод к такой программе неприменим.

Таким образом, в идеале, вычислительная программа должна создавать загрузку CPU на 100%. Если же программа гибридная, и наряду с вычислениями на CPU используются вычисления на ускорителе, то в случае, когда ускоритель обрабатывает выделенные ему данные медленнее, чем CPU – выделенные ему, программа будет простаивать в точках синхронизации в ожидании окончания работы ускорителя. Вследствие этого загрузка программой CPU будет меньше 100%. Получение программой данных о том, насколько она потребляет выделенный ей ресурс CPU, делается очень легко и быстро. В OS семейства UNIX это, например, делается с помощью системного вызова **clock_gettime(...)**. Вызов с параметром **CLOCK_REALTIME** даёт общее системное время, а с параметром **CLOCK_PROCESS_CPUTIME_ID** – процессорное время, потреблённое выполняющейся программой. Если засечь эти два параметра за какой-то период, то величина загрузки CPU данной программой за этот период может быть вычислена по следующей формуле:

$$\text{CPU}_{\text{load}} = t_{\text{CLOCK_PROCESS_CPUTIME_ID}} / N_{\text{thread}} / t_{\text{CLOCK_REALTIME}} * 100\%,$$

где t – соответствующие времена, а N_{thread} – количество выполняющихся программных потоков на CPU.

Далее рассмотрим, как трактовать полученную величину загрузки CPU. Как говорилось выше, в идеале у вычислительной программы загрузка CPU должна

быть 100%. Но она также будет 100% в случае, если ускоритель обрабатывает свои данные быстрее, чем CPU, и программа не простаивает в ожидании готовности ускорителя. Чтобы иметь возможность диагностировать такую ситуацию, а также чтобы позволить программе проводить некоторое время в точках синхронизации с другими процессами, предлагается считать нормальной загрузкой CPU эмпирическое значение 95%. То есть допускается, чтобы CPU простаивал в ожидании ускорителя, но простаивал совсем немного, не более 5%. Если значение загрузки CPU больше того, что предложено считать нормой, значит, надо уменьшить его зону (и, соответственно, увеличить зону ускорителя). Если наоборот, меньше нормального – то увеличить зону CPU.

Но сдвиг границы зон влечёт за собой запуск процесса перераспределения данных, и выполнение его может занимать значительное время. Поэтому крайне желательно избежать ситуации в виде скачков постоянного то уменьшения, то увеличения размеров зон. Для этого нормальной загрузкой CPU предлагается считать не конкретное значение, а небольшой диапазон, например, от 85% до 95%.

Таким образом, периодически (например, на каждом n -ом шаге итерации) оценивая величину загрузки CPU за истёкший с момента предыдущей оценки интервал времени, можно принимать решение о том, оставить ли текущее распределение данных (если загрузка CPU находится в желаемом диапазоне), или увеличить зону CPU (если загрузка CPU ниже диапазона), или уменьшить зону CPU (если загрузка CPU выше диапазона). Также необходимо определить, насколько сильно надо сдвигать границу зон. Очевидно, что чем больше значение загрузки CPU отличается от желаемого, тем больше должно быть изменение границы зон. С другой стороны, двигать границу зон, когда размер одной зоны значительно превышает размер другой, надо с осторожностью, так как даже небольшой сдвиг границы может в процентном отношении существенно изменить объём обрабатываемых данных, что в свою очередь может в корне изменить баланс вычислительной загрузки между CPU и ускорителем. Резкое изменение соотношения размера зон может привести на следующем шаге к необходимости изменения в обратную сторону. И если оно также будет резким, то получатся только постоянные скачки то в одну, то в другую сторону, которых, как объяснялось выше, необходимо избегать. Поэтому вблизи крайних значений относительного размера зон алгоритм изменения грани-

цы зон должен вести себя достаточно аккуратно, сдвигая границу на небольшие значения. В результате алгоритм, определяющий величину изменения границы зон, может быть описан функцией от двух переменных – отклонение от желаемой загрузки CPU и текущий относительный размер зон.

Предложенный метод может быть алгоритмизирован, подходит для широкого класса вычислительных задач и не зависит от характеристик конкретной вычислительной системы. Следовательно, он может быть применён для автоматической балансировки вычислительной нагрузки. Планируется реализовать его в компиляторе программ на языке FORTRAN [10] так, чтобы компилятор автоматически генерировал весь необходимый код для определения загрузки CPU, принятия решения о сдвиге границы зон и перераспределения обрабатываемых данных. Пока же изложенный метод реализован «вручную» в одной программе из области газодинамики, и в следующей главе рассказано о результатах его применения.

РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ ПРЕДЛОЖЕННОГО МЕТОДА

Метод автоматической балансировки вычислительной нагрузки был реализован в гибридной программе, решающей задачу из области газодинамики. Программа построена с использованием технологий MPI для задействования нескольких узлов вычислительного комплекса, OpenMP для вычислений на CPU внутри одного узла и NVIDIA CUDA для вычислений на GPU.

Приведённые ниже результаты были получены на вычислительном кластере K-100 [11] с использованием компиляторов Intel версии 15.0.0, nvcc версии 6.5 и Intel MPI Library 5.0 Update 1. В программе использовались 4 MPI процесса, каждый из которых выполнялся на своём узле вычислительного кластера, имеющего 12 ядер CPU и 3 GPU. Испытания проводились с разным количеством GPU, и метод исправно работал вне зависимости от количества GPU. Но, так как данная программа хорошо подходит для выполнения на GPU, доля вычислений CPU получалась очень маленькой. Поэтому далее приведены данные для запусков с использованием только 1 GPU, чтобы вклад CPU в общую долю вычислений был более заметным, а происходящие процессы были более наглядными. Программа также запускалась на вычислительном кластере K-60 [11], имеющем более мощные GPU. Метод автоматической балансировки вычислительной нагрузки также исправно работал и

там, но доля вычислений на CPU получалась ещё меньше – 4% при использовании всего 1 GPU.

На рис. 2 представлена диаграмма времён выполнения программы при различных значениях размера зоны GPU. Первые 10 столбцов соответствуют запускам с фиксированной границей зон, без использования автоматической балансировки вычислительной нагрузки, и зона GPU задаётся от 100% (когда CPU совсем не используется) до 83%. Последние 4 столбца – запуски с использованием автоматической балансировки вычислительной нагрузки, с различным значением первоначального размера зоны GPU: 100%, 75%, 50% и 0%.

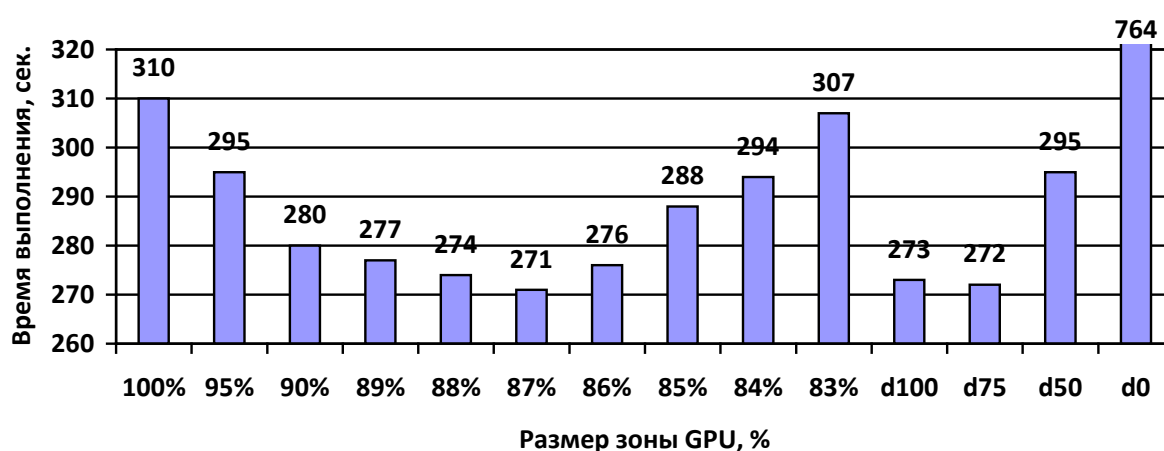


Рисунок 2. Время выполнения программы в зависимости от размера зоны GPU.

Как видно на диаграмме, наименьшее время выполнения программы достигается при размере зоны GPU 87% (и зоны CPU 13% соответственно). Затем, при небольшом увеличении зоны CPU, время выполнения программы начинает быстро расти – фактически на столько же в процентном отношении, на сколько растёт зона CPU, так как именно время работы CPU начинает определять время работы всей программы.

Особый интерес представляют столбцы, соответствующие запускам с использованием автоматической балансировки вычислительной нагрузки. Они показывают, что если первоначальное распределение вычислительной нагрузки было выбрано, хоть и приблизительно, правильно (d100 – начальная зона GPU 100% и d75 – начальная зона GPU 75%), то и время работы всей программы получается близким к идеальному. Но если первоначальное распределение было выбрано неверно

(d50 и, в особенности, d0), то программа тратит значительное время для выхода на своё правильное распределение.

На рис. 3 приведены графики изменения размеров зон в зависимости от шага итерации для запусков с использованием автоматической балансировки вычислительной нагрузки. Значения вычислительной нагрузки анализировались и корректировались на каждом 100-м шаге итерации.

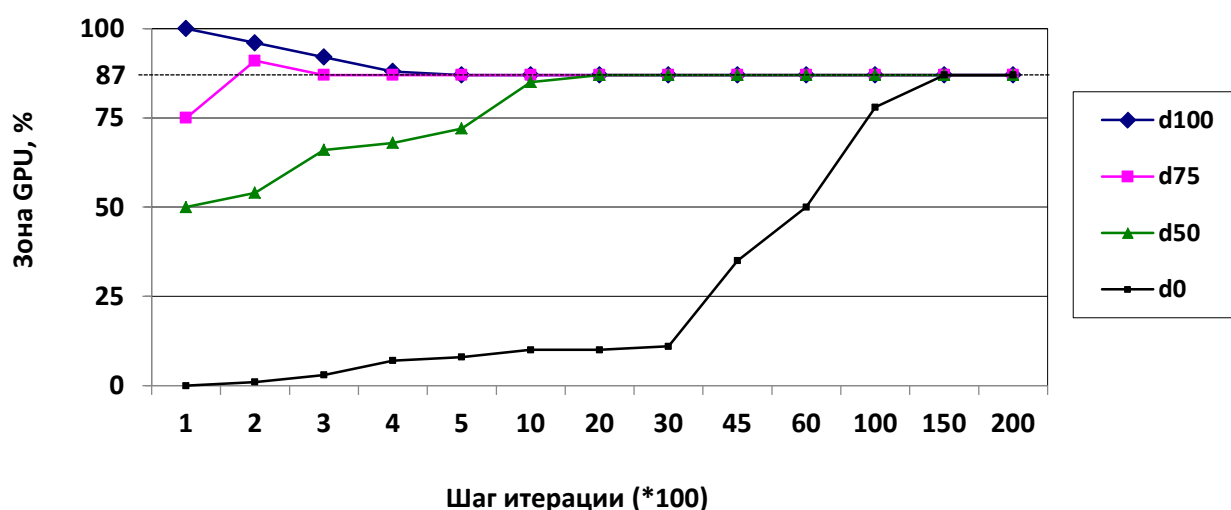


Рисунок 3. Изменение зоны GPU в зависимости от шага итерации.

Как видно на приведённом графике, запуски d100 и d75 уже на 4-й корректировке выходят на распределение, идеальное для данной задачи на данной аппаратуре: 87% для зоны GPU. Запуск d50 тратит на это уже больше шагов корректировки, и, как следствие, общее время его выполнения получается заметно больше. Запуск d0 очень долгое время осторожно отходит от нулевого размера зоны GPU, средние значения распределения зон проходятся уже значительно быстрее, и в итоге он также выходит на то же самое идеальное распределение, 87% для зоны GPU. Но для этого ему потребовалось 150 корректировок, и значительное время было упущено.

ЗАКЛЮЧЕНИЕ

Предложенный метод автоматической балансировки вычислительной нагрузки, несмотря на свою простоту, успешно может быть использован при реше-

нии расчётных задач на гибридных вычислительных системах. Испытания показали, что применение данного метода успешно выводит задачу на её идеальное распределение вычислительной нагрузки, а в случае небольшого изменения нагрузки способно быстро реагировать на это и корректировать такие изменения. Приведённый метод не зависит ни от аппаратных особенностей гибридной вычислительной системы, ни от средств программирования, выбранных для решения прикладной задачи.

СПИСОК ЛИТЕРАТУРЫ

1. Воеводин В.В. Математические модели и методы в параллельных процессах. М.: Наука, 1986. 296 с.

2. Воеводин В.В., Воеводин Вл.В. Параллельные вычисления. С.-Петербург: БХВ-Петербург, 2002. 608 с.

3. Информационно-аналитический центр по параллельным вычислениям. URL: <http://parallel.ru>

4. OpenACC. URL: <http://openacc.org>

5. DVM-система. URL: <http://www.keldysh.ru/dvm>

6. Система НОРМА. URL: <http://www.keldysh.ru/pages/norma>

7. Андрианов А.Н., Баранова Т.П., Бугеря А.Б., Гладкова Е.Н., Ефимкин К.Н. Язык НОРМА // Препринты ИПМ им. М.В.Келдыша. ISSN 2071-2898 (Print). ISSN 2071-2901 (Online). 2019. № 132. 48 с. doi:10.20948/prepr-2019-132. URL: <http://library.keldysh.ru/preprint.asp?id=2019-132>.

8. Андрианов А.Н., Баранова Т.П., Бугеря А.Б., Ефимкин К.Н. Распределение вычислений в гибридных вычислительных системах при трансляции программ на языке НОРМА // Вычислительные методы и программирование. ISSN 1726-3522. М.: НИВЦ МГУ им. М.В. Ломоносова, 2019. Т. 20, № 3. С. 224–236. DOI: 10.26089/NumMet.v20r321.

URL: http://num-meth.srcc.msu.ru/zhurnal/tom_2019/pdf/v20r321.pdf

9. Андрианов А.Н., Баранова Т.П., Бугеря А.Б., Ефимкин К.Н. Методы распределения вычислений при автоматическом распараллеливании непроцедурных спецификаций // Суперкомпьютерные дни в России: Труды международной конференции. 23–24 сентября 2019 г., г. Москва. Под. ред. Вл.В. Воеводина.

М.: МАКС Пресс, 2019. ISBN 978-5-317-06007-7. e-ISBN 978-5-317-06244-6. С. 59–70. DOI: 10.29003/m680.RussianSCDays.

URL: <http://russianscdays.org/files/2019/pdf/59.pdf>

10. Андрианов А.Н., Бугеря А.Б., Ефимкин К.Н., Колударов П.И. Модульная архитектура компилятора языка Норма+ // М.: Препринт ИПМ им. М.В. Келдыша РАН, 2011. № 64. 16 с.

URL: http://keldysh.ru/papers/2011/prep64/prep2011_64.pdf

11. Центр коллективного пользования ИПМ им. М.В. Келдыша РАН.
URL: <http://ckp.kiam.ru/?hard>

COMPUTATIONAL LOAD BALANCING METHOD FOR HYBRID COMPUTING SYSTEMS

T. P. Baranova¹, A. B. Bugerya², K. N. Efimkin³

¹⁻³*Keldysh Institute of Applied Mathematics, Moscow*

¹bart1950@yandex.ru, ²shurabug@yandex.ru, ³bigcrocodile@yandex.ru

Abstract

The paper considers the issues of the computations distributing within one node of a hybrid computing system for applied programs with computation-intense operations. A method is proposed for static distribution of computations, as well as a method for automatic balancing of the computational load during program execution, which is based on periodic analyzing the CPU load by the executed program and making decision to redistribute computational load if necessary. The proposed methods are implemented in an applied program that solves a gas dynamic problem using the computing resources of the multicore central processor and graphics accelerators. The results of program execution with various data distributions were obtained and analyzed, both with and without the mechanism for automatic balancing of the computational load.

Keywords: *parallel programming, programming automation, computational load balancing, hybrid computing architectures, NORMA language, automatic program generation*

REFERENCES

1. Voevodin V.V. Matematicheskie modeli i metody v parallelnykh protsessakh. M.: Nauka, 1986. 296 s.
2. Voevodin V.V., Voevodin V.I. Parallelnye vychisleniia. S.-Peterburg: BKhV-Peterburg, 2002. 608 s.
3. Informational Analytical Center. URL: http://parallel.ru/index_eng.html
4. OpenACC. URL: <http://openacc.org>
5. DVM-system. URL: <http://www.keldysh.ru/dvm>
6. Sistema NORMA. URL: <http://www.keldysh.ru/pages/norma>
7. Andrianov A.N., Baranova T.P., Bugerya A.B., Gladkova E.N., Efimkin K.N. Iazyk NORMA. // Preprinty IPM im. M.V. Keldysha. ISSN 2071-2898 (Print). ISSN 2071-2901 (Online). 2019. № 132. 48 s. doi:10.20948/prepr-2019-132. URL: <http://library.keldysh.ru/preprint.asp?id=2019-132>.
8. Andrianov A.N., Baranova T.P., Bugerya A.B., Efimkin K.N. Raspredelenie vychislenii v gibridnykh vychislitelnykh sistemakh pri transliatsii programm na iazyke NORMA // Vychislitelnye metody i programmirovaniye. ISSN 1726-3522. M.: NIVTs MGU im. M.V. Lomonosova, 2019. Vol. 20, № 3. S. 224–236. DOI: 10.26089/NumMet.v20r321. URL: http://num-meth.srcc.msu.ru/zhurnal/tom_2019/pdf/v20r321.pdf
9. Andrianov A.N., Baranova T.P., Bugerya A.B., Efimkin K.N. Metody raspredeleniia vychislenii pri avtomaticheskom rasparallelivanii neprotsedurnykh spetsifikatsii // Superkompiuternye dni v Rossii: Trudy mezhdunarodnoi konferentsii. 23-24 sentiabria 2019 g., g. Moskva. Pod. red. V.I.V. Voevodina. M.: MAKS Press, 2019. P. 59–70. ISBN 978-5-317-06007-7. e-ISBN 978-5-317-06244-6. DOI: 10.29003/m680.RussianSCDays. URL: <http://russianscdays.org/files/2019/pdf/59.pdf>

10. *Andrianov A.N., Bugerya A.B., Efimkin K.N., Koludarov P.I.* Modulnaia arkhitektura kompiliatora iazyka Norma+ // М.: Preprint IPM im. M.V. Keldysha RAN, 2011. № 64. 16 s. URL: http://keldysh.ru/papers/2011/prep64/prep2011_64.pdf

11. Centr kollektivnogo pol'zovaniya IPM im. M.V. Keldysha RAN. URL: <http://ckp.kiam.ru/?hard>

СВЕДЕНИЯ ОБ АВТОРАХ



БАРАНОВА Татьяна Петровна – научный сотрудник ИПМ им. М.В. Келдыша РАН;

Tat'yana Petrovna BARANOVA – Keldysh Institute of Applied Mathematics, researcher.

email: bart1950@yandex.ru

ORCID: 0000-0002-2003-3737



БУГЕРЯ Александр Борисович – старший научный сотрудник ИПМ им. М.В. Келдыша РАН, к. ф.-м. н.

Alexander Borisovich BUGERYA – Keldysh Institute of Applied Mathematics, senior researcher, Ph.D.

email: shurabug@yandex.ru

ORCID: 0000-0002-9698-458X



ЕФИМКИН Кирилл Николаевич – старший научный сотрудник ИПМ им. М.В. Келдыша РАН, к. ф.-м. н.

Kirill Nikolaevich EFIMKIN – Keldysh Institute of Applied Mathematics, senior researcher, Ph.D.

email: bigcrocodile@yandex.ru

ORCID: 0000-0002-1087-7645

Материал поступил в редакцию 25 ноября 2020 года

УДК 519.6 + 519.2

СЕЗОННАЯ И МНОГОЛЕТНЯЯ ИЗМЕНЧИВОСТЬ АТМОСФЕРНОГО ДАВЛЕНИЯ В АРКТИКЕ, СТАТИСТИЧЕСКИЙ И ВРЕМЕННОЙ АНАЛИЗ

К. П. Беляев^{1[0000-0003-2111-2709]}, Г. М. Михайлов^{2[0000-0002-4535-7180]},

А. Н. Сальников^{3[0000-0001-8669-9905]}, Н. П. Тучкова^{4[0000-0001-6518-5817]}

^{1,2,3,4}Вычислительный центр им. А.А. Дородницына ФИЦ ИУ РАН, г. Москва

¹Институт океанологии им. П.П. Ширшова РАН, г. Москва

³ФОУ ВПО «Московский государственный университет имени М.В.Ломоносова», факультет ВМК, г. Москва

¹kosbel55@gmail.com, ²gmickail@ccas.ru, ³salnikov@angel.cs.msu.ru, ⁴natalia_tuchkova@mail.ru

Аннотация

Проведен статистический и временной анализ сезонной и многолетней изменчивости поля атмосферного давления в Арктическом регионе России. В качестве основного метода исследования использованы схемы частотных характеристик вероятностей переходов при построении диффузионных процессов, описывающих изменчивость полей давления. На основе заданных рядов длиной в 60 лет с 1948 по 2008 годы вычислялись такие параметры диффузионных процессов, как среднее (снос процесса) и дисперсия (диффузия процесса), и строились их карты и временные кривые. Изучена сезонная и многолетняя изменчивость полученных полей, а также их зависимость от интервалов и величины разбиения. Эти характеристики проанализированы, и проведена их геофизическая интерпретация. В частности, выявлены известные циклы солнечной активности в 11 и 22 года, а также известный из литературы квазидвухлетний период. Численное моделирование выполнено на суперкомпьютере Ломоносов-2 Московского государственного университета имени М.В. Ломоносова.

Ключевые слова: анализ временных рядов, диффузионные случайные процессы, сезонная и многолетняя изменчивость атмосферного давления

ВВЕДЕНИЕ

Анализ временных рядов (АВР) – одно из наиболее хорошо разработанных и широко применяемых направлений в математической статистике. Методы АВР успешно применяются в геофизике, экономике, технике и других работах, связанных с исследованиями массивов данных. Например, одним из первых применений методов АВР был анализ данных по урожаю в Англии в 18 веке [1], связанный со сбором зерна, который разделялся на многолетний тренд сезонных составляющих и нерегулярных компонентов, зависящих от текущих событий (погодные условия, инфляционный ценовой всплеск и пр.). В последующем АВР стал использоваться при анализе финансового рынка [2], анализе многолетней изменчивости геофизических характеристик, таких как температура воздуха или воды [3], в более сложных моделях и схемах [4]. На современном этапе анализа АВР часто используются схемы авторегрессии и скользящего среднего (ARIMA) [5]. Отметим, что этот анализ требует использования большого объема вычислительных мощностей, компьютерного времени и памяти, решения задач визуализации результатов и многих сопутствующих проблем. В последние годы, благодаря значительному прогрессу в области вычислительных систем и численного моделирования, накопления и обработки больших данных, эксперименты по АВР становятся доступными многим исследовательским группам и отдельным пользователям, принадлежащим научному сообществу. Это, в свою очередь, способствует дальнейшему развитию численного моделирования и возможностей анализа модельных данных, полученных результатов и их последующего сравнения.

Исследования по АВР широко используются также непосредственно в теории вероятностей и математической статистике. Одним из методов анализа является представление ряда в виде цепи Маркова и/или Марковского процесса. Поскольку литература по Марковским процессам весьма обширна, отметим только несколько из наиболее известных работ в данной области, в которых изложены все необходимые теоретические положения и практические методы расчета характеристик, необходимых для дальнейшего исследования. Например, в работе [6] описаны все нужные в настоящем исследовании теоретические

сведения о том, как определять приведенные в дальнейшем параметры процесса, а в работе [7] изложены конкретные примеры таких процессов.

В данной работе моделируется поведение поля атмосферного давления на основе Марковского диффузионного процесса. Такие процессы хорошо описывают поведение характеристик полей, которые изменяются под воздействием двух сил – короткопериодной, называемой диффузией процесса, и длиннопериодной, называемой дрейфом или сносом. Эти модели обобщают приведенные ранее в литературе разложения временного ряда на тренд, периодическую и случайную составляющие, упомянутые в работах [1], [3, 4]. В теории вероятностей такие процессы описываются стохастическими дифференциальными уравнениями [6, 7], а их плотности вероятностей задаются решениями уравнения Фоккера–Планка–Колмогорова [8].

Характеристики этих процессов адекватно описываются такими моделями, если выполнены два основных условия. Во-первых, инкремент (то есть разность между двумя соседними моментами времени) должен быть много меньше, чем общая длина ряда. Во-вторых, поле этих характеристик должно быть достаточно однородно, то есть поведение в соседних точках пространства не сильно отличается друг от друга, особенно если это поведение рассматривать на длительных интервалах времени. Для поля атмосферного давления в относительно небольшом регионе, которое мы рассматриваем, эти условия выполнены. Длина ряда составляет 60 лет, при этом шаг по времени, то есть инкремент, составляет одни сутки. И размер циклонического атмосферного образования, которое в основном и формирует поле давления, сопоставим с размерами всей рассматриваемой области, то есть внутри области для одного образования оно не сильно изменяется. При этом важно исследовать, насколько результат зависит от разбиения реально наблюдаемого интервала давления (то есть разности максимума и минимума давления во всей области) на отдельные под-интервалы, которые используются для расчета частотных (статистических) характеристик при анализе изменчивости этого поля.

Методы диффузионных случайных процессов применялись ранее для различных задач, в том числе задач усвоения данных, в методиках, предложенных

авторами [9, 10]. Однако для описания сезонного и многолетнего поведения атмосферных процессов этот метод широко не применялся.

В работе сделано следующее:

- построены характеристики моделей, описаны их особенности, в частности, особенности сезонного и многолетнего хода, проведен анализ особенностей;
- построены временные графики и пространственные карты этих характеристик, проведен их анализ;
- проведен анализ устойчивости к разбиению всего интервала значений давления (максимума минус минимума поля по всей области) на частотные под-интервалы

1. ВЕРОЯТНОСТНАЯ МОДЕЛЬ

Изменчивость случайного процесса, в нашем случае это поле давления, представляется в виде

$$dX = a(t, X)dt + b(t, X)dW, \quad (1)$$

где X – значение (давление) поля (в момент времени t и в точке с заданными координатами, явно не показано), t – время, dW – стандартное обозначение гауссова «белого шума» – обобщенного случайного процесса с нулевым средним значением и дисперсией, равной единице, при этом его ковариационная функция равна дельта-функции, то есть $E dW(t)dW(\tau) = \delta(t - \tau)$, здесь и далее $\delta(t - \tau) = 1$, если $t = \tau$, и нулю, если нет; $a(t, x)$, $b(t, x)$ – некоторые функции, которые вычисляем по формулам из монографии [6]. Выражение (1) понимается в интегральном смысле, то есть

$$X(t + \Delta t) - X(t) = \int_t^{t+\Delta t} a(u, X)du + \int_t^{t+\Delta t} b(u, X)[W(u + du) - W(u)]. \quad (2)$$

В формуле (2) выражение $W(u + du) - W(u)$ представляет собой гауссову случайную величину с нулевым средним и дисперсией, равной du . Теория стохастического интеграла и все определения, необходимые для понимания формул (1) и (2), содержатся в работах [6, 7].

Согласно работе [6], для определения коэффициентов $a(t, x)$ и $b^2(t, x)$ используются следующие формулы

$$a(t, x) = (dt)^{-1} \int_t^{t+dt} (y - x) p(y | x) dy, \quad (3)$$

$$b^2(t, x) = (dt)^{-1} \int_t^{t+dt} (y - x)^2 p(y | x) dy, \quad (4)$$

где использованы следующие обозначения:

x и y – значения процесса $X(t)$ в моменты времени t и $t + dt$, соответственно;

$p(y | x)dt$ – вероятность (условная вероятность) события $X(t + dt) = y$ при условии $X(t) = x$, то есть $p(y | x)dt = P(X(t + dt) = y | P(t) = x)$.

Решается задача – вычислить эти коэффициенты и произвести анализ полученных характеристик.

Для статистического определения условной вероятности $p(y | x)dt$ нужно иметь выборку наблюдений (значений) x и y в фиксированной точке пространства. Однако, так как рассматриваемая область однородная, как отмечалось выше, можно отметить точки с этими значениями по всей этой области. Именно, техника определения этих вероятностей следующая: на шаге t отмечаются все точки в области, где $X(t) = x$ ($x_{\min} < x < x_{\max}$). Для простоты значения x_{\min} , x_{\max} можно считать одинаковыми для всех t . Пусть таких точек будет $n(x)$. Далее, на шаге $t + dt$ в тех и только в тех точках, где $X(t) = x$ выбираются все те точки, где $X(t + dt) = y$. Пусть таких точек будет $m(y)$. Тогда $p(y | x)dt = m(y)/n(x)$. Очевидно, что $0 \leq p(y | x)dt \leq 1$, то есть это действительно вероятность. Далее, расчет коэффициентов $a(t, x)$ и $b^2(t, x)$ осуществляется по формулам (3) и (4). Такой метод определения коэффициентов $a(t, x)$ и $b^2(t, x)$ ранее был опубликован для несколько иной задачи в работах [9, 10]. При этом, как видно из описания метода, результат зависит от числа под-интервалов, на которые разбивается весь интервал изменчивости поля давления (то есть максимум минус минимум значе-

ния поля давления по области). Были проведены эксперименты по разбиению на 20 и 60 интервалов.

2. ДАННЫЕ НАБЛЮДЕНИЙ И РЕЗУЛЬТАТЫ РАСЧЕТОВ

Рассмотрим поле атмосферного давления в области, ограниченной координатами 62°с.ш.–80°с.ш. и 30°в.д.–90°в.д., то есть область России от побережья Балтики близ С.-Петербурга до Северной Земли и Енисея в Сибири. Эта область достаточно широка, чтобы пренебречь локальными особенностями атмосферных процессов, и достаточно однородна, так как размеры крупных атмосферных образований сопоставимы с размерами всей области. Данные по давлению записаны с 1 января 1948 г по 31 декабря 2008 г. ежедневно в одноградусной сетке. Данные получены в Гидрометцентре РФ и использовались ранее в некоторых работах (см. [11]). На рис. 1(а–г) представлены поля давления в ограниченной области на 1 января с интервалом 20 лет (от 1948 до 2007 гг.).

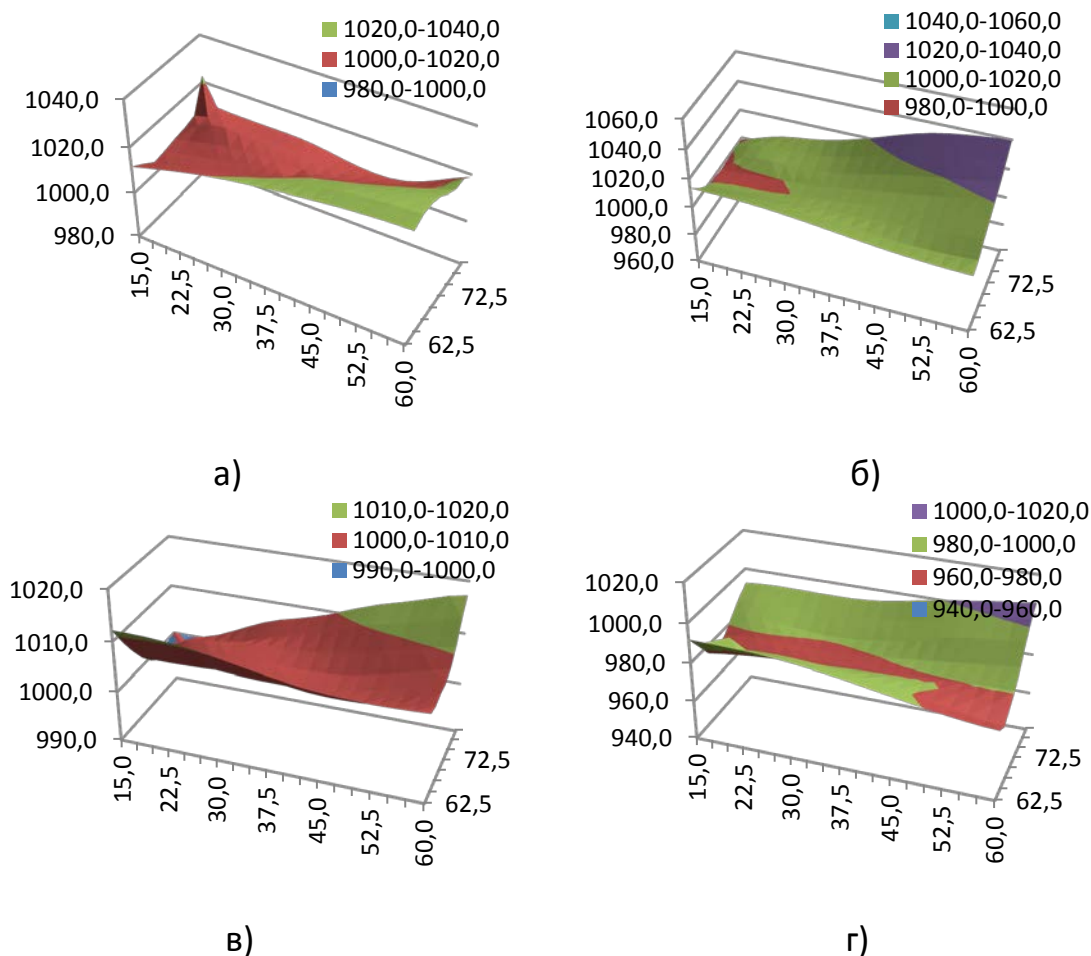


Рисунок 1. Наблюдаемые поля давления Северной зоны РФ на 1 января:
а) 1948 год; б) 1967 год; в) 1987 год; г) 2007 год.

Из рис. 1 видно, что поля давления имеют многолетнюю изменчивость, но достаточно однородны по пространству, области примерно равного давления занимают большую часть пространственной области. Некоторое исключение составляет 1987 г, но и в этом году пространственный градиент давления невелик, составляет приблизительно 10 гпк на 1000 км. Поэтому с большой степенью точности поле давления можно считать однородным, и к нему применимы методы, описанные в предыдущем параграфе.

По приведенным данным наблюдений по формулам (3) и (4) строились коэффициенты $a(t, x)$ и $b^2(t, x)$ для каждого X – значения давления в рассматриваемой области на конкретные сутки. При этом разбиения всей области значений давления осуществлялись на 20 и 60 интервалов.

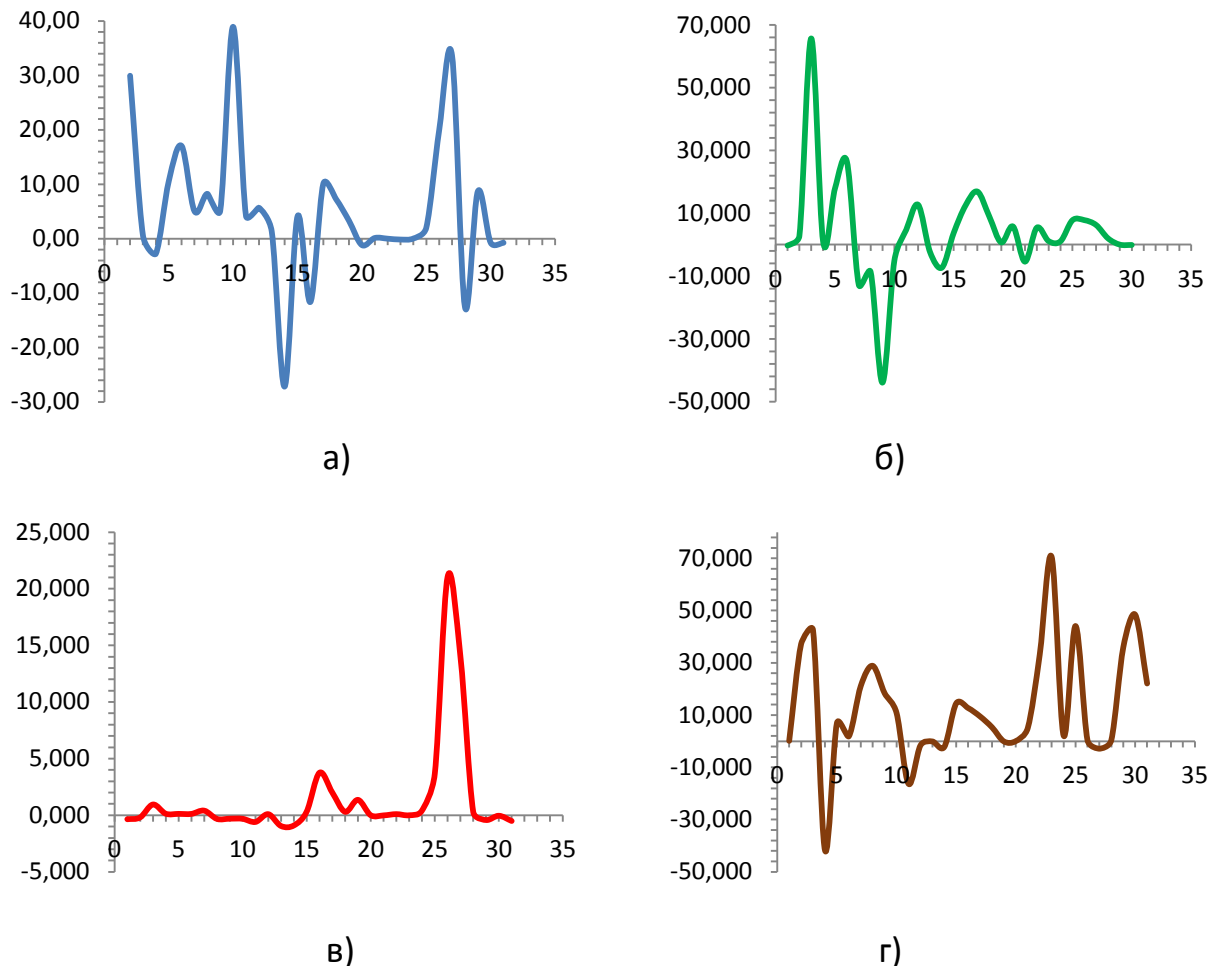


Рисунок 2. Поведение среднего коэффициента $a(t, x)$ 20 интервалов за 2007 год: а) январь; б) апрель; в) июль; г) октябрь.

В представленных иллюстрациях продемонстрированы поведение среднего значения коэффициентов $a(t, x)$ и $b^2(t, x)$ их сезонная и межгодовая изменчивость.

Рисунки 2(а–г) демонстрируют поведение среднего коэффициента $a(t, x)$ за 2007 год при разбиении на 20 интервалов. Из формулы (1) видно, что средним для процесса dX будет $a(t, x) = 0$. Однако выборочное среднее $a(t, x)$ с теоретическим средним может не совпадать, и это отличие необходимо проанализировать. 2007 год взят как показательный, в другие годы картина похожая. На этих рисунках обращает на себя внимание то, что разброс коэффициента $a(t, x)$ около среднего значения, равного нулю, очень мал летом, в июле и достаточно велик в переходные месяцы – апрель и октябрь, особенно в октябре, причем в основном отклонение от нуля положительное. Это значит, что атмосферные образования (циклоны и антициклоны) в рассматриваемом районе в основном изменяются в сторону повышения давления, то есть приходящий циклон (а их большинство) не углубляется, то есть давление не понижается.

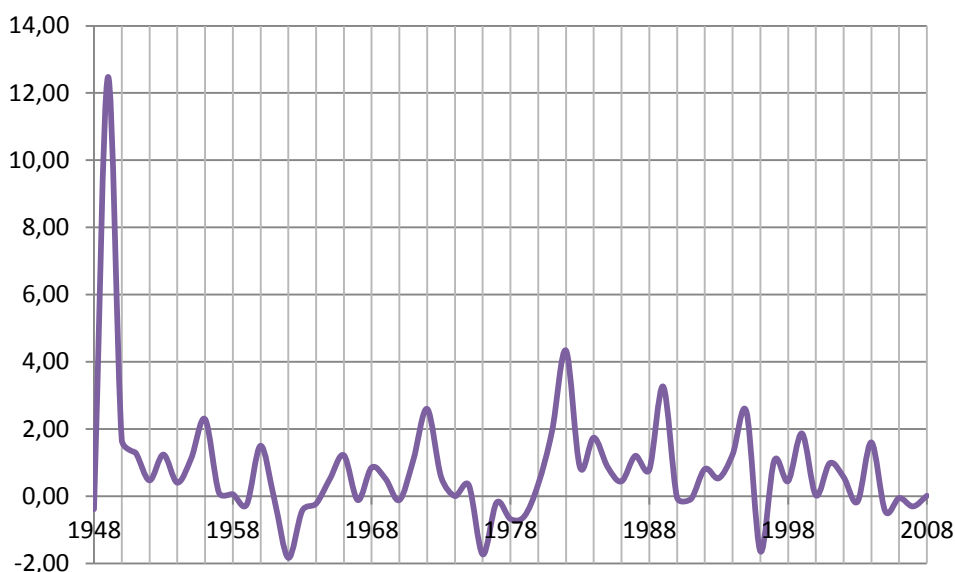


Рисунок 3. Кривая изменения коэффициента $a(t, x)$ с 1948 по 2008 годы (измерение в 0 часов 15 января).

На рис. 3 показан межгодовой ход для коэффициента $a(t, x)$ для 15 января в период с 1948 по 2008 годы. Изменение давления, выраженное в значениях величин $a(t, x)$, взятое на 15-е число каждого месяца, показывает, что есть квазипериодические многолетние колебания, с довольно хорошо выраженным 11-

летним циклом и несколько хуже с квазидвухлетним циклом. Эти циклы хорошо известны в геофизике и связаны как с колебаниями солнечной активности, так и двухлетним колебанием ветра и давления, отмеченным ранее многими авторами, например, [12]. Отметим, что в начале расчета, 1948 г график содержит резкий всплеск в значениях коэффициента, а потом таких всплесков больше не наблюдается. Это объясняется тем фактом, что в начале архива за 1948 г данные наблюдений были недостаточно гладкие, содержали много шумов и недостоверных значений. Однако затем эти данные существенно вычищались и сглаживались.

Рис. 4(а–г) иллюстрирует поведение среднего коэффициента $b^2(t, x)$ за 2007 год, а рис. 5 – межгодовой ход коэффициента $b^2(t, x)$ для 15 января соответственно.

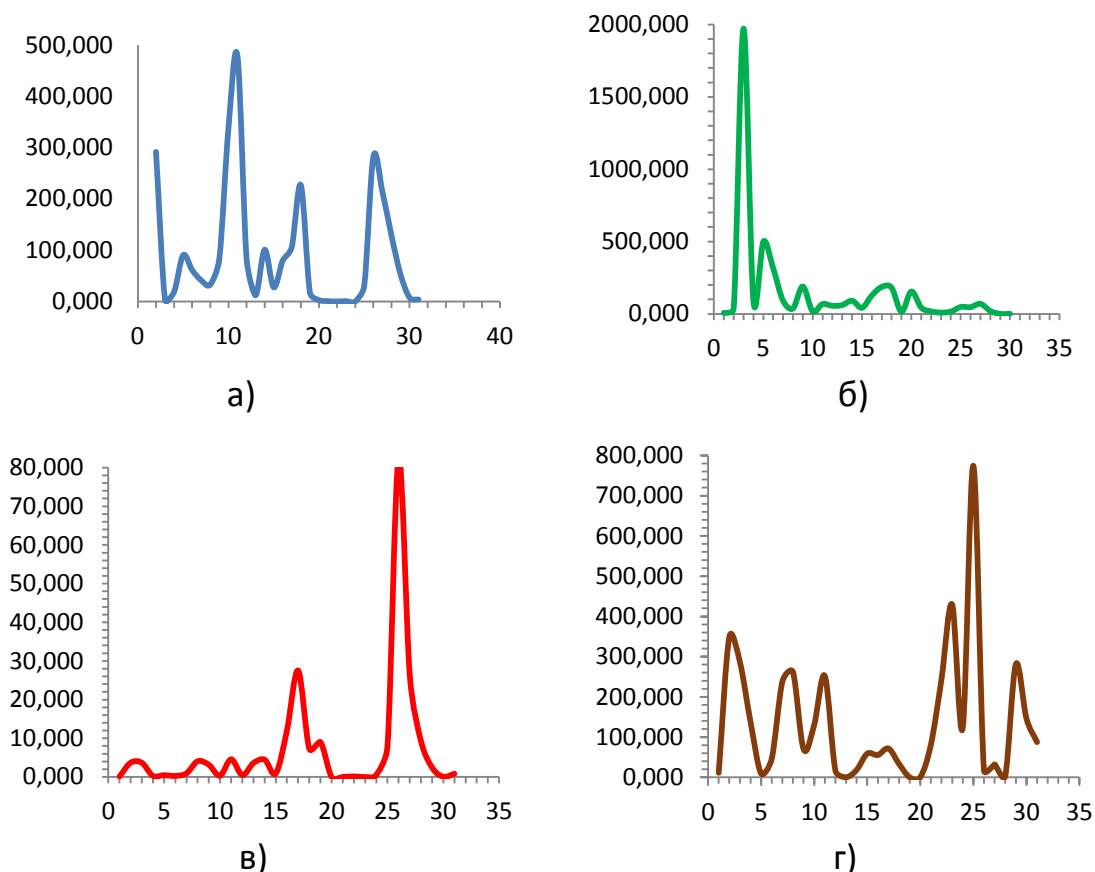


Рисунок 4. Поведение среднего коэффициента $b^2(t, x)$, 20 интервалов за 2007 год: а) январь; б) апрель; в) июль; г) октябрь.

Из рис. 4 и 5 видно, что коэффициент $b^2(t, x)$ достаточно хорошо соответствует коэффициенту $a(t, x)$, но с некоторыми отличиями. Так, из рис. 5 видно, что сезонный ход для коэффициента $b^2(t, x)$ выражен слабее, например, его почти не видно в апреле или июле, а межгодовой ход хуже отражает 11-летний цикл (хотя он тоже есть) и лучше – квазидвухлетний цикл, чем коэффициент $a(t, x)$. Каких-либо линейных трендов ни коэффициент $a(t, x)$, ни коэффициент $b^2(t, x)$ не содержат. Также наблюдается сильный всплеск в 1948 г, его объяснения даны выше.

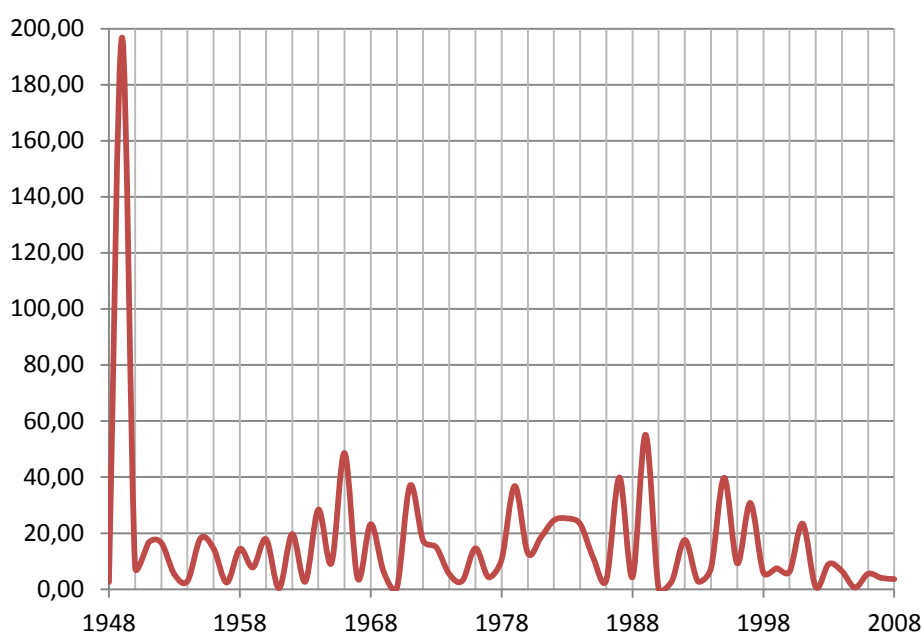


Рисунок 5. Кривая изменения коэффициента $b^2(t, x)$ с 1948 по 2008 годы (измерение в 0 часов 15 января).

На рис. 6 приведено пространственное расположение коэффициента $a(X)$, всего 4 значения на 15 января 1948, 1968, 1988 и 2008 гг. (через 20 лет).

Видно, что скачки значений этого коэффициента достаточно локализованы и не превышают 5 гпк/сутки с разным знаком. Пространственное расположение этого коэффициента по области равномерное, заметных зон локализации и выбросов не наблюдается. Еще следует отметить, что изолинии значений коэффициента $a(t, x)$ достаточно локальны по сравнению с размерами самой области (показано разными цветами), что говорит о локальном характере изменений давления в крупном масштабе.

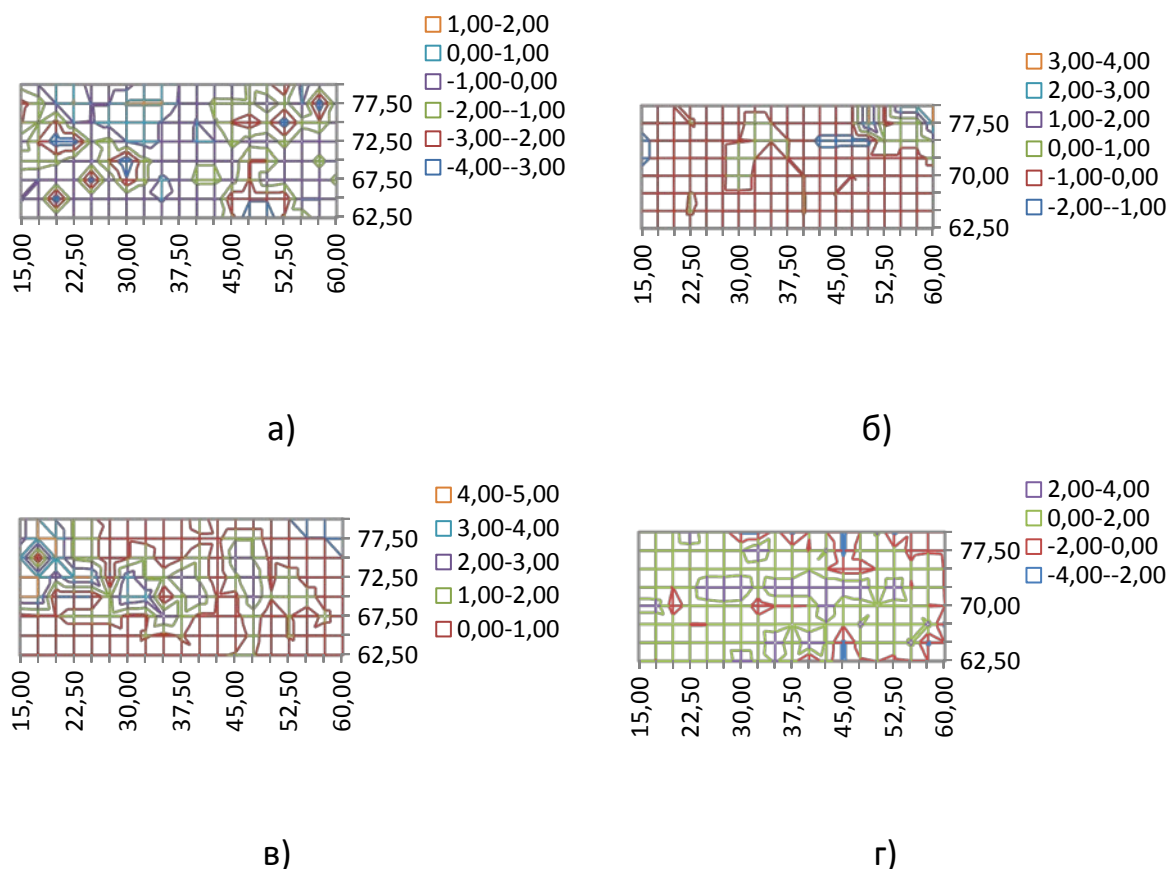


Рисунок 6. Поля коэффициентов $a(t, x)$ на 15 января за: а) 1948; б) 1968; в) 1988; г) 2008 годы.

Коэффициент $b^2(t, x)$ (рис. 7) более хаотичен, его распределение по пространству не столь гладко, как для коэффициента $a(X)$. Кроме того, можно сделать вывод, что пространственное расположение коэффициента $b^2(t, x)$ более локализовано и сосредоточено сплошными зонами (кроме значений за 2008 г.). По амплитуде размах этого коэффициента существенно больше, чем для коэффициента $a(t, x)$, локально же он занимает меньший размер всей области.

Еще можно отметить, что при увеличении разбиения на число интервалов графики 3 и 5 становятся более гладкими, рис. 4 и 6 – более выраженными, но качественно они совпадают. Поэтому результаты работы качественно не зависят от числа разбиений. Однако для корректности и надежности расчетов требуется, чтобы при разбиении каждый интервал содержал хотя бы одно наблюдение,

чтобы можно было корректно рассчитывать условные вероятности по формулам (3) и (4).

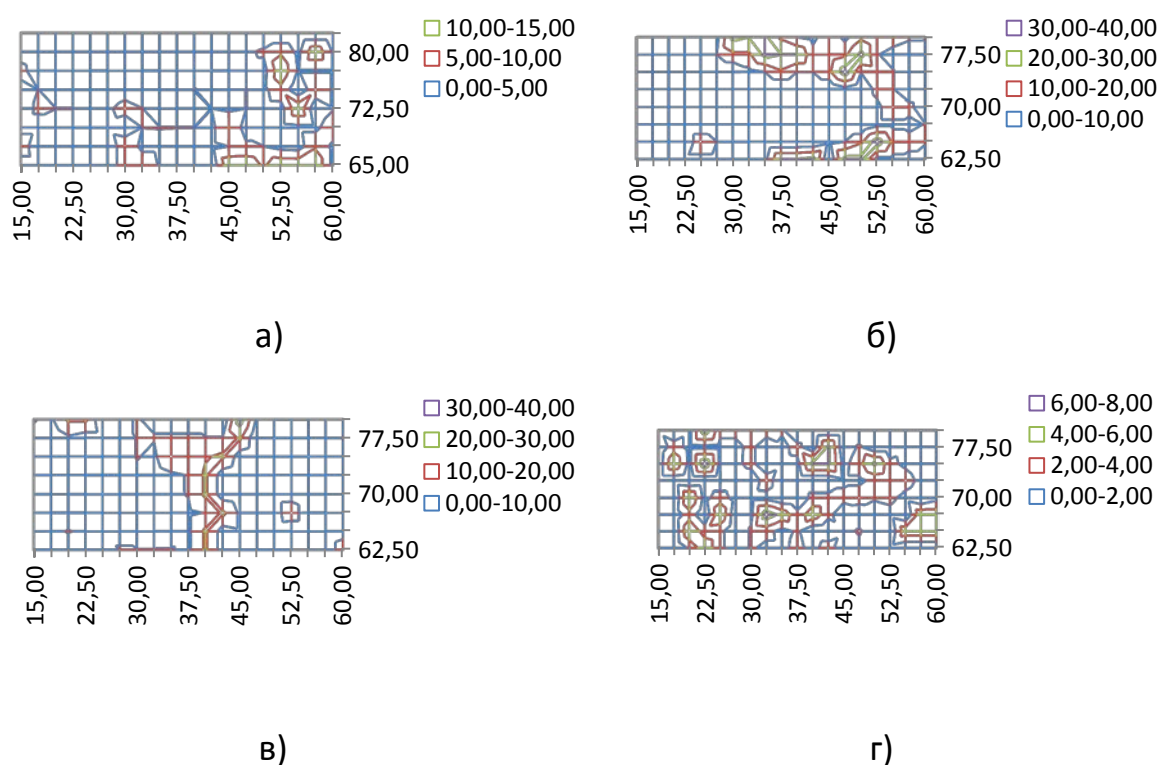


Рисунок 7. Поля коэффициентов $b^2(t, x)$ на 15 января: а) 1948; б) 1968; в) 1988; г) 2008 годы.

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

В работе получены характеристики, отражающие как долговременное, так и кратковременное поведение инкремента давления в Северной области России за 60 лет. Знание таких характеристик очень полезно при среднесрочных и долгосрочных прогнозах погоды и изменений климата, а также моделировании динамики течений в Северных морях России, особенно при расчетах проводки судов по Северному морскому пути. Кроме того, знание и прогноз характеристик, полученных в работе, позволит рассчитать и определить доверительные границы возможных значений давления, значит, и ряда производных от этой величины, например, геострофического ветра, что позволит применить эти знания при

определении экстремальных величин, таких как сильные ветра, экстремальные волны и ряда других характеристик.

БЛАГОДАРНОСТИ

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований, проект 18-29-10085мк, и в рамках тем Министерства науки и высшего образования РФ 0149-2019-0004 и «Математические методы анализа данных и прогнозирования».

СПИСОК ЛИТЕРАТУРЫ

1. *Kendall M., Stuart A., Ord J.K.* The Advanced Theory of Statistics. Vol. 3: Design and Analysis, and Time-Series. Fourth edition Hardcover. March 13, 1983.

2. *Murphy J.* Technical analysis of the futures markets. A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, 1986. 556 p.

3. *Привальский В.Е.* Статистическая предсказуемость средней годовой температуры воздуха северного полушария // Докл. АН СССР. 1981. Т. 257. № 6. С. 1342–1345.

4. *Беляев К.П., Музыченко А.С., Селеменов К.М.* Статистические характеристики формирования аномалий поверхностной температуры воды // Сб. Статистические закономерности климатической изменчивости океанов, ред. Лаппо С.С. Л: Гидрометиздат, 1988. С. 65–72.

5. *Balasmeh O., Babbar R., Karmaker T.* Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan // Arabian Journal of Geosciences. 2019. V. 12, 27. <https://doi.org/10.1007/s12517-018-4205-z>

6. *Гихман И., Скороход А.* Введение в теорию случайных процессов. М.: Наука, 1965. 655 с.

7. *Назаров А., Терпунов А.* Теория вероятностей и случайных процессов. Изд-во Томского госуниверситета, 2010. 204 с.

8. *Risken H.* The Fokker–Planck Equation: Methods of Solutions and Applications. Springer. 1984. 452 p.

9. *Беляев К., Кулешов А., Танажура К., Тучкова Н.* Метод коррекции расчетов динамической модели данными наблюдений и его применение к анализу

динамики Атлантического океана // Матем. моделирование. 2015. Т. 27. № 2. С. 20–32.

10. *Tanajura C.A.S., Belyaev K.* On the oceanic impact of a data-assimilation method in a coupled ocean-land-atmosphere model // *Ocean Dynamics*. 2002. Vol. 52. № 3. P. 123–132.

11. *Попов С.К.* Влияние морского льда на приливные колебания уровня моря и скорости течений в Баренцевом и Белом морях // *Труды Гидрометцентра РФ, Гидрометеорологические исследования и прогнозы*. 2018. № 4 (370). С. 137–155.

12. *Девятова Е.В., Мордвинов В.И.* Квазидвухлетнее колебание ветра в низкоширотной стратосфере и волновая активность атмосферы зимой в северном полушарии // *Изв. РАН, ФАО*. 2011. Т. 47. № 5. С. 608–621.

SEASONAL AND DECADEAL VARIABILITY OF ATMOSPHERE PRESSURE IN ARCTIC, ITS STATISTICAL AND TEMPORAL ANALYSIS

K. P. Belyaev^{1[0000-0003-2111-2709]}, G. M. Mickailov^{2[0000-0002-4535-7180]},

A. N. Salnikov^{3[0000-0001-8669-9905]}, N. P. Tuchkova^{4[0000-0001-6518-5817]}

^{1,2,3,4}*Dorodnicyn Computing Center FRC CSC of RAS, Vavilov str., 40, 11933, Moscow*

¹*Shirshov Institute of Oceanology of RAS, Nahimovskiy pr., 36, 117218, Moscow*

³*Lomonosov Moscow State University, GSP-1, Leninskie Gory, 11999, Moscow*

¹kosbel55@gmail.com, ²gmickail@ccas.ru, ³salnikov@angel.cs.msu.ru, ⁴natalia_tuchkova@mail.ru

Abstract

The paper analyzes the statistical and temporal seasonal and decadal variability of the atmospheric pressure field in the Arctic region of Russia. Schemes for the frequency analysis of probability transitions for characteristics of stochastic-diffusion processes were used as the main research method. On the basis of the given series of 60 years long from 1948 to 2008, such parameters of diffusion processes as the mean (drift process) and variance (diffusion process) were calculated and their maps and time curves were constructed. The seasonal and long-term variability of calculated fields was studied as well as their dependencies on a discretization of the frequency

intervals. These characteristics were analyzed and their geophysical interpretation was carried out. In particular, the known cycles of solar activity in 11 and 22 years were revealed. Numerical calculations were performed on the Lomonosov-2 supercomputer of the Lomonosov Moscow State University.

Keywords: *time series analysis, random diffusion processes, seasonal and long-term variability of atmospheric pressure*

REFERENCES

1. Kendall M., Stuart A., Ord J.K. The Advanced Theory of Statistics. Volume 3: Design and Analysis, and Time-Series. Fourth edition Hardcover – March 13, 1983.
2. Murphy J. Technical analysis of the futures markets. A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, 1986. 556 p.
3. Prival'skij V.E. Statisticheskaya predskazuemost' srednej godovoj temperatury vozduha severnogo polushariya // Dokl. AN SSSR. 1981. Vol. 257. № 6. S. 1342–1345.
4. Belyaev K.P., Muzychenko A.C., Selemenov K.M. Statisticheskie harakteristiki formirovaniya anomalij poverhnostnoj temperatury vody // Sb. Statisticheskie zakonomernosti klimaticheskoy izmenchivosti okeanov, red. Lappo S.S. L: Gidrometizdat, 1988. S. 65–72.
5. Balasmeh O., Babbar R., Karmaker T. Trend analysis and ARIMA modeling for forecasting precipitation pattern in Wadi Shueib catchment area in Jordan // Arabian Journal of Geosciences. 2019. V. 12, 27. <https://doi.org/10.1007/s12517-018-4205-z>.
6. Gihman I., Skorohod A. Vvedenie v teoriyu sluchajnyh processov. M.: Nauka, 1965. 655 s.
7. Nazarov A., Terpunov A. Teoriya veroyatnostej i sluchajnyh processov. Izd-vo Tomskogo Gosuniversiteta. 2010. 204 s.
8. Risken H. The Fokker–Planck Equation: Methods of Solutions and Applications. Springer. 1984. 452 p.
9. Belyaev K., Kuleshov A., Tanajura K., Tuchkova N. Metod korrekcii raschetov dinamicheskoy modeli dannymi nablyudenij i ego primenenie k analizu dinamiki Atlanticheskogo okeana // Matem. Modelirovanie. 2015. Vol. 27. № 2. S. 20–32.

10. *Tanajura C.A.S., Belyaev K.* On the oceanic impact of a data-assimilation method in a coupled ocean-land-atmosphere model // Ocean Dynamics. 2002. Vol. 52. № 3. P. 123–132.

11. *Popov S.K.* Vliyanie morskogo l'da na prilivnye kolebaniya urovnya morya i skorosti techenij v Barencevom i Belom moryah // Trudy Gidrometcentra RF, Gidrometeorologicheskie issledovaniya i prognozy. 2018. № 4 (370). S. 137–155.

12. *Devyatova E.V., Mordvinov V.I.* Kvazidvuhletnee kolebanie vetra v nizkoshirotnoj stratosfere i volnovaya aktivnost' atmosfery zimoj v severnom polusharii // Izv. RAN, FAO. 2011. Vol. 47. № 5. S. 608–621.

СВЕДЕНИЯ ОБ АВТОРАХ



БЕЛЯЕВ Константин Павлович – ведущий научный сотрудник Института океанологии им. П.П. Ширшова РАН и ФИЦ ИУ, доктор физ.-мат. наук, профессор кафедры теории вероятностей и статистики МГУ им. М.В. Ломоносова. Сфера научных интересов – математическое моделирование и усвоение данных наблюдений, статистический анализ натуральных данных.

Konstantin Pavlovich BELYAEV – leading scientist of Shirshov Institute of Oceanology, Russian Academy of Science. Doctor of science, professor of Dept. of Applied Math and Cybernetics, Lomonosov Moscow State University. Research interests – math. modelling and data assimilation, statistical analysis of natural data.

email: kosbel55@gmail.com



МИХАЙЛОВ Гурий Михайлович – ведущий научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук. Сфера научных интересов – архитектура вычислительных систем и сетей, вычислительные и информационные технологии.

Gury Mickailovich MICKAILOV – leading scientist of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree. Research interests include architecture of computing systems and networks, computing and information technology.

email: gmickail@ccas.ru



САЛЬНИКОВ Алексей Николаевич – ведущий научный сотрудник кафедры математической физики факультета ВМиК МГУ им. М.В. Ломоносова, кандидат физ.-мат. наук. Сфера научных интересов – параллельное программирование, биоинформатика, суперкомпьютеры.

Alexey Nikolaevich SALNIKOV – leading researcher Dept. of Applied Math. and Cybernetics, Lomonosov Moscow State University, PhD in physics with a math degree. Research interests include bioinformatics, parallel and supercomputing programming

email: salnikov@angel.cs.msu.ru



ТУЧКОВА Наталья Павловна – старший научный сотрудник Вычислительного центра им. А.А. Дородницына ФИЦ ИУ РАН, кандидат физ.-мат. наук, окончила ВМиК МГУ им. М.В. Ломоносова. Специалист в области алгоритмических языков и информационных технологий.

Natalia Pavlovna TUCHKOVA – senior researcher of Dorodnicyn computing center FRC SCS RAS, PhD in physics with a math degree, graduated from CS Faculty of Lomonosov MSU. The expert in the field of algorithmic languages and information technologies.

email: natalia_tuchkova@mail.ru

Материал поступил в редакцию 25 ноября 2020 года

УДК 004

РОССИЙСКАЯ НАУЧНАЯ ПУБЛИКАЦИЯ ПЕРИОДА COVID-19

М. М. Горбунов-Посадов^[0000-0002-7044-8287]

Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»

gorbunov@keldysh.ru

Аннотация

Влияние пандемии COVID-19 на мир научных публикаций: стремительная публикация и упрощенный доступ для статей о вирусе, открытый доступ на время самоизоляции в мире и в России. Впечатляющая разница между числом читателей российских научных статей в открытом и платном доступах, политика журналов Российской академии наук. Впечатляющий рост читательской аудитории КиберЛенинки. Онлайнное заседание диссертационного совета. Новости eLibrary. Комплексный балл публикационной активности от Минобрнауки.

Ключевые слова: научная публикация, COVID-19, открытый доступ

COVID-19

По-видимому, наиболее значимые события этого года в области научных публикаций связаны с пандемией COVID-19, прежде всего, энергичный отклик на возникшую потребность в оперативном обмене научными данными о вирусе. Практически все публикации о COVID-19 стремительно попадали в открытый доступ, независимо от политики, проводимой изданиями. Во всем мире сроки рецензирования научных сообщений о COVID-19 сократились более чем вдвое (при весьма незначительном замедлении рецензирования других статей) [1].

Новое дыхание обрели препринты, где авторы публиковали наиболее срочные научные сообщения. Число выпускаемых препринтов выросло многократно. Ставшая таким образом доступной оперативная информация оказалась чрезвычайно востребованной: количество обращений к отдельным актуальным препринтам доходило до 4 млн. Отношение издателей к препринтам давно и устойчиво теплеет [2–4], коронавирус лишь ускорил этот процесс.

За последние несколько лет ведущие научные журнальные издательства открыли для своих авторов сервера препринтов: Springer – Research Square [5], Wiley – Under Review [6], Elsevier – First Look [7]. В 2020 году сервер препринтов появился у IEEE – TechRxiv [8]. В 2016 году специализированное обслуживание препринтов было предложено в Crossref [9]. В частности, было разрешено членство в Crossref для серверов препринтов, а также реализованы настраиваемые метаданные, которые отражают рабочие процессы исследователя от препринта до официальной публикации.

Поскольку во время карантина сотрудники научных и учебных учреждений потеряли возможность доступа в свои библиотеки, подписанные на множество журналов, издатели журналов на это время открывали свободный онлайн-доступ к своим изданиям. Однако обещанный согласно плану "S" на 2021 год переход мира научных публикаций ЕС ко всеобщему открытому доступу в очередной раз отложен по крайней мере на год, а скорее до 2024 года [10, 11].

Не осталась в стороне от массового движения к открытому доступу, обусловленного COVID-19, и Российская академия наук. На время режима самоизоляции, объявленного весной, был открыт свободный доступ к 138 журналам, издаваемым РАН [12]. Подобное краткосрочное открытие доступа уже случалось в конце 2018 года, когда Счетная палата сделала жесткое замечание РАН, получающей из бюджета финансирование на издание своих журналов и тем не менее не открывающей к ним бесплатного доступа.

ОТКРЫТЫЙ ДОСТУП К ЖУРНАЛАМ В РАН

Здесь в очередной раз наглядно проявилась ущербность журнальной политики РАН, по сей день ориентирующейся на платный доступ к полным текстам статей. Проиллюстрирую это на примере судьбы трех своих недавних статей на родственные темы.

Первая статья [13] вышла в 2019 году в академическом журнале «Программирование». За полгода пребывания в платном доступе по цене 220 руб. за скачивание ее приобрели два читателя. К моменту появления коронавируса статья уже отчасти устарела, и за время предоставленного открытого доступа ее

прочитали лишь 16 человек. Далее к статье по непонятным причинам и вовсе закрыли доступ, сейчас ее нельзя получить даже за деньги.

Вторая статья [14] вышла в журнале «Вестник РАН» в 2020 году, накануне появления коронавируса. Тем самым бесплатный доступ к ней был открыт практически сразу, и ее успели бесплатно прочитать 50 человек. Сейчас статья доступна к прочтению по цене 253 руб., и, разумеется, число прочитавших ее практически не растет.

Наконец, третья (хронологически первая) статья [15] вышла в 2018 году в издании «Троицкий вариант – наука», не принадлежащем РАН, и была размещена в открытом доступе. Ее судьбу можно уверенно отнести к благополучным – 190 тыс. посещений.

Почему не читают журналы, издаваемые РАН? Дело, конечно, не только и не столько в нежелании расстаться с 250 руб. – сумма не такая уж и страшная. Отпугивает процесс перечисления денег, ведь по ходу придется еще и сообщать какие-нибудь свои персональные данные. Отметим также, что и во время бесплатного доступа перед читателем возникало множество препятствий. Мешали, в частности, не только вялое оповещение о кратковременном открытии доступа, но и обязательность регистрации, а также невозможность сохранить прямую ссылку на полный текст статьи. Таким образом, вполне объяснимо то, что число читателей, сумевших добраться к полным текстам статей в академических изданиях, оказалось в 10 000 раз или в 100 000 раз меньше, чем прочитавших подобную статью в открытом доступе.

Остаются непонятными приоритеты РАН. Недоступность «Программирования» еще можно попытаться объяснить привычным для РАН пренебрежительным отношением к IT-отрасли. Но «Вестник РАН» — флагманское издание Академии. Для чего потребовалось сделать его материалы практически недоступными для массового читателя? В полном недоумении оказываешься, если вспомнить, что все это происходит на фоне регулярных заседаний Президиума РАН, посвященных популяризации науки. К сожалению, и дальнейшая судьба открытого доступа к академическим изданиям внушает серьезные опасения: в академических верхах продолжают не замечать существование европейского плана "S".

А ведь популяризация научных результатов широко востребована в нашей стране. КиберЛенинка в 2019 году показала взрывной рост аудитории, которая достигла 52 млн человек — на 80% больше, чем годом ранее. Всего за год пользователями КиберЛенинки было прочитано 250 млн статей, что говорит о массовом запросе на свежее общедоступное научное знание [16]. Чрезвычайно досадно, что издания РАН по сей день упорно не используют главный козырь, определивший успех КиберЛенинки, — открытый доступ к полным текстам научных публикаций.

ЗАЩИТА ДИССЕРТАЦИЙ В ОНЛАЙН

И еще одно событие в научном мире, вызванное COVID-19. В мае вышло постановление правительства РФ о порядке проведения защиты диссертаций в период пандемии [17]. На это время разрешается проведение защиты в режиме онлайн, где до двух третей членов диссертационного совета участвуют в заседании удаленно. Голосование в таком случае проводится открытое: как всегда, IT-сектор в науке отстает — ведь в стране уже имеется успешный опыт проведения ответственного онлайн-ого тайного голосования по проекту внесения изменений в конституцию РФ.

Конечно, если часть членов диссертационного совета тайно голосует очно, а часть дистанционно, то во время подсчета голосов становятся известными предпочтения и той, и другой частей, что уже только с некоторой натяжкой можно было бы отнести к классическому тайному голосованию. Однако никто не мешает всем, включая очных участников, тайно проголосовать, скажем, посредством механизма, предоставляемого Zoom [18], и такое тайное голосование уже ни в чем не уступает очному. Непонятно, почему чиновники не решились легализовать такую несложную схему тайного голосования, а вынудили диссертационные советы столь основательно поломать ключевую процедуру проведения защиты.

eLIBRARY

Несколько событий прошедшего года в сфере научных публикаций связано с eLibrary. Осенью 2019 года рецензируемые издания, индексируемые в РИНЦ,

получили сообщение об открывшейся возможности передачи рецензий в eLibrary. Изданию была предоставлена свобода выбора режима: по его решению передаваемые рецензии могли публиковаться или не публиковаться в eLibrary с указанием или без указания рецензента. Открытые рецензии — мощный современный тренд, и, если инициатива eLibrary в такой форме приживется на российской почве, мир российских научных публикаций станет богаче, интереснее.

СПИСОК СТАТЕЙ, ЦИТИРУЮЩИХ ДАННУЮ ПУБЛИКАЦИЮ

Цитируемая публикация: **ИНТЕРНЕТ-АКТИВНОСТЬ КАК ОБЯЗАННОСТЬ УЧЕНОГО**
Горбунов-Посадов М.М.
Информационные технологии и вычислительные системы. 2007. № 3. С. 88-93.

Журнал: _____

Автор: _____ Год: _____ Название: _____

Сортировка: по числу цитирований Порядок: по убыванию Очистить Выборка

	Публикация	Цит.
1	КИБЕРСОЦИАЛИЗАЦИЯ ЧЕЛОВЕКА. ОТ НОМО SAPIENSA ДО НОМО CYBERUSA Плешаков В.А. Москва, 2012.	101
2	УПРАВЛЕНИЕ ЖИЗНЕННЫМ ЦИКЛОМ ЭЛЕКТРОННЫХ ПУБЛИКАЦИЙ В ИНФОРМАЦИОННОЙ СИСТЕМЕ НАУЧНОГО ЖУРНАЛА Елизаров А.М., Зуев Д.С., Липачёв Е.К. Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2014. № 4. С. 81-88.	13
3	ИНФОРМАЦИОННЫЕ СИСТЕМЫ И СЕРВИСЫ КОМПЛЕКСНОЙ ПОДДЕРЖКИ ПЕРИОДИЧЕСКИХ НАУЧНЫХ ИЗДАНИИ Ахметов Д.Ю., Елизаров А.М., Липачев Е.К. В сборнике: Научный сервис в сети Интернет. труды XVII Всероссийской научной конференции. ИПМ им. М.В.Келдыша. 2015. С. 16-25.	11
4	ОЦЕНИВАНИЕ И РАНЖИРОВАНИЕ ВЕБ-САЙТОВ. ВЕБОМЕТРИЧЕСКИЕ РЕЙТИНГИ Поляк Ю.Е. Научный редактор и издатель. 2017. Т. 2. № 1. С. 19-29.	10
5	НАУЧНЫЕ ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ: АКТУАЛЬНЫЕ ЗАДАЧИ И СОВРЕМЕННЫЕ ПУТИ ИХ РЕШЕНИЯ Семячкин Д.А., Кисляк Е.В., Сергеев М.А. Научная периодика: проблемы и решения. 2013. № 2 (14). С. 20-29.	9
6	ИНФОРМАЦИОННОЕ ПРОСТРАНСТВО РОССИЙСКОЙ АКАДЕМИЧЕСКОЙ НАУКИ КАК ИНТЕГРИРОВАННАЯ ПРОФЕССИОНАЛЬНО-ИНТЕЛЛЕКТУАЛЬНАЯ СРЕДА: ПЕРСПЕКТИВЫ ФОРМИРОВАНИЯ И УПРАВЛЕНИЯ Соколова М.Е. В сборнике: Концепция "общества знания" в современной социальной теории. Сб. науч. тр. Сер. "Теория и история социологии" Центр социал. науч.-информ. исслед. Отд. социологии и социал. психологии; Отв. ред. Д.В. Ефременко. МОСКВА, 2010. С. 134-158.	8

Рисунок 1. Список цитирующих публикаций в eLibrary

Затем eLibrary объявила о появлении Application Programming Interface (API) [19] — интерфейса программирования приложений, предназначенного для реализации автоматических запросов и извлечения информации из базы данных РИНЦ. К сожалению, возможности API оказались небогатыми: с его помощью

можно узнать только библиографические показатели отдельного автора, отдельного журнала и отдельной публикации.

В библиографические показатели публикации, доступные посредством API, не попал список работ, цитирующих эту публикацию, имеется только общее их количество. А ведь такой список — один из наиболее востребованных читателями видов библиографической информации, он позволяет прицельно получить самую свежую информацию о рассматриваемом научном направлении. Список этот eLibrary формирует, он доступен на соответствующей странице сайта для авторизованного посетителя eLibrary (рис. 1), но в API список почему-то не вошел.

Это обстоятельство заметно обедняет содержание публикационной карточки российской научной статьи. Сейчас там приходится ограничиваться информацией, предоставляемой зарубежными библиографическими базами.

Так, например, если в карточку статьи включена гиперссылка вида

<https://scholar.google.com/scholar?q=link:https://keldysh.ru/gorbunov/duty.htm>

где часть, записываемая вслед за префиксом "link:", — адрес (URL) цитируемой статьи, то у читателя появляется возможность в один клик перейти на список публикаций, цитирующих данную статью, согласно данным Google Scholar (рис. 2). Аналогичная конструкция существует и для библиографической базы Web of Science, но там надо указать в качестве параметра не URL, а номер ("Accession Number"), присвоенный статье в этой базе. Несколько сложнее, но все же можно составить (посредством специального API) такой список для библиографической базы Crossref. И только eLibrary почему-то эту полезную возможность в свой открывшийся API не включила.

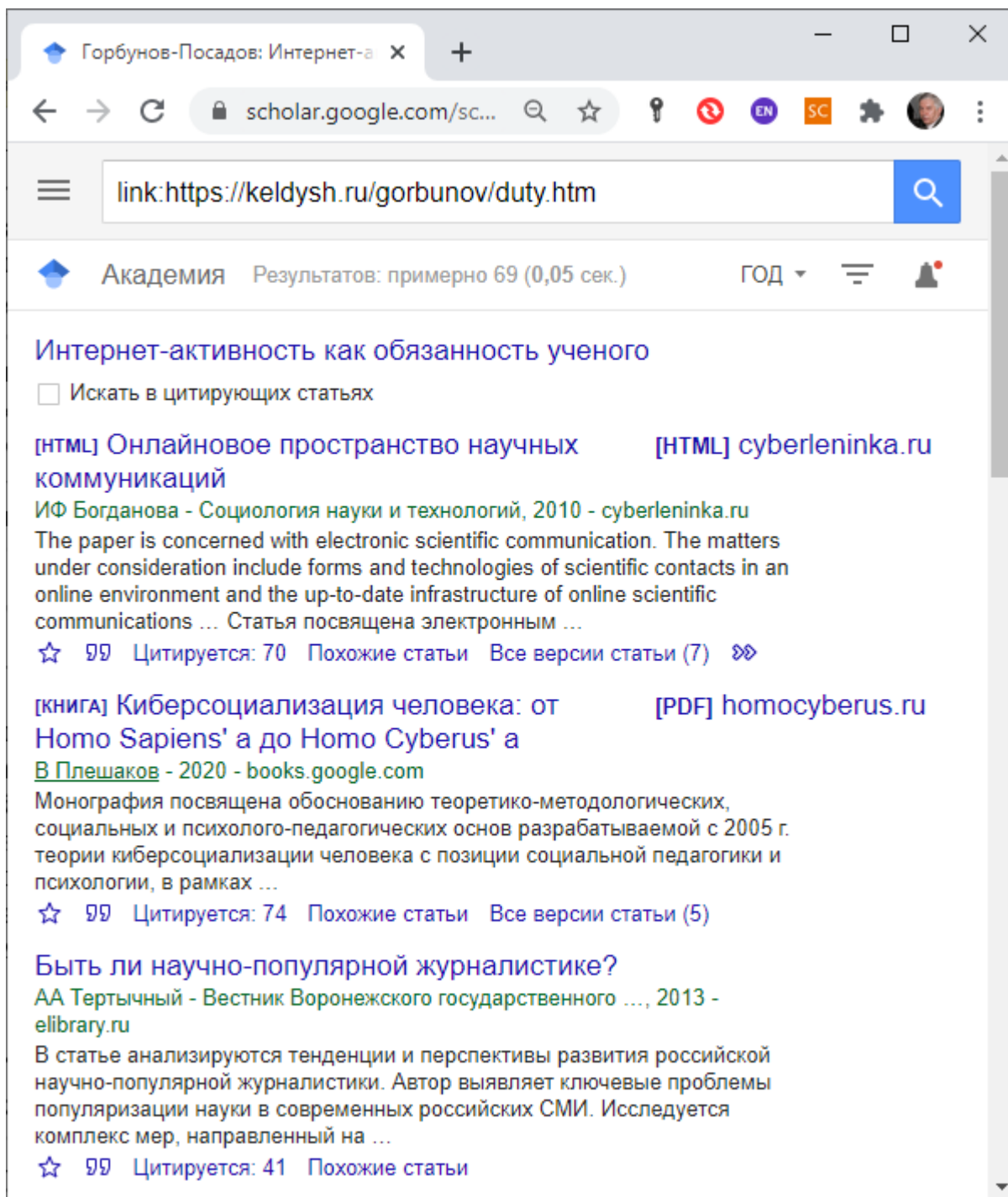


Рисунок 2. Генерируемый Google Scholar список публикаций, цитирующих указанную статью

КОМПЛЕКСНЫЙ БАЛЛ ПУБЛИКАЦИОННОЙ РЕЗУЛЬТАТИВНОСТИ

И, наконец, несколько слов о наделавшем много шума [20] письме [21] Минобрнауки «О методике расчета комплексного балла публикационной ре-

зультативности» (КБПР). Наиболее существенны два аргумента многочисленных критиков этой методики.

Во-первых, методика приводит к удушению российских научных журналов, в силу очевидной инерции имеющих не соответствующие их реальному научному весу позиции в рейтинге Web of Science. Наиболее сильные авторы вынуждены отказываться от публикации в таких журналах из-за мизерных баллов КБПР, однако без сильных статей журналам никогда не пробиться в гранды рейтинга. А ведь помочь российским журналам встать на ноги можно легко: достаточно было бы, сохранив действующую методику подсчета КБПР, считать квартиль Q российского журнала по формуле $Q=W-2$, где W — квартиль журнала в рейтинге Web of Science. Однако предложение даже более слабой формулы $Q=W-1$ поддержки чиновников, принимающих решения, не получило.

Во-вторых, оценка монографии, приравнивающая ее к публикации статьи в рядовом журнале из Scopus, звучит просто нелепо и унижительно для серьезного ученого. К сожалению, российские чиновники уже много лет третируют монографии, либо вообще не признавая их в качестве научного результата, либо начисляя за монографию ничтожные баллы, никак не отражающие ни труд, затраченный на ее написание, ни приносимую ею пользу науке. Результаты такой политики плачевны. Так, например, сотрудники ИПМ им. М.В. Келдыша РАН в 2006 году выпустили более 20 монографий, а в 2019 году — лишь две.

Существует и еще один, редко упоминаемый аргумент критиков КБПР. Самые актуальные направления науки время от времени перестают довольствоваться сложившимся сообществом журналов, и тогда создается новый журнал, куда сразу устремляются наиболее мощные работы в новом актуальном направлении и куда в первую очередь будет обращен читательский интерес. Однако российский ученый оказывается отрезан от этого чрезвычайно востребованного потока, от такого журнала: ведь формальные высокие позиции в рейтинге WoS, требуемые Министерством, журнал, согласно существующим правилам, завоеует лишь через несколько лет, и все эти годы КБПР будет бездумно-механически мешать ему там публиковаться.

Благодарности

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 19-01-00069 А.

СПИСОК ЛИТЕРАТУРЫ

1. *P. Smart*. Publishing during pandemic: Innovation, collaboration, and change // *Learned publishing*. 2020. V. 33, No. 3. P. 194–197.
URL: <https://doi.org/10.1002/leap.1314>
2. *Отношение к препринтам давно и устойчиво теплеет*. URL: <https://indicator.ru/humanitarian-science/springer-nature-preprint-16-05-2019.htm>
3. *In Review. Journal-integrated preprint sharing from Springer Nature and Research Square*. URL: <https://www.springernature.com/gp/authors/campaigns/in-review>
4. *Peter Coles*. The Age of Preprints. September 23, 2020.
URL: <https://astro.theoj.org/post/674-the-age-of-preprints>
5. Research Square. An expanded suite of services and tools to help promote your research. URL: <https://www.researchsquare.com/>
6. Under Review. Preprint your article with Under Review, Powered by Authorea. URL: <https://authorservices.wiley.com/author-resources/Journal-Authors/submission-peer-review/wiley-under-review.html>
7. First Look program to share your newest latest research with a wider audience faster than ever before. URL: <https://www.elsevier.com/solutions/ssrn/institutional-value/first-look>
8. TechRxiv preprint server. URL: Powered by IEEE. <https://www.techrxiv.org/>
9. Preprints are go at Crossref!
URL: <https://www.crossref.org/blog/preprints-are-go-at-crossref/>
10. *Судьба плана S открытого доступа*.
URL: <https://indicator.ru/humanitarian-science/open-access-bitva.htm>
11. *План S отодвинут с 2020 на 2021 год, но и это под большим вопросом*. URL: <https://indicator.ru/humanitarian-science/realizaciyu-plan-s-reshili-otlozhit-31-05-2019.htm>

12. *О свободном бесплатном доступе к журналам РАН.* URL: <http://www.ras.ru/news/shownews.aspx?id=a18fbb23-d2cf-430e-9df6-05ffa9fe31ce>

13. *М.М. Горбунов-Посадов, Т.А. Полилова.* Инструменты поддержки онлайн-публикации // Программирование. 2019. № 3. С. 38–42. URL: <https://doi.org/10.1134/S013234741903004X>

14. *М.М. Горбунов-Посадов.* Научная публикация в России: для чего и как? // Вестник РАН. 2020. Т. 90. № 1. С. 35–39.

URL: <https://doi.org/10.31857/S0869587320010053>.

URL: <https://sciencejournals.ru/view-article/?j=vestnik&y=2020&v=90&n=1&a=Vestnik2001005GorbunovPosado>

15. *М.М. Горбунов-Посадов.* Цифровая наука в РАН // Троицкий вариант — наука. 2018. № 5. С. 14. URL: <https://trv-science.ru/2018/03/13/cifrovaya-nauka-v-ran/>

16. *Популяризация науки показывает взрывной рост* // Ассоциация «Открытая наука». 18/02/2020. URL: <https://open-science.ru/2020/02/popularization-of-science-shows-high-growth.html#more-1160>

17. *Об особенностях проведения заседаний советов.* URL: <http://publication.pravo.gov.ru/Document/View/0001202005270020?index=0&rangeSize=1>

18. Video Conferencing, Web Conferencing, Webinars, Screen Sharing - Zoom. URL: <https://zoom.us/>

19. *Интерфейс программирования приложений API.*

URL: https://www.elibrary.ru/projects/api/api_info.asp

20. *Обсуждение Методики расчета комплексного балла публикационной результативности (КБПР).*

URL: <http://www.sib-science.info/ru/ras/predlozhennaya-09022020>

21. *О методике расчета комплексного балла публикационной результативности (КБПР).*

URL: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=EXP&n=421632&dst=100001#025803375598292577>

RUSSIAN SCIENTIFIC PUBLICATION FOR THE COVID-19 PERIOD

M. M. Gorbunov-Posadov¹

¹*Keldysh Institute of Applied Mathematics, Moscow*

¹gorbunov@keldysh.ru

Abstract

The impact of the COVID-19 pandemic on the world of scientific publications: rapid publication and simplified access for articles about the virus, open access for a time of self-isolation in the world, and in Russia. There is an impressive difference between the number of readers for Russian scientific articles in open access and in paid access. The policy of journals of the Russian Academy of Sciences. Impressive growth in readership of CyberLeninka. Online meeting of the dissertation council. eLibrary news. Comprehensive publication activity score from the Ministry of education and science.

Keywords: *scientific publication, COVID-19, open access*

REFERENCES

1. P. Smart. Publishing during pandemic: Innovation, collaboration, and change // *Learned publishing*. 2020. Volume 33, Issue 3. P. 194–197.
URL: <https://doi.org/10.1002/leap.1314>
2. *Otnoshenie k preprintam davno i ustoichivo tepleet*. URL: <https://indicator.ru/humanitarian-science/springer-nature-preprint-16-05-2019.htm>
3. *In Review. Journal-integrated preprint sharing from Springer Nature and Research Square*. URL: <https://www.springernature.com/gp/authors/campaigns/in-review>
4. Peter Coles. The Age of Preprints. September 23, 2020.
URL: <https://astro.theoj.org/post/674-the-age-of-preprints>
5. *Research Square. An expanded suite of services and tools to help promote your research*. URL: <https://www.researchsquare.com/>
6. *Under Review. Preprint your article with Under Review, Powered by Authorea*.

URL: <https://authorservices.wiley.com/author-resources/Journal-Authors/submission-peer-review/wiley-under-review.html>

7. *First Look program to share your newest latest research with a wider audience faster than ever before.*

URL: <https://www.elsevier.com/solutions/ssrn/institutional-value/first-look>

8. *TechRxiv preprint server.* URL: Powered by IEEE. <https://www.techrxiv.org/>

9. *Preprints are go at Crossref!*

URL: <https://www.crossref.org/blog/preprints-are-go-at-crossref/>

10. *Sudba plana S otkrytogo dostupa.* URL: <https://indicator.ru/humanitarian-science/open-access-bitva.htm>

11. *Plan S otodvinut s 2020 na 2021 god, no i eto pod bolshim voprosom.* URL: <https://indicator.ru/humanitarian-science/realizaciyu-plan-s-reshili-otlozhit-31-05-2019.htm>

12. *O svobodnom besplatnom dostupe k zhurnalom RAN.* URL: <http://www.ras.ru/news/shownews.aspx?id=a18fbb23-d2cf-430e-9df6-05ffa9fe31ce>

13. *M.M. Gorbunov-Posadov, T.A. Polilova. Instrumenty podderzhki onlainovoi nauchnoi publikatsii // Programmirovaniye. 2019. № 3. S. 38–42.* URL: <https://doi.org/10.1134/S013234741903004X>

14. *M.M. Gorbunov-Posadov. Nauchnaia publikatsiia v Rossii: dlia chego i kak? // Vestnik RAN. 2020. T. 90. № 1. S. 35–39*

URL: <https://doi.org/10.31857/S0869587320010053>

<https://sciencejournals.ru/view-article/?j=vestnik&y=2020&v=90&n=1&a=Vestnik2001005GorbunovPosado>

15. *M.M. Gorbunov-Posadov. Tsifrovaia nauka v RAN // Troitskii variant – nauka. 2018. № 5. S. 14.* URL: <https://trv-science.ru/2018/03/13/cifrovaya-nauka-v-ran/>

16. *Populiarizatsiia nauki pokazyvaet vzryvnoi rost // Assotsiatsiia «Otkrytaia nauka». 18/02/2020.* URL: <https://open-science.ru/2020/02/popularization-of-science-shows-high-growth.html#more-1160>

17. *Ob osobennostiakh provedeniia zasedaniy sovetov.*

URL:

<http://publication.pravo.gov.ru/Document/View/0001202005270020?index=0&rangeSize=1>

18. *Video Conferencing, Web Conferencing, Webinars, Screen Sharing – Zoom.*

URL: <https://zoom.us/>

19. *Interfeis programmirovaniia prilozhenii API.*

URL: https://www.elibrary.ru/projects/api/api_info.asp

20. *Obsuzhdenie Metodiki rascheta kompleksnogo balla publikatsionnoi rezultativnosti (KBPR).*

URL: <http://www.sib-science.info/ru/ras/predlozhennaya-09022020>

21. *O metodike rascheta kompleksnogo balla publikatsionnoi rezultativnosti (KBPR).*

URL: <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=EXP&n=421632&dst=100001#025803375598292577>

СВЕДЕНИЯ ОБ АВТОРЕ



ГОРБУНОВ-ПОСАДОВ Михаил Михайлович – главный научный сотрудник ИПМ им. М.В. Келдыша РАН, д. ф.-м. н.

Mikhail Mikhailovich GORBUNOV-POSADOV – Keldysh Institute of Applied Mathematics, chief researcher.

gorbunov@keldysh.ru

ORCID: 0000-0002-7044-8287

Статья поступила в редакцию 21 ноября 2020 года

УДК 004

НЕКОТОРЫЕ ИНСТРУМЕНТЫ ДЛЯ РАБОТЫ С БИБЛИОГРАФИЧЕСКИМИ ССЫЛКАМИ

А. В. Ермаков^[0000-0002-6054-0813]

*Институт прикладной математики им. М.В. Келдыша
Российской академии наук, г. Москва*

Ermakov@Keldysh.ru

Аннотация

Быстрый количественный рост научных публикаций подтолкнул развитие различных интернет-инструментов для поиска необходимой информации. Не меньшее значение, чем поиск необходимой информации, имеет статистика по ее изучению и использованию (цитированию).

В данной статье рассматриваются инструменты для поддержки цитирования, предоставляемые Google-академией, Crossref и другими разработчиками.

В Российской Федерации НЭИКОН (национальный консорциум российских организаций образования, науки, и культуры) оказывает техническую и методическую поддержку по работе с DOI и другими сервисами Crossref.

В статье показано, как рассматриваемые инструменты встраиваются в интерфейсы пользователей электронной библиотеки ИПМ. Рассматриваются дополнительные возможности, предоставляемые указанными инструментами.

Ключевые слова: *электронная библиотека, библиографическая ссылка, обратная библиографическая ссылка, набор инструментов.*

ВВЕДЕНИЕ

Стремительное развитие интернета за последние 20 лет привело к значительному росту накопленной научно-технической информации, резкому увеличению количества электронных библиотек и их объема.

Быстрый количественный рост научных публикаций подтолкнул развитие различных интернет-инструментов для поиска необходимой информации (например, Google-академия). Следует отметить, что Google-академия сотрудни-

чает с издателями, чтобы индексировать рецензируемые научные статьи и другие академические материалы, делая их доступными для поиска в Google.

Поисковая база Google-Академии включает журналы, материалы конференций, авторефераты, диссертации, научные книги, препринты, технические отчеты и другую научную литературу из различных областей знаний. В Google-Академии можно найти публикации многих научных издательств, профессиональных сообществ, а также исследовательские статьи из различных интернет-источников.

С точки зрения реальных показателей цитируемости для русскоязычных авторов, Google-Академия представляет возможно больший интерес, чем WebofScience или Scopus, поскольку в него включено максимальное количество научных журналов на русском языке.

Не меньшее значение, чем поиск необходимой информации, имеет статистика по ее изучению и использованию (цитированию) другими читателями. Эти данные безусловно интересны и самому автору научной публикации, и пользователям сети, заинтересовавшимся статьей.

Одним из разработчиков специализированных инструментов для поддержки цитирования стала ассоциация Crossref, занимающаяся разработкой и поддержкой всемирной высокотехнологичной инфраструктуры научных коммуникаций.

Использованию инструментов, предоставляемых Google-академией и Crossref, для развития технологических возможностей электронной библиотеки ИПМ им. М.В. Келдыша посвящена данная статья.

АССОЦИАЦИЯ CROSSREF

Развитие электронных библиотек как источника информации неразрывно связано с историей Интернет. Особое место в ряду электронных библиотек занимают научно-технические и научно-образовательные библиотеки. Мировой опыт показывает, что образовательные и научные электронные ресурсы формируются, как правило, самими научными и образовательными организациями, в которых работают авторы этих публикаций, и ориентированы они прежде всего на читате-

лей, также тесно связанных с научной и образовательной деятельностью (научных сотрудников, преподавателей, аспирантов, студентов).

Работа научных сотрудников в рамках определенного научного направления предполагает активное изучение публикаций других авторов в данной тематике и легальное использование чужих материалов (цитирование) при обсуждении в рамках журнальных статей, научных докладов и т. п.

Ассоциация Crossref [1], членом которой с 2016 г. является ИПМ им. М.В. Келдыша, поддерживает совместную всемирную службу взаимной цитируемости, функционирующую как своеобразный шлюз между электронными платформами издателей. Эта служба не хранит полные тексты научных публикаций, но она заносит в свою базу данных информацию о связи публикаций с помощью технологии DigitalObjectIdentifier (DOI), а также метаданные опубликованных научных материалов.

Разрабатываемые инструменты Crossref (и некоторых других организаций, таких как GoogleScholar, Scopus и WebofScience, которые используют различные источники для своих данных цитирования) облегчают как автору публикации, так и читателям, поиск, цитирование, оценку, повторное использование результатов научных исследований.

ПОЛУЧИТЬ ПОСТОЯННЫЕ ССЫЛКИ

Библиографическая ссылка – это совокупность библиографических сведений о цитируемом, рассматриваемом или упоминаемом в тексте документе, необходимых и достаточных для общей характеристики, идентификации и поиска документа. От того, насколько точно автор публикации указал в ссылке информацию о цитируемом документе, зависит возможность потенциального читателя найти этот документ и ознакомиться с ним [6, 10].

И первый инструмент Crossref, с которым мы хотели бы познакомить читателей – «получить постоянные ссылки»:

<https://doi.crossref.org/simpleTextQuery>

← → ↻ <https://doi.crossref.org/simpleTextQuery>

← [Back to the main Crossref website](#)

Crossref

We enhanced Simple Text Query - as you requested BLOG POST

Get persistent links for your reference list or bibliography.
Copy and paste the list, we'll match with our metadata and return the links.
Members may also [deposit reference lists](#) here too.

Enter text in the box below:

Include PubMed IDs in results. List all possible DOIs per reference.

We now provide space to match 1,000 references per submission
To learn about other services and interfaces see our [MetaData Delivery](#) page.
Please [contact us](#) if you have any questions.
Tips:

Рисунок 1. Шаблон Crossref для поиска ссылок DOI.

Если скопировать и вставить в предлагаемое окно список литературы, то Crossref попытается сопоставить указанным публикациям их метаданные и вернет ссылки с DOI.

В качестве примера мы возьмем первые 4 элемента списка литературы из препринта ИПМ № 16 (2020 г.) [8]:

1. Иванов Д.С. и др. Лабораторный стенд для моделирования движения микроспутников // Известия РАН. Теория и системы управления. 2018. № 1. С. 125–140.
2. Ivanov D. et al. Determination of disturbances acting on small satellite mock-up on air bearing table // Acta Astronaut. 2018. Vol. 142, p. 265–276.
3. Биндель Д. и др. Лабораторный стенд для верификации алгоритмов управления группировкой спутников // Известия РАН. Теория и системы управления. 2009. Vol. 48, № 5. С. 109–117.

4. Ovchinnikov M.Y. et al. Development, integrated investigation, laboratory and in-flight testing of Chibis-M microsatellite ADCS // *Acta Astronaut.* 2014. Vol. 93. P. 23–33.

Выполнив указанные выше действия, получим:

1. Иванов Д.С. и др. Лабораторный стенд для моделирования движения микроспутников // *Известия РАН. Теория и системы управления.* 2018. № 1. С. 125–140. <https://doi.org/10.7868/S0002338818010122>

2. Ivanov D. et al. Determination of disturbances acting on small satellite mock-up on air bearing table // *Acta Astronaut.* 2018. Vol. 142, p. 265–276. <https://doi.org/10.1016/j.actaastro.2017.11.010>

3. Биндель Д. и др. Лабораторный стенд для верификации алгоритмов управления группировкой спутников // *Известия РАН. Теория и системы управления.* 2009. Vol. 48, № 5. С. 109–117.

4. Ovchinnikov M.Y. et al. Development, integrated investigation, laboratory and in-flight testing of Chibis-M microsatellite ADCS // *Acta Astronaut.* 2014. Vol. 93. P. 23–33. <https://doi.org/10.1016/j.actaastro.2013.06.030>

Следует отметить, что если бы мы использовали транслитерацию (как это обычно делается для зарубежных изданий), то не получили бы ссылку для первой позиции этого списка.

Таким образом, дополнив свой список литературы ссылками DOI, автор статьи может, во-первых, сам убедиться в корректности этих ссылок, а, во-вторых, поможет читателю найти за один клик цитируемую статью, если она его заинтересовала.

Есть еще несколько инструментов, для получения постоянных ссылок, например: <https://search.crossref.org/references>.

При реализации этих инструментов рабочей группой Crossref разработаны и протестированы четыре подхода (алгоритма) поиска DOI, индексирующего данную библиографическую ссылку:

- традиционный подход, основанный на синтаксическом анализе ссылок;
- алгоритм SBM (BasedMatchingwithValidation) с простым порогом, который ищет ссылочную строку в поисковой системе и возвращает первое совпаде-

ние, у которого оценка релевантности превышает предварительно определенный порог;

- алгоритм SBM с нормализованным порогом, который выполняет поиск эталонной строки в поисковой системе и возвращает первое совпадение, если его оценка релевантности, деленная на длину строки, превышает предварительно определенный порог;
- алгоритм SBMV (Search-Based Matching with Validation), который сначала применяет алгоритм SBM с нормализованным порогом для выбора ряда элементов-кандидатов, а для выбора окончательного целевого элемента используется отдельная процедура проверки.

В различных инструментах Crossref (при загрузке контента, при определении ссылки и т. д.) в зависимости от требований простоты-сложности и оперативности-точности используется тот или иной алгоритм. Но в любом случае, автору или сотруднику издательства рекомендуется самостоятельно проверить корректность результата распознавания библиографических записей и точность полученной ссылки.

ОБРАТНАЯ БИБЛИОГРАФИЧЕСКАЯ ССЫЛКА

Обратная библиографическая ссылка – это библиографическая ссылка на публикацию, ссылающуюся на данную статью. Инструменты для построения обратного библиографического списка разрабатывает не только Crossref, но и другие. Например, Google Академия (Googlescholar). В соответствующем инструменте Google достаточно указать в качестве параметра запроса URL интересующей вас статьи и можно получить список ссылающихся на нее публикаций.

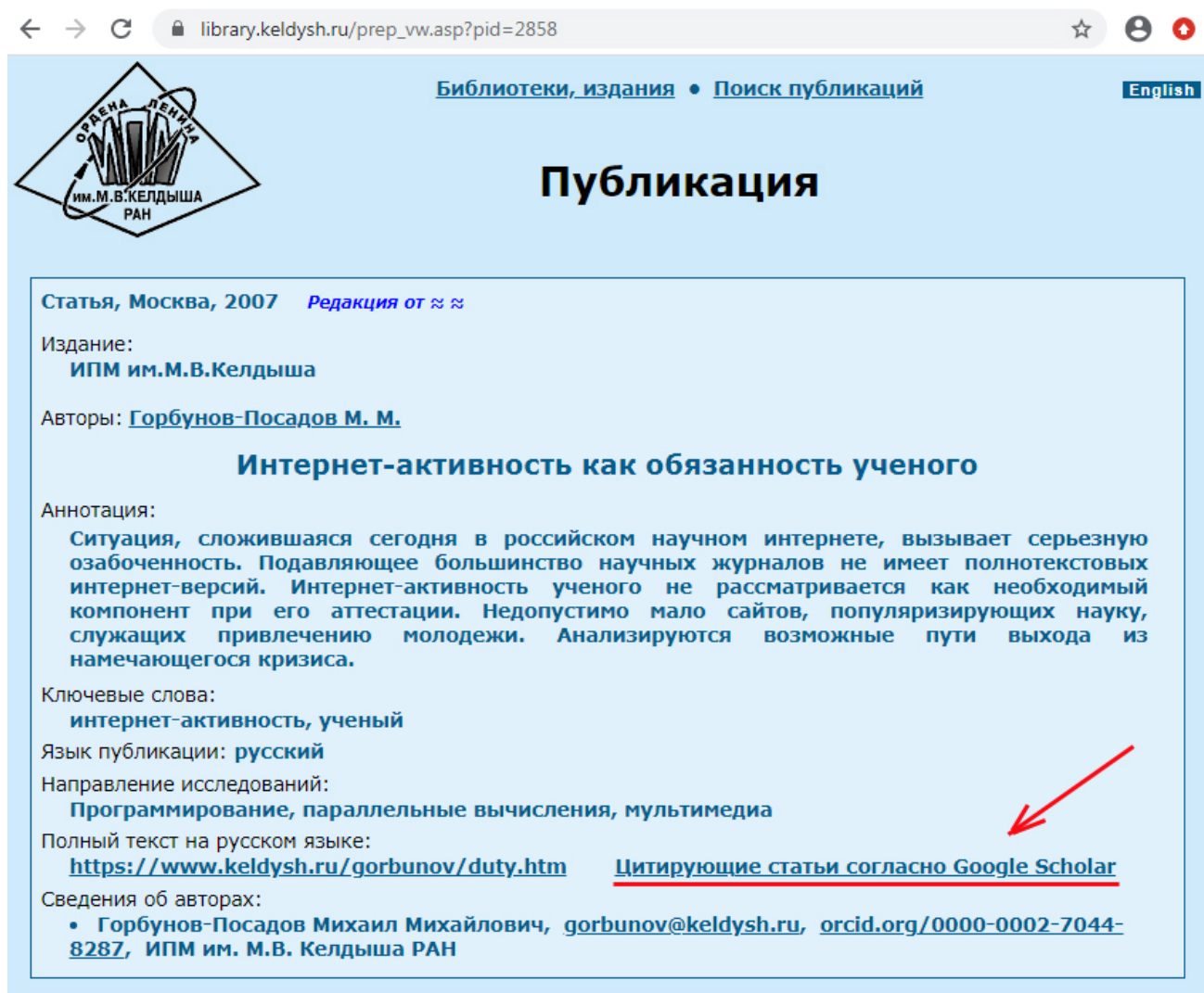
Например, на статью М.М. Горбунова-Посадова «Интернет-активность как обязанность ученого» (<https://keldysh.ru/gorbunov/duty.htm>) ссылаются 69 публикаций (рис. 3), список которых можно получить с помощью запроса

<https://scholar.google.com/scholar?q=link:https://www.keldysh.ru/gorbunov/duty.htm>

The screenshot shows a Google Scholar search interface. The search bar contains the URL `link:https://www.keldysh.ru/gorbulnov/duty.htm`. The search results are filtered to show articles. The first result is titled "Интернет-активность как обязанность ученого" by ИФ Богданова, published in 2010. The second result is a book titled "Киберсоциализация человека: от Homo Sapiens' a до Homo Cyberus' a" by В Плешаков, published in 2020. The third result is an article titled "Быть ли научно-популярной журналистике?" by АА Тертыйный, published in 2013. The fourth result is a book titled "Управление жизненным циклом электронных публикаций в информационной системе научного журнала" by АМ Елизаров, ДС Зуев, and ЕК Липачёв, published in 2014. The interface includes filters for time range (from 2020 to 2016), relevance, date, and options to include patents and show quotes.

Рисунок 2. Список обратного цитирования согласно GoogleScholar.

Сотрудниками ИПМ, занимающимися поддержкой и развитием функциональных возможностей электронной библиотеки ИПМ [7–9], эта возможность – получение обратного библиографического списка с помощью GoogleScholar – встроена в электронную библиографическую карточку публикации (рис. 3).



The screenshot shows a web browser window with the URL library.keldysh.ru/prep_vw.asp?pid=2858. The page header includes the logo of the Institute of Problems in Mechanics (IPM) named after M.V. Keldysh, the text "Библиотеки, издания • Поиск публикаций", and an "English" language selector. The main heading is "Публикация".

The article information is as follows:

- Статья, Москва, 2007 Редакция от ≈ ≈
- Издание: ИПМ им.М.В.Келдыша
- Авторы: [Горбунов-Посадов М. М.](#)

Интернет-активность как обязанность ученого

Аннотация:
Ситуация, сложившаяся сегодня в российском научном интернете, вызывает серьезную озабоченность. Подавляющее большинство научных журналов не имеет полнотекстовых интернет-версий. Интернет-активность ученого не рассматривается как необходимый компонент при его аттестации. Недопустимо мало сайтов, популяризирующих науку, служащих привлечению молодежи. Анализируются возможные пути выхода из намечающегося кризиса.

Ключевые слова:
интернет-активность, ученый

Язык публикации: **русский**

Направление исследований:
Программирование, параллельные вычисления, мультимедиа

Полный текст на русском языке:
<https://www.keldysh.ru/gorbunov/duty.htm> [Цитирующие статьи согласно Google Scholar](#)

Сведения об авторах:
• Горбунов-Посадов Михаил Михайлович, gorbunov@keldysh.ru, orcid.org/0000-0002-7044-8287, ИПМ им. М.В. Келдыша РАН

A red arrow points to the "Цитирующие статьи согласно Google Scholar" link.

Рисунок 3. Ссылка для обратного библиографического списка.

Один из инструментов Crossref, позволяющий строить обратный библиографический список (Citedby), называется getForwardLinks и имеет формат:

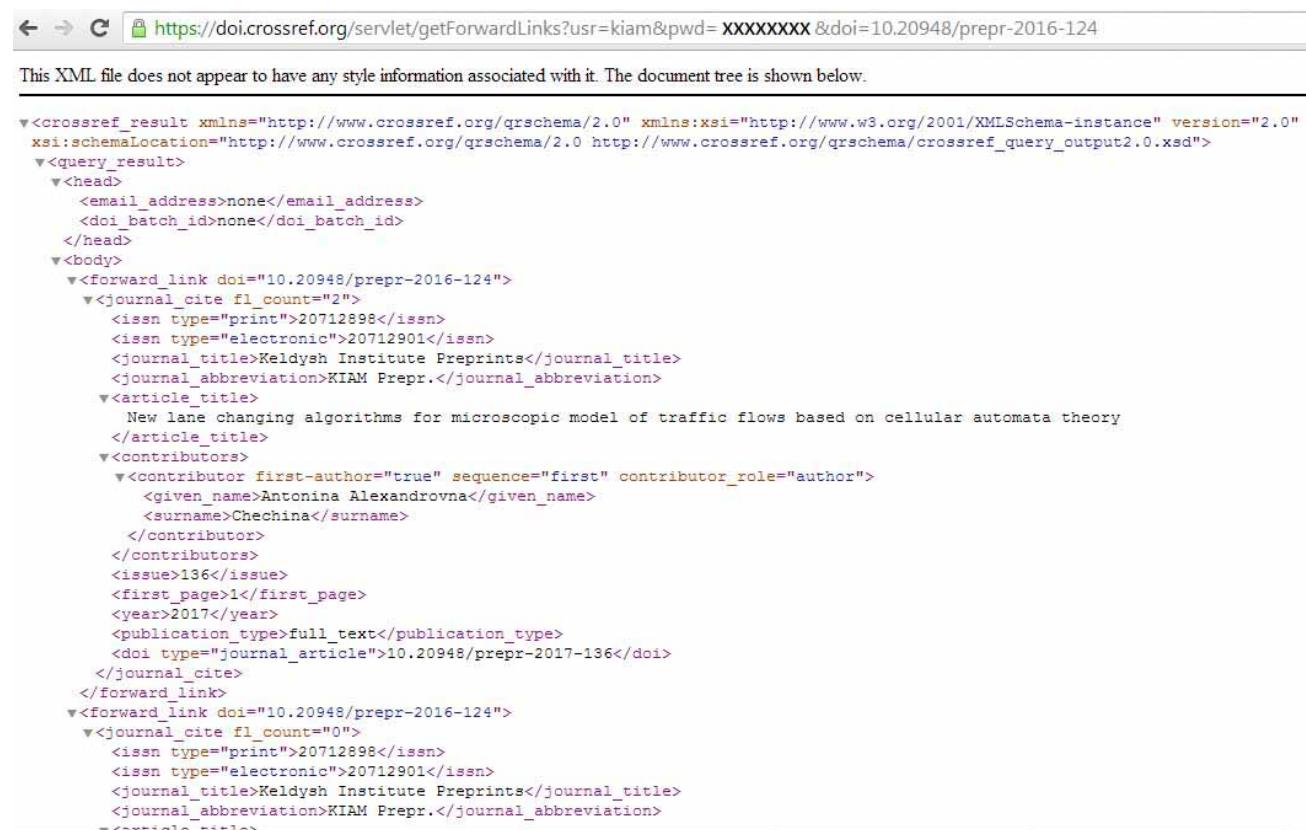
[https://doi.crossref.org/servlet/getForwardLinks?usr=username&pwd=password&doi=\[DOI\]](https://doi.crossref.org/servlet/getForwardLinks?usr=username&pwd=password&doi=[DOI])

где указываются логин и пароль издателя (логин и пароль в базе данных Crossref), а также DOI интересующей публикации.

Например:

<https://doi.crossref.org/servlet/getForwardLinks?usr=kiam&pwd=XXX&doi=10.20948/prepr-2016-124>

Результат запроса возвращается в XML-формате. Для вышеуказанного примера:



```
<?xml version="1.0" encoding="UTF-8" standalone="no" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.crossref.org/qrschema/2.0" xsi:schemaLocation="http://www.crossref.org/qrschema/2.0 http://www.crossref.org/qrschema/crossref_query_output2.0.xsd">
  <query_result>
    <head>
      <email_address>none</email_address>
      <doi_batch_id>none</doi_batch_id>
    </head>
    <body>
      <forward_link doi="10.20948/prepr-2016-124">
        <journal_cite fl_count="2">
          <issn type="print">20712898</issn>
          <issn type="electronic">20712901</issn>
          <journal_title>Keldysh Institute Preprints</journal_title>
          <journal_abbreviation>KIAM Prepr.</journal_abbreviation>
          <article_title>
            New lane changing algorithms for microscopic model of traffic flows based on cellular automata theory
          </article_title>
          <contributors>
            <contributor first-author="true" sequence="first" contributor_role="author">
              <given_name>Antonina Alexandrovna</given_name>
              <surname>Chechina</surname>
            </contributor>
          </contributors>
          <issue>136</issue>
          <first_page>1</first_page>
          <year>2017</year>
          <publication_type>full_text</publication_type>
          <doi type="journal_article">10.20948/prepr-2017-136</doi>
        </journal_cite>
      </forward_link>
      <forward_link doi="10.20948/prepr-2016-124">
        <journal_cite fl_count="0">
          <issn type="print">20712898</issn>
          <issn type="electronic">20712901</issn>
          <journal_title>Keldysh Institute Preprints</journal_title>
          <journal_abbreviation>KIAM Prepr.</journal_abbreviation>
          <article_title>
```

Рисунок 4. Результат запроса в формате XML.

Другой инструмент Crossref также позволяет получить информацию о ссылающихся на указанную статью публикациях в формате JSON:

<http://api.crossref.org/works/10.20948/prepr-2016-124>

тели) могут узнать, растет ли интерес к публикации, повлияли ли внесенные изменения (или какие-то события в научном мире) на цитируемость указанной статьи.

[https://doi.crossref.org/servlet/getForwardLinks?usr=\[username&pwd=\[password&doi=\[prefix\]&startDate=YYYY-MM-DD&endDate=YYYY-MM-DD](https://doi.crossref.org/servlet/getForwardLinks?usr=[username&pwd=[password&doi=[prefix]&startDate=YYYY-MM-DD&endDate=YYYY-MM-DD)

ЗАКЛЮЧЕНИЕ

Рассмотренные в данной работе инструменты поддержки цитирования используются для развития технологических возможностей электронной библиотеки ИПМ, частично встраиваются в интерфейс работы с библиотекой.

Благодарности

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты 19-01-00069, 20-07-00528, 18-07-01292, 18-07-00841.

СПИСОК ЛИТЕРАТУРЫ

1. Google-академия. <https://scholar.google.ru/schhp?hl=ru>.
2. Ассоциация Crossref. <https://www.crossref.org/about/>
3. НЕИКОН – Национальный консорциум российских организаций образования, науки, и культуры. <https://neicon.ru/>.
4. Электронная библиотека ИПМ. https://library.keldysh.ru/prep_qf.asp.
5. Crossref Labs. <https://www.crossref.org/labs/>.
6. Горбунов-Посадов М.М. Библиографическая ссылка на онлайн-публикацию // Препринты ИПМ им. М.В. Келдыша. 2020. № 11. 12 с. <https://library.keldysh.ru/preprint.asp?id=2020-11;> <https://doi.org/10.20948/prepr-2020-11>.
7. Слепенков М.И. Материалы конференций в онлайн-библиотеке ИПМ им. М.В. Келдыша // Препринты ИПМ им. М.В. Келдыша. 2020. № 18. 16 с. [http://doi.org/10.20948/prepr-2020-18;](http://doi.org/10.20948/prepr-2020-18) <http://library.keldysh.ru/preprint.asp?id=2020-18>
8. Автоматическая балансировка макета микроспутника на аэродинамическом подвесе с помощью системы грузов / Д.С. Иванов [и др.] // Препринты ИПМ им. М.В. Келдыша. 2020. № 16. 26 с. [http://doi.org/10.20948/prepr-2020-16;](http://doi.org/10.20948/prepr-2020-16) <http://library.keldysh.ru/preprint.asp?id=2020-16>

9. Ревякин Ю.Г. Возможности web-аналитики для оценки эффективности научных публикаций // Препринты ИПМ им. М.В. Келдыша. 2020. № 50. 42 с.

<http://doi.org/10.20948/prepr-2020-50>;

<http://library.keldysh.ru/preprint.asp?id=2020-50>.

10. Полилова Т.А. Инфраструктура научных публикаций // Препринты ИПМ им. М.В. Келдыша. 2009. № 15. 30 с.

<http://library.keldysh.ru/preprint.asp?id=2009-15>.

SOME BIBLIOGRAPHIC REFERENCES TOOLS

A. V. Ermakov¹

¹*Keldysh Institute of Applied Mathematics, Moscow*

¹Ermakov@Keldysh.ru

Abstract

The rapid quantitative growth of scientific publications has pushed the development of various Internet tools for finding the necessary information. Statistics on its study and use (citation) are no less important than the search for the necessary information. This article discusses the citation support tools provided by Google Academy, Crossref, and other developers. In the Russian Federation, NEICON (a national consortium of Russian educational, scientific, and cultural organizations) provides technical and methodological support for working with DOI and other Crossref services. The article shows how the considered tools are built into the user interfaces of the IPM electronic library. Additional features provided by the specified tools are considered.

Keywords: *e-library, bibliographic link, toolbox.*

REFERENCES

1. Google Scholar. <https://scholar.google.ru/schhp?hl=ru>.
 2. Crossref association. <https://www.crossref.org/about/>.
 3. NEIKON is a National Consortium of Russian Educational, Scientific and Cultural Organizations. <https://neicon.ru/>.
-

4. KIAM digital library. https://library.keldysh.ru/prep_qf.asp.
5. Crossref Labs. <https://www.crossref.org/labs/>.
6. *Gorbunov-Possadov M.M.* Online bibliographic reference // KIAM Preprint № 11, Moscow, 2020. <https://library.keldysh.ru/preprint.asp?id=2020-11>;
<https://doi.org/10.20948/prepr-2020-11>.
7. *Slepenkov M.I.* Conference proceedings in the on-line library of KeldyshInstitute // KIAM Preprint No. 18, Moscow, 2020. <http://doi.org/10.20948/prepr-2020-18>; <http://library.keldysh.ru/preprint.asp?id=2020-18>.
8. *Ivanov D.S., Ivlev N.A., Ivanova T.A., Roldugin D.S.* Microsatellite Mock-Up Automatic Balancing on the Air-Bearing Laboratory Facility // KIAM Preprint No. 16, Moscow, 2020. <http://doi.org/10.20948/prepr-2020-16>;
<http://library.keldysh.ru/preprint.asp?id=2020-16>.
9. *Revyakin Y.G.* The opportunities of web-analysis to estimate the effectiveness of scientific publications // KIAM Preprint № 50, Moscow, 2020. <http://doi.org/10.20948/prepr-2020-50>;
https://keldysh.ru/papers/2020/prep2020_50.pdf.
10. *Polilova T.A.* Infrastructure of scientific publications // KIAM Preprint No. 15, Moscow, 2009. <http://library.keldysh.ru/preprint.asp?id=2009-15>.

СВЕДЕНИЯ ОБ АВТОРЕ



ЕРМАКОВ Алексей Викторович – старший научный сотрудник Института прикладной математики им. М.В. Келдыша РАН, кандидат физико-математических наук.

Alexey Viktorovich ERMAKOV – senior researcher of the Institute of Keldysh Institute of Applied Mathematics Russian Academy of Sciences.

email: Ermakov@Keldysh.ru

Материал поступил в редакцию 19 ноября 2020 года

УДК 025.4: [004.65:579]

ТЕХНОЛОГИЯ НАПОЛНЕНИЯ ПРЕДМЕТНЫХ ОНТОЛОГИЙ ПРОСТРАНСТВА НАУЧНЫХ ЗНАНИЙ

Н. Е. Каленов^[0000-0001-5269-0988]

*Межведомственный суперкомпьютерный центр (МСЦ) РАН – филиал ФГУ ФНЦ
Научно-исследовательский институт системных исследований (НИИСИ) РАН*

nkalenov@jssc.ru

Аннотация

Под предметной онтологией в контексте этой статьи понимается совокупность ключевых понятий, относящихся к некоторой области науки, с их семантическими связями, дополненная индексами различных классификационных систем, описывающих данную научную область. Предметные онтологии являются необходимой составляющей каждого подпространства, входящего в Единое цифровое пространство научных знаний (ЕЦПНЗ). В данной статье приводятся результаты исследований, связанных с построением предметных онтологий на базе созданной автоматизированной системы поддержки терминологических словарей и предлагается методология выделения новых ключевых терминов отдельной области науки. Предлагаемая методология базируется на использовании существующих классификационных систем в совокупности с базами данных цитирования (БДЦ), такими как Web of Science и Scopus для англоязычных публикаций и Российский индекс цитирования (РИНЦ) – для русскоязычных. Методология предполагает разбиение научной области на ряд разделов в соответствии с выбранной классификационной системой, выделение из БДЦ ядра статей, относящихся к каждому разделу, а из статей – новых авторских ключевых терминов, которые и должны составлять, в совокупности с соответствующими разделами классификационных систем, основу предметной онтологии данной научной области.

Ключевые слова: *пространство научных знаний, предметная онтология, базы данных цитирования, ключевые термины, тезаурус, классификационные системы.*

ВВЕДЕНИЕ

Одним из важных направлений современной информатики, связанных с сохранением и распространением достижений науки, является создание единого цифрового пространства научных знаний (ЕЦПНЗ). В этом пространстве должны быть отражены достоверные знания, полученные в различных областях науки; цель создания пространства – предоставление пользователям различных категорий многоаспектной информации как внутри отдельных научных направлений, так и на стыке наук. В соответствии с концепцией, отраженной в [1–3], ЕЦПНЗ представляет собой совокупность разнородных информационных ресурсов, сгруппированных в тематические подпространства, объединяемые единой онтологией. Под единой онтологией понимаются общие для всех подпространств принципы их построения – единые подходы к хранению и предоставлению информации, формированию классов объектов и связей между ними, профилей метаданных и атрибутов, пользовательских интерфейсов и т. д. Онтология и поддерживающая ее программная оболочка ЕЦПНЗ должны обеспечивать развитый многоаспектный поиск разнородной информации, удобную ее визуализацию и навигацию по связанным ресурсам. Основой для тематического поиска информации в каждом подпространстве должна служить его предметная онтология – максимально полный набор терминов, отражающих все аспекты научного направления, с установленными между ними связями. По определению, ЕЦПНЗ должно содержать разнообразные ресурсы, в том числе, извлекаемые из существующих баз данных и каталогов библиотек. Тематический (предметный) поиск информации в этих базах данных, как правило, базируется на той или иной классификационной системе (КС). Если говорить про российские политематические библиографические ресурсы, то они базируются, в подавляющем большинстве, на одной из таких КС, как ГРНТИ, УДК, ББК, МКИ. Для обеспечения точного и полного импорта данных из внешних библиографических систем в состав предметной онтологии ЕЦПНЗ должны входить индексы этих КС. Иными словами, предметная онтология подпространства ЕЦПНЗ должна представлять

собой тематический тезаурус по данному научному направлению, дополненный индексами различных классификационных систем, выступающих в роли дескрипторов. Очевидно, что проблема формирования предметных онтологий тесно связана с традиционными задачами формирования тематических тезаурусов. Этим задачам посвящено достаточно много исследований – как зарубежных, так и отечественных [4–6]. Разработаны стандартные формы представления тезаурусов в машинном виде [7], программные средства для их формирования и встраивания в электронные библиотеки [8].

Теоретические разработки, связанные с проблемами построения и представления тезаурусов в цифровой форме, создают определенную основу для формирования предметных онтологий ЕЦПНЗ. Однако единая методология их практической реализации для различных научных областей отсутствует. Один из возможных подходов к решению этой задачи представлен в данной статье.

1. СИСТЕМА «ТЕРМИН»

В 2017–2019 гг., при поддержке РФФИ (проект 17-03-12013-ОГН) коллективом специалистов с участием автора проводились исследования в области создания прототипа предметной онтологии на основе использования существующих информационных ресурсов [9–12]. Результатом этих исследований явилось создание системы терминологических словарей «Термин» [13], включающей термины и их определения, соответствующие понятиям, отраженным в ГРНТИ [14]. Информационной основой построения системы послужили терминологические словари, сформированные в печатном виде в ВИНТИ.

Эта система в первоначальном виде включала более 12 000 терминов, относящихся к 69 тематическим научным направлениям, и определения терминов с активными ссылками на их источники в интернете. Система обеспечивала возможность ввода и редактирования данных, поиска, просмотра и навигации по ее элементам. На рис. 1 представлена информация о термине «волны в плазме», относящемся к словарю «Физика». Здесь представлено название словаря, к которому относится термин, индексы ББК (в данном примере отсутствует), индексы УДК (в данном случае 533.95), код ГРНТИ (29.27.29), ссылки на определение термина.

Показать	Редактировать	История	Удалить
волны в плазме / Физика			
Эта запись ссылается на записи			
Словарь	• Физика / ✓		
Раздел ББК	[Отсутствуют]		
Раздел УДК	• Динамика плазмы / 533.95		
См.	[Отсутствуют]		
На эту запись ссылаются			
Определение термина	Посмотреть (1)		
Связь терминов (как source)	Посмотреть (9)		
Атрибуты			
Термин	волны в плазме		
Код ГРНТИ	29.27.29		

Рисунок 1. Пример метаданных термина

Переход по ссылке «Посмотреть» в строке «Определение термина» открывает окно (рис. 2), в котором приведены определение термина и активная ссылка на его источник.

Показать	Редактировать	История	Удалить
Эта запись ссылается на записи			
Термин	• волны в плазме		
Источник	• Физическая энциклопедия femto.com.ua>articles/part_1/0569.html		
На эту запись ссылаются			
Связь терминов (как target)	Посмотреть (8)		
Атрибуты			
Определение	<p>Волны в плазме (плазменные волны) — электромагнитные волны, самосогласованные с коллективным движением заряж. частиц плазмы. Специфика плазмы, в частности, её отличие от нейтрального газа, связана с волновыми процессами. Существует много типов В. в п., определяемых её состоянием, зависящим от наличия или отсутствия внеш. магн. полей и от конфигурации плазмы и полей. Классификация В. в п. производится, прежде всего, по величине амплитуды. При больших амплитудах волновые движения наз. нелинейными волнами; они могут быть регулярными, например, солитоны, либо хаотическими, например, бесстолкновительные ударные волны. Общее решение задачи о нелинейных волнах отсутствует. Задачу о волнах малой амплитуды удаётся решить до конца в общем виде, линеаризовав уравнения, описывающие состояние плазмы. Обычно под термином "В. в п." понимаются именно такие линейные волны.</p>		

Рисунок 2. Окно определения термина

Когда разработка и наполнение первой версии системы были завершены, возникла идея выявления связей терминов путем выявления присутствия каждого термина в определении других терминов, как внутри «своего» словаря, так и во внешних словарях. Был разработан и реализован соответствующий про-

граммный алгоритм, в результате применения которого в системе было сформировано более 300 000 пар терминов, связанных между собой в указанном смысле, и обеспечена возможность просмотра и редактирования этих связей. Полученные связи, относящиеся к нескольким словарям, были отредактированы специалистами в данных предметных областях – каждой связи было присвоено одно из значений пяти типов: тождественны, близки, содержит, содержится, пересекаются. В системе, наряду со словесными обозначениями типа связей, для наглядности используются, соответственно, символы – «=», «~», «>», «<», «><».

На страницах метаданных терминов присутствуют ссылки на связи данного термина с другими. На основной странице (рис. 1) имеется ссылка на связи с теми терминами, в определение которых входит данное понятие. На странице, содержащей определение термина (рис. 2), размещена ссылка на связи с терминами, которые входят в определение данного. В примере, приведенном на рис. 1, таких ссылок 9, на рис. 2 – 8. Переход по ссылке «Посмотреть 9» открывает окно с конкретизацией связей (рис. 3). Как видим, система предлагает не только 7 связей терминов внутри словаря «Физика», но и связи термина «волны в плазме» из словаря «Физика» с терминами «Течение плазмы» из словаря «Механика» и «Солнечный ветер» из словаря «Астрономия».

Терминологические словари

Пользователь [+] Источник [+] Определение термина [+] Термин [+] Словарь [+] ББК [+] Связь терминов [+]

УДК [+]

admin Выйти

Список Структура

- волны в плазме (Физика) < течения плазмы (Механика)
- волны в плазме (Физика) > ленгмюровские волны (Физика)
- волны в плазме (Физика) > магнитозвуковые волны (Физика)
- волны в плазме (Физика) > ударные волны в плазме (Физика)
- волны в плазме (Физика) >< нелинейные явления в плазме (Физика)
- волны в плазме (Физика) >< солнечный ветер (Астрономия)
- волны в плазме (Физика) >< турбулентность плазмы (Физика)
- волны в плазме (Физика) >< Шредингера уравнение нелинейное (Физика)
- волны в плазме (Физика) < волны (Физика)

Найдено 9 записей

Выбор

source
волны в плазме / Физика

name

Применить

Рисунок 3. Пример связей терминов (слова данного термина входят в определение других).

Каждая связь является активной ссылкой, при переходе по которой открывается окно с определением термина, связанного с данным (рис. 4). Авторизованный пользователь, имеющий соответствующие права, может редактировать тип связи или ее удалить.

Аналогичное окно открывается при переходе по ссылке «Посмотреть 8» на рис. 2. Здесь мы видим связи термина «волны в плазме» с терминами из словарей «Геофизика», «Ядерная техника», «Космические исследования» и «Математика».

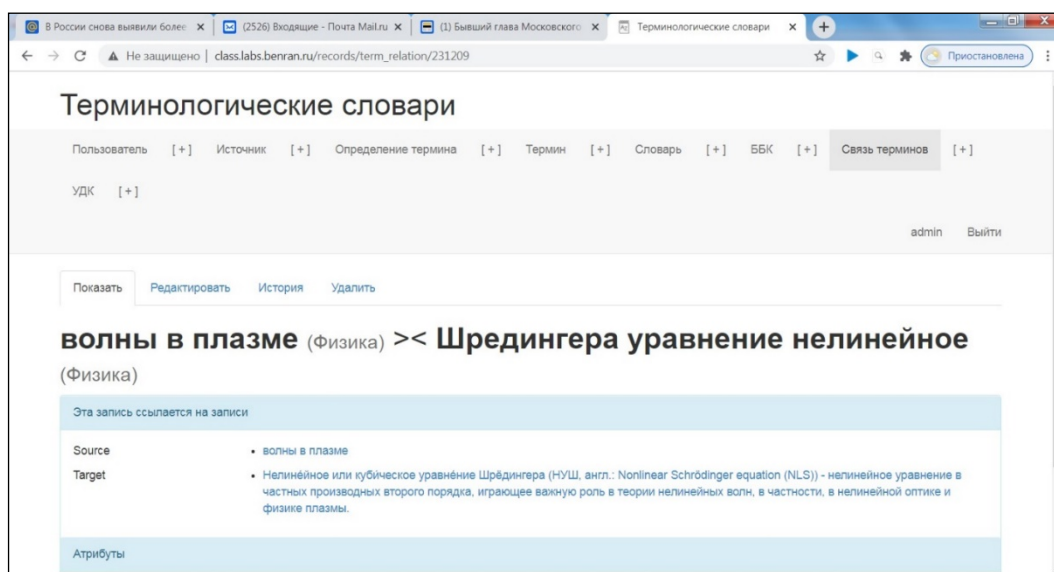


Рисунок 4. Окно просмотра деталей связи

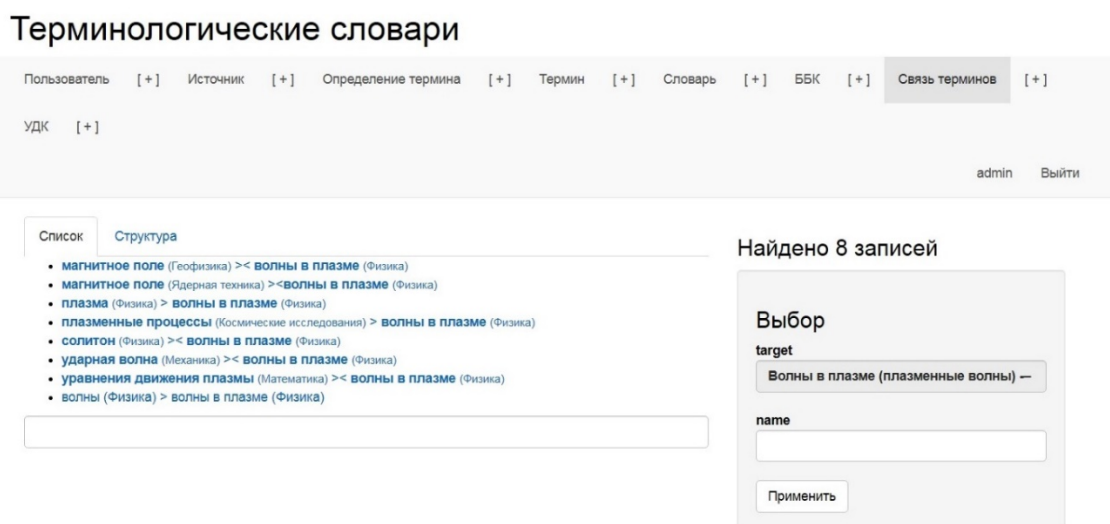


Рисунок 5. Пример связей терминов (термины, входящие в определение данного).

2. МЕТОДОЛОГИЯ ФОРМИРОВАНИЯ ПРЕДМЕТНОЙ ОНТОЛОГИИ

Для развития системы «Термин» до полноценной версии предметной онтологии необходимо дополнить ее ключевыми терминами (КТ), достаточно детально описывающими выбранные научные направления. Эту задачу предлагается решать, используя ключевые слова, которые указывают авторы в своих научных статьях. Для автоматизации этого процесса и отбора качественных публикаций могут быть использованы базы данных цитирования, такие как Web of Science (WoS) и Scopus (для англоязычных публикаций) и Российский индекс цитирования (РИНЦ) – для русскоязычных.

Предлагаемая методика включает следующие процессы.

1. Разбиение области науки на отдельные разделы. Степень детализации такого разбиения определяется учеными совместно с информационными работниками на основе анализа существующих классификационных систем (КС), таких как ГРНТИ, УДК, МКИ и др., достаточно хорошо проработанных для данной научной области.

2. Установление связей между выделенными разделами и индексами выбранных КС, относящимися к данной области науки. По отношению к каждому разделу в каждой КС выбираются индексы, в той или иной мере связанные с данным разделом. Формируются пары «раздел – индекс» и внутри пар устанавливаются связи одного из 4-х типов: тождественны, содержит, содержится, пересекаются.

3. По каждому из выделенных разделов формулируются запросы к БД цитирования (WoS, Scopus, РИНЦ), в соответствии с которыми выбираются публикации за определенный интервал годов, зависящий от научного направления. Метаданные каждой статьи, полученной по запросу, имеют атрибуты, содержащие перечень ключевых терминов и ссылки на публикации, цитирующие данную статью.

Полученные данные могут служить основой для формирования предметной онтологии. При этом возможны различные подходы к отбору ключевых терминов, подлежащих включению в нее. Можно выделить и проанализировать все ключевые термины, полученные на этом этапе; определить частотность встречаемости каждого и выбрать наиболее часто встречаемые. Второй вариант

предусматривает фильтрацию публикаций путем выделения наиболее цитированных в качестве источников ключевых терминов.

Технически отбор ключевых терминов из WoS и Scopus не представляет сложности – обе эти системы могут обрабатывать запросы в автоматическом режиме и предоставляют возможность получения информации в различных структурированных форматах, позволяющих легко выделять авторские ключевые термины и количество цитирующих работ.

Несколько сложнее обстоит дело с РИНЦ, в котором не предусмотрены ни автоматическая обработка запросов, ни выдача информации в каком-либо структурированном формате. Система выдает данные в виде текстовых записей. Для извлечения необходимой информации потребуется разработка программы, обрабатывающей HTML-страницы, содержащие найденные публикации.

В результате обработки полученной информации по каждому выделенному разделу данной научной области формируется массив русскоязычных (РИНЦ) и англоязычных терминов (WoS, Scopus) с указанием частоты их встречаемости в течение любого заданного интервала времени.

В первом приближении можно считать, что все выделенные ключевые термины входят в соответствующие разделы данного научного направления, которые, в свою очередь, связаны установленным ранее видом связи с индексами различных КС. Очевидно, что полученные наборы ключевых терминов требуют редактирования со стороны специалистов в данном научном направлении, но эта работа существенно проще, чем поиск ключевых терминов.

3 МОДЕЛЬНАЯ РЕАЛИЗАЦИЯ

Предлагаемая методика была опробована в 2019 году на примере моделирования предметной онтологии по микробиологии [15]. В этом научном направлении было выделено 42 раздела. Для каждого раздела была установлена связь одного из 5-ти вышеуказанных типов с ГРНТИ и УДК. По каждому из них на основании обработки статей, полученных по запросам к БД WoS, программным образом были выбраны ключевые термины.

Всего было обработано 5865 статей, из которых выделено 22715 различных ключевых англоязычных терминов. После смысловой обработки (отсева КТ, не связанных с микробиологией) общее количество уникальных КТ составило

7346. Эти термины были переведены на русский язык и вместе с данными по микробиологии, выгруженными из системы терминологических словарей, загружены в отдельную базу данных, представляющую собой упрощенную модель предметной онтологии. Обмен данными между системами осуществлялся на основе формата описания тезаурусов, предложенного в рамках концепции Semantic Web (рекомендации SKOS) [16].

На рис. 6 показана форма выдачи терминов, связанных с понятием «фотосинтез». Элементы предметной онтологии, расположенные под заголовком «тождественны», показывают, что «фотосинтез» является базовым термином одного из терминологических словарей и заголовком разделов рубрикаторов ВИНТИ и ГРНТИ.

Поиск связанных терминов

Термины, рубрикаторы, ключевые слова

фотосинтез

Искать

Тождественны

- + Фотосинтез
- + ВИНТИ 341.31.17
- + ГРНТИ 34.31.17

Близки

- + фотосинтезирующие бактерии
- + фотосинтезирующие микроорганизмы
- + anoxygenic photosynthesis
- + mechanism of photosynthesis

Шире

- + Обмен веществ и транспорт
- + УДК 581.13
- + ВИНТИ 341.31
- + ВИНТИ 341
- + ГРНТИ 34.31
- + ГРНТИ 34
- + УДК 581.11/13
- + Plant physiology
- + УДК 581.1

Уже

- + УДК 581.11/13
- + Катаболизм: дыхание и гликолиз
- + Обмен других веществ
- + Транспорт. Обмен неорганических веществ
- + ВИНТИ 341.31.19
- + ВИНТИ 341.31.21
- + ГРНТИ 34.31.23

Связаны

- + Физиология микроорганизмов
- + ВИНТИ 341.27.17.09

Рисунок 6. Пример визуализации терминов, связанных с «фотосинтезом»

В списках «Шире» и «Уже» присутствуют как соответствующие рубрики трех рубрикаторов (ГРНТИ, УДК и ВИНТИ), так и разделы, выделенные микробиологами. В список «Близки» входят ключевые термины, выбранные из WoS, в названия которых входит основа слова «фотосинтез»; в список «Связаны» – разделы микробиологии, с которыми была установлена связь, в терминологические названия которых входит «фотосинтез» (в данном примере – одна из рубрик ВИНТИ).

ЗАКЛЮЧЕНИЕ

Предложенная методология показала свою работоспособность. Работы по ее развитию будут продолжены в МСЦ РАН в рамках государственного задания 0580–2021–0016 и при поддержке РФФИ (проект № 20-07-00103). Ближайшая перспектива – расширение модели предметной онтологии на примере микробиологии – отработка технологии пополнения набора ключевых терминов из русскоязычных баз данных (в первую очередь, из РИНЦ), ввод в систему англоязычных терминов с установлением синонимии и других отношений с выделенными из РИНЦ русскоязычными терминами; объединение всех КТ с рубриками дополнительных классификационных систем (МКИ, рубрикатор ВАК и др.) в единой базе данных.

Предлагаемая технология позволяет не только формировать предметные онтологии, но и может служить основой для наукометрического анализа развития отдельных направлений той или иной области науки. На основании сопоставительного анализа набора и частотности ключевых терминов, полученных в разные хронологические периоды, можно делать выводы о динамике развития различных областей науки.

СПИСОК ЛИТЕРАТУРЫ

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89. № 7. С. 728–735.
2. Антопольский А.Б. и др. Принципы построения и структура единого цифрового пространства научных знаний (ЕЦПНЗ) // Научно-техническая информация. Сер. 1. 2020. № 4. С. 9–17.
3. Kalenov N., Sobolevskaya I., Sotnikov A. Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // CEUR Workshop Proceedings. 2020. V. 2543. P. 391–398.
4. M. Mercedes Martínez-González, María Luisa Alvite Díez. The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance // Comput. Stand. Interfaces. 2019. Iss. 65. P. 79–91.
5. Roche C., Costa R., Carvalho S., Almeida B. Knowledge-based terminological e-dictionaries: The EndoTerm and al-Andalus Pottery projects // Terminology, Inter-

national Journal of Theoretical and Applied Issues in Specialized Communication. 2019. No. 2. P. 259–290.

6. Белоозеров В.Н., Гуревич И.Б., Трусова Ю.О. Тезаурус по анализу изображений в сети терминологических словарей // Перспективные направления исследований и критические технологии в классификационных системах: материалы конф. Москва, 2017. С. 35–36.

7. UNESCO Thesaurus <https://skos.um.es/unescothes/> (дата обращения 28.04.2020).

8. Атаева О.М., Серебряков В.А. Персональная открытая семантическая цифровая библиотека LibMeta. Конструирование контента. Интеграция с источниками LOD // Информатика и её применения. 2017. Т. 11, Вып. 2. С. 85–100.

9. Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей // Информационные ресурсы России. 2017. № 5 (159). С. 2–7.

10. Antopolskiy A.B. and others. The Development of a Semantic Network of Keywords Based on Definitive Relationships // Scientific and Technical Information Processing. 2017. Vol. 44, No. 4. P. 261–265.

11. Антопольский А.Б., Белоозеров В.Н., Каленов Н.Е., Маркарова Т.С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информационные ресурсы России. 2018. № 5 (165). С. 22–30.

12. Белоозеров В.Н., Шабурова Н.Н. О разработке классификационно-тезаурусной онтологии для предметной области физики и радиоэлектроники // Информационное обеспечение науки: новые технологии: сб. науч. тр. Екатеринбург, 2018. С. 75–86.

13. Kalenov N.E., Senko A.M. Interactive system of terminological dictionaries as one of the elements in the ontology of scientific knowledge // Software Journal: Theory and Applications (electronic Journal). 2019. Iss. 4. <http://swsys-web.ru/en/interactive-system-of-terminological-dictionaries.html> (дата обращения 28.04.2020)

14. Государственный рубрикатор научно-технической информации. <http://grnti.ru> (дата обращения 28.04.2020).

15. Цветкова В.А., Харыбина Т.Н., Мохначева Ю.В., Бескаравайная Е.В., Митрошина И.Ю. Особенности совмещения классификационных систем и формирования массива ключевых слов для определения пространства знаний по микробиологии // Научные и технические библиотеки. 2019. № 11. С. 25–43.

16. Marcia Lei Zeng, Philipp Mayr. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review // International Journal on Digital Libraries. 2019. Vol. 20, Issue 3. P. 209–230.

TECHNOLOGY FOR FILLING SUBJECT ONTOLOGIES OF THE SCIENTIFIC KNOWLEDGE SPACE

N. E. Kalenov

Joint Supercomputer Center of the Russian Academy of Sciences – JSCC

nkalenov@jssc.ru

Abstract

Subject ontology in the context of this article is understood as a set of key concepts related to a certain field of science, with their semantic connections, supplemented by indexes of various classification systems describing this scientific field. Subject ontologies are a necessary component of each subspace that is part of the Unified digital space of scientific knowledge (DSSK). This article presents the results of research related to the construction of subject ontologies based on the created automated system for supporting terminological dictionaries and suggests a methodology for identifying new key terms in a particular field of science. The proposed methodology is based on the use of existing classification systems in conjunction with citation databases, such as Web of Science and Scopus for English-language publications and the Russian citation index for Russian-language publications. The methodology involves dividing the scientific field into a number of sections in accordance with the selected classification system, extracting from the CSB the core of articles related to each section, and from the articles - new author's keywords, which should constitute, in combination with the corresponding sections of classification systems, the basis of the subject ontologies of this scientific field.

Keywords: *scientific digital space, subject ontology, citation databases, keywords, thesaurus, classification systems.*

REFERENCES

1. *Antopolskij A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N.* O edinom cifrovom prostranstve nauchnyh znaniy // *Vestnik Rossijskoj akademii nauk.* 2019. Iss. 89. № 7. S. 728–735.

2. *Antopolskij A.B. and others.* Principy postroeniya i struktura edinogo cifrovogo prostranstva nauchnyh znaniy (ECPNZ) // *Nauchno-tehnicheskaya informaciya.* Ser. 1. 2020. № 4. S. 9–17.

3. *Kalenov N., Sobolevskaya I., Sotnikov A.* Mathematical modeling of the processes of interdisciplinary collections formation in the digital libraries environment // *CEUR Workshop Proceedings.* 2020. Vol. 2543. P. 391–398.

4. *M. Mercedes Martínez-González, María Luisa Alvite Díez.* The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance. *Comput. Stand. Interfaces,* 2019. Iss. 65. P. 79–91.

5. *Roche C., Costa R., Carvalho S., Almeida B.* Knowledge-based terminological e-dictionaries: The EndoTerm and al-Andalus Pottery projects // *Terminology, International Journal of Theoretical and Applied Issues in Specialized Communication.* 2019. No. 2. P. 259–290.

6. *Beloozerov V.N., Gurevich I.B., Trusova Yu.O.* Tezaurus po analizu izobrazhenij v seti terminologicheskikh slovarej // *Perspektivnye napravleniya issledovanij i kriticheskie tehnologii v klassifikacionnyh sistemah : materialy konf. Moskva,* 2017. S. 35–36.

7. UNESCO Thesaurus <https://skos.um.es/unescothes/> (28.04.2020).

8. *Ataeva O.M., Serebryakov V.A.* Personalnaya otkrytaya semanticheskaya cifrovaya biblioteka LibMeta. Konstruirovanie kontenta. Integraciya s istochnikami LOD // *Informatika i eyo primeneniya,* 2017, tom 11. vyp 2. S. 85–100.

9. *Antopolskij A.B., Beloozerov V.N., Markarova T.S.* O razrabotke ontologii na osnove klassifikatorov nauchnoj informacii i terminologicheskikh slovarej // *Informacionnye resursy Rossii.* 2017. № 5 (159). S. 2–7.

10. *Antopolskiy A.B. and others.* The Development of a Semantic Network of Keywords Based on Definitive Relationships // Scientific and Technical Information Processing. 2017. Vol. 44. No. 4. S. 261–265.

11. *Antopolskiy A.B., Beloozerov V.N., Kalenov N.E., Markarova T.S.* O razvitii terminologicheskoy bazy dannyh v vide kompleksa otraslevykh informacionno-poiskovykh tezaurusov // Informacionnye resursy Rossii. 2018. № 5 (165). S. 22–30.

12. *Beloozerov V.N., Shaburova N.N.* O razrabotke klassifikacionno-tezaurusnoj ontologii dlya predmetnoj oblasti fiziki i radioelektroniki // Informacionnoe obespechenie nauki: novye tehnologii: sb. nauch. tr. Ekaterinburg, 2018. S. 75–86.

13. *Kalenov N.E., Senko A.M.* Interactive system of terminological dictionaries as one of the elements in the ontology of scientific knowledge // Software Journal: Theory and Applications (electronic Journal), 2019. Iss. 4. <http://swsys-web.ru/en/interactive-system-of-terminological-dictionaries.html> (28.04.2020)

14. Gosudarstvennyj rubrikator nauchno-tehnicheskoy informacii. <http://grnti.ru> (28.04.2020).

15. *Cvetkova V.A., Harybina T.N., Mohnacheva Yu.V., Beskaravajnyaya E.V., Mitroshina I.Yu.* Osobennosti sovmesheniya klassifikacionnykh sistem i formirovaniya massiva klyuchevykh slov dlya opredeleniya prostranstva znanij po mikrobiologii // Nauchnye i tehnicheckie biblioteki. 2019. № 11. S. 25–43.

16. *Marcia Lei Zeng, Philipp Mayr.* Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review // International Journal on Digital Libraries. 2019. Vol. 20, Issue 3. P. 209–230.

СВЕДЕНИЯ ОБ АВТОРЕ



КАЛЕНОВ Николай Евгеньевич – главный научный сотрудник Межведомственного Суперкомпьютерного Центра РАН – филиала Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», доктор технических наук, профессор.

Nikolay Evgenievich KALENOV – Chief Researcher of Joint Super Computer Center of the Russian Academy of Sciences – Branch of Federal State Institution «Scientific Research Institute for System Analysis of the Russian Academy of Sciences», Doctor of Technical Sciences, Professor.

email: nkalenov@jsc.ru; ORCID: 0000-0001-5269-0988

Материал поступил в редакцию 18 ноября 2020 года

УДК 001.812 + 004.032.6 + 004.051 + 004.057.5 + 37.026

ВАРИАНТЫ ВИДЕОЛЕКЦИЙ

Ф. О. Каспаринский^{1[0000-0002-1048-9212]}

¹ ООО «МАСТЕР-МУЛЬТИМЕДИА», г. Москва, Шоссе Энтузиастов 98-3-274,
Москва, 111531

¹ felix@kasparinsky.pro

Аннотация

Статья посвящена типологии основных вариантов видеолекций (<https://vimeo.com/showcase/6184718>), отличающихся дидактической ценностью и трудоёмкостью создания. Для каждого варианта видеолекций определены ключевые условия эффективного формирования и использования. Особое внимание уделено вопросам оптимизации аппаратного обеспечения для создания различных вариантов видеолекций в условиях очного и дистанционного обучения.

В качестве базового варианта рассматривается видеозапись демонстрации презентации PowerPoint с речевыми и графическими комментариями слайдов. Стандартный профильный вариант — отснятый слева экран (доска) с лектором в профиль — рекомендуется использовать при дефиците места в аудитории. При наличии видеокамеры с оптикой, обеспечивающей сохранение детализации изображений наглядных материалов, предпочтительным вариантом видеолекции является стандартный фронтальный, отснятый из дальнего конца аудитории. Дидактически оптимизированный монтаж видеолекции позволяет комбинировать высококачественную запись наглядных материалов базового варианта с любыми видеозаписями действий преподавателя. Интегративные варианты видеолекций создаются посредством оверлейной вставки видеозаписи действий преподавателя поверх специально предназначенного инвариантного места в дизайне слайдов базового варианта видеолекции. Предложены варианты адаптации дизайна презентаций PowerPoint для формирования базового и интегративного варианта видеолекций. Наивысшее соотношение «дидактическое качество/ресурсоёмкость» достигается при создании интегративных видеолекций

непосредственно в процессе очных и дистанционных занятий.

Ключевые слова: видеолекция, интегративная видеолекция, создание учебных видео, дистанционное обучение, очное обучение, рабочая среда, Microsoft PowerPoint, HyperCam.

ВВЕДЕНИЕ

Интенсивное развитие сетевых сервисов в течение второго десятилетия XXI века создало условия для модернизации способов создания и использования видеолекций. Современная видеолекция — аудиовизуальный ресурс с записью процесса трансляции знаний преподавателем, обеспечивающий дидактические функции первичного восприятия, повторения, дополнения и закрепления информации [1, 2]. Видеолекции помогают учащимся получить знания в случае пропуска занятий. Структурированная видеолекция с оглавлением [3, 4] способствует повторению пройденного материала. QR-коды и гиперссылки на дополнительные материалы, содержащиеся в видеолекции, можно использовать для расширения кругозора. Возможность мгновенного перехода к нужному месту видеолекции по URL-адресу и сетевой функционал разметки любого места на кадре с организацией коллективного обсуждения удобны в процессах закрепления, систематизации и обобщения существенных сведений.

Технологическая эволюция аппаратно-программных средств формирования, трансформации и трансляции аудиовизуальных ресурсов привела к изменению набора главных проблем процесса создания профессиональной видеолекции [1, 2]. Обеспечение технического качества записи и организация распространения видеолекций перестали быть лимитирующими факторами. В третьем десятилетии XXI века главной проблемой становится обеспечение дидактической полноценности видеолекции, создаваемой в условиях академической мобильности и дефицита времени в период бесшовной интеграции очной и дистанционной форм обучения.

1. ИСТОРИЧЕСКИЕ ВЕХИ СТАНОВЛЕНИЯ ВИДЕОЛЕКЦИЙ КАК ЖАНРА УЧЕБНОГО ПРОИЗВЕДЕНИЯ

Профессиональное использование видеолекций в отечественном образовательном пространстве началось более четырёх десятков лет назад. В 1974 году

лекции ведущих преподавателей Северо-Западного государственного заочного технического университета (СЗТУ, г. Санкт-Петербург) стали транслироваться Ленинградским телевидением. Благодаря технологиям, разработанным Научно-исследовательским институтом телевизионного обучения, Современный гуманитарный университет (г. Москва) в 1998 году начал поточное производство видеолекций в аналоговой форме, которые распространялись по системам спутниковых связей и на видеокассетах формата VHS. В это же время в Университете Южной Флориды (Тампа) стартовали эксперименты по созданию видеолекций в цифровой форме при помощи инструментария мультимедийных и сетевых технологий. Своевременные инвестиции в организацию жизненного цикла видеолекций обеспечили формирование богатейшей учебной видеотеки Массачусетского технологического института.

На рубеже тысячелетий в результате технологической эволюции варианты распространения видеолекций дополнились спутниковыми каналами, аналоговыми видеокассетами, цифровыми физическими носителями и интернет-трансляцией. С 2001 года в Московском государственном университете имени М.В. Ломоносова по инициативе Ф.О. Каспаринского начались научные исследования по адаптации эволюционирующего программно-аппаратного инструментария для создания и распространения профессиональных цифровых видеолекций на CD и DVD, а также в интернете [1, 2, 4–8]. В 2005 году возникли видеохостинги, совершенствование функционала которых к настоящему времени достигло уровня, обеспечивающего дидактически полноценное использование сетевых вариантов видеолекций с интерактивными оглавлениями [4], тизерами и целевыми коллективными заметками (<https://vimeo.com>). Глобальное распространение и усовершенствование информационно-коммуникационных технологий в начале второго десятилетия XXI века создало условия для доминирования интернет-форм видеолекций как наиболее утилитарной формы видеометода обучения [2].

2. ДИДАКТИЧЕСКАЯ ПОЛНОЦЕННОСТЬ ВИДЕОЛЕКЦИЙ

Дидактически полноценная видеолекция должна обеспечивать эффективность первичного восприятия, повторения, дополнения и закрепления информации [5]. Достоверность информации профессиональной видеолекции сертифици-

фицируется авторством лектора-первоисточника. Ценность личностного примера лектора и информационного содержимого видеолекции для воспитания личностных качеств и формирования компетентностей из суммы компетенций предопределяет привлечение и удержание внимания ученика. Практический опыт показывает, что постановочные видеолекции, как правило, не обеспечивают высокого уровня эмоционального воздействия преподавателя на учащихся, характерного для синхронного процесса общения с живой аудиторией.

2.1. НАГЛЯДНЫЕ МАТЕРИАЛЫ ВИДЕОЛЕКЦИЙ

Техническое исполнение видеолекции должно обеспечивать когнитивный комфорт и надежную доступность в произвольной программно-аппаратной среде. Наглядные материалы должны быть сформированы таким образом, чтобы обеспечивать комфортное восприятие информации (текстов, схем, фотографий, анимаций и видеозаписей) на экранах любых устройств, от электронных досок до планшетов и смартфонов. В рамках парадигмы BYOD (Bring Your Own Device) целесообразно дополнять наглядные материалы QR-кодами с гиперссылками на релевантные первоисточники информации. Благодаря QR-кодам учащиеся получают возможность оперативного ознакомления с заинтересовавшими их сведениями. С учётом вероятного использования наглядных материалов не только в процессе синхронных очных и дистанционных занятий, но и в форме сетевых PDF-публикаций для асинхронного самостоятельного обучения, с изображениями QR-кодов целесообразно ассоциировать чувствительные к нажатию гиперссылки на соответствующие ресурсы.

Эффективность создания контента учебного курса была увеличена посредством универсализации представления текста и иллюстраций [9] в первично преподносимых наглядных материалах синхронных занятий и в распространяемом варианте учебного курса для асинхронного изучения в дистанционном формате. Для создания и проведения презентаций рекомендуется программа PowerPoint из набора Microsoft 365. Следует отметить, что устаревшая версия PowerPoint из набора Microsoft Office 2016, установленная в электронных досках и преподавательских компьютерах госбюджетных учреждений, отличается существенно урезанным функционалом (нет панели выбора графических инструментов и возможности создавать интегративные видеолекции).

Для обеспечения оптимального размера слайдов рекомендуется использовать не стандартные параметры, а специально подобранные для утилитарной демонстрации на экранах произвольного размера и в адаптивных контейнерах страниц веб-сайтов [10]. Для этого в пункте «Конструктор» верхнего меню PowerPoint следует выбрать кнопку «Размер слайда», из ниспадающего меню выбрать строку «Настроить размер слайдов ...», после чего во всплывающем окне установить ширину 27,093 см. при высоте 20,32 см., что соответствует классическому соотношению сторон кадра 4:3, оптимальному для демонстрации графики и видео на любых экранах [10].

ОПРЕДЕЛЕНИЕ ЭНЕРГИИ

Термин «энергия» введён Аристотелем в трактате «Φυσικὴ ἀκρόασις» («Физика»), как **характеристика деятельности человека**.

Томас Юнг в «A Course of Lectures on Natural Philosophy and the Mechanical Arts» интерпретировал энергию как **способность производить работу**.

Ἀριστοτέλης
384 до н.э – 322 до н.э.

Thomas Young
1773 – 1829

Энергия (E, др.-греч. ἐνέργεια — действие, деятельность, сила, мощь) — скалярная физическая величина, являющаяся единой мерой различных форм движения и взаимодействия форм материи. Энергия количественно характеризует способность системы производить действия (работу).

© 2019. Авторские права, дизайн и адаптация графики: Ф.О. Каспаринский, 2019 | info@kasparynsky.com | Авторский сайт: kasparynsky.com

Рисунок 1. Дизайн слайда PowerPoint в классическом формате 4:3

По периметру слайда рекомендуется разместить декоративную рамку шириной 3–5% от размера слайда (см. рис. 1). Эта рамка позволяет избежать проблем восприятия расположенного по периферии контента в случае использования режима «overscan», при смещении кадра на устаревшей проекционной аппаратуре, а также в случае визуализации оверлейных периферийных панелей навигации в системах видеоконференций.

В верхней части слайда под рамкой рекомендуется выделить место для

фризовой панели (10–15% высоты слайда), содержащей заголовков и тематическую символику. В левой части фриза размещается символика курса или относящегося к слайду раздела группировки информационных элементов. Такое оформление позволяет эффективно идентифицировать контекстную привязку контента. Правая часть фриза отведена для символа темы занятия. Заголовочный текст (рубленый шрифт, кегль 28–44) размещается в 1–2 строки по центру фриза (60–70% ширины слайда).

ОПРЕДЕЛЕНИЕ ЭНЕРГИИ

Термин «энергия» введён Аристотелем в трактате «*Φυσική ἀκρόασις*» («Физика»), как **характеристика деятельности человека**.

Томас Юнг в «*A Course of Lectures on Natural Philosophy and the Mechanical Arts*» интерпретировал энергию как **способность производить работу**.

Аристотέλης
384 до н.э – 322 до н.э.

Thomas Young
1773 – 1829

Энергия (E, др.-греч. ἐνέργεια — действие, деятельность, сила, мощь) — скалярная физическая величина, являющаяся единой мерой различных форм движения и взаимодействия форм материи. Энергия количественно характеризует способность системы производить действия (работу).

Автор лекции:
Ф.О. Каспаринский

Рисунок 2. Широкоэкранный дизайн слайда PowerPoint 16:9

Нижнюю часть слайда с обрамлением (относительная высота 25–30%) рекомендуется резервировать для размещения толкования терминов (рубленый шрифт, кегль 20–24) и ссылок на первоисточники в форме расположенного слева QR-кода с дополнительной гиперссылкой. Опыт показал утилитарность дублирования ссылочной информации, поскольку на очных занятиях учащиеся используют свои устройства с камерами (смартфоны, планшеты) для оперативного перехода к первоисточникам посредством распознавания QR-кода (утилитарный размер – 20% высоты слайда), а при использовании PDF-вариантов презентаций предпочитают применять нажатие гиперссылки, ассоциированной с изображением QR-кода. Основная область слайда между фризом и зоной терминов используется для размещения иллюстраций (слева) и текста (справа). Размещение

иллюстраций и текста в основной зоне слайда варьируется в зависимости от специфики контента. Альтернативный дизайн презентации 16:9 (см. рис. 2) оптимизирован для интегративной записи демонстрации презентации PowerPoint, в которой графические и речевые комментарии преподавателя дополнены видеозаписью его действий (см. рис. 3). В этом случае используется широкоэкранный шаблон презентации (16:9) с шириной 36,124 см. при высоте 20,32 см. Все рекомендации по дизайну слайдов актуализуются с учётом увеличения ширины на 25% при выборе режима масштабирования «Подбор размера».



Рисунок 3. Экранные снимки идентичных кадров презентации в процессе демонстрации и по окончании записи интегративной видеолекции

Для размещения видеоокна с действующим преподавателем резервируется 25% правой части обрамлённой нижней зоны всех слайдов (содержит QR-код и толкование терминов, занимает 25% общей высоты слайда). В зарезервированном прямоугольнике целесообразно размещать фотографию автора презентации с подписью (ФИО) и QR-кодом на релевантное портфолио (см. рис. 2 и рис. 3).

2.2. ГРАФИЧЕСКИЕ КОММЕНТАРИИ НАГЛЯДНЫХ МАТЕРИАЛОВ ВИДЕОЛЕКЦИЙ

По мнению учащихся, содержимое презентации запоминается лучше, если речевые комментарии дополняются графическими (выделение ключевой информации, схематизация и дополнение изображений, эмоциональные маркеры).

В современных условиях (аудитория с электронной доской и потенциальной сетевой трансляцией событий) использование материальной лазерной указки неэффективно: световое пятно слабого лазера (5–200 мВт) не выделяется на ярком экране электронной доски, а мощный лазер (2 Вт и более) повреждает экран и опасен для зрения.

Распространённый способ использования курсора для фокусировки внимания не оптимален, поскольку на экране передающей системы он плохо заметен (даже в форме виртуальной лазерной указки), а на экранах принимающих систем в некоторых системах видеоконференций (Teams) курсор заменяется на крупный красный крест, который может заслонять существенные для восприятия учащихся детали наглядных материалов. Когда нет необходимости акцентировать внимание на определённом участке слайда презентации, преподавателю приходится своевременно выводить курсор за границу окна презентации, чтобы он не превращался в дистрактор внимания. Перемещение курсора можно осуществлять посредством сенсорных панелей (графический планшет, сенсорный экран преподавательского компьютера, тачпад беспроводной клавиатуры, такой как IPazzPort 19s).

После перехода в режим «Записать слайд шоу» становится доступным дополнительный функционал графического комментирования: имитация лазерной указки, перо и маркер. Использование этих инструментов посредством мыши или графического планшета удобно при дистанционной форме обучения, когда преподаватель сидит за столом. При очной форме преподавания лектору, стоящему у доски, удобнее использовать сенсорную панель портативной клавиатуры или стилус электронной доски.

Использование сенсорной панели в сочетании с имитатором лазерной указки требует особого навыка оперативного удаления указателя за границы

кадра презентации в момент, когда он перестаёт быть нужен, поскольку присутствие указателя в нерелевантном месте слайда контрпродуктивно. Инструмент «маркер» позволяет выделять нужные фрагменты кадра, но непригоден для письма и изображения сложных схем. Наиболее утилитарным вариантом графического комментирования является непосредственное использование стилуса электронной доски в сочетании с инструментом «перо». Функционал графического комментирования презентации в версии PowerPoint из комплекта Microsoft 365 значительно богаче и удобнее (выбор инструментов и палитры), чем в штатной версии приложения электронной доски из устаревшего комплекта Office 2016. По этой причине эффективнее использовать электронную доску в режиме графического планшета, воспроизводящего контент с подключённого USB-кабелем компьютера, имеющего новейшую версию PowerPoint.

Для создания графических комментариев в режиме «Записать слайд-шоу» не обязательно активировать процесс записи, при котором запрещено перемещение по ряду слайдов в обратном направлении. Если процесс записи был активирован, после завершения демонстрации слайдов и остановки записи вариант презентации с графическими комментариями следует сохранить в ручном режиме для возможности последующего экспорта видео. Размер файла презентации с записью демонстрации может увеличиться на порядок, что необходимо учитывать при выборе носителя. К примеру, 40-минутная презентация из 40 слайдов после сохранения записи демонстрации с речевыми и графическими комментариями без включённой видеокамеры может увеличиться в размере с 10 Мб до 100 Мб. При включённой видеокамере расход места в хранилище увеличивается еще на порядок. Рекомендуется до начала демонстрации презентации отключать её автосохранение в облако, поскольку этот процесс может потребовать много времени и не вписаться в рамки расписания очной формы обучения.

Для организации видеоконференций рекомендуется использовать сервисы Cisco Webex, Skype, Teams, Zoom и Discord, при работе с которыми преподаватель видит на экране содержимое своего рабочего стола в реальном времени, а не после его обработки, как в OBS Studio. В последнем случае преподаватель наблюдает появление собственных графических комментариев с дискомфортной задержкой (1–2 сек), что ингибирует процесс общения. Вещание рабочего

стола во время демонстрации презентации при включённой вебкамере позволяет ученикам одновременно наблюдать за работой преподавателя и появлением графических комментариев на слайде.

2.3. АУДИОВИЗУАЛЬНАЯ ЗАПИСЬ ДЕЙСТВИЙ ПРЕПОДАВАТЕЛЯ

Видеозапись демонстрации слайдов с голосовыми комментариями не является дидактически полноценным вариантом видеолекции, поскольку не помогает акцентировать внимание учащихся на важных элементах наглядных материалов. Функционал «Запись слайд-шоу» программы для создания и проведения презентаций PowerPoint позволяет осуществлять видеорегистрацию графических комментариев синхронно с речевым сопровождением. Экспорт такой записи в формат формат mp4-видео позволяет получить базовый вариант видеолекции (см. п. 3.1).

Для формирования дидактически полноценного варианта видеолекции необходима запись действий преподавателя, дополняющая повествование жестиком, эмоциональными акцентами мимики и возможной наглядной демонстрацией материальных объектов. Стандартный вариант видеолекции – техническая аудиовизуальная запись действий лектора, адресно комментирующего наглядные материалы, в которой источником звука является направленный или петличный микрофон. Ракурс съёмки определяет вариант стандартной видеолекции — профильный (см. п. 3.2, пример: <https://vimeo.com/204236018>) или фронтальный (см. п. 3.3, пример: <https://vimeo.com/204236043>). Создание стандартных вариантов видеолекций сопряжено со значительными затратами ресурсов (профессиональная аппаратура [11], операторская работа). Наиболее бюджетный вариант – интегративная видеолекция, обеспечивающая сочетание видеозаписи действий преподавателя с графическими и речевыми комментариями наглядных материалов (см. п. 3.5, пример: <https://vimeo.com/204235168>).

2.4. ОБРАБОТКА АУДИОВИЗУАЛЬНЫХ ЗАПИСЕЙ ВИДЕОЛЕКЦИЙ

Базовые варианты видеолекций курса «Естествознание» [11] — экспортированные в mp4-формат записи процесса демонстрации презентаций PowerPoint с речевыми и графическими комментариями слайдов — непосредственно публиковались на интернет-видеохостинге Vimeo.com в рамках подпис-

ки «Pro» (квота 20 Гб/неделя) или использовались для последующего монтажа дидактически оптимизированных вариантов видеолекций (см. рис. 4).

Видеозаписи очных занятий в формате MTS и медиаданные онлайн-занятий обрабатываются посредством графического редактора и экспортируются с разрешением 1920x1080 в формат mp4-видеозаписей, после чего публикуются на видеохостинге или используются для монтажа дидактически оптимизированных вариантов видеолекций (см. п. 3.4).

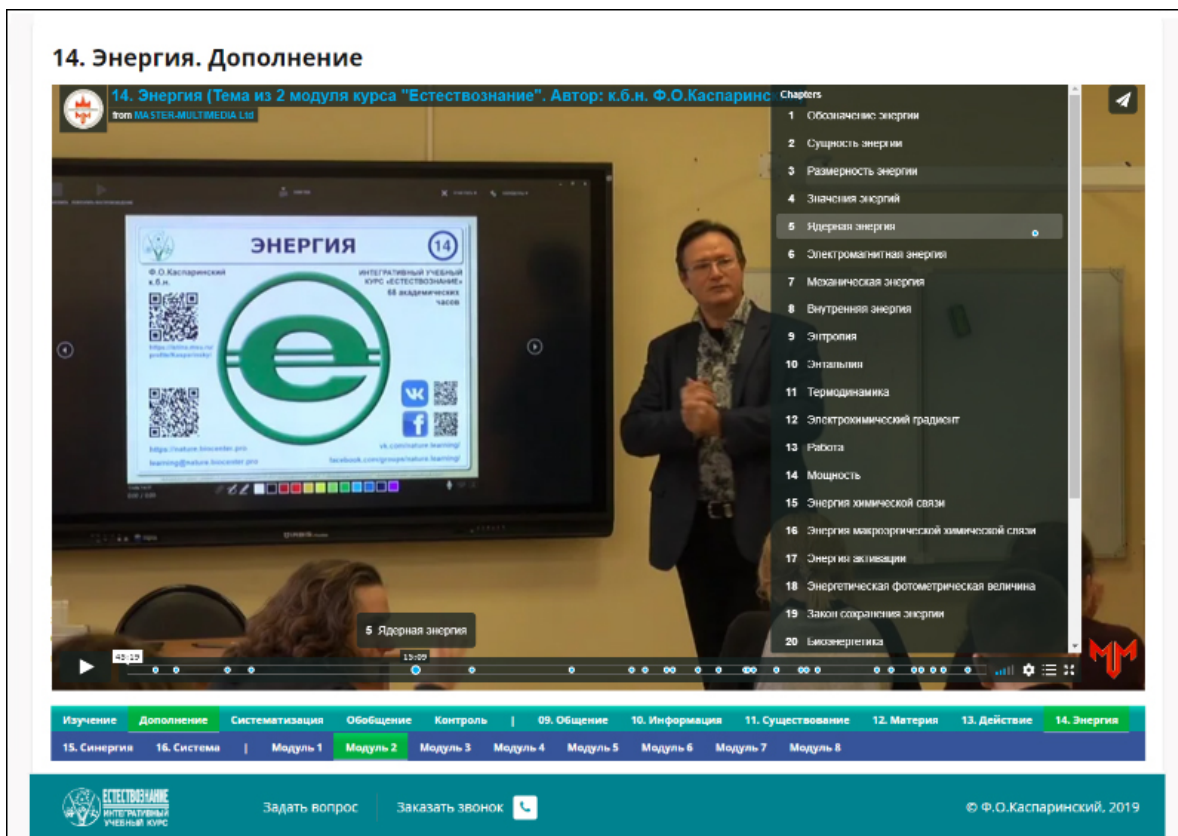


Рисунок 4. Экранный снимок фрагмента веб-страницы с внедрённой видеолекцией, содержащей навигационные меню

2.5. ИНФОРМАЦИОННО-НАВИГАЦИОННЫЙ СЕРВИС ВИДЕОЛЕКЦИЙ

Видеолекции можно использовать с различным сервисным функционалом. Все аудиовизуальные материалы специально адаптированы для онлайн-просмотра на любых пользовательских устройствах. Плеер видеолекции, размещенной на видеохостинге Vimeo, можно вставить в веб-страницу учебного портала кодом, поддерживающим адаптивность шаблона сайта.

Пример кода вставки видеоплеера для видеолекции с соотношением сто-

рон кадра 16x9: `<div style='padding:56% 0 0 0; position:relative;'> <iframe src='https://player.vimeo.com/video/398614701' allowfullscreen frameborder='0' style='position:absolute; top:0; left:0; width:100%; height:100%;'></iframe></div>`. При вставке видеолекции с классическим соотношением сторон кадра (4x3) корректируется параметр `'padding:75%`.

После внедрения видеолекции кодом в веб-страницу становится доступен функционал оверлейных оглавлений и меток полосы прокрутки со всплывающими названиями, позволяющими без перерыва просмотра нелинейно выбирать для последующего ознакомления наиболее интересный тематический фрагмент (см. рис. 4. и пример: <https://nature.biocenter.pro/videolecture/>, пункт 2).

2.6. КОЛЛЕКТИВНОЕ ИСПОЛЬЗОВАНИЕ ВИДЕОЛЕКЦИЙ

На видеохостинге Vimeo доступен просмотр видеолекций в режиме создания целевых заметок и коллективного обсуждения. Любой пользователь имеет возможность акцентировать всеобщее внимание на любом элементе кадра видеоряда посредством установки хорошо заметной круглой голубой метки. При необходимости метку можно перемещать в поле кадра. Во время создания метки формируется соответствующий пункт интерактивного оглавления аудиовизуального ряда, нажатие на который впоследствии обеспечивает переход к нужному кадру и визуализацию маркера. Длина имени и количество пунктов такого оглавления не ограничено. Пользователи имеют возможность комментировать текст, сопровождающий метку, что создаёт замечательную возможность для коллективного обсуждения помеченного содержимого видеоряда в рамках научно-образовательной, деловой, общественной и любой другой деятельности. В заметку может быть вставлено любое количество гиперссылок на сторонние ресурсы и прочие заметки аудиовизуальных рядов сервиса, ссылки на которые автоматически формируются в адресной строке браузера.

Всем участникам обсуждения по электронной почте рассылаются уведомления о появлении новых заметок и пользовательских реакциях на них. Для сохранения своих целевых заметок и получения уведомлений при реакции на них не зарегистрированные на сервисе Vimeo пользователи должны предоставить адрес своего электронного почтового ящика (email).

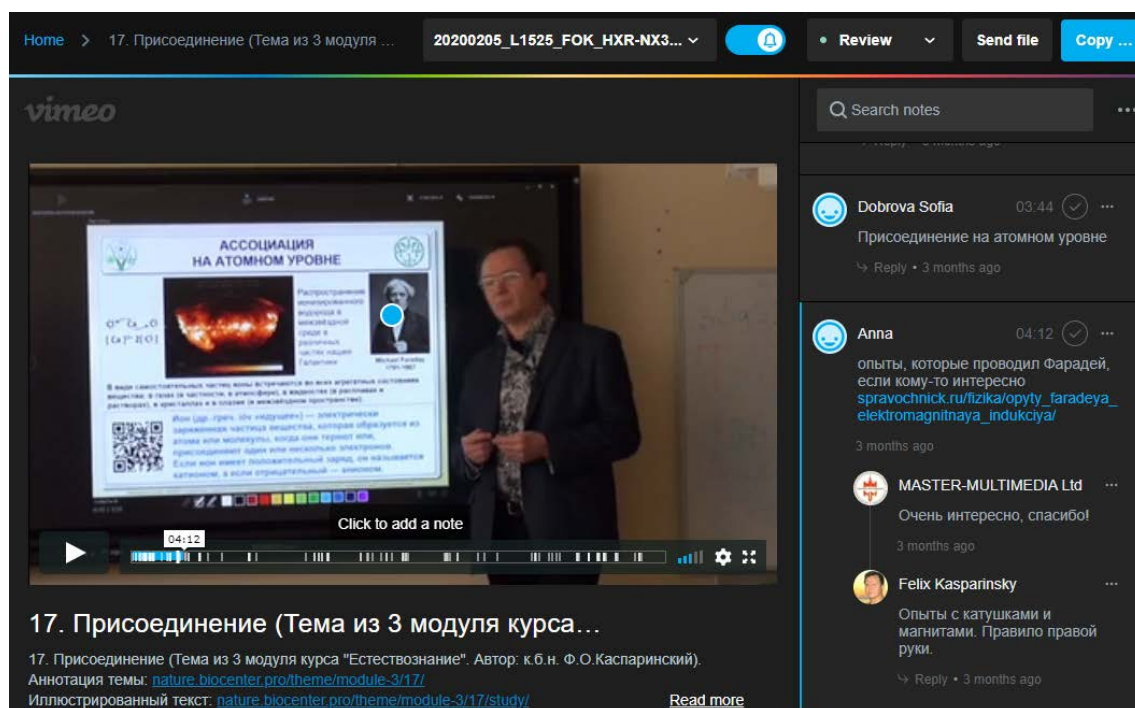


Рисунок 5. Экранный снимок фрагмента веб-страницы с целевыми заметками видеолекции и коллективным обсуждением

Каждый участник обсуждения может вставить в любой кадр аудиовизуального ряда произвольное количество позиционных меток, в связи с каждой из которых может формироваться собственное дерево заметок и ответов на них. Эта возможность позволяет дополнить использование видеолекций дидактическими функциями повторения, закрепления и контроля.

Обновление версии видеолекции приводит к созданию нового набора целевых заметок. Для каждой версии видеолекции создаётся собственный набор коллективных заметок. Эта особенность функционала позволяет посредством повторной загрузки на видеохостинг многократно использовать один аудиовизуальный ряд для организации новых дискуссий каждый учебный период. После очистки дискуссионного пространства предыдущие наборы дискуссий сохраняются и могут использоваться администратором после выбора для просмотра соответствующей версии видеолекции.

3. ВАРИАНТЫ ВИДЕОЛЕКЦИЙ

В настоящее время главным препятствием для использования дидактически полноценных видеолекций в практике повседневной педагогической деятельности является ресурсоёмкость их создания. В течение 2019–2020 годов в процессе создания интегративного курса «Естествознание» было апробировано 4 классических (см. рис. 6) и два интегративных варианта создания видеолекций, отличающихся дидактической ценностью и трудоёмкостью (см. примеры: <https://vimeo.com/showcase/6184718>).

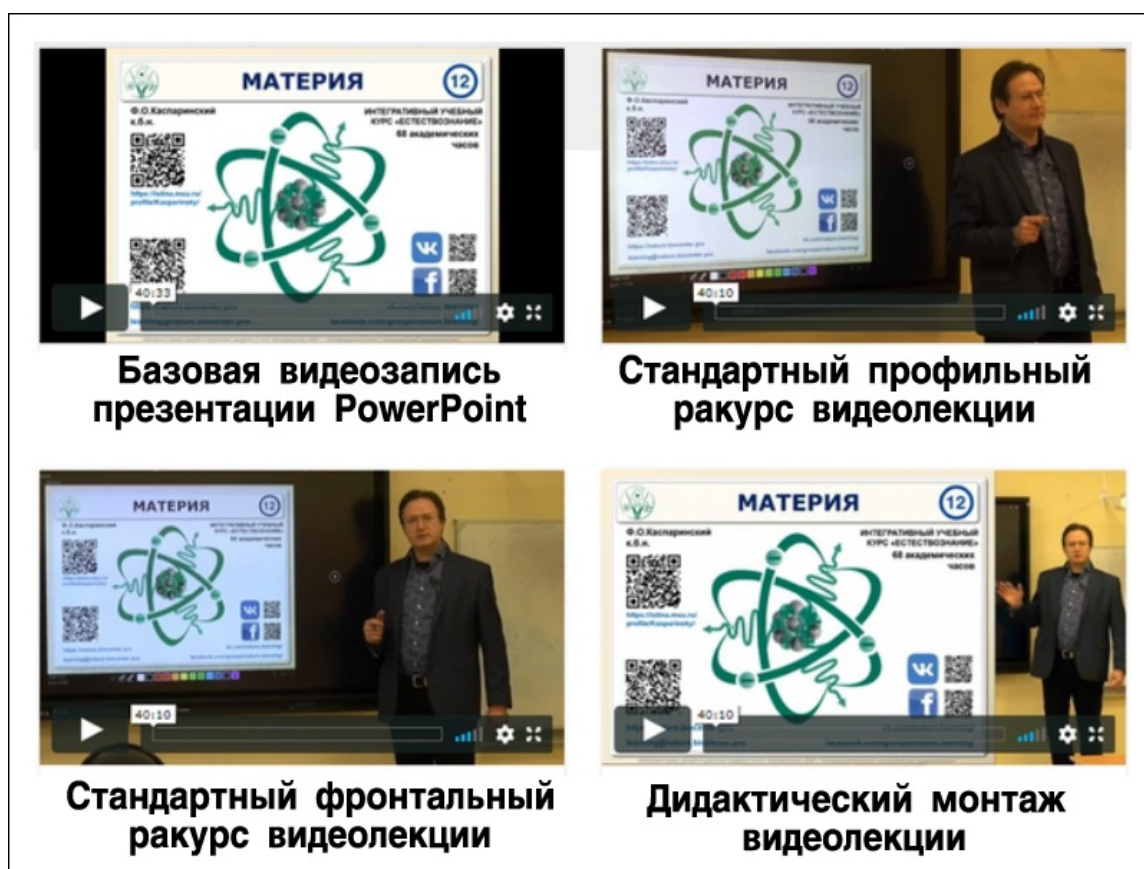


Рисунок 6. Классические варианты видеолекций

Практический опыт показал эффективность последовательного использования версий видеолекций: первым к видеохостингу подключается базовый вариант видеолекции, создаваемый с наименьшими трудозатратами. По мере создания совершенных вариантов видеолекций ими замещаются более примитивные, создаются интерактивные меню (см. рис. 4) и аннотации с гиперссылка-

ми на релевантные ресурсы. Следует отметить, что код вставки и меню при замене вариантов видеолекций сохраняются неизменными.

3.1. БАЗОВЫЙ ВАРИАНТ ВИДЕОЛЕКЦИЙ

В качестве базового варианта видеолекций можно рассматривать видеозапись процесса презентации PowerPoint с речевыми и графическими комментариями слайдов. Базовый вариант видеолекций привлекателен минимальными затратами для пользователей Microsoft PowerPoint, но ограничивает информационное пространство только упомянутым средством проведения презентаций.

Высокое визуальное качество представления наглядных материалов в базовых вариантах видеолекций обуславливает их предпочтительное использование при отсутствии специальной аппаратуры и программного обеспечения для создания иных вариантов видеолекций. Из аппаратуры потребуется только преподавательский компьютер (ноутбук, микрокомпьютер и т. п.) с микрофоном. Для улучшения качества звука рекомендуется использовать микрофон с кардиоидной диаграммой направленности.

Технология создания базовых вариантов видеолекций посредством инструментария PowerPoint отличается простотой (видеоинструкция по созданию для приложения PowerPoint из набора Office2016: <https://youtu.be/odjX5cf1B8U>). При переходе в режим записи процесса демонстрации презентации PowerPoint из комплекта Microsoft 365 (последовательность команд: «Слайд-шоу» / «Записать слайд-шоу») программа автоматически переходит в полноэкранный режим, в котором центрированный слайд обрамляется навигационно-сервисными элементами. По сторонам располагаются кнопки перемещения вперед/назад, сверху размещены кнопки «Записать», «Остановить», «Повторить воспроизведение»; «Заметки», «Выход из режима записи» и «Выбор источников видео- и аудиозаписи». При активации последней упомянутой кнопки появляется ниспадающее меню, из которого можно выбрать оптимальные устройства. Снизу находятся счётчик хронометража, кнопки выбора графических инструментов (ластик, перо и маркер), палитра, кнопки включения/выключения микрофона, камеры и отображения видео с камеры). При доступности клавиатуры в процессе работе с презентацией графические инструменты можно переключать сочетанием клавиш: ластик – Ctrl+E, перо – Ctrl+P, маркер – Ctrl+I. Следует принимать во внимание,

что в PowerPoint из комплекта Microsoft Office 2016 нет инструментов верхней и нижней панелей.

Непосредственно перед началом демонстрации презентации рекомендуется активировать один из графических инструментов, поскольку после включения записи переход на предыдущий слайд запрещается, а любое нажатие на изображение без задействованного графического инструмента приводит к смене слайда на последующий. Следует принимать во внимание, что по умолчанию микрофон находится во включённом состоянии, а видеочасть – в выключенном. Если запись демонстрации презентации проводится в дистанционном формате, то микрофон и видеочасть не должны совпадать с устройствами, используемыми для вещания.

При организации видеозаписи рекомендуется обращать внимание на то, что большие экраны (электронная доска и т. п.) имеют разрешение Ultra HD 4K (2880x2160), и при подключении к ним презентация будет демонстрироваться и записываться с избыточным размером. Во избежание неэффективного расхода ресурсов рекомендуется перед началом презентации переключать экран в режим FullHD 1080p и экспортировать презентацию с таким же размером.

Для создания видеолекции следует экспортировать запись презентации из файла PowerPoint в формат mp4 (вызывается последовательностью команд меню: «Файл» / «Экспорт» / «Создать видео»). Для слайдов с дизайном, выдержанным в соответствии с приведёнными выше рекомендациями (см. п. 2.1), содержимое производного видеоряда хорошо воспринимается при размере кадра, соответствующем стандарту FullHD 1080p (1080 строк по вертикали с прогрессивной развёрткой). Если размер слайдов (высота x ширина) соответствует пропорциям 4x3 (см. рис. 1), то при их демонстрации на экране в альбомной ориентации с размером 1980x1080, видеоряд формируется с размером 1440x1080 при частоте кадров 30/сек.

При планировании действий необходимо учитывать, что процесс формирования видео из презентации PowerPoint происходит в несколько раз медленнее, чем обработка обычных видеорядов, что существенно снижает привлекательность базового варианта видеолекций.

3.2. СТАНДАРТНЫЙ ПРОФИЛЬНЫЙ ВАРИАНТ ВИДЕОЛЕКЦИЙ

Для создания стандартных вариантов видеолекций требуются наличие высококачественных видеокамер, дополнительного звукового оборудования и времени для обработки данных в видеоредакторе. Стандартные варианты видеолекций привлекательны отсутствием привязки к программе демонстрации презентаций, что даёт возможность учителю использовать в качестве наглядных материалов локальные и сетевые медиаресурсы, а также материальные объекты (модели, препараты и т. п.). Следует иметь в виду, что размещение и подключение стандартного комплекта профессиональной аппаратуры (см. рис. 7) требуют дополнительного времени, что может оказаться неприемлемым в условиях высокой академической мобильности. Для получения высококачественной видеолекции рекомендуется использование профессиональных видеокамер в режиме Full HD.

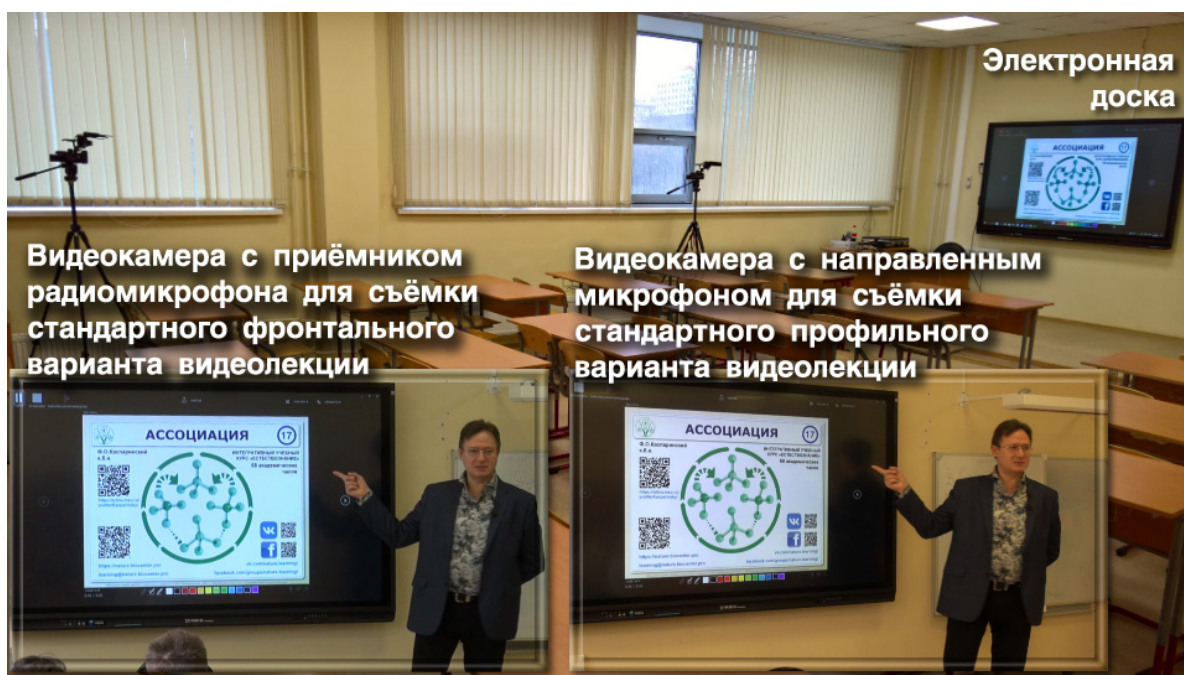


Рисунок 7. Размещение видеокамер для профильной и фронтальной съёмки лекции в аудитории с электронной доской

Стандартные профильные варианты видеолекций целесообразно создавать при дефиците места для размещения аппаратуры и в условиях высокой вероятности возникновения помех, таких как перемещение учащихся по аудитории. Недостатком профильных вариантов видеолекций является трапецевид-

ное искажение изображения наглядных материалов.

Для формирования стандартного профильного варианта видеолекции видеокамера размещается не более чем в 5 метрах от преподавателя слева от доски под углом 45° (см. рис. 7) и используется с направленным кардиоидным микрофоном. В процессе формирования видеолекций курса «Естествознание» [11] использовалась видеокамера SONY HXR-NX30P с направленным кардиоидным электретным микрофоном SONY ECM-XM1.

3.3. СТАНДАРТНЫЙ ФРОНТАЛЬНЫЙ ВАРИАНТ ВИДЕОЛЕКЦИЙ

Стандартный фронтальный вариант видеолекции выгодно отличается от профильного отсутствием трапециевидных искажений изображения наглядных материалов. Для формирования стандартного фронтального варианта видеолекции необходима видеокамера с высококачественным оптическим зумом, которая размещается в дальнем торце аудитории напротив электронной доски (см. рис. 7) и получает аудиоданные с петличного микрофона посредством радиосистемы из передатчика и приёмника. В соответствии с результатами опросов учащихся, стандартный фронтальный вариант видеолекции воспринимается лучше, чем профильный.

В процессе формирования видеолекций курса «Естествознание» [11] были апробированы разные варианты передачи голоса преподавателя с петличных (лавальер, или лапель) микрофонов в систему: через передатчики Bluetooth (OXLasers OX-304C) или UHF (ALLOYSEED, CORKT T1R1, SONY UWP-D16), связанные с соответствующими приёмниками, подключёнными к аналогово-цифровому преобразователю (АЦП) портативной звуковой USB-карты (Vention CDKHB), присоединённой к микрокомпьютеру или электронной доске (см. рис. 8). Последний вариант оказался предпочтительным, поскольку позволял увеличить громкость записи и избавить звук от электромагнитных помех, усиливающихся по мере приближения к компьютеру или устройствам с Bluetooth-интерфейсами, использующими диапазон 2,4 ГГц (клавиатуры, мыши).



Рисунок 8. Подключение радиоприёмника, звуковой карты и USB-носителя данных к электронной доске

Выбор радиосистем может определяться качеством звукопередачи в соответствии со спецификой преподаваемой дисциплины и количеством доступных каналов связи с учётом расписания территориальной организации учебного процесса в организации (комплекты одновременно работающей аппаратуры в радиусе 30–100 м должны быть настроены на разные частоты). При настройке UHF- (УКВ-) радиосистем на территории России необходимо оставаться в пределах свободного диапазона от 506 МГц. По этой причине использование радиокомплекта ALLOYSEED (200 МГц–270 МГц) сопряжено с потенциальными проблемами.

Системы эконом-класса (\$10–\$20) имеют небольшое количество выделенных частот (4–16), что затрудняет их использование в больших педагогических коллективах. Качество звукопередачи радиосистемы CORKT T1R1 (620–690 МГц) оказалось приемлемым для записи занятий с несложной звуковой динамикой. Недорогие радиосистемы для улучшения качества передачи сигнала в ограниченной полосе канала радиочастот используют аналоговые компандеры речи, которые глушат шипящие звуки и создают неприятные на слух артефакты при передаче быстро меняющихся сигналов (например, хлопок в ладоши, звук при

глотке воды или звон колокольчика). При общении на языках с избытком шипящих согласных аналоговые компрессоры формируют иллюзию дефекта речи. Дорогие радиосистемы (\$200–\$2000) используют цифровые аудиопроцессоры (DSP, Digital Sound Processor) с компрессорами, учитывающими сложную динамику звука как в передатчике, так и в приёмнике. Микрофонный выход радиоприёмника присоединяется TRS-кабелем к микрофонному входу компьютера (звуковой карты) непосредственно или через разветвитель AUX Robotsky, позволяющий осуществлять звукозапись на внешнее устройство с целью повышения надёжности формирования комплекта медиаданных (см. рис. 9).



Рисунок 9. Схема подключения звукового оборудования для компьютерной записи видеолекций в аудитории

Для страховки рекомендуется осуществлять дополнительную запись звука внешним диктофоном в формате wav (PCM, стерео, 48 КГц) через микрофонный вход, подключённый к разветвителю AUX Robotsky непосредственно или с устройством для подавления обратной акустической связи микрофона. Резервные аудиоданные используются для замещения искажённых компьютерных звукозаписей (в случае сбоев системы или ошибок при выборе громкости звука).

В больших аудиториях возникает необходимость организации звукоусиления голоса преподавателя. Усиливающая звук акустическая система (колонка) может подключаться к выходу для наушников радиоприёмника или диктофона (см. рис. 9). В первом случае для подавления обратной акустической связи микрофона с колонкой необходимо использовать специальный микрофон (Audio-Technica ATR3350 или Sennheiser ME 4-N). При отсутствии такого микрофона используется второй вариант подключения колонки, при котором между микрофонным выходом радиоприёмника и входом диктофона присоединяется активный фильтр для подавления обратной акустической связи (SnapVox FBX-3 Feedback Howl Eliminator и т. п.).

Выбор петличных микрофонов может определяться особенностями аудитории (тихая, шумная или со звукоусилением) и стилем обращения преподавателя к аудитории. Проблема аудиторного шума звукозаписей всенаправленных микрофонов последовательно минимизируется при переходе к модельному ряду направленных (кардиоидных, суперкардиоидных и гиперкардиоидных микрофонов). Однако постоянное изменение ориентации головы преподавателя (по периметру аудитории и направлению к доске) относительно фиксированного направленного микрофона вызывает существенные изменения громкости записанного звука. Эта проблема может решаться использованием двух направленных микрофонов, укрепленных по разным сторонам воротника. Два микрофона можно присоединить к одному радиопередатчику кабелем с V-образным разветвлением, но такой способ снижает общую громкость звукозаписи вследствие некомпенсированной конкуренции микрофонов за энергоснабжение. Для преодоления проблем с электропитанием при объединении звука двух микрофонов существуют профессиональные микшеры-предусилители сигнала (Comica CVM-D03; Comica AD3; Saramonic-SR-PAX2) или радиокомплекты из двух и более передатчиков, связанных с одним приёмником. Стоимость микшеров (\$50–\$150) сопоставима с радиосистемами бюджетного класса, такими как XTUGA X-CM012 (диапазон частот 550–580 МГц). Среди радиосистем с несколькими передатчиками в среднем ценовом классе (\$250–\$500) по соотношению цена/качество выделяется Mailada WM-10 (диапазон частот 538–594 МГц).

Лучшими петличными микрофонами по соотношению цена/качество оказались всенаправленный микрофон из радиокомплекта CORKT T1R1, а также

кардиоидные лавальеры MiCWL BC03 и Xgwh L777. При выборе комплектов радиопередатчиков и вариантов разъемов микрофонов (TS – 2-контактный, TRS – 3-контактный, TRRS – 4-контактный) важно проверять их совместимость с микрофонным портом компьютера или звуковой карты (TRS или TRRS). Разъемы TRS (Tip-Ring-Sleeve) совместимы со звуковыми картами компьютеров, фото- и видеокамерами, а TRRS-интерфейсы (Tip-Ring-Ring-Sleeve) адаптированы для смартфонов.

В процессе формирования видеолекций курса «Естествознание» [11] в качестве эталона использовалась трансляция звука на видеокамеру SONY HXR-NX30P с петличного суперкардиоидного электретного микрофона SONY через радиосистему премиум-класса SONY UWP-D16 (диапазон частот 794–805 МГц). Микрофоны (см. пп. 3.2 и 3.3) присоединялись к камере посредством адаптера XLR-A1M с фантомным питанием. Оба варианта обеспечивали реалистичное качество передачи и записи структурно различных аудиоданных, от речи до звона караталов.

3.4. ДИДАКТИЧЕСКИ ОПТИМИЗИРОВАННЫЙ ВАРИАНТ ВИДЕОЛЕКЦИЙ

При отсутствии видеокамер с высококачественной оптикой, обеспечивающей сохранение детализации наглядных материалов, можно использовать технологию монтажа дидактически оптимизированных вариантов видеолекций. Для формирования исходных монтажных материалов требуется создание базового варианта видеолекции одновременно с видеозаписью действий преподавателя одной или двумя видеокамерами с разных ракурсов.

Целенаправленная дидактическая адаптация видеолекции осуществляется посредством контекстного монтажа кадров ближнего и дальнего планов видеозаписи с базовым вариантом видеолекции и дополнительными медиаматериалами (см. рис. 10, пример: <https://vimeo.com/204236066>). Монтаж должен обеспечивать акцентирование внимания учащихся на процессе графического комментирования, эмоциональной реакции и жестикуляции лектора. Для облегчения синхронизации аудиовизуальных и звуковых рядов при их монтаже перед началом демонстрации презентации рекомендуется формировать сигнал, хорошо различимый на аудиограмме (двойной хлопок). Главная проблема при создании дидактически оптимизированных вариантов видеолекций – трудоём-

кость их создания (10-кратный расход времени, исходя из общего хронометража записи).

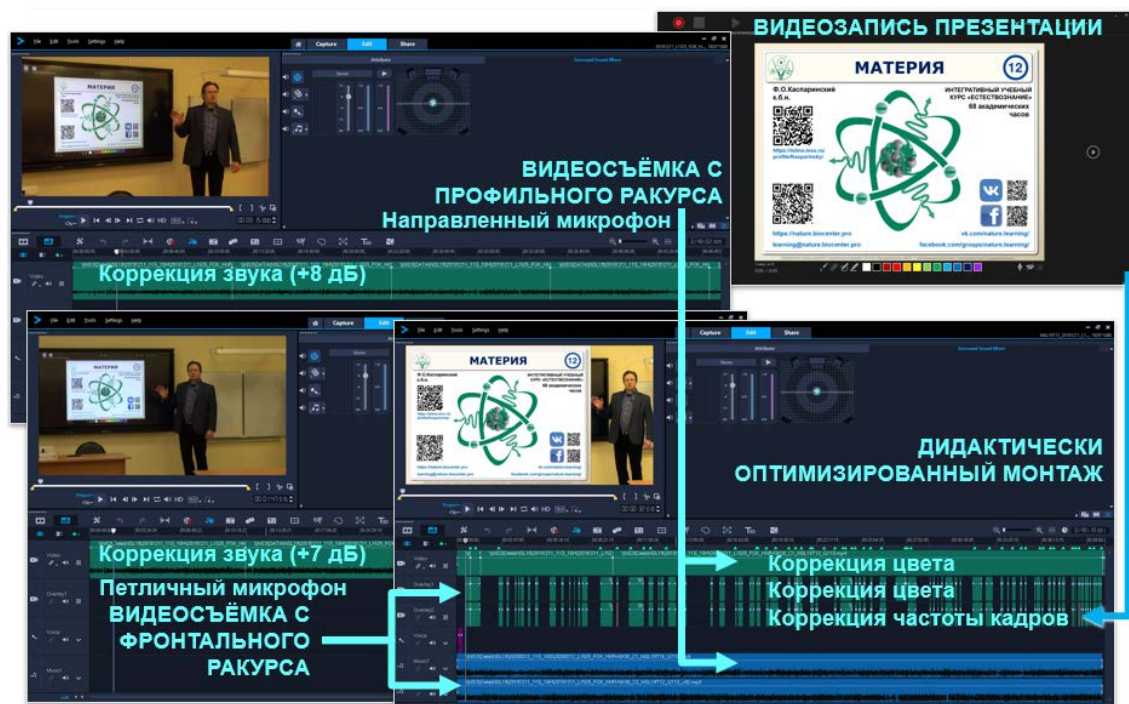


Рисунок 10. Схема рабочих процессов при создании дидактически оптимизированных видеолекций

3.5. ИНТЕГРАТИВНЫЙ ВАРИАНТ ВИДЕОЛЕКЦИЙ

Интегративные варианты видеолекций (совмещение видеоокна с записью действий преподавателя и презентации с графическими и голосовыми комментариями) создаются посредством USB-видеокамеры (вебкамеры), подключённой к презентационному компьютеру. При очной форме преподавания вебкамера позиционируется на штативе таким образом, чтобы в поле кадра попадали преподаватель и вся электронная доска (см. рис. 11). Звукоряд для записи может поступать как с вебкамеры, так и с приёмника радиосистемы от петличного микрофона (оптимальный вариант, см. п. 3.3). Во втором варианте перед началом записи необходимо скорректировать источник звукозаписи в настройках записи слайд-шоу PowerPoint (см. рис. 11).

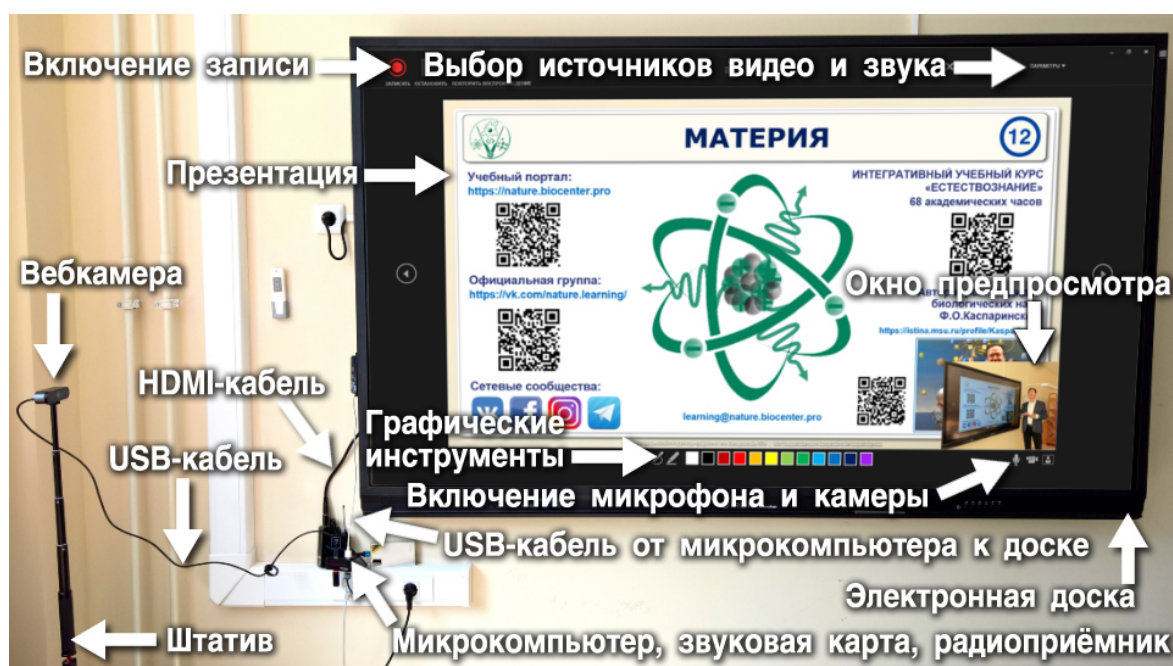


Рисунок 11. Организация видеозаписи презентации PowerPoint (Microsoft 365) на электронной доске и действий преподавателя (опционально)

При включении видеокамеры в режиме записи процесса демонстрации презентации PowerPoint из комплекта Microsoft 365 появляется окно с видеокадром, заслоняющее содержимое нижней правой части слайда (см. рис. 11). Видео записывается с размером 640x480 пикселей, но отображается в прямоугольнике с соотношением сторон 360x270 пикселей при полноэкранном режиме просмотра с высотой 1080 пикселей. Кнопкой отключения отображения видеоокна можно избавиться от присутствия видеоокна во время записи демонстрации презентации при очной форме преподавания. Во время онлайн-занятий видеоокно с изображением преподавателя целесообразно оставлять включённым и транслировать ученикам не окно презентации PowerPoint, а рабочий стол преподавателя. В этом случае учащиеся получают возможность видеть и слышать преподавателя одновременно с наблюдением за формированием графических комментариев на слайдах. Результаты видеозахвата графических комментариев и изображения действующего преподавателя независимы друг от друга и не изменяются при включении или выключении режима просмотра кадра вебкамеры.

После выхода из режима записи слайд-шоу видеоокно с записью действий

преподавателя получает возможность перемещения и масштабирования. Если в дизайне презентации предусмотрена интеграция видеозаписи действий преподавателя на стадии экспорта mp4-файла из PowerPoint в специально зарезервированной зоне слайда (см. п. 2.1, рис. 11 и рис. 12), то следует учитывать, что по умолчанию видеоокно с размером 360x270 автоматически размещается в правой нижней части слайда (рис. 11) и громкость аудиоряда уменьшается до 75%. Перед экспортом видеолекции позицию окна с записью действий преподавателя и громкость аудиоряда на каждом слайде необходимо скорректировать.



Рисунок 12. Экранный снимок кадра интегративной видеолекции с графическими комментариями презентации и действующим преподавателем

Каждый слайд с интегративной видеозаписью может быть экспортирован в отдельный видеоряд. Слайды с интегративными видеозаписями можно комбинировать в любой последовательности в зависимости от дидактических целей. С учётом этого обстоятельства перед переключением слайдов целесообразно выдерживать акустическую паузу.

Главные недостатки способа создания интегративных видеолекций посредством функционала «Запись слайд-шоу» PowerPoint – блокировка обратного направления навигации по слайдам, невозможность демонстрации наглядных

материалов помимо презентации и длительный процесс экспорта mp4-видео.

Альтернативный способ формирования интегративных видеолекций позволяет освободиться от жесткой привязки к презентациям PowerPoint и в реальном времени совмещать видеоряд вебкамеры в произвольной зоне с любым содержимым экрана посредством специализированной программы видеозахвата экрана HyperCam, разработанной в компании Solveig Multimedia из Томска. При использовании HyperCam преподаватель получает возможность демонстрировать слайды PowerPoint в любой последовательности и переключаться на другие приложения, что предоставляет лектору беспрецедентную свободу использования любых наглядных материалов. По окончании записи занятия mp4-файл сразу направляется в редактор HyperCam Media Editor, позволяющий обрезать ненужные концы записи и сохранить интегративную видеолекцию без длительного перекодирования, за несколько секунд. Для предотвращения рассинхронизации аудиального и визуального рядов целесообразно производить обрезку строго по ключевым кадрам. Программа HyperCam выгодно отличается от аналогов малой ресурсоёмкостью и высоким качеством компрессии аудиовизуальных рядов. К примеру, интегрированная видеолекция с размером кадра 1920x1080 (Full HD), сформированная на уроке продолжительностью 45 минут, помещается в mp4-файле размером 500 Мб. Для бесперебойной работы рекомендуется компьютер с оперативной памятью не менее 8 Гб и твердотельным накопителем данных (SSD).

Опыт показывает, что в процессе очного обучения подключение комплекта аппаратуры, загрузка операционной системы, презентации и программы видеозахвата занимают несколько минут, что обеспечивает преподавателю возможность поддерживать академическую мобильность в соответствии с современными требованиями и минимальными затратами ресурсов организовать оперативное создание собственных видеолекций во время занятия. Таким образом, наивысшее соотношение «дидактическое качество/ресурсоёмкость» достигается при формировании интегративных вариантов видеолекций, которые готовы к размещению на сетевых сервисах уже через десяток секунд по окончании занятия, в процессе которого велась видеозапись.

Одновременный захват видео в PowerPoint и HyperCam осуществлять не

рекомендуется по причине высокой вероятности возникновения сбоев системы.



Рисунок 13. Экранный снимок рабочего стола преподавателя в процессе онлайн-занятия с демонстрацией презентации PowerPoint

В случае применения high-touch формы дистанционного обучения [6] использование полноэкранного режима демонстрации PowerPoint неприемлемо, поскольку в процессе занятия необходим постоянный визуальный контакт с учащимися одновременно с показом и графическим комментированием презентации. При формировании видеолекции с использованием материалов онлайн-занятий появляется еще одна проблема – отсутствие или малая величина окна, отображающего действия преподавателя в приложении для видеоконференций. Решить проблемы комплексно (обеспечить проведение high-touch занятия одновременно с формированием высококачественного контента видеолекции для последующего использования в асинхронном режиме high-tech) позволяет использование двойного комплекта аудиовизуальной аппаратуры в сочетании с оконным режимом работы PowerPoint и видеозахватом рабочего стола приложением HyperCam. Для работы следует использовать презентацию PowerPoint с дизайном слайдов 4:3, переключать её с полноэкранного на оконный режим, масштабировать окно при полной высоте на 60% ширины (1200 пикселей) экрана с разрешением FullHD (1920x1080) и располагать его слева. Правая часть

экрана занимается окном приложения для видеоконференций с окнами участников онлайн-занятия (см. рис. 13).

Следует иметь в виду, что в режиме записи слайдов при вызове оконного режима презентации происходит дублирование в окне при сохранении полноэкранный «зеркала», маскирующего собой окно приложения видеоконференций. Для обнаружения окна видеоконференции и его перемещения поверх полноэкранный «зеркала» презентации следует использовать клавиатурную комбинацию клавиш Alt+Tab. Запись занятия целесообразно осуществлять посредством полноэкранный видеозахвата рабочего стола специальным приложением HyperCam (версия от 5 и выше). В этом приложении следует выбрать микрофон и видеокамеру, отличные от задействованных в видеоконференции, и в режиме предварительного просмотра устанавливать расположение окна видеозаписи действий преподавателя поверх средней части окна видеоконференций (720x405 или 720x540, в зависимости от настроек камеры). Окно видеозаписи действий преподавателя после активации захвата становится невидимым и в фоновом режиме совмещается с видеозаписью демонстрации рабочего стола в файле формата mp4.

При необходимости, для соблюдения требований ФЗ-152 от 27.07.2006 «О персональных данных», в видеоредакторе можно установить расфокусировку изображения на остающихся видимыми фрагментах окна видеоконференции с лицами участников занятия. При использовании эффектов и фонового звукового сопровождения звуковой карты звукоряда видеозаписи и трансляции отличаются, поэтому при итоговом монтаже видеолекции захваченный звукоряд следует замещать записью с внешнего записывающего устройства, получающего аудиоданные со звуковой карты. В результате экспорта из видеоредактора получается mp4-файл дидактически полноценной видеолекции, созданной из медиакомпонентов в процессе дистанционного занятия. Использование записей видеоконференционных систем в качестве источника медиаресурсов для видеолекций оказалось неэффективным вследствие несоответствия стандартам качества.

4. СОЗДАНИЕ ИНТЕГРАТИВНЫХ ВАРИАНТОВ ВИДЕОЛЕКЦИЙ В ПРОЦЕССЕ ОНЛАЙН-ПРЕПОДАВАНИЯ

Наиболее простой способ формирования интегративной видеолекции в процессе дистанционных занятий – использование режима вещания содержимого рабочего стола компьютера преподавателя при записи демонстрации оконной или полноэкранный презентации PowerPoint, в дизайне которой предусмотрена вставка видеокна с записью действий преподавателя (см. рис. 3). Следует помнить, что для дидактически полноценного онлайн-преподавания рекомендуется включать отображение кадра вебкамеры в настройках презентации PowerPoint при включенном отображении кадра вебкамеры. Режим отображения кадра вебкамеры можно изменять в процессе записи презентации без каких-либо последствий для качества интегративной видеолекции.

При онлайн-вещании окна презентации PowerPoint на компьютеры пользователей в режиме «Запись-слайд-шоу» передаётся только область контента.

Необходимо учитывать, что сетевое вещание поступающих с одного устройства видео- и аудиоданных одновременно с их записью на локальный компьютер вызывает нестабильность его работы на больших временах. Для одновременного вещания и высококачественной записи онлайн-занятия следует использовать два комплекта аудиовизуальной аппаратуры. При отсутствии второй вебкамеры необходимо отключать от вещания вебкамеру до активации режима «Запись слайд-шоу».

Основные критерии при выборе видеокамер для онлайн-вещания (вебкамер) в процессе дистанционного обучения – возможность записи видео в формат H.264 с размером 1920x1080 и частотой 30 кадров/сек, автофокусировка с коррекцией баланса белого цвета, наличие системы активного шумоподавления с двумя микрофонами, присутствие стандартного штативного крепления 1/4", возможность замены USB-кабеля или выбора его подходящей длины для подключения к компьютеру с позиции, оптимальной для съёмки. Если предполагается съёмка действий преподавателя на фоне яркого экрана электронной доски без дополнительного освещения (прожекторов), видеокамеры должны поддерживать режим записи с широким динамическим диапазоном освещённости (HDR, High Dynamic Range или более дорогой вариант WDR, Wide Dynamic Range).

Для работы с видеоданными при создании интегративных видеолекций курса «Естествознание» [11] использовались вебкамеры Logitech BR10 и Logitech C920 (см. рис. 14) с поддержкой потоковой передачи данных, подключённые к одному из двух задействованных компьютеров, сконфигурированных [18] для работы с аудиовизуальными материалами (8 Гб оперативной памяти, дискретная графическая карта, монитор 27"–28" с MVA-матрицей). Было установлено, что направление взгляда преподавателя, формирующего графические комментарии слайдов, кажется направленным прямо на аудиторию, если основной монитор расположен выше вебкамер (см. рис. 14). Мониторы Iiyama X2783HSU-B3 (верхний, основной, отображает формирование транслируемого контента) и X2888HS-B2 (нижний, вспомогательный, служит для мониторинга качества приёма) располагались в альбомной ориентации по вертикали на кронштейне KRON D421FS. Оперативное переключение аудиовизуальных сигналов между мониторами удобно осуществлять посредством матричного HDMI-коммутатора 2x2 Navceker ZY-HM22.



Рисунок 14. Оборудование рабочего места преподавателя для онлайн-преподавания

В зазоре между мониторами на уровне глаз преподавателя укреплялись

вебкамеры на шарнирном креплении Fasdga SZ-GTFS-I020833 с горизонтальной расширительной планкой Ulanzi PT-2 или Neewer 10082556 с двойным креплением. Альтернативный способ размещения вебкамер – посредством кубического крепления Wellchosen и пары рычагов Smallrig 2157 или Jadkinsta WA0196 с двумя шаровыми головками (см. рис. 14). Этот способ позволяет наклонять вебкамеры для обеспечения обзора физического рабочего стола преподавателя, что позволяет демонстрировать материальные объекты в процессе обучения. В режиме многокамерной трансляции посредством программы Logitech Capture возможно использовать оперативное переключение источника вещания с целью смены ракурса или объекта демонстрации. При однокамерной трансляции функции вещания и записи контента целесообразно контролировать во избежание возникновения конфликтного использования одного ресурса. Вебкамера Logitech C920 использовалась для видеотрансляции, а Logitech BRIO применялась для формирования окна с видеозаписью действий преподавателя (см. рис. 14), комбинируемого с видеозахватом содержимого экрана посредством программы HyperCam, разработанным компанией SolveigMultimedia из Томска. Было установлено, что звукоряд, формируемый микрофоном вебкамеры Logitech BRIO, обеднён в низкочастотной области спектра, в связи с чем для создания качественных видеолекций целесообразно использование специального микрофона.

В качестве альтернативных способов ввода аудиоданных использовались внешние микрофоны, укрепленные на пантографах Neewer NB-35 и NB-39 с амортизаторами «паук» (см. рис. 14). Студийный конденсаторный кардиоидный микрофон Shure PG42-USB со встроенным АЦП передавал данные непосредственно в компьютер. Использование студийного микрофона оказалось неэффективным вследствие его чрезмерной чувствительности к бытовым шумам. Для минимизации влияния бытовых шумов на качество звука при проведении дистанционных занятий целесообразно использовать направленные кардиоидные и суперкардиоидные микрофоны.

Направленный суперкардиоидный микрофон Comica CVM-V30 LITE, непосредственно подключаемый к аудиовходу компьютера без фантомного питания, оказался непригоден для создания видеолекций профессионального уровня вследствие низкой чувствительности. Направленный кардиоидный высокочув-

ствительный микрофон SONY ECM-XM1 с низким уровнем собственных шумов требует фантомного питания (48 В), что препятствует его непосредственному использованию с компьютером. Присоединение микрофона SONY ECM-XM1 через микшер с фантомным питанием (Muslady D6), подключённый к микрофонному входу компьютера, оказалось неэффективным вследствие высокого уровня электромагнитных помех, которые снижались включением шумоподавителя (Alctron NM-2) в цепь между микшером и компьютером.

Подключение микрофона SONY ECM-XM1 к микрофонному входу компьютера через портативный предусилитель (Comica AD2, Comica AD3 или Saramonic SR-PAX2), обеспечивающий фантомное питание за счет автономного источника энергии (батарея 9В), позволило снизить шумы и увеличить чувствительность звукозаписи. Недостатком этого варианта является необходимость своевременной замены элемента питания. Оптимальный выбор предусилителя с автономным питанием – Saramonic SR-PAX2, в котором присутствует индикация оставшегося заряда батареи.

Наиболее практичным вариантом для стационарных условий оказалось подключение микрофона SONY ECM-XM1 к микрофонному входу компьютера через обеспечивающий фантомное питание предусилитель (Alctron RD501) с электроснабжением от сети 220В. Увеличение качества и громкости звукозаписи оказалось возможным во всех случаях при направлении аудиоданных не в микрофонный вход компьютера, а во внешнюю звуковую карту с собственным АЦП и последующей передачей данных через USB-порт (см. рис. 15).

В процессе записи голоса направленными микрофонами использовать динамики для формирования звукового фона не рекомендуется вследствие неизбежного искажения его амплитудно-частотных характеристик. При онлайн-вещании в режиме трансляции рабочего стола можно демонстрировать аудитории аудиовизуальные ресурсы, но использовать их для фонового оформления нецелесообразно вследствие отсутствия эффективных способов оперативного управления громкостью воспроизведения, вариациями тембра и пр.

Выбор звуковой карты для онлайн-преподавания определяется спецификой предмета. Для улучшения качества микрофонной звукозаписи (снижение уровня шума до -50 дБ) оказалось достаточно использовать устройство базового

ценового уровня (\$ 10) с надёжным экранированием от электромагнитных помех и функцией интеллектуального шумоподавления (Vention CDKHB).



Рисунок 15. Схема подключения звукового оборудования для онлайн-преподавания

Применение специализированных для онлайн-вещания USB-звуковых карт с АЦП в ценовой категории \$50 (ALLOYSEED v8; iTENCHOR 2880486; K-Song Live F007; MiCWL-M10; SZKOSTON 802575-S8) обеспечивает дополнительное снижение уровня базового шума (менее -70 дБ), управление микшированием звука двух микрофонов и фонового музыкального сопровождения (см. рис. 15). Опыты показали, что карта K-Song Live F007 обеспечивает активную фильтрацию фоновых шумов при минимальном искажении звукопередачи голоса (эффекты KTV+Chat). Если в работе требуются дополнительные звуковые эффекты высокого качества, рекомендуется использовать звуковые карты в ценовой категории \$100 и выше (Tenlamp G3; TKL R1).

При выборе оптимальной модели звуковой карты следует обращать внимание на эргономичность регулирования уровня звука микрофонов, фонового звукового сопровождения и наушников (используются для прослушивания звуков онлайн-аудитории и фонового сопровождения), анализировать набор звуко-

вых эффектов (аплодисменты, смех, эмоциональные возгласы) и их качество (громкость и аккуратное исполнение – плавное начало и завершение), а также отсутствие искажений в режиме записи оригинального голоса. Фоновое звуковое сопровождение с внешнего аудиоплеера, компьютера или смартфона вводится в звуковую карту через выделенный TRS порт (см. рис. 15). Звуковые карты HiFi-уровня позволяют управлять громкостью звукового сопровождения в ручном или автоматическом режиме (Dodge-эффект: обратимое снижение громкости при появлении звука с микрофона); отсекают вокал и изменять амплитудно-частотные характеристики звукоряда.

После подключения звуковой карты к компьютеру рекомендуется в компьютерной Панели управления звуком (активируется последовательностью команд: «Параметры» / «Звук» / «Сопутствующие параметры» / «Панель управления звуком» / «Запись» / «КАРТА» / «Свойства» / «Дополнительно» / «Формат по умолчанию») перевести карту в режим записи с разрядностью и частотой дискретизации, соответствующим параметрам записи звука при видеозаписи (2 канала, 16 бит, 48000 Гц), чтобы избежать несовпадения при монтаже видеолекций.

Мониторинг звука сетевой аудитории осуществлялся через наушники (см. рис. 15), присоединённые к соответствующему выходу внешней звуковой карты, позволявшей аппаратным способом оперативно варьировать громкость звука по мере необходимости. Современные звуковые карты имеют один цифровой (USB-B) и два аналоговых звуковых линейных TRRS-выхода, специализированных для альтернативного подключения к смартфонам Android и iPhone, а также TRS- и TRRS-выходы для наушников и гарнитуры, соответственно. Звук целесообразно одновременно выводить через USB-B и один из линейных TRRS-портов. Через порт, специализированный для смартфонов Android, можно выводить звуковой сигнал на TRS-вход внешних устройств, к примеру, целесообразно осуществлять дополнительную запись звука диктофоном Tascam DR-40 в формате wav (PCM, стерео, 48 КГц).

Для оперативного мониторинга оптимальной громкости и выбора параметров эквалайзеров микрофона и звукового сопровождения рекомендуется использовать визуализаторы общего уровня громкости (Kaisaya Dual 40 Spectrum

VU Meter) и амплитудно-частотного спектра (Kaisaya OLED Music Spectrum Display). В этом случае визуализатор спектра и диктофон можно подключать к трёхполосному разветвителю акустических сигналов (MuxBOXS MB-SH03DA), который присоединяется к линейному TRS-порту или к TRRS-выходу для подключения смартфонов Android (см. рис. 15). Оптимальный вариант подключения индикатора уровня громкости – через выход звуковой карты для гарнитуры.

В связи с вероятностью возникновения акустических помех при включении и работе вспомогательного электрооборудования (источники питания, лампы дневного света, климатическая техника и пр.) для сохранения высокого качества звука передача аналоговых данных осуществлялась посредством экранированных кабелей Vochara, GOLLE YLION и KUYIONIFI.

Для управления компьютерами целесообразно использовать один комплект клавиатуры и мыши, поддерживающий технологию Logitech Flow с переключением фокуса управления при пересечении чувствительных зон на краях экранов. Надёжность и комфорт при длительной работе обеспечивают сборные комплекты «клавиатура и мышь» с приёмниками Logitech Unifying: Logitech K375s и Logitech MX Anywhere 2S (мобильный вариант), а также Logitech Craft и Logitech MX Master 2S (стационарный вариант). Графические комментарии интегративных онлайн-видеолекций курса «Естествознание» [11] создавались в процессе презентации посредством планшета Wacom Intuos A3.

Эксперименты с дополнительным освещением (4 светодиодных прожектора Yongnuo YN-300III) в разных режимах показали целесообразность их использования со светорассеивающими фильтрами на минимальной яркости (1–3%). Этого достаточно для формирования интегративных онлайн-видеолекций с частотой кадров 30 кадров в секунду. Режим освещенности, обеспечиваемой двумя компьютерными экранами (27”–28”) при использовании светлой темы интерфейса приложений, неприемлем для создания интегративных видеолекций. В этих условиях частота регистрации видеокадров начинает варьировать от 25 до 12 кадров в секунду в зависимости от содержания наглядных материалов и положения лектора, что приводит к дидактически неприемлемой порче качества интегративных видеолекций вследствие рассинхронизации аудиального и визуального рядов.

ЗАКЛЮЧЕНИЕ

В условиях академической мобильности, предусматривающей бесшовную интеграцию очной и дистанционной форм образования, целесообразно использовать базовые и интегративные варианты видеолекций. Дидактически полноценными являются интегративные варианты видеолекций, в которых графические и звуковые комментарии демонстрации наглядных материалов дополнены видеозаписью действий преподавателя, вставленной в специально предназначенное место базового варианта видеолекции.

Наиболее надёжным приложением для формирования интегративных видеолекций является средство презентаций PowerPoint из комплекта Microsoft 365 с режимом «Запись слайд-шоу». Наилучшее решение для создания интегративных видеолекций в условиях дефицита времени и при использовании разнообразных приложений – использование специализированная программа видеозахвата экрана HyperCam.

СПИСОК ЛИТЕРАТУРЫ

1. Каспаринский Ф.О., Полянская Е.И. Видеолекция как жанр // Качество дистанционного образования: концепции, проблемы, решения (DEQ–2013). Межвузовский сборник научных трудов. М.: МГИУ, 2013. С. 80–92.

URL: <https://istina.msu.ru/publications/article/3559014/>.

2. Каспаринский Ф.О., Полянская Е.И. Профессиональные интернет-видеолекции // Научный сервис в сети Интернет: многообразие суперкомпьютерных миров. Труды Международной суперкомпьютерной конференции (22–27 сентября 2014 г., г. Новороссийск). М.: Издательство МГУ, 2014. С. 432–435.

URL: <https://agora.guru.ru/abrau2014/pdf/432.pdf>.

3. Gallardo D. Easy navigation for your videos with Chapters // Vimeo blog, 29.01.2020. URL: <https://vimeo.com/blog/post/easy-navigation-for-your-videos-with-chapters/>.

4. Каспаринский Ф.О., Полянская Е.И. Информационно-навигационный сервис сетевых аудиовизуальных ресурсов // Научный сервис в сети Интернет: труды XX Всероссийской научной конференции (17–22 сентября 2018 г., г. Новороссийск). М.: ИПМ им. М.В. Келдыша, 2018. С. 284–294.

URL: <https://doi.org/10.20948/abrau-2018-42>.

5. *Каспаринский Ф.О., Полянская Е.И.* Дидактически целенаправленное использование информационного инструментария // Качество дистанционного образования: концепции, проблемы, решения (DEQ–2012). Межвузовский сборник научных трудов. М.: МГИУ, 2012. С. 74–85.

URL: <https://istina.msu.ru/publications/article/1769099/>.

6. *Каспаринский Ф.О., Полянская Е.И.* Организация high-touch формы дистанционного обучения посредством Skype-видеоконференций // Качество дистанционного образования: концепции, проблемы, решения (DEQ–2015). Материалы Международной конференции 11 декабря 2015 г. г. Жуковский: АНО ВО «Международный институт менеджмента ЛИНК», 2016. С. 42–45.

URL: https://www.elibrary.ru/download/elibrary_27225408_77128884.pdf .

7. *Kasparinsky F.* The Organization of the Use of Audiovisual Recordings of Synchronous Lessons in the Process of Distance Learning // CEUR Workshop Proceedings. 2020. Vol. 2543. P. 172–184.

URL: <http://ceur-ws.org/Vol-2543/rpaper16.pdf>.

8. *Каспаринский Ф.О.* Аудиовизуальная запись синхронных занятий при очном и дистанционном обучении // Электронные библиотеки. 2020. Т. 23, вып. 3. Тематический выпуск «Научный сервис в с-ети Интернет». Часть 1, 2020. С. 451–472. URL: <https://elbib.ru/article/view/598/693>.

9. *Каспаринский Ф.О.* Универсальная адаптация онлайн-учебников к разнообразным пользовательским устройствам // Качество открытого дистанционного образования: концепции, проблемы, решения (DEQ–2017). Молодежь и наука. Материалы XIX международной научно-практической конференции. г. Жуковский: Международный институт менеджмента ЛИНК, 2018. С. 94–98. URL: <https://elibrary.ru/item.asp?id=32696165>.

10. *Каспаринский Ф.О.* Публикация интернет-ресурсов дистанционного обучения в соответствии со стандартом BYOD // Качество открытого дистанционного образования: концепции, проблемы, решения (DEQ–2017). Молодежь и наука. Материалы XIX международной научно-практической конференции. г. Жуковский: Международный институт менеджмента ЛИНК, 2018. С. 89–94.

URL: <https://elibrary.ru/item.asp?id=32696164>.

11. *Каспаринский Ф.О.* Интернет-сервис интегративного учебного курса

«Естествознание» // Научный сервис в сети Интернет: труды XXII Всероссийской научной конференции (21–25 сентября 2020 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2020. С. 307–361. URL: <https://doi.org/10.20948/abrau-2020-19>.

VARIETIES OF VIDEO LECTURES

F. O. Kasparinsky¹[0000-0002-1048-9212]

¹ MASTER-MULTIMEDIA LLC, Entuziastov Shosse 98-3-274, Moscow, 111531

¹ felix@kasparinsky.pro

Abstract

The article is devoted to the typology of the main variants of video lectures (<https://vimeo.com/showcase/6184718>), which differ in didactic value and laboriousness of creation. For each option of video lectures, the key conditions for effective formation and use are determined. Particular attention is paid to the optimization of hardware for the creation of various options for video lectures in full-time and distance learning.

As a basic option, a video recording of a PowerPoint presentation with speech and graphic comments on the slides is considered. The screen (electronic board) videorecording on the left with the lecturer in profile, named as standard profile version, is recommended for use when there is a shortage of space in the classroom. In the presence of a video camera with optics that ensure the preservation of the detail of images of visual materials, the preferred option for video lecture is a standard frontal one, filmed from the far end of the audience. Didactically optimized editing of video lectures allows you to combine high-quality recording of visual materials of the basic version with mediocre video recordings of the teacher's actions.

Integrative versions of video lectures are created by overlaying the video recording of the teacher's actions over a specially designated invariant place in the design of the slides of the basic version of the video lecture. The options for adapting the design of PowerPoint presentations to form a basic and integrative version of video lectures are proposed. The highest ratio of "didactic quality / labor costs" is achieved when creating integrative video lectures directly in the process of face-to-face and distance teaching.

Keywords: *video lecture, integrative video lecture, training video creation, distance learning, face-to-face training, work environment, Microsoft PowerPoint, HyperCam.*

REFERENCES

1. *Kasparinsky F.O., Polyanskaya E.I.* Videolektsiia kak zhanr // Kachestvo distantsionnogo obrazovaniia: kontseptsii, problemy, resheniia (DEQ–2013). Mezhvuzovskii sbornik nauchnykh trudov. M.: MGIU, 2013. S. 80–92.

URL: <https://istina.msu.ru/publications/article/3559014/>.

2. *Kasparinsky F.O., Polyanskaya E.I.* Professionalnye internet-videolektsii // Nauchnyi servis v seti Internet: mnogoobrazie superkompiuternykh mirov. Trudy Mezhdunarodnoi superkompiuternoii konferentsii (22–27 sentiabria 2014 g., g. Novorossiisk). M.: Izdatelstvo MGU, 2014. S. 432–435.

URL: <https://agora.guru.ru/abrau2014/pdf/432.pdf>.

3. *Gallardo D.* Easy navigation for your videos with Chapters // Vimeo blog, 29.01.2020. URL: <https://vimeo.com/blog/post/easy-navigation-for-your-videos-with-chapters/>.

4. *Kasparinsky F.O., Polyanskaya E.I.* Informatsionno-navigatsionnyi servis setevykh audiovizualnykh resursov // Nauchnyi servis v seti Internet: trudy XX Vserossiiskoi nauchnoi konferentsii (17–22 sentiabria 2018 g., g. Novorossiisk). M.: IPM im. M.V. Keldysha, 2018. S. 284–294. URL: <https://doi.org/10.20948/abrau-2018-42>.

5. *Kasparinsky F.O., Polyanskaya E.I.* Didakticheski tselenapravlennoe ispolzovanie informatsionnogo instrumentariia // Kachestvo distantsionnogo obrazovaniia: kontseptsii, problemy, resheniia (DEQ–2012). Mezhvuzovskii sbornik nauchnykh trudov. M.: MGIU, 2012. S. 74–85.

URL: <https://istina.msu.ru/publications/article/1769099/>.

6. *Kasparinsky F.O., Polyanskaya E.I.* Organizatsiia high-touch formy distantsionnogo obucheniia posredstvom Skype-videokonferentsii // Kachestvo distantsionnogo obrazovaniia: kontseptsii, problemy, resheniia (DEQ–2015). Materialy Mezhdunarodnoi konferentsii 11 dekabria 2015 g. g. Zhukovskii: ANO VO «Mezhdunarodnyi institut menedzhmenta LINK», 2016. S. 42–45.

URL: https://www.elibrary.ru/download/elibrary_27225408_77128884.pdf.

7. *Kasparinsky F.* The Organization of the Use of Audiovisual Recordings of Syn-

chronous Lessons in the Process of Distance Learning // CEUR Workshop Proceedings. 2020. Vol. 2543. P. 172–184.

URL: <http://ceur-ws.org/Vol-2543/rpaper16.pdf>.

8. *Kasparinsky F.O.* Audiovizualnaia zapis sinkhronnykh zaniatii pri ochnom i distantsionnom obucheniiakh // *Elektronnye biblioteki*. 2020. T. 23, No. 3. Tematicheskii vypusk «Nauchnyi servis v seti Internet». Chast 1, 2020. S. 451–472.

URL: <https://elbib.ru/article/view/598/693>.

9. *Kasparinsky F.O.* Universalnaia adaptatsiia onlain-uchebnikov k raznoobraznym polzovatelskim ustroistvam // *Kachestvo otkrytogo distantsionnogo obrazovaniia: kontseptsii, problemy, resheniia (DEQ–2017)*. Molodezh i nauka. Materialy XIX mezhdunarodnoi nauchno-prakticheskoi konferentsii. Zhukovskii: Mezhdunarodnyi institut menedzhmenta LINK, 2018. S. 94–98.

URL: <https://elibrary.ru/item.asp?id=32696165>.

10. *Kasparinsky F.O.* Publikatsiia internet-resursov distantsionnogo obucheniiia v sootvetstvii so standartom BYOD // *Kachestvo otkrytogo distantsionnogo obrazovaniia: kontseptsii, problemy, resheniia (DEQ–2017)*. Molodezh i nauka. Materialy XIX mezhdunarodnoi nauchno-prakticheskoi konferentsii. g. Zhukovskii: Mezhdunarodnyi institut menedzhmenta LINK, 2018. S. 89–94.

URL: <https://elibrary.ru/item.asp?id=32696164>.

11. *Kasparinsky F.O.* Internet-servis integrativnogo uchebnogo kursa "Estestvoznaznie" // *Nauchnyi servis v seti Internet: trudy XXII Vserossiiskoi nauchnoi konferentsii (21–25 sentiabria 2020 r., online)*. M.: IPM im. M.V. Keldysha, 2020. S. 307–361.

URL: <https://doi.org/10.20948/abrau-2020-19>.

СВЕДЕНИЯ ОБ АВТОРЕ



КАСПАРИНСКИЙ Феликс Освальдович – кандидат биологических наук, основатель и научный руководитель Лаборатории мультимедийных технологий Биологического факультета МГУ имени М.В. Ломоносова, учредитель и Генеральный директор ООО «МАСТЕР-МУЛЬТИМЕДИА» Сфера научных интересов – формирование информационной среды, дидактически целенаправленное использование мультимедийных технологий.

Felix Oswaldovich KASPARINSKY – Candidate of Biological Sciences, Founder and Scientific Director of Multimedia Technologies Laboratory (Biological Faculty, M.V. Lomonosov Moscow State University), Founder and General Director of MASTER-MULTIMEDIA LLC. Research interests include creating an information environment and didactically targeted use of multimedia technologies.

email: felix@kasparinsky.pro; ORCID: 0000-0002-1048-9212

Материал поступил в редакцию 27ноября 2020 года

УДК 004

СРЕДСТВА ИНТЕРАКТИВНОГО ВЗАИМОДЕЙСТВИЯ С ПОЛЬЗОВАТЕЛЕМ В СИСТЕМЕ SAPFOR

Н. А. Катаев^[0000-0002-7603-4026]

ИПМ им. М.В. Келдыша РАН, г. Москва;

kataev_nik@mail.ru

Аннотация

Автоматизация параллельного программирования затрагивает различные этапы в разработке параллельной программы, начиная от профилирования исходной программы, ее преобразования и приведения к виду, допускающему эффективное распараллеливание, и заканчивая построением параллельной версии программы и ее последующей оптимизацией. Немалое значение имеет выбор целевой модели параллельного программирования, с одной стороны, позволяющей задействовать разнообразие существующих на данный момент аппаратных ресурсов, а с другой, упрощающей разработку автоматизированных средств и позволяющей пользователю изучить решения, принимаемые системой автоматизированного распараллеливания. Система SAPFOR (System FOR Automated Parallelization) объединяет различные подходы, направленные на автоматизацию программирования, и позволяет пользователю принимать активное участие в процесс распараллеливания программ. Кроме того, распараллеливание выполняется в модели DVMH, позволяющей разрабатывать эффективные параллельные программы для гетерогенных вычислительных кластеров.

В настоящей статье рассмотрен подход к автоматизированному распараллеливанию программ, реализованный в системе SAPFOR. Отдельное внимание уделено архитектуре системы и реализации подсистемы интерактивного взаимодействия с пользователем. Рассмотрено применение интерактивной оболочки в процессе распараллеливания и приведены результаты распараллеливания некоторых программ из набора NAS Parallel Benchmarks 3.3.1 с ручным распараллеливанием, выполненным с помощью OpenCL.

Ключевые слова: анализ программ, преобразование программ, автоматизация распараллеливания, графический интерфейс пользователя, SAPFOR, DVM, LLVM

ВВЕДЕНИЕ

Многочисленные исследования, проводимые в направлении разработки средств автоматизации параллельного программирования, а также разнообразие существующих и разрабатываемых инструментов показывают, что для снижения сложности параллельного программирования и минимизации ошибок, связанных с распараллеливанием программ, необходимы комплексный подход и разработка целого набора программных средства. Разрабатываются как полностью автоматически распараллеливающие компиляторы [1–3], так и инструменты, накладывающие ограничения на применение стандартных конструкций последовательных языков программирования и опирающиеся на дополнительные указания пользователя о свойствах последовательной программы [4, 5]. Большое внимание уделяется инструментам, которые позволяют исследовать распараллеливаемую программу, но принятие решений о распараллеливании оставляют за пользователем [6, 7]. Поэтому при разработке системы автоматизированного распараллеливания SAPFOR (System FOR Automated Parallelization) [8] мы ведем исследования по трем ключевым направлениям:

- Исследование характеристик и свойств распараллеливаемой программы. Помимо выявления участков кода, которые могут быть выполнены параллельно, и обнаружения препятствующих этому зависимостей по данным необходимо определять потенциал от возможного распараллеливания участков программы, в том числе выполняя профилирование программы на различных наборах входных данных. Кроме того, необходимо понимать взаимосвязь тех или иных участков кода, уметь выявлять фрагменты программы, которые связаны с выполнением разных задач и, более того, не могут выполняться одновременно на одних и тех же данных. Такого рода проблемы многовариантности [9] программных комплексов, приводят к тому, что одновременное распараллеливание разных участков кода

может требовать принятия противоречивых решений, например, при построении распределения данных, и найти компромиссное решение, учитывающее все варианты, может быть невозможно. Но в связи с невозможностью одновременного выполнения тех или иных путей в программе это, в конечном счете, и не требуется.

- Автоматическое распараллеливание «хорошо» написанных потенциально параллельных программ. Возможности существующих автоматически распараллеливающих компиляторов по-прежнему сильно ограничены. Распараллеливание в первую очередь выполняется на системы с общей памятью (CPU, GPU). При этом достаточно успешные и активно развивающиеся подходы, такие как модель многогранников (полиэдральная модель), накладывают существенные ограничения на структуру распараллеливаемых участков программы [1–3]. Зависимость выполнения программы от входных данных, сложный граф потока управления, использование косвенной адресации и адресной арифметики существенно затрудняют применение данных подходов и могут сделать невозможным эффективное распараллеливание в контексте всей программы. Проблема стоит особенно остро, если цель распараллеливания – гетерогенный вычислительный кластер, и требуется выполнить не только распределение вычислений, но и распределение данных, используемых в вычислениях. Одним из способов преодоления возникающих сложностей является следование идеям неявного параллелизма [10] при разработке последовательных программ и создании соответствующих автоматически распараллеливающих компиляторов. В этом случае, если программа написана на некотором подмножестве стандартного языка параллельного программирования и, возможно, содержит в исходном коде специального вида указания, описывающие ее свойства, не поддающиеся автоматическому анализу, сложность автоматического распараллеливания снижается [4, 5].
- Автоматизированное приведение программ к потенциально параллельному виду. Часто, даже имея возможность аккуратно проанализи-

зировать программу и точно выявить и описать все присущие ей свойства, влияющие на распараллеливание, распараллелить программу в исходном виде оказывается невозможным либо такое распараллеливание оказывается неэффективным. Могут потребоваться как преобразования, связанные с ограничениями, накладываемыми теми или иными средствами параллельного программирования (например, требование канонической формы циклов, накладываемое OpenMP или требование необходимости тесной вложенности циклов при объединении их пространств итераций и распараллеливании гнезда циклов целиком), так и преобразования, связанные с устранением зависимостей по данным в программе. При этом стоит отметить, что для распараллеливания программ часто требуется выполнять похожие преобразования [11], в то же время выбор порядка их выполнения оказывает значительное влияние на возможность дальнейшего распараллеливания программы [12]. Однако компилятор не всегда способен автоматически определить требуемые преобразования и выстроить их в правильную последовательность. В этом случае решением может быть предоставление пользователю системы набора автоматически выполняемых преобразований и возможности управлять процессом их применения к распараллеливаемой программе. Это позволяет значительно снизить трудоемкость приведения программы к виду, допускающему автоматическое распараллеливание.

При разработке программных комплексов, направленных на автоматизацию распараллеливания, не менее важным является выбор целевого языка параллельного программирования, в терминах которого создается параллельная программа. Система SAPFOR в качестве целевой модели параллельного программирования опирается на модель DVMH [13, 14].

Данная модель является высокоуровневым расширением стандартных последовательных языков программирования C и Fortran. Так как для задания спецификаций параллельного программирования используются невидимые для стандартных компиляторов директивы (специального вида комментарии и

прагмы), то параллельная программа остается в то же время последовательной, что существенно облегчает ее дальнейшее сопровождение, а также делает возможным проанализировать решения, принятые системой. Исследовать исходный код автоматически сгенерированных параллельных программ часто оказывается затруднительно, даже если выходным является язык высокого уровня (сгенерированный код может существенно отличаться от исходного) [3], не говоря уже о ситуации, когда компилятор генерирует код в терминах низкоуровневого представления [1, 2].

Более того, DVMH-языки описывают параллелизм на нескольких уровнях. Параллелизм уровня целого кластера задает распределение данных и вычислений между узлами кластера. Параллелизм внутри узла распределяет данные и вычисления как между ядрами центральных процессорных устройств, так и позволяет задействовать ускорители, имеющиеся в узле.

Кроме того, DVM система, реализующая DVMH модель, включает целый ряд инструментов, облегчающих пользователю работу с параллельной программой, в том числе инструменты, направленные на исследование эффективности и корректности параллельной программы.

Система SAPFOR ориентирована на активное участие пользователя в процессе распараллеливания и наряду со средствами анализа и преобразования программы должна включать средства организации интерактивного взаимодействия с пользователем.

Данные средства должны позволять как описывать свойства программы, которые не удастся выявить автоматически, так и управлять процессом распараллеливания в целом, задавая глобальные опции анализа программы и выбирая необходимые преобразования. При этом пользователь, в первую очередь, должен иметь возможность управлять преобразованиями программы на уровне исходного кода, так как явное конфигурирование последовательности оптимизационных проходов, например, так, как предлагается в [15], может оказаться для прикладного программиста слишком трудоемким. Инструмент, описываемый в статье [15], нацелен на оптимизацию программ для выполнения на встраиваемых системах, которые выпускаются огромными партиями и для которых

тонкая настройка последовательности оптимизационных проходов, применяемой к низкоуровневому представлению программы, является оправданной.

В данной статье рассматривается архитектура системы SAPFOR с точки зрения ее ориентированности на решение задач, стоящих перед системой. Отдельное внимание уделено организации интерактивного взаимодействия с пользователем и возможности создания различных систем визуализации на основе ядра системы SAPFOR, предоставляющего единый интерфейс для управления процессом распараллеливания.

Статья организована следующим образом. В разделе 2 рассмотрены архитектура системы SAPFOR и организация процесса распараллеливания с использованием подсистемы интерактивного взаимодействия с пользователем. Раздел 3 посвящен обзору возможностей интерактивной оболочки системы SAPFOR. В разделе 4 рассмотрены особенности реализации компонент системы SAPFOR и особенности взаимодействия между компонентами системы. Раздел 5 посвящен практическому применению системы SAPFOR при распараллеливании трех программ из пакета NAS Parallel Benchmarks 3.3.1 [16], также выполнено сравнение с ручным распараллеливанием данных программ с помощью OpenCL. Выводы приведены в заключении.

АРХИТЕКТУРА СИСТЕМЫ SAPFOR

Систему SAPFOR можно условно разделить на три части: ядро системы, отвечающее за анализ, преобразование и распараллеливание программы; средства динамического анализа программы; средства интерактивного взаимодействия с пользователем.

Ядро системы включает:

- фроненды с поддерживаемых языков высокого уровня, генерирующие внутреннее представление программы в системе,
- средства анализа и преобразования внутреннего представления программы,
- средства для выполнения преобразований на уровне исходного кода программы,

- средства отображения результатов анализа, выполненного над внутренним представлением программы на конструкции высокоуровневого языка программирования.

Отображение результатов анализа между разными уровнями абстракции в системе SAPFOR особенно важно, так как для повышения качества и полноты анализа используется низкоуровневое представление программы в виде LLVM IR [17]. В то же время система должна отображать результаты своей работы в терминах исходного высокоуровневого языка программирования. Более того, использование высокоуровневого представления наряду с низкоуровневым позволяет учитывать некоторые особенности исходного языка (например, ограничения на допустимость пересечений по памяти между изменяемыми параметрами процедур в языке Fortran), которые могут теряться при переходе к LLVM IR.

В статье [18] подробно описан механизм сопоставления разных уровней представления программы, а также объяснен выбор LLVM в качестве базовой инфраструктуры для разработки SAPFOR.

Средства динамического анализа программ в первую очередь включают средства для анализа зависимостей по данным, которые не могут быть доказаны или опровергнуты во время статического анализа. В этом случае динамический анализ опирается на инструментацию программы на уровне LLVM IR, что позволяет, во-первых, выполнять выборочную инструментацию, а, во-вторых, выполнять ряд предварительных оптимизаций программы, не теряя при этом полноту проводимого анализа [19]. Свойства программы, выявленные в результате запуска ее инструментированной версии, передаются в ядро системы и учитываются в процессе всех анализов и преобразований исходной программы.

Для исследования свойств программы также могут быть полезны средства профилирования, входящие в состав DVM системы, но на данный момент они интегрированы в состав SAPFOR не полностью, и пользователю приходится вручную определять наиболее времяемкие участки кода, требующие распараллеливания в первую очередь.

Подсистема интерактивного взаимодействия с пользователем предназначена не только для того, чтобы отобразить результаты анализа программы и исследовать ее информационную структуру. В ее задачу также входит обеспечить

пользователя механизмом удобного задания дополнительных свойств программы и управления преобразованиями программы, выполняемыми на уровне исходного кода. В крайнем случае, пользователь может получить возможность вносить изменения в программу вручную.

С точки зрения архитектуры системы SAPFOR и организации взаимодействия между компонентами системы подсистема интерактивного взаимодействия является независимой и взаимодействует с ядром системы на основе фиксированного интерфейса. Фактически управление процессом распараллеливания опирается на организацию клиент-серверного взаимодействия, в котором ядро системы выступает в роли сервера и определяет интерфейс, доступный клиенту. Обмен информацией осуществляется через передачу сообщений в JSON формате. Таким образом, становится возможна реализации различных интерактивных сред в зависимости от требуемой функциональности и окружения, в котором должна функционировать система. Это могут быть как отдельное приложение с графическим интерфейсом, так и расширение для некоторой среды разработки.

С использованием интерактивной оболочки процесс распараллеливания программы организован следующим образом:

1. Пользователь подготавливает программу к анализу и преобразованию средствами системы SAPFOR, задает опции, необходимые для синтаксического разбора программы (например, пути к заголовочным файлам, значения макросов). Текущая реализация интерактивной оболочки ограничена возможностью анализа только одного файла с исходным кодом (допускается использование нескольких заголовочных файлов, подключаемых с помощью директив include). Данное ограничение относится в первую очередь к выполнению преобразований программ на уровне исходного кода, так как в этом случае требуется представление всей программы в виде единого дерева разбора. Также нежелательно использование макроконструкций в области преобразования, так как дерево разбора Clang, используемое для преобразования, не содержит данные конструкций в явном виде, и оценить влияние макросов на допустимость преоб-

разования часто оказывается невозможным. Например, макрос с одним и тем же именем может иметь разное определение в разных точках программы, и межпроцедурные преобразования, такие как подстановка функций, приведут к генерации некорректной программы. В этом случае SAPFOR принимает консервативное решение отказаться от выполнения преобразования.

2. Пользователь выполняет профилирование программы с целью определить наиболее времяемкие участки кода, которые требуется распараллелить в первую очередь. Для этих целей могут быть использованы механизм интервалов, доступный в системе DVM, либо сторонние инструменты, такие как `gprof`. Дальнейшие работы по развитию интерактивной оболочки и системы SAPFOR в целом предполагают, что информация, полученная в результате профилирования программы, будет визуализирована и соотнесена с результатами последующего анализа программы на предмет возможности распараллеливания найденных времяемких фрагментов. На этом этапе также может быть выполнен динамический анализ программы, если пользователь считает, что возможностей статического анализа недостаточно.
3. Пользователь использует интерактивную оболочку, чтобы выполнить статический анализ программы (возможно дополнив его ранее полученными результатами динамического анализа, указав глобальные опции анализа или вручную задав свойства отдельных переменных и циклов программы) и изучить потенциал параллелизма, доступного в программе и обнаруженного SAPFOR.
4. На основе результатов анализа, профилирования и рекомендаций системы пользователь выбирает фрагменты программного кода, которые должны быть распараллелены в первую очередь. Если распараллеливание данных фрагментов требует дополнительного преобразования программы, пользователь может разметить исходный код, чтобы задать преобразования, которые требуется выполнить. Система SAPFOR автоматически преобразует программу по запросу

пользователя, после чего процесс распараллеливания переходит снова к шагам (2) или (3). Если возможностей системы недостаточно для преобразования программы, то может потребоваться выполнение некоторых преобразований вручную.

5. Если программа приведена к виду, допускающему автоматическое распараллеливание, то пользователь может разметить области программы, которые требуется распараллелить с помощью директивы `SAPFOR region`, и система `SAPFOR` автоматически сгенерирует параллельный вариант программы либо с использованием `OpenMP`, либо в модели `DVMH`. В случае отсутствия директив `region` распараллеливание выполняется для всей программы.

ПОДСИСТЕМА ИНТЕРАКТИВНОГО ВЗАИМОДЕЙСТВИЯ С ПОЛЬЗОВАТЕЛЕМ

В данный момент интерактивная оболочка реализована в виде расширения для кроссплатформенного редактора `Microsoft Visual Studio Code` [20] и позволяет пользователю исследовать доступный в программе параллелизм и наглядно показывает проблемы, препятствующие ее автоматическому распараллеливанию.

Редактор `Microsoft Visual Studio Code` предоставляет удобный механизм для разработки расширений и имеет открытые исходные коды. Кроме того, он позволяет запускать расширения на удаленной рабочей станции, если к ней возможен доступ с использованием `SSH`. В этом случае, даже при работе напрямую на рабочей станции, становится возможным получить преимущества от интерактивного взаимодействия с системой `SAPFOR`.

Так как `SAPFOR` ориентирована на извлечение параллелизма уровня циклов, интерактивная оболочка обеспечивает навигацию по циклам в программе.

Для каждого цикла в программе доступна следующая информация: является ли поток управления внутри тела цикла безопасным для распараллеливания, находится ли цикл в канонической форме [21] и может ли автоматически распараллеливающий компилятор системы `SAPFOR` распараллелить данный цикл, в том числе устранив имеющиеся в нем зависимости по данным.

Безопасность потока управления означает отсутствие побочных эффектов у вызовов функций из тела цикла, в том числе отсутствие операций ввода/вывода,

а также отсутствие нескольких выходов из тела цикла. По запросу пользователя может быть построен подграф графа вызовов и отображена последовательность вызовов, приводящая к небезопасному потоку управления.

Пример, показывающий сводные результаты анализа для циклов модельной программы и решающий систему уравнений в частных производных методов переменных направлений (ADI), приведен на рис. 1. Зеленым подсвечены циклы программы, допускающие автоматическое распараллеливание. Видно, что итерационный цикл не может быть распараллелен в том числе из-за наличия в нем дополнительных операторов выхода из цикла и операторов ввода вывода.

Functions and Loops	Parallel	Canonical	Perfect	Exit	IO	Readonly	Unsafe CFG
- main at <code>Adi.func.c:21:1 - Adi.func.c:44:1</code>	-			1	✓	-	✓
for loop in main at <code>Adi.func.c:29:3 - Adi.func.c:34:3</code>	-	✓	✓	2	✓	-	✓
- init at <code>Adi.func.c:46:1 - Adi.func.c:57:1</code>	✓			1	-	-	-
+ for loop in init at <code>Adi.func.c:48:3 - Adi.func.c:56:24</code>	✓	✓	✓	1	-	-	-
- iter at <code>Adi.func.c:59:1 - Adi.func.c:79:1</code>	✓			1	-	-	-
+ for loop in iter at <code>Adi.func.c:62:3 - Adi.func.c:65:58</code>	✓	✓	✓	1	-	-	-
+ for loop in iter at <code>Adi.func.c:66:3 - Adi.func.c:69:58</code>	✓	✓	✓	1	-	-	-
+ for loop in iter at <code>Adi.func.c:70:3 - Adi.func.c:77:7</code>	✓	✓	✓	1	-	-	-

Рисунок 1. Сводная информация по циклам программы ADI

На рис. 2 приведен пример графа вызовов для всей программы ADI. Зеленым выделены функции, содержащие параллельные циклы. Серым цветом показаны библиотечные функции, о которых у SAPFOR может не быть достаточно информации, если их реализация не доступна в виде исходных кодов, и которые не могут быть преобразованы на уровне исходного кода, так как расположены за пределами программы пользователя.

С другой стороны, SAPFOR использует информацию о таких функциях, которую гарантирует реализация стандартной библиотеки. Это позволяет в определенных ситуациях гарантировать отсутствие побочного эффекта у вызываемой функции или быть уверенным, что возвращаемый функцией указатель уникален и не может указывать на память, ранее выделенную в программе.

При выборе дуги графа вызовов, справа отображается список всех вызовов функции, в которую ведет выбранная дуга.

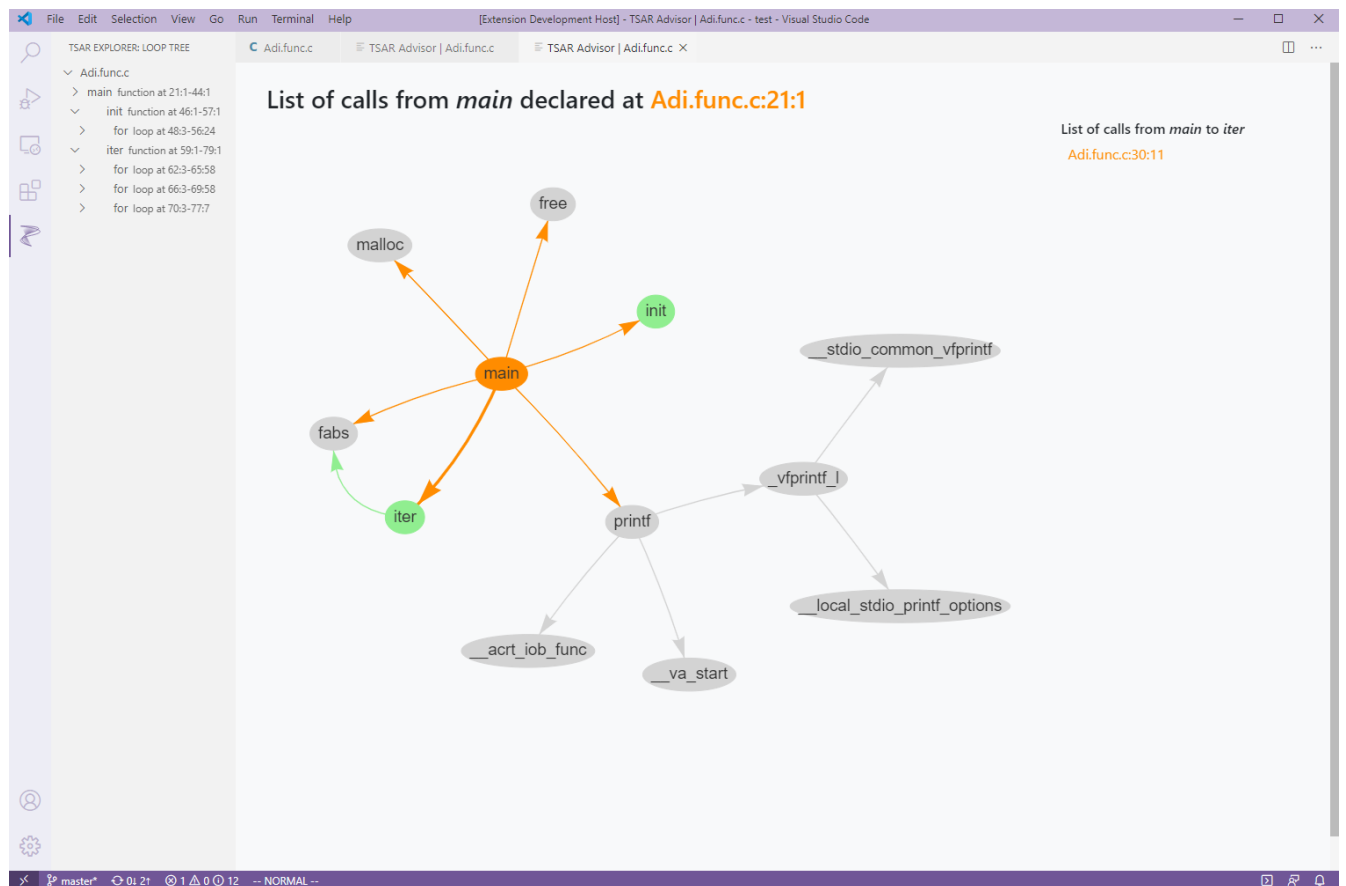


Рисунок 2. Граф вызовов для программы ADI

Для каждого цикла в отдельности можно исследовать, присутствуют ли в нем обращения к памяти, порождающие зависимости по данным. Зависимости по данным между различными итерациями цикла будут классифицированы в соответствии с возможностями их устранения (редукционные и индуктивные переменные, переменные, зависимость по которым может быть устранена за счет приватизации данных или за счет организации конвейерного выполнения цикла). Описание обнаруженных зависимостей доступно в отдельной вкладке, отображающей используемую в программе память в виде дерева псевдонимов [18].

При необходимости уточнить свойства отдельных переменных интерактивная оболочка помогает пользователю сформировать описание данных свойств в JSON-формате, которое затем может быть использовано в процессе анализа. Кроме того, системой могут быть учтены свойства программы, обнаруженные в результате динамического анализа.

На рис. 3 приведен пример отображения свойств одного из циклов программы ADI. Красным цветом выделены переменные (в данном случае массив A), которые вызывают зависимость в данном цикле. При этом справа отображается информация о найденной зависимости: ее тип (anti и flow), степень уверенности анализатора в наличии зависимости (must), причины ее возникновения (запись(store)/чтение(load)) и расстояние зависимости. В данном случае зависимость является регулярной и может быть устранена за счет организации конвейера, таким образом, цикл может быть распараллелен автоматически. Соответствующая информация будет отображена в сводном окне по всем циклам программы (свойство Parallel для цикла на рис. 1).

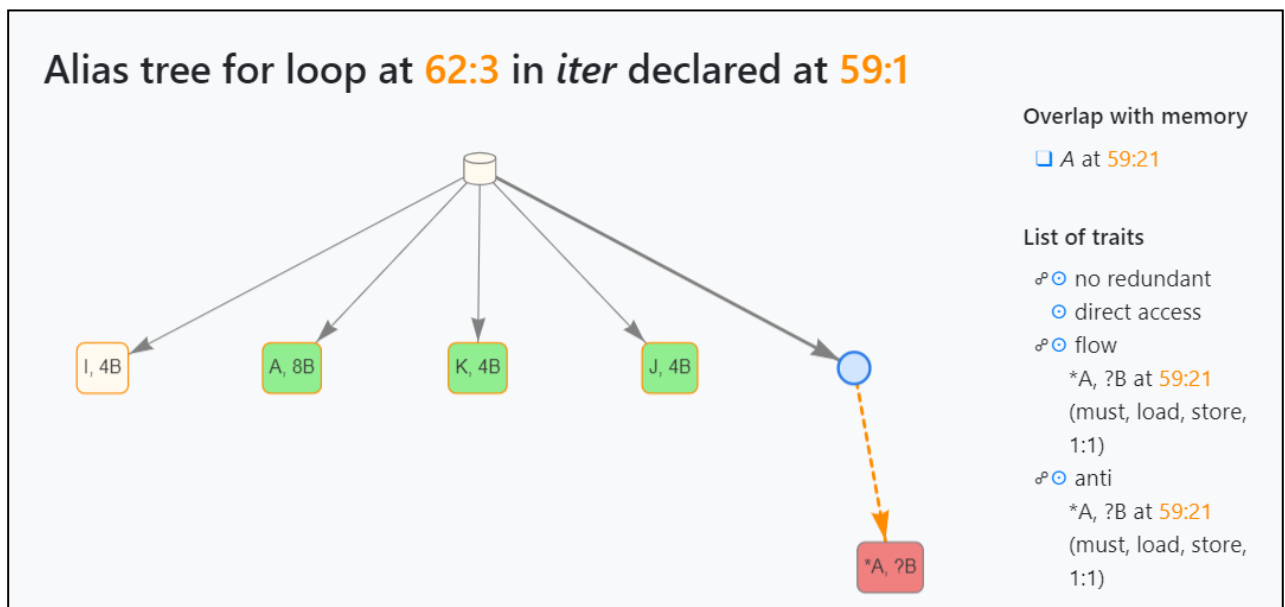


Рисунок 3. Свойства переменных одного из циклов программы ADI

Некоторые проблемы анализа могут быть устранены за счет задания свойств, влияющих на анализ всей программы в целом и описывающих общие факты, известные пользователю о программе. Примером одного из таких свойств является отсутствие побочного эффекта у математических функций стандартной библиотеки языка C, которые в обычной ситуации сохраняют признак возникновения ошибки в специальной глобальной переменной. Если такое поведение не требуется при выполнении программы, то соответствующий побочный эффект может быть проигнорирован, в противном случае это препятствует распараллеливанию циклов, использующих математические операции.

Задать глобальные свойства анализа, а также дополнительные опции компиляции программы и указать файлы, содержащие результаты динамического анализа, можно в начале обработки программы системой SAPFOR (рис. 4).

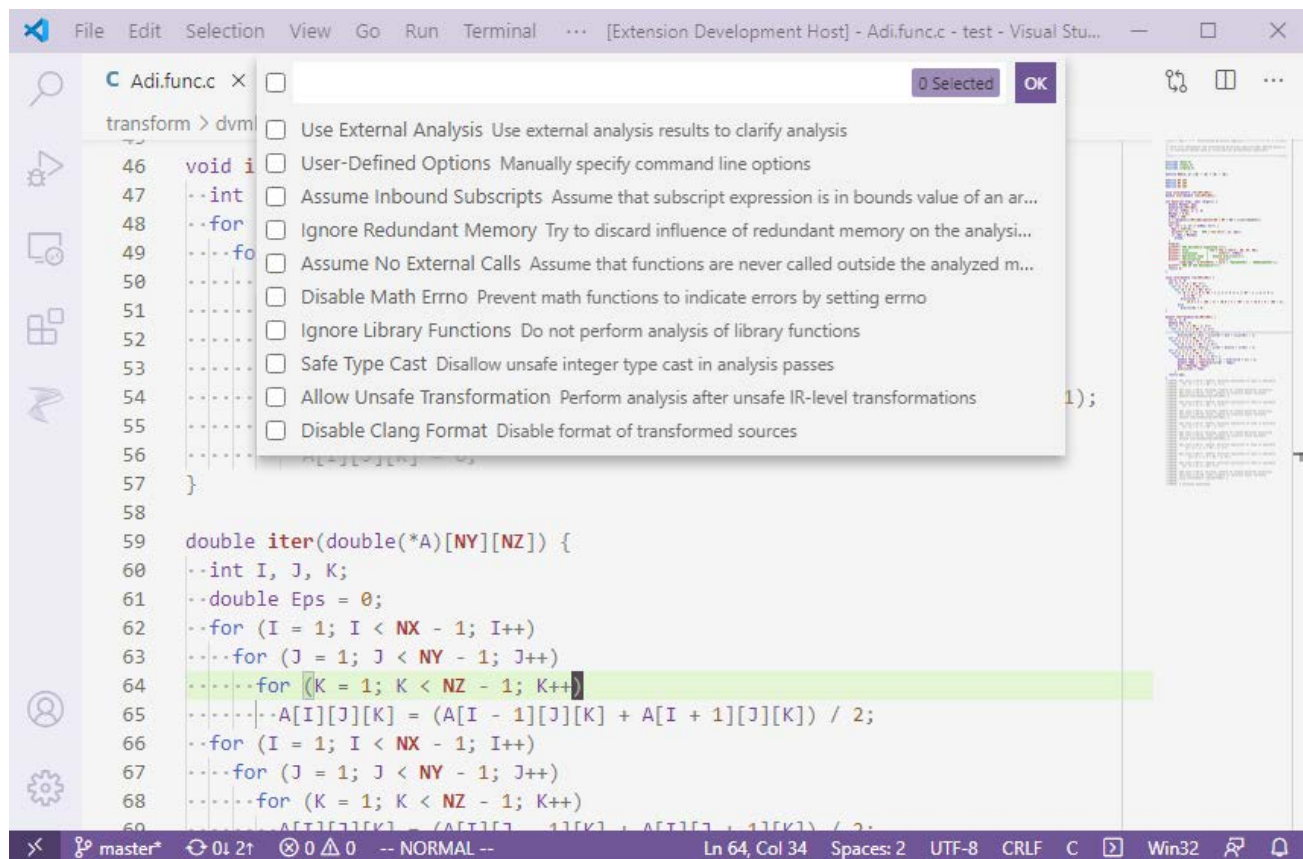


Рисунок 4. Окно задания глобальных свойств анализа программы

ОРГАНИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ МЕЖДУ КОМПОНЕНТАМИ SAPFOR

Ядро системы SAPFOR основано на инфраструктуре компиляторов LLVM, в данный момент используется LLVM версии 11. Абстрактное синтаксическое дерево, генерируемое компилятором Clang (Clang AST), используется для выполнения преобразований на уровне исходного кода, а также для получения свойств программы, зависящих от конкретного языка программирования (например, структура организации циклов программы и факт их тесной вложенности).

Интерактивная подсистема взаимодействия с пользователем разрабатывается на языке TypeScript в виде расширения для Visual Studio Code.

Схема процесса анализа и распараллеливания программы, а также взаимодействия ядра системы с подсистемой интерактивного взаимодействия изображена на рис. 5.

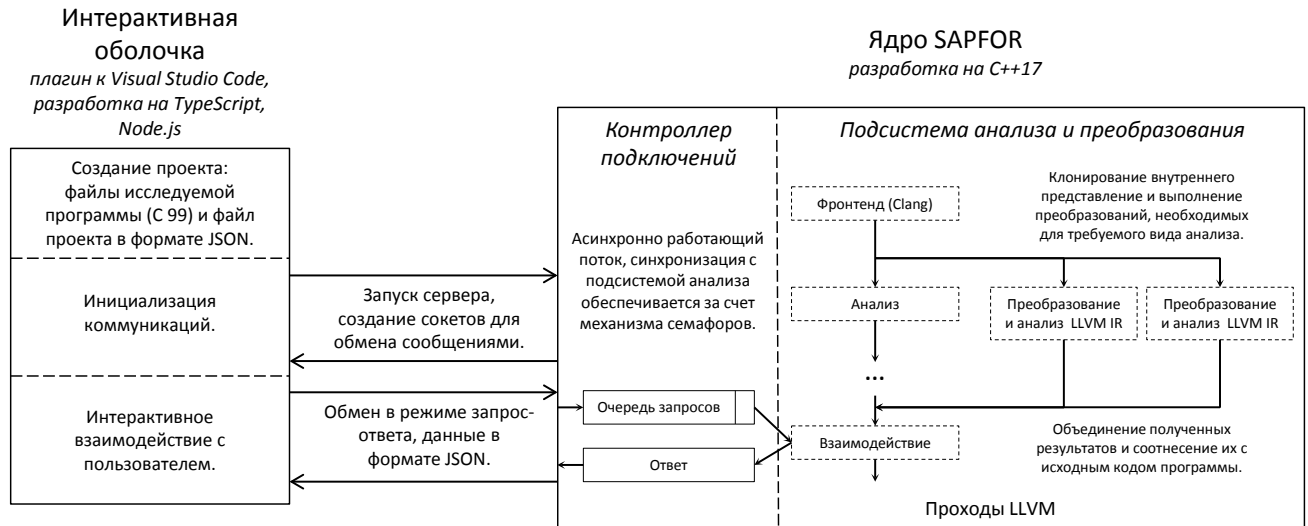


Рисунок 5. Схема взаимодействия компонент системы SAPFOR

Перед началом анализа и преобразования программы пользователь создает проект, содержащий файлы с исходным кодом, а также файл в формате JSON, описывающий правила компиляции каждой отдельной единицы трансляции. После того, как будет установлено соединение с ядром системы, обмен информацией между интерактивной оболочкой и ядром системы будет осуществляться в виде передачи сообщений в формате JSON. Для этого будут использованы механизм сокетов и соответствующие его реализации, доступные в Node.js и C++.

Получив файлы с исходными кодами программы, ядро системы SAPFOR строит Clang AST и LLVM IR, а также устанавливает соответствие между представлениями анализируемых объектов программы на данных уровнях абстракции.

Проводимый статический анализ должен принимать консервативные решения и не допускать преобразований, которые потенциально могут привести к некорректному поведению программы. Система SAPFOR предполагает наличие зависимостей по данным в циклах или пересечений по памяти между указателями, если не может быть доказано обратное.

Чтобы сократить количество ситуаций, не поддающихся точному анализу, предварительно выполняются дополнительные преобразования LLVM IR [18]. Это также позволяет задействовать доступные в LLVM анализы, такие как определение редукционных переменных и переменных индукции. Дополнительные преобразования LLVM также необходимы для делинеаризации массивов, линейризованных в LLVM IR [22]. Делинеаризация массивов необходима для реализации анализа зависимостей по данным.

Система SAPFOR позволяет одновременно запустить несколько потоков анализа программы и тем самым выполнять разные последовательности преобразований низкоуровневого LLVM IR, наиболее подходящие для определения разных свойств программы. Для этого создаются копии LLVM IR исходной программы, а после завершения анализа каждой преобразованной копии обобщенная информация о полученных результатах отображается сначала на исходный LLVM IR, а затем на объекты исходной программы, представленные в виде Clang AST.

РАСПАРАЛЛЕЛИВАНИЕ ПРОГРАММ ИЗ НАБОРА NAS PARALLEL BENCHMARKS 3.3.1

Мы воспользовались системой SAPFOR и разработанной интерактивной оболочкой, чтобы выполнить распараллеливание программ CG, BT, EP из пакета NAS Parallel Benchmarks 3.3.1 [16, 23]. Для оценки ускорения, полученного после распараллеливания данных программ в модели DVMH, был использован 6-ядерный процессор Intel Xeon CPU E5-1660 v2, 3.70 GHz с включенным Hyper Threading (2 потока на ядро, всего 12 потоков) и отключенным Turbo Boost. В качестве ускорителя была использована GPU GeForce GTX 1660 Ti. Компиляция программ выполнялась с помощью Intel Compiler 19.0.2.187 и NVIDIA Compiler 10.2, использовалась опция оптимизации -O3. На рис. 6 приведено ускорение программ относительно исходных последовательных программ. Автоматическое распараллеливание выполнялось для преобразованных последовательных версий программ. Также приведено ускорение программ, вручную написанных с помощью OpenCL [23]. Результаты были получены для трех классов A, B и C, определяющих размер (в порядке увеличения) используемых данных для каждой программы [16].

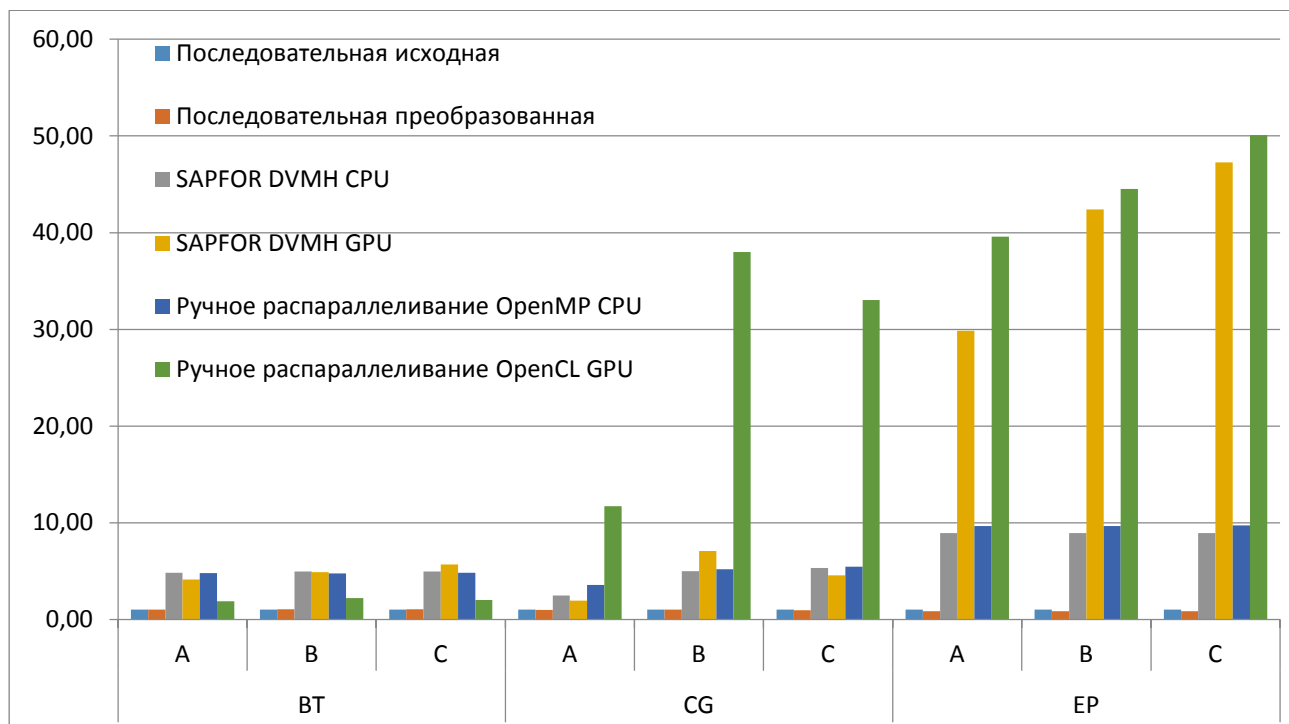


Рисунок 6. Ускорение выполнения программ из пакета NAS Parallel Benchmarks (NPB)

Стоит отметить, что программы, разработанные вручную с помощью OpenCL, существенно отличаются от исходных программ, что в общем случае вызывает дополнительные трудности по сопровождению и модификации программ. В то же время программы, распараллеленные в модели DVMH, остаются написанными в последовательном стиле, а параллелизм в них выражен за счет высокоуровневых спецификаций, заданных в виде директив. Существенное преимущество ручного распараллеливания программы CG над автоматическим вызвано в первую очередь дополнительной векторизацией некоторых внутренних циклов и использованием общей памяти ускорителя.

Основным преобразованием, которое потребовалось для распараллеливания программ, является подстановка функций. Это связано с тем, что текущая реализация межпроцедурного анализа в SAPFOR достаточно грубо оценивает участки памяти, используемые и изменяемые внутри вызовов функций. Чтобы обнаружить приватные массивы в программах BT и EP, был дополнительно применен динамический анализ.

Программа EP также потребовала некоторых ручных преобразований, связанных с устранением редуцированного массива фиксированного размера (замены его на несколько скалярных переменных) и устранением приватного массива большого размера (объединение цикла инициализации массива и использования массива в один цикл с перевычислением требующихся элементов на каждой итерации цикла). Редуцированные массивы в данный момент не поддерживаются в языке C DVMH в случае использования ускорителей. Суммарный размер всех копий массива после его приватизации превышал объем памяти, доступной на GPU, что делало невозможным выполнение программы. В Таблицах 1 и 2 приведен фрагмент кода до и после преобразования соответственно.

Таблица 1. Фрагмент программы EP, требующий ручного преобразования.

```
double x[2*NK]; // NK is 65536, NP depends on an input
#pragma dvm parallel(1) private(x) ...
for (k = 1; k <= NP; k++) { ...
  for (i = 0; i < NK; i++) { ...
    x[i] = r46 * (*x4);
  }
  for (i = 0; i < NK; i++) {
    x1 = 2.0 * x[2 * i] - 1.0;
    x2 = 2.0 * x[2 * i + 1] - 1.0;
    t1 = x1 * x1 + x2 * x2;
    ...
  }
  ...
}
```

Таблица 2. Фрагмент программы EP после ручного преобразования.

```
#pragma dvm parallel(1) ...
```

```
for (k = 1; k <= NP; k++) { ...
```

```
  for (i = 0; i < NK; i++) {
```

```
    double x_2i, x_2i1;
```

```
    { ...
```

```
      x_2i = r46 * (*x4);
```

```
    }
```

```
    { ...
```

```
      x_2i1 = r46 * (*x4);
```

```
    }
```

```
    x1 = 2.0 * x_2i - 1.0;
```

```
    x2 = 2.0 * x_2i1 - 1.0;
```

```
    ...
```

```
  }
```

```
  ...
```

```
}
```

Директивы языка C DVMH, приведенные в Таблицах 1 и 2, были добавлены в исходный код программы EP системой SAPFOR автоматически.

ЗАКЛЮЧЕНИЕ

В данной работе рассмотрены архитектура системы SAPFOR, а также ключевые моменты, которые были приняты во внимание при проектировании системы. Система автоматизированного распараллеливания SAPFOR разрабатывается с целью добиться снижения трудоемкости ручного распараллеливания программ и уменьшить количество возникающих при этом ошибок, вносимых программистом. В этой связи перед системой ставятся три основные задачи: исследование свойств распараллеливаемой программы, автоматическое распаралле-

ливание некоторого класса программ, относящихся к так называемым «хорошо» написанным потенциально параллельным программам, и разработка средств автоматизации выполнения преобразований программ для приведения программы к потенциально параллельному виду.

Система SAPFOR ориентирована на тесное взаимодействие с пользователем в процессе анализа и преобразования программ. Поэтому наряду со средствами как статического, так и динамического анализа система включает интерактивную подсистему, реализующую взаимодействие с пользователем в процессе распараллеливания. Использование фиксированного интерфейса, предоставляемого ядром системы, упрощает разработку интерактивной оболочки и делает возможным ее реализацию в виде, определяемым, в том числе, средой функционирования системы. Таким образом, могут быть реализованы как расширение для существующих сред разработки программного обеспечения, так и отдельное приложение с графическим интерфейсом пользователя.

В данной работе рассмотрен пример реализации системы интерактивного взаимодействия в виде расширения кроссплатформенного редактора Microsoft Visual Studio Code. Разработанная интерактивная оболочка упрощает участие пользователя в процессе распараллеливания, позволяя ему изучить решения, принимаемые системой, и при необходимости дать соответствующие рекомендации системе. Пользователь может влиять как на выполнение анализа программы, подсказывая системе наличие тех или иных свойств программы, так и на выполнение преобразований исходного кода, необходимых для распараллеливания программы. Полученная в итоге программа на последовательном языке программирования может быть автоматически распараллелена системой SAPFOR.

В статье представлены результаты успешного применения предложенного подхода к распараллеливанию нескольких программ на языке C из набора программ NAS Parallel Benchmarks 3.3.1. Приведены результаты распараллеливания в модели DVMH для систем с общей памятью (многоядерные процессоры и ускорители) и их сравнение с ручным распараллеливанием с помощью OpenCL и OpenMP.

Исходные коды системы SAPFOR доступны на GitHub [24].

СПИСОК ЛИТЕРАТУРЫ

1. *Grosser T., Groesslinger A., Lengauer C.* Polly – performing polyhedral optimizations on a low-level intermediate representation // *Parallel Processing Letters*. 2012. Vol. 22, No. 04. 1250010. <https://doi.org/10.1142/S0129626412500107>.
2. *Grosser T., Hoefler T.* Polly-ACC Transparent compilation to heterogeneous hardware // *ICS '16: Proceedings of the 2016 International Conference on 373 Supercomputing June 2016*. P. 1–13. <https://doi.org/10.1145/2925426.2926286>.
3. *Bondhugula U., Hartono A., Ramanujam J., Sadayappan P.* A practical automatic polyhedral parallelizer and locality optimizer // *ACM SIGPLAN Notices*. 2008. Vol. 43, No. 6. P. 101–113. <https://doi.org/10.1145/1379022.1375595>.
4. *Vandierendonck H., Rul S., De Bosschere K.* The Paralax Infrastructure: Automatic Parallelization with a Helping Hand // *Proceedings of the 19th international conference on Parallel architectures and compilation techniques (PACT'10)*. 2010. P. 389–400. <https://doi.org/10.1145/1854273.1854322>.
5. *Baghdadi R., Beaunon U., Cohen A., Grosser T., Kruse M., Reddy C., Verdoolaege S., Betts A., Donaldson A.F., Ketema J., Absar J., Haastregt S., Kravets A., Lokhmotov A., David R., Hajiyev E.* Pencil: A platform-neutral compute intermediate language for accelerator programming // *Proceedings of the 2015 International Conference on Parallel Architecture and Compilation (PACT), PACT'15*. IEEE Computer Society. Washington, DC, USA, 2015. P. 138–149. <https://doi.org/10.1109/PACT.2015.17>.
6. *Kim M., Kim H., Luk C.-K.* Prospector: A Dynamic Data-Dependence Profiler To Help Parallel Programming // *2nd USENIX Workshop on Hot Topics in Parallelism (HotPar'10)*, 2010. P. 1–6.
7. Intel Parallel Studio. URL: <https://software.intel.com/en-us/parallel-studioxe>.
8. *Клинов М.С., Крюков В.А.* Автоматическое распараллеливание Фортран-программ. Отображение на кластер // *Вестник Нижегородского университета им. Н.И. Лобачевского*. 2009. № 2. С. 128–134.
9. *Бахтин В.А., Жукова О.Ф., Катаев Н.А., Колганов А.С., Крюков В.А., Кузнецов М.Ю., Поддерюгина Н.В., Притула М.Н., Савицкая О.А., Смирнов А.А.* Распараллеливание программных комплексов. Проблемы и перспективы // *Труды XX Всероссийской научной конференции «Научный сервис в сети Интернет»*, Но-

вороссийск, Россия, 17–22 сентября 2018 г. М.: ИПМ им. М.В. Келдыша, 2018. С. 63–72.

URL: <http://keldysh.ru/abrau/2018/theses/33.pdf>. <https://doi.org/10.20948/abrau-2018-33>

10. *Hwu W.-m., Ryoo S., Ueng S.-Z., Kelm J.H., Gelado I., Stone S.S., Kidd R.E., Baghsorkhi S.S., Mahesri A.A., Tsao S.C., Navarro N., Lumetta S.S., Frank M.I., Patel S.J.* Implicitly parallel programming models for thousand-core microprocessors // Proceedings of the 44th annual Design Automation Conference (DAC '07), ACM, New York, NY, USA. 2007. P. 754–759. <https://doi.org/10.1145/1278480.1278669>.

11. *Blume W., Eigenmann R.* Performance analysis of parallelizing compilers on the Perfect Benchmarks programs // IEEE Transactions on Parallel and Distributed Systems. 1992. Vol. 3, Issue 6. P. 643–656. <https://doi.org/10.1109/71.180621>.

12. *Wolfe M.* Scalar vs. parallel optimizations // CSETech. 210. 1990. URL: <https://classes.cs.uoregon.edu/16S/cis410parallel/Documents/scalar-paralleloptimizations-wolfe.pdf>

13. *Konovalov N.A., Krukov V.A., Mikhajlov S.N., Pogrebtsov A.A.* Fortan DVM: a Language for Portable Parallel Program Development // Programming and Computer Software. 1995. V. 21, No. 1. P. 35–38.

14. *Бахтин В.А., Клинов М.С., Крюков В.А., Поддерюгина Н.В., Притула М.Н., Сазанов Ю.Л.* Расширение DVM-модели параллельного программирования для кластеров с гетерогенными узлами // Вестник Южно-Уральского государственного университета, серия «Математическое моделирование и программирование». 2012. №18 (277), выпуск 12. Челябинск: Издательский центр ЮУрГУ. С. 82–92.

15. *Kulkarni P., Zhao W., Moon H., Cho K., Whalley D., Davidson J., Bailey M., Paek Y., Gallivan K.* Finding effective optimization phase sequences // Proceedings of the 2003 ACM SIGPLAN Conference on Languages, Tools, and Compilers for Embedded Systems. 2003. P. 12–23. <https://doi.org/10.1145/780731.780735>.

16. NAS Parallel Benchmarks. URL: <https://www.nas.nasa.gov/publications/npb.html>.

17. *Lattner C., Adve V.* LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation // Proc. of the 2004 International Symposium on Code Generation and Optimization (CGO'04), Palo Alto, California. 2004.

<https://doi.org/10.1109/CGO.2004.1281665>.

18. *Kataev N.A.* Application of the LLVM Compiler Infrastructure to the Program Analysis in SAPFOR // Voevodin V., Sobolev S. (eds). Supercomputing. RuSCDays 2018. Communications in Computer and Information Science. 2018. V. 965. Springer, Cham. P. 487–499. doi:10.1007/978-3-030-05807-4_41.

19. *Катаев Н.А., Смирнов А.А., Жуков А.Д.* Определение зависимостей по данным средствами динамического анализа системы SAPFOR // Электронные библиотеки. Тематический выпуск «Научный сервис в сети Интернет». Часть 1. 2020. Том 23. № 3. С. 473–493. <https://doi.org/10.26907/1562-5419-2020-23-3-473-493>.

20. Visual Studio Code. URL: <https://code.visualstudio.com/>, last accessed 2020/11/25.

21. OpenMP Application Programming Interface. <https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-1.pdf>, last accessed 2020/11/25.

22. *Катаев Н.А., Василькин В.Н.* Восстановление многомерной формы обращений к линеаризованным массивам в системе SAPFOR // Электронные библиотеки. Тематический выпуск «Научный сервис в сети Интернет». Часть 2. 2020. Том 23. № 4. С. 770–787. <https://doi.org/10.26907/1562-5419-2020-23-4-770-787>.

23. *Seo S., Jo G., Lee J.* Performance Characterization of the NAS Parallel Benchmarks in OpenCL // 2011 IEEE International Symposium on. Workload Characterization (IISWC), 2011. P. 137–148. <https://doi.org/10.1109/IISWC.2011.6114174>.

24. SAPFOR. <https://github.com/dvm-system>.

INTERACTION WITH THE USER IN THE SAPFOR SYSTEM

N. A. Kataev

Keldysh Institute of Applied Mathematics RAS

kataev_nik@mail.ru

Abstract

Automation of parallel programming is important at any stage of parallel program development. These stages include profiling of the original program, program transformation, which allows us to achieve higher performance after program parallelization, and, finally, construction and optimization of the parallel program. It is also important to choose a suitable parallel programming model to express parallelism available in a program. On the one hand, the parallel programming model should be capable to map the parallel program to a variety of existing hardware resources. On the other hand, it should simplify the development of the assistant tools and it should allow the user to explore the parallel program the assistant tools generate in a semi-automatic way. The SAPFOR (System FOR Automated Parallelization) system combines various approaches to automation of parallel programming. Moreover, it allows the user to guide the parallelization if necessary. SAPFOR produces parallel programs according to the high-level DVMH parallel programming model which simplify the development of efficient parallel programs for heterogeneous computing clusters. This paper focuses on the approach to semi-automatic parallel programming, which SAPFOR implements. We discuss the architecture of the system and present the interactive subsystem which is useful to guide the SAPFOR through program parallelization. We used the interactive subsystem to parallelize programs from the NAS Parallel Benchmarks in a semi-automatic way. Finally, we compare the performance of manually written parallel programs with programs the SAPFOR system builds.

Keywords: *program analysis, program transformation, automated parallelization, graphical user interface, SAPFOR, DVM, LLVM*

REFERENCES

1. *Grosser T., Groesslinger A., Lengauer C.* Polly - performing polyhedral optimizations on a low-level intermediate representation // *Parallel Processing Letters*. 2012. Vol. 22, No. 04. 1250010. <https://doi.org/10.1142/S0129626412500107>.
2. *Grosser T., Hoefler T.* Polly-ACC Transparent compilation to heterogeneous hardware // *ICS'16: Proceedings of the 2016 International Conference on 373 Supercomputing June 2016*. P. 1–13. <https://doi.org/10.1145/2925426.2926286>.
3. *Bondhugula U., Hartono A., Ramanujam J., Sadayappan P.* A practical automatic polyhedral parallelizer and locality optimizer // *ACM SIGPLAN Notices*. 2008. Vol. 43, No. 6. P. 101–113. <https://doi.org/10.1145/1379022.1375595>.
4. *Vandierendonck H., Rul S., De Bosschere K.* The Paralax Infrastructure: Automatic Parallelization with a Helping Hand // *Proceedings of the 19th International Conference on Parallel architectures and compilation techniques (PACT'10)*. 2010. P. 389–400. <https://doi.org/10.1145/1854273.1854322>.
5. *Baghdadi R., Beaunon U., Cohen A., Grosser T., Kruse M., Reddy C., Verdoolaege S., Betts A., Donaldson A.F., Ketema J., Absar J., Haastregt S., Kravets A., Lokhmotov A., David R., Hajiyev E.* Pencil: A platform-neutral compute intermediate language for accelerator programming // *Proceedings of the 2015 International Conference on Parallel Architecture and Compilation (PACT), PACT'15*. IEEE Computer Society. Washington, DC, USA, 2015. P. 138–149. <https://doi.org/10.1109/PACT.2015.17>.
6. *Kim M., Kim H., Luk C.-K.* Prospector: A Dynamic Data-Dependence Profiler To Help Parallel Programming // *2nd USENIX Workshop on Hot Topics in Parallelism (HotPar '10)*, 2010. P. 1–6.
7. Intel Parallel Studio. URL: <https://software.intel.com/en-us/parallel-studio>
8. *Klinov M.S., Kriukov V.A.* Avtomaticheskoe rasparallelivanie Fortran-programm. Otobrazhenie na klaster // *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo*. 2009. No. 2. S. 128–134.
9. *Bakhtin V.A., Zhukova O.F., Kataev N.A., Kolganov A.S., Kriukov V.A., Kuznetsov M.Iu., Podderiugina N.V., Pritula M.N., Savitskaia O.A., Smirnov A.A.* Rasparallelivanie programmnykh kompleksov. Problemy i perspektivy // *Trudy XX Vserossiiskoi nauchnoi konferentsii «Nauchnyi servis v seti Internet»*, Novorossiisk, Rossiia, 17–22 sentiabria 2018 g. M.: IPM im. M.V. Keldysha, 2018. S. 63–72.

<http://keldysh.ru/abrau/2018/theses/33.pdf>. <https://doi.org/10.20948/abrau-2018-33>.

10. *Hwu W.-m., Ryou S., Ueng S.-Z., Kelm J.H., Gelado I., Stone S.S., Kidd R.E., Baghsorkhi S.S., Mahesri A.A., Tsao S.C., Navarro N., Lumetta S.S., Frank M.I., Patel S.J.* Implicitly parallel programming models for thousand-core microprocessors // Proceedings of the 44th annual Design Automation Conference (DAC'07), ACM, New York, NY, USA. 2007. P. 754–759. <https://doi.org/10.1145/1278480.1278669>.

11. *Blume W., Eigenmann R.* Performance analysis of parallelizing compilers on the Perfect Benchmarks programs // IEEE Transactions on Parallel and Distributed Systems. 1992. Vol. 3, Issue 6. P. 643–656. <https://doi.org/10.1109/71.180621>.

12. *Wolfe M.* Scalar vs. parallel optimizations // CSETech. 210. 1990. URL: <https://classes.cs.uoregon.edu/16S/cis410parallel/Documents/scalar-paralleloptimizations-wolfe.pdf>.

13. *Konovalov N.A., Krukov V.A., Mikhajlov S.N., Pogrebtsov A.A.* Fortran DVM: a Language for Portable Parallel Program Development // Programming and Computer Software. 1995. V. 21. No. 1. P. 35–38.

14. *Bakhtin V.A., Klinov M.S., Kriukov V.A., Podderiugina N.V., Pritula M.N., Sazanov Iu.L.* Rasshirenije DVM-modeli paralelnogo programirovaniia dlia klasterov s geterogennymi uzlami // Vestnik Iuzhno-Uralskogo gosudarstvennogo universiteta, seriia "Matematicheskoe modelirovanie i programirovanie". 2012, No. 18 (277), vypusk 12. S. 82–92.

15. *Kulkarni P., Zhao W., Moon H., Cho K., Whalley D., Davidson J., Bailey M., Paek Y., Gallivan K.* Finding effective optimization phase sequences // Proceedings of the 2003 ACM SIGPLAN Conference on Languages, Tools, and Compilers for Embedded Systems. 2003. P. 12–23. <https://doi.org/10.1145/780731.780735>.

16. NAS Parallel Benchmarks. <https://www.nas.nasa.gov/publications/npb.html>.

17. *Lattner C., Adve V.* LLVM: A Compilation Framework for Lifelong Program Analysis & Transformation // In: Proc. of the 2004 International Symposium on Code Generation and Optimization (CGO'04), Palo Alto, California. 2004. <https://doi.org/10.1109/CGO.2004.1281665>.

18. *Kataev N.A.* Application of the LLVM Compiler Infrastructure to the Program Analysis in SAPFOR // Voevodin V., Sobolev S. (eds) Supercomputing. RuSCDays 2018. Communications in Computer and Information Science, 2018. V. 965. Springer, Cham. P. 487–499. doi:10.1007/978-3-030-05807-4_41.

19. *Kataev N.A., Smirnov A.A., Zhukov A.D.* Opredelenie zavisimosti po dannym sredstvami dinamicheskogo analiza sistemy SAPFOR // Elektronnye biblioteki. Tematicheskii vypusk «Nauchnyi servis v seti Internet». Chast 1. 2020. Tom 23. № 3. S. 473–493. <https://doi.org/10.26907/1562-5419-2020-23-3-473-493>.

20. Visual Studio Code. <https://code.visualstudio.com/>, last accessed 2020/11/25.

21. OpenMP Application Programming Interface.
URL: <https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-1.pdf>, last accessed 2020/11/25.

22. *Kataev N.A., Vasilkin V.N.* Vosstanovlenie mnogomernoi formy obrashchenii k linearizovannym massivam v sisteme SAPFOR // Elektronnye biblioteki. Tematicheskii vypusk «Nauchnyi servis v seti Internet». Chast 2. 2020. Tom 23. № 4. S. 770–787. <https://doi.org/10.26907/1562-5419-2020-23-4-770-787>.

23. *Seo S., Jo G., Lee J.* Performance Characterization of the NAS Parallel Benchmarks in OpenCL // 2011 IEEE International Symposium on. Workload Characterization (IISWC), 2011. P. 137–148. <https://doi.org/10.1109/IISWC.2011.6114174>.

24. SAPFOR. URL: <https://github.com/dvm-system>.

СВЕДЕНИЯ ОБ АВТОРЕ



КАТАЕВ Никита Андреевич – научный сотрудник ИПМ им. М.В. Келдыша, специалист в области системного программирования. Сфера научных интересов – компиляторные технологии, автоматизация распараллеливания программ.

Nikita Andreevich KATAEV – Researcher of KIAM RAS, a specialist in system programming. Research interests include compiler technologies, semi-automatic program parallelization.

email: kataev_nik@mail.ru; ORCID: 0000-0002-7603-4026.

Материал поступил в редакцию 30 ноября 2020 года

УДК 519.178: 004.9

О НЕКОТОРЫХ СВОЙСТВАХ ГРАФОВ СОТРУДНИЧЕСТВА УЧЁНЫХ В MATH-NET.RU

А. А. Печников¹ [0000-0002-0683-0019], Д. Е. Чебуков² [0000-0001-9738-8707]

¹Институт прикладных математических исследований – обособленное подразделение ФИЦ «Карельский научный центр Российской академии наук», г. Петрозаводск

² Математический институт им. В.А. Стеклова Российской академии наук, г. Москва

¹pechnikov@krc.karelia.ru, ²tche@mi-ras.ru

Аннотация

Проведено исследование двух графов научного сотрудничества, построенных на основе соавторства и цитирования по данным Общероссийского математического портала Math-Net.Ru. Граф научного сотрудничества на основе цитирования представляет собой ориентированный граф без петель и кратных ребер, вершинами которого являются авторы публикаций, а дуги связывают их, когда имеется хотя бы одна публикация первого автора, цитирующая публикацию второго автора. Граф соавторства – это неориентированный граф, в котором вершинами являются авторы, а ребра фиксируют соавторство двух авторов хотя бы в одной статье. Проводится традиционное исследование основных характеристик обоих графов: диаметр и среднее расстояние, компоненты связности и кластеризация. В обоих графах мы наблюдаем схожую структуру связности – наличие гигантской компоненты и большое количество маленьких компонент. Отмечается сходство и различие научного сотрудничества через соавторство и цитирование.

Ключевые слова: научное сотрудничество, цитирование, соавторство, граф, математический портал Math-Net.Ru

ВВЕДЕНИЕ

Социальные сети активно изучаются в течение, по крайней мере, шестидесяти лет из-за присущего интереса в повторяющихся элементах человеческого

взаимодействия, и потому, что их структура имеет большое значение для распространения информации. Сети научного сотрудничества являются примерами таких сетей. В контексте данной статьи термин «сеть» понимается не как веб-инструмент сотрудничества типа ResearchGate (www.researchgate.net), а как виртуальная «социальная сеть», включающая некий круг знакомых индивидуума и социальных связей между людьми.

Научная работа, как один из наиболее интеллектуально требовательных и сложных видов человеческой деятельности, не может быть просто оценена по единой универсальной схеме. Изучение научных связей ученых, структуры и динамики развития научных сообществ дает возможность оценивать тенденции развития различных научных направлений, идентифицировать персоны, научные центры и научные школы, выявлять взаимосвязи.

В работе [1] отмечается, что «... изучение сетей цитирования как окна в науку представляет собой проверенную временем традицию». Достаточно условно можно обозначить три аспекта исследований, а именно, – собственно сети цитирования/соавторства, структура сложных сетей и динамика сложных сетей, – ведущиеся на фоне анализа социальных сетей [2] и библиометрии [3], имеющих долгую совместную историю, и которые часто невозможно отделить друг от друга.

Одними из часто исследуемых сетей научного сотрудничества являются сети соавторства, возможно, из-за изначальной очевидности определения «знакомства» учёных. Приведем полностью определение отношения соавторства по Ньюману: «Мы изучаем сети ученых, в которых два ученых считаются связанными, если они совместно написали статью. Это кажется разумным определением научного знакомства: большинство людей, которые написали статью вместе, будут хорошо знать друг друга. Это – умеренно строгое определение, так как есть много ученых, которые знают друг друга до некоторой степени, но никогда не сотрудничали в написании статьи» [4].

Вопросы соавторства и цитирования часто рассматриваются для конкретных электронных хранилищ, что очень важно, когда после этапов построения и исследования формальной модели реальной сети сотрудничества переходят к этапу содержательной интерпретации полученных результатов. Например, в

упомянутой работе [4] изучается граф сотрудничества, построенный на основе соавторства в Los Alamos e-Print Archive, Networked Computer Science Technical Reference Library и ряде других электронных баз. Определяются такие характеристики построенных графов, как количество авторов, размерность гигантской компоненты, среднее расстояние и диаметр, коэффициент кластеризации, показатели распределения степеней вершин.

Несмотря на большой объем проделанной работы, тема представляется неисчерпаемой, никакой единой теории сетей научного сотрудничества не построено, исследования конкретных электронных библиотек приносят новые, порой неожиданные результаты, расширяется спектр применяемых теоретических подходов и методов, появляются новые технические и программные возможности. Хороший обзор публикаций на английском языке дается в работе [5].

В данной работе исследуются графы научного сотрудничества учёных, построенные на основе сведений из базы данных Общероссийского математического портала Math-Net.Ru (<http://www.mathnet.ru>), известного веб-ресурса, содержащего богатую коллекцию полнотекстовых архивов ведущих российских математических журналов и информацию об их авторах. По данным на 30 марта 2020 года на портале зарегистрировано 119139 авторов, 238733 научных статьи из 135 журналов (периодических изданий). Ключевым элементом реляционной базы данных (используется СУБД MSSQL) является таблица, содержащая метаданные статей. Изначально база создавалась таким образом, что авторы и аффилиации содержались в отдельных таблицах, и таблица статей соединялась с таблицами авторов и аффилиаций посредством связи «один ко многим». Каждый автор статьи и каждая аффилиация являются уникальными элементами базы данных, которые объединяются в таблицы персоналий и организаций. Данный подход позволяет для каждого автора выбрать список его публикаций по его коду (`author_id`), а не поиском по фамилии, как реализовано в Web of Science и некоторых других библиографических ресурсах. База персоналий с их привязкой к статьям легко позволяет выбрать авторов одних и тех статей – соавторов, а также авторов, работающих индивидуально, то есть не имеющих соавторов.

Информационная система Math-Net.Ru также индексирует списки литературы и хранит их в базе данных в структурированном виде [6]. Списки литературы всех публикаций объединены в одну таблицу базы данных, в которой в от-

дельных колонках хранятся данные об авторе, названии, годе, томе, страницах цитируемой публикации. Каждой индивидуальной ссылке соответствует одна запись в таблице. Такой подход облегчает задачу автоматического проставления гиперссылок на библиометрические базы данных, решает задачу поиска обратных ссылок, а также позволяет автоматически экспортировать ссылки в разные форматы: PDF, XML, HTML. Среди гиперссылок с элементов списков литературы есть и ссылки на статьи, индексированные в базе публикаций Math-Net.Ru. Таким способом осуществляется связь между цитирующей и цитированной статьями.

Перенесем утверждение Ньюмана [4] на цитирование и будем считать, что один учёный «знает» (не обязательно лично) другого ученого, если в своей статье он сделал ссылку на работу этого ученого (обратное неверно). Исходя из сказанного, в нашем случае в графах научного сотрудничества, построенных на основе соавторства и цитирования, дуги (рёбра) имеют кратность 1, поскольку мы не ставим задачу изучения силы сотрудничества в зависимости от количества статей, написанных в соавторстве или ссылок от одного автора на статьи другого.

В работе построены два вида графов научного сотрудничества на основе данных о цитировании и соавторстве в Math-Net.ru. Проводится традиционное исследование основных характеристик обоих графов, таких как степени и значимость вершин, диаметр и среднее расстояние, компоненты связности и кластеризация, и обсуждаются их сходство и различие.

1. ГРАФ НА ОСНОВЕ ЦИТИРОВАНИЯ

По данным Math-Net.Ru мы можем сформировать базу данных цитирования с записями (в несколько упрощенном виде):

`<autor_citing><paper_citing><paper_cited><autor_cited>`,

то есть цитирующий автор и публикация, в которой сделана ссылка, и, соответственно, публикация, на которую сделана ссылка, и ее автор. На рис. 1 а) приведен граф цитирования, где буквами «а» с номером обозначены авторы, а буквой «р» с номером – публикации. К примеру, авторы a_1 и a_2 являются соавторами статьи p_1 , в которой сделана ссылка на статью p_3 , авторами которой являются a_1 и a_3 . Таким образом, из графа цитирования получаем 5 ссылок, связывающих ученых: $a_1 \rightarrow a_1$, $a_1 \rightarrow a_3$, $a_2 \rightarrow a_1$, $a_2 \rightarrow a_3$ и еще раз $a_2 \rightarrow a_3$. Самоцитирование нас

далее не интересует, как не интересуют и кратные дуги, не влияющие на связность графа, поэтому граф на рис. 1 б) содержит три вершины и три дуги.

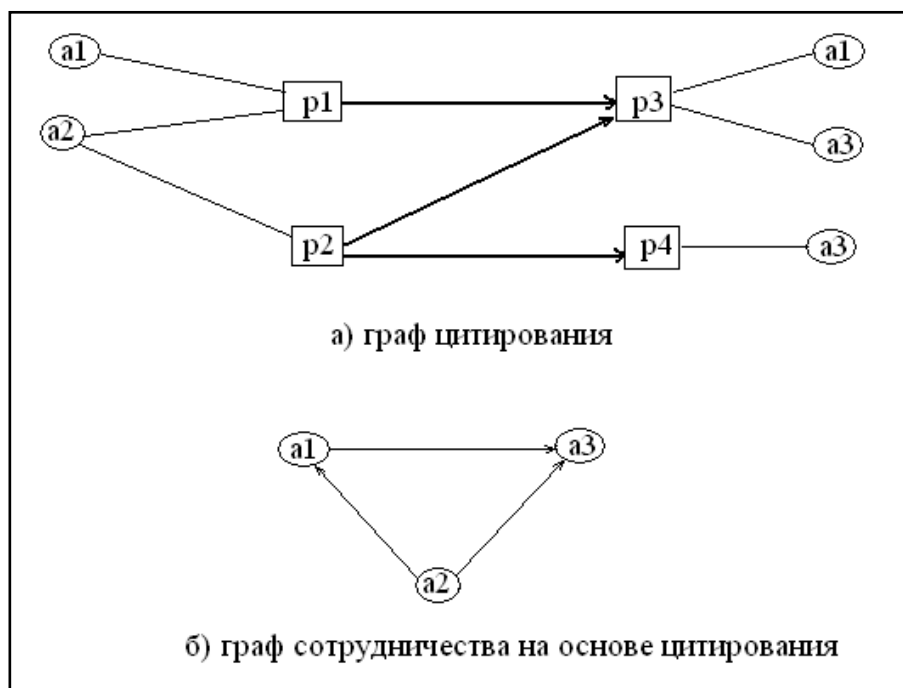


Рисунок 1. Граф цитирования и построенный на его основе граф сотрудничества

База данных цитирования Math-Net.Ru содержит более 1 млн. записей, из них 862476 не являются самоцитированиями. Из 120 тысяч авторов, зарегистрированных в Math-Net.Ru, 52728 ссылаются в своих работах на работы коллег или имеют такие ссылки от коллег на свои работы, проиндексированные на портале. Речь идет о более 12 тысячах статей, проиндексированных в Math-Net.Ru и написанных индивидуально или в соавторстве авторами, зарегистрированными на портале, в которых имеются ссылки на более чем 18 тысяч статей, также проиндексированных в Math-Net.Ru и написанных индивидуально или в соавторстве. С учетом пересечения двух указанных множеств статей (цитирующих и цитируемых) всего обработано 19729 статей.

Построенный граф сотрудничества, который обозначим $G\text{-cit}$, содержит 52728 вершин и (убрав кратные) 388654 дуги. Очевидно по построению, что $G\text{-cit}$ – ориентированный граф без петель и кратных дуг.

Связность является важной характеристикой графа. Сначала на некоторое время забудем об ориентации дуг. В таком неориентированном графе обнаруживается гигантская компонента (максимальная компонента связности – КС), со-

держащая 50931 вершину и 385951 ребро, а вторая по размерности КС содержит только 24 вершины. Всего же граф содержит 504 КС. Диаметр для неориентированного случая равен 17, а средняя длина неориентированного пути (маршрута) – всего 5,515. То есть имеются гигантская компонента и много-много маленьких компонент, не связанных с ней и между собой. Самые удаленные друг от друга вершины гигантской компоненты расположены очень далеко, но средняя длина маршрута среди ее вершин не велика, что свидетельствует о большом количестве коротких маршрутов.

Рассмотрим одну из маленьких компонент связности, состоящую из четырех вершин: 145585 – Наяк Хамса, India; 67904 – Молхаси Али, Iran; 145586 – Кунчам Сям, India; 145587 – Кедукоди Бабушри С, India. Оказывается, первый, третий и четвертый авторы в своей совместной публикации сослались на публикацию второго автора.

Вернемся к ориентированному графу G-cit и рассмотрим только гигантскую компоненту, найденную в неориентированном случае, поскольку остальные компоненты связности не слишком влияют на основные характеристики сильной связности G-cit. Максимальная компонента сильной связности (КСС) этого графа содержит 21108 вершин и 264690 дуг. Диаметр 14, средняя длина пути 5,715.

Следующая по размерности КСС содержит лишь 19 вершин. Всего имеем 643 КСС размерности от 2 до 19, причем больше половины из них (382) – это пары вершин, а 141 – триады.

Несколько проверенных КСС из четырех-шести вершин показывают, что, как правило, это авторы, работающие в одном институте, написавшие две и более совместных статьи, у которых в последующих статьях есть ссылка на хотя бы одну предыдущую.

Интересная особенность реальных сетей заключается в наличии у них свойства кластеризации или структуры в виде сообществ, в соответствии с которым топология графа организована в сообщества (также называемые модулями или кластерами) [7]. Один из теоретически хорошо обоснованных подходов для неориентированных графов к формализации этой проблемы называется кластеризацией на основе плотности связей. Мера модулярности показывает, насколько данное разбиение качественно в том смысле, что существует много дуг, ле-

жащих внутри сообществ, и мало дуг, лежащих вне сообществ, но соединяющих их между собой. Модулярность часто используется для определения качества разбиения графа на сообщества. Мы здесь используем определение меры модулярности Q из [8]. Значение Q лежит в интервале $[-1,1]$ и разбиение считается хорошим, если значение Q больше 0.7.

Для графа G -cit алгоритм, предложенный в [9], дает значение $Q=0.722$. При этом максимальное сообщество содержит 12796 вершин, общее количество сообществ равно 605, самые маленькие сообщества содержат по 2 вершины. Содержательные объяснения по поводу организации маленьких сообществ те же самые, что и в случае КС и КСС.

Если в G -cit оставить только подграф, состоящий из максимальной КС, то значение меры модулярности остается почти тем же, $Q_{CC}=0.724$, но их количество уменьшается до 90, максимальное сообщество содержит 12045 вершин. Самые маленькие сообщества содержат по 3 вершины, их количество равно 10.

Если же в G -cit оставить только подграф, состоящий из максимальной КСС, то получаем $Q_{SCC}=0.687$ и 24 сообщества размерности от 3 до 5472 вершин. Сообщество из трех вершин только одно, все его участники работают в одной организации.

Следующее по величине сообщество содержит 11 вершин, и тут получаем 8 участников из Молдовы, 2 из Испании и 1 из Канады. Содержательно все ссылки сделаны из статей, опубликованных в журнале «Buletinul Academiei de Stiinta a Republicii Moldova. Matematica», на статьи из этого же журнала. Большие сообщества не поддаются содержательной интерпретации с такой же легкостью.

2. ГРАФ НА ОСНОВЕ СОАВТОРСТВА

Обозначим граф научного сотрудничества, построенный на основе данных о соавторстве G -aut. По данным о соавторстве из Math-Net.Ru получаем более 105 тысяч авторов и более 340 тысяч случаев соавторства, то есть пар авторов в том случае, если они совместно написали хотя бы одну статью. Таким образом, имеем неориентированный граф G -aut без петель и кратных ребер, не имеющий изолированных вершин, который содержит 105327 вершин и 340643 ребра.

Основные характеристики графа: плотность очень мала (0.00006), диаметр 24, средняя длина пути 19.062, а средняя степень вершины 6.468.

Степени вершин показывают количество соавторов данного автора. У академика А.М. Прохорова наибольшая степень вершины в этом графе (798), у авторов данной статьи – 9 и 4 соответственно.

Максимальная КС G-aut содержит 79517 вершин, а вторая по размерности компонента содержит 78 вершин. Всего же граф содержит 7939 компонент связности. Как и в случае G-cit, имеются гигантская компонента и еще больше маленьких компонент, не связанных с ней и между собой. При этом количество компонент из двух вершин равно 4115, трёх – 1833, четырёх – 849, пяти – 408, и так далее по убыванию. В сумме «маленькие» КС содержат почти четверть всех вершин графа.

Рассмотрим компоненту из 78 вершин. Оказывается, большую ее часть составляют вершины, идентифицирующие коллег из Самарского государственного университета при участии небольшого числа ученых из других институтов и вузов, ведущих исследования в области органической химии и публикующихся в «Вестнике Самарского университета. Естественнонаучная серия», входящем в раздел «Журналы» Math-Net.Ru.

Несколько произвольно выбранных КС из четырех-шести вершин показывают, что наиболее часто это авторы, работающие в одном институте, имеющие статью в соавторстве. Хотя этот пример нельзя и абсолютизировать, встречаются примеры авторских групп из разных вузов одного города, а иногда (очень редко) из разных городов и стран.

Для графа G-aut алгоритм [9] дает значение $Q=0.857$. Очевидно, столь высокое значение Q объясняется почти восемью тысячами компонент связности: граф просто «распадается» на несвязные части.

Максимальное сообщество содержит 11502 вершины, общее количество сообществ равно 8229, самые маленькие сообщества количеством 4115 содержат по 2 вершины, еще 1833 сообщества по 3 вершины и т. д. Более 20 сообществ имеют от тысячи до почти четырех тысяч вершин. Для содержательной интерпретации представляют интерес сообщества, которые образуются в максимальной КС.

Для максимальной компоненты связности графа G-aut $Q_{cc}=0.842$, то есть тенденция к разбиению на сообщества остается большой. Максимальное сооб-

щество осталось большим (10346 вершины), но общее количество сообществ уменьшилось до 294. При этом исчезли все сообщества размерности 2 и 3, сообществ из четырех вершин оказалось 27, из пяти – 24, а из шести – 19.

3. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В работе [10] очень точно сказано, что «... наиболее достоверным источником информации, посвященной любому научному направлению, являются собственно научные публикации, составляющие основное наполнение таких порталов. Эти данные нуждаются в средствах анализа, которые могли бы способствовать упрощению их понимания и оптимизации научного менеджмента».

Исследования математического портала Math-Net.Ru позволяют построить две возможных модели научного сотрудничества на основе данных о цитировании и соавторстве как графов со специфическими свойствами. Представленные модели и некоторые результаты их анализа позволяют упростить процедуры обнаружения сообществ ученых и установления взаимосвязей между сообществами. Графы G-cit и G-aut построены на основании разных подходов, первый отражает сотрудничество через цитирование, а второй – сотрудничество через соавторство. Граф G-cit является ориентированным, а G-aut нет. Графы существенно различаются по мощности множеств вершин и дуг (ребер).

Тем не менее, оба графа имеют схожие черты, такие как очень маленькую плотность, большой диаметр и высокую степень модулярности. Наблюдается схожая структура связности – это наличие гигантской компоненты, содержащей десятки тысяч вершин, вторую по размерности компоненту, в тысячу раз меньшую гигантской, и большое количество маленьких компонент связности. Естественным объяснением этого в том и другом случаях служит то, что «знакомство по цитированию» и «знакомство по соавторству» часто проявляются в небольших коллективах авторов.

Отсюда же следует и объяснение больших значений коэффициента модулярности: графы с такой структурой связности, очевидно, показывают тенденцию к кластеризации в большое количество слабо- или несвязных сообществ.

Обнаружение сообществ как соавторов, так и цитирующих друг друга авторов, на самом деле являющихся соавторами более ранних работ, является хорошим подспорьем для оптимизации научного менеджмента. Как справедливо

заметил Д. Кеннеди, редактор журнала Science, «... команда есть команда, и ее члены должны разделять и заслуги, и упреки» [11].

Установление взаимосвязей между сообществами ученых, занимающихся различными направлениями математики, может способствовать принятию рациональных решений по поощрению междисциплинарной кооперации, в том числе и за счёт грантовой поддержки.

СПИСОК ЛИТЕРАТУРЫ

1. *Kas M., Carley K.M., Carley L.R.* Trends in science networks: understanding structures and statistics of scientific networks // *Social Network Analysis and Mining*. 2012. No. 2. P. 169–187.
2. *Евин И.А., Хабибуллин Т.Ф.* Социальные сети // *Компьютерные исследования и моделирование*. 2012. Т. 4, № 2. С. 423–430.
3. *Дадалко В.А., Дадалко С.В.* Метрические исследования как форма анализа научной продуктивности // *Аналитика и научное проектирование*. 2019. №2. С. 125–135.
4. *Newman M.E.J.* The structure of scientific collaboration networks // *Proceedings of the National Academy of Sciences of the USA*. 2001. No. 98 (2). P. 404–409.
5. *Kas M., Carley K.M., Carley L.R.* Trends in science networks: understanding structures and statistics of scientific networks // *Social Network Analysis and Mining*. 2012. No. 2. P. 169–187.
6. *Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A.* Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // *Lecture Notes in Comput. Sci.* 2013. Vol. 7961. P. 344–348.
7. *Malliaros F.D., Vazirgiannis M.* Clustering and community detection in directed networks: A survey // *Physics Reports*. 2013. Vol. 533, Iss. 4. P. 95–142.
8. *Newman M.E., Girvan M.* Finding and evaluating community structure in networks // *Physical Review E*. 2004. Vol. 69(2). P 026113.
9. *Blondel V.D., Guillaume J-L., Lambiotte R., Lefebvre E.* Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment*. 2008. P 10008.

10. Апанович З.В., Винокуров П.С. Методы и средства визуализации сетей соавторства и сетей цитирования больших научных порталов // Электронные библиотеки. 2011. Т. 14, Вып. 3. <https://elbib.ru/article/view/308>.

11. Kennedy D. Multiple Authors, Multiple Problems // Science. 2003. Vol. 301(5634). P. 733.

ON SOME PROPERTIES OF COLLABORATION GRAPHS OF SCIENTISTS IN MATH-NET.RU

A. A. Pechnikov¹, D. E. Chebukov²

¹ *Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk*

² *Steklov Mathematical Institute of RAS, Moscow*

¹pechnikov@krc.karelia.ru, ²tche@mi-ras.ru

Abstract

A study of two graphs of scientific cooperation based on co-authorship and citation according to the all-Russian mathematical portal was conducted Math-Net.Ru. A citation-based scientific collaboration graph is a directed graph without loops and multiple edges, whose vertices are the authors of publications, and arcs connect them when there is at least one publication of the first author that cites the publication of the second author. A co-authorship graph is an undirected graph in which the vertices are the authors, and the edges record the co-authorship of two authors in at least one article. The customary study of the main characteristics of both graphs is carried out: diameter and average distance, connectivity components and clustering. In both graphs, we observe a similar connectivity structure – the presence of a giant component and a large number of small components. The similarity and difference of scientific cooperation through co-authorship and citation is noted.

Keywords: *scientific collaboration, citation, co-authorship, graph, mathematical portal Math-Net.Ru*

REFERENCES

1. *Kas M., Carley K.M., Carley L.R.* Trends in science networks: understanding structures and statistics of scientific networks // *Social Network Analysis and Mining*. 2012. No. 2. P. 169–187.
2. *Yevin I.A., Khabibullin T.F.* Social networks // *Computer Research and Modeling*. 2012. Vol. 4, No. 2. P. 423–430 (in Russian).
3. *Dadalko V.A., Dadalko S.V.* Metricheskie issledovanija kak forma analiza nauchnoi produktivnosti // *Analitika I nauchnoe proektirovanie*. 2019. №2. S. 125–135 (in Russian).
4. *Newman M.E.J.* The structure of scientific collaboration networks // *Proceedings of the National Academy of Sciences of the USA*. 2001. No. 98 (2). P. 404–409.
5. *Kas M., Carley K.M., Carley L.R.* Trends in science networks: understanding structures and statistics of scientific networks // *Social Network Analysis and Mining*. 2012. No. 2. P. 169–187.
6. *Chebukov D., Izaak A., Misyurina O., Pupyrev Yu., Zhizhchenko A.* Math-Net.Ru as a digital archive of the Russian mathematical knowledge from the XIX century to today // *Lecture Notes in Comput. Sci.* 2013. Vol. 7961. P. 344–348.
7. *Malliaros F.D., Vazirgiannis M.* Clustering and community detection in directed networks: A survey // *Physics Reports*. 2013. Vol. 533, Iss. 4. P. 95–142.
8. *Newman M.E., Girvan M.* Finding and evaluating community structure in networks // *Physical Review E*. 2004. Vol. 69(2). P 026113.
9. *Blondel V.D., Guillaume J-L, Lambiotte R., Lefebvre E.* Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment*. 2008. P 10008.
10. *Apanovich Z.V., Vinokurov P.S.* Visualization tools for co-author networks and citation networks of large scientific portals// *Russian Digital Libraries Journal*. 2011. Vol. 14, Iss. 3. <https://elbib.ru/article/view/308/353>.
11. *Kennedy D.* Multiple Authors, Multiple Problems // *Science*. 2003. Vol. 301(5634). P. 733.

СВЕДЕНИЯ ОБ АВТОРАХ



ПЕЧНИКОВ Андрей Анатольевич – главный научный сотрудник Института прикладных математических исследований – обособленного подразделения ФИЦ «Карельский научный центр Российской академии наук», д. т. н. Сфера научных интересов – математическое моделирование, дискретная оптимизация, вебометрика, наукометрия.

Andrey Anatolievich PECHNIKOV – Chief Research Associate, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences, Doctor (DSc) of Technics. Research interests include mathematical modeling, discrete optimization, webometrics, scientometrics.

email: pechnikov@krc.karelia.ru,

ORCID: 0000-0002-0683-0019



ЧЕБУКОВ Дмитрий Евгеньевич – зав. информационно-издательским сектором Математического института им. В.А. Стеклова Российской академии наук, к. х. н. Сфера научных интересов – библиометрия, наукометрия, электронные библиотеки.

Dmitry Evgen'evich CHEBUKOV – Head of Information and Publishing Sector, Steklov Mathematical Institute of Russian Academy of Sciences, Candidate Chem. Sci. Research interests include bibliometrics, scientometrics, digital libraries.

email: tche@mi-ras.ru,

ORCID: 0000-0001-9738-8707

Материал поступил в редакцию 28 ноября 2020 года