

Материалы секции
ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ



12-23 апреля 2021
НОВОСИБИРСК

СИБИРСКОЕ ОТДЕЛЕНИЕ РОССИЙСКОЙ АКАДЕМИИ НАУК
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

МНСК-2021

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Материалы
59-й Международной научной студенческой конференции

12–23 апреля 2021 г.

Новосибирск
2021

УДК 004
ББК 32.81
И 741

Председатель секции — канд. техн. наук Б. Н. Пищик

Ответственный секретарь секции — доц. М. А. Держо

Экспертный совет секции

канд. физ.-мат. наук Д. С. Мигинский, канд. физ.-мат. наук
Е. С. Черемушкин, Ю. В. Вяткин, канд. физ.-мат. наук Т. В. Батура,
канд. техн. наук Ю. А. Загорulyко, канд. техн. наук Г. Б. Загорulyко,
канд. физ.-мат. наук Е. А. Сидорова, канд. физ.-мат. наук С. Р. Шакиров,
С. С. Журавлев, канд. техн. наук М. С. Тарков, канд. техн. наук
Ю. Г. Медведев, канд. физ.-мат. наук Д. А. Мигов, д-р физ.-мат. наук
Д. Е. Пальчунов, д-р филос. наук, канд. биол. наук А. Н. Савостьянов,
канд. физ.-мат. наук Г. Э. Яхьяева, д-р техн. наук В. Б. Барахнин,
д-р экон. наук Н. Л. Казначеева, канд. экон. наук А. В. Костин, канд. техн.
наук К. Ф. Лысаков, И. Г. Таранцев, канд. физ.-мат. наук Л. В. Городняя,
Н. А. Иванчева, Т. А. Андреева, Р. А. Пермяков, В. В. Селифанов,
В. В. Реутов, Д. В. Иртегов, Е. М. Толкачева, канд. техн. наук А. А. Власов

И 741 Информационные технологии : Материалы 59-й Междунар. науч.
студ. конф. 12–23 апреля 2021 г. / Новосиб. гос. ун-т. — Новоси-
бирск : ИПЦ НГУ, 2021. — 236 с.

ISBN 978-5-4437-1173-7

Данное издание представляет собой публикации тезисов 59-й
Международной научной студенческой конференции 2021 года
(МНСК-2021) по информационным технологиям. Сборник включает
219 тезисов.

Материалы конференции представляют интерес для студентов,
аспирантов, преподавателей, научных работников, сотрудников обра-
зовательных учреждений.

УДК 004
ББК 32.81

© СО РАН, 2021
© Новосибирский государственный
университет, 2021

ISBN 978-5-4437-1173-7

SIBIRIAN BRANCH OF RAS
MINISTRY OF SCIENCE AND HIGHER EDUCATION
OF THE RUSSIAN FEDERATION
NOVOSIBIRSK STATE UNIVERSITY

ISSC-2021

INFORMATION TECHNOLOGY

Proceedings
of the 59th International Students Scientific Conference

April 12–23, 2021

Novosibirsk
2021

УДК 004
ББК 32.81
И 741

Section head — Cand. Eng. B. N. Pischik

Responsible secretary — Assoc. Prof. M. A. Derzho

Section scientific committee

Cand. Phys.-Math. D. S. Miginskiy, Cand. Phys.-Math. E. S. Cheryomushkin,
Yu. V. Vyatkin, Cand. Phys.-Math. T. V. Batura, Cand. Tech. Yu. A. Zagorulko,
Cand. Tech. G. B. Zagorulko, Cand. Phys.-Math. E. A. Sidorova, Cand.
Phys.-Math. S. R. Shakirov, S. S. Zhuravlev, Cand. Tech. M. S. Tarkov, Cand.
Tech. Yu. G. Medvedev, Cand. Phys.-Math. D. A. Migov, Dr. Phys.-Math.
D. E. Palchunov, Dr. Phyl., Cand. Biol. A. N. Savost'yanov, Cand. Phys.-Math.
G. E. Yakh'yaeva, Dr. Tech. V. B. Barakhnin, Dr. Econ. N. L. Kaznacheeva,
Cand. Econ. A. V. Kostin, Cand. Tech. K. F. Lysakov, I. G. Tarantsev,
Cand. Phys.-Math. L. V. Gorodnyaya, N. A. Ivancheva, T. A. Andreeva,
R. A. Permyakov, V. V. Selifanov, V. V. Reutov,
D. V. Irtegov, E. M. Tolkacheva, Cand. Tech. A. A. Vlasov

И 741 Information technology : Proceedings of the 59th International Students Scientific Conference. April 12–23, 2021 / Novosibirsk State University. — Novosibirsk : IPC NSU, 2021. — 236 p.

ISBN 978-5-4437-1173-7

This edition represents the publications of the 59th International Scientific Student Conference 2020 (ISSC-2020) theses in information technology. The collection includes 219 theses.

These Conference materials can be of interest for students, Ph.D. candidates, professors, scientists, and members of educational institutions.

УДК 004
ББК 32.81

ISBN 978-5-4437-1173-7

© SB RAS, 2021
© Novosibirsk State University, 2021

ЗАЩИТА ИНФОРМАЦИИ И КИБЕРБЕЗОПАСНОСТЬ

УДК 004.056.53

Разработка алгоритма проведения аудита кибербезопасности

В. Р. Ан, В. А. Табакаева

Новосибирский государственный технический университет
Сибирский государственный университет геосистем
и технологий, Новосибирск

В современных условиях, когда государственные информационные системы (ГИС) проникают во все сферы деятельности системы государственной власти и органов местного самоуправления, их взаимодействие с сетью Интернет оказывает негативное влияние с точки зрения информационной безопасности (ИБ), так как становятся доступными для реализации угроз внутренними и внешними нарушителями разного потенциала. В настоящее время не уделяется достаточного внимания выполнению работ, связанных с аудитом кибербезопасности ГИС. Это связано прежде всего с отсутствием необходимой нормативной правовой базы и методик в области проведения аудита кибербезопасности.

Аудит кибербезопасности является одной из важнейших составляющих для решения проблемы обеспечения защищенности информации. В результате аудита кибербезопасности ожидается отчет по состоянию уровня безопасности ГИС для дальнейшего анализа и определения перспектив развития с точки зрения менеджмента ИБ. Одной из промежуточных целей при разработке метода проведения аудита кибербезопасности ГИС является обоснованный выбор конкретных критериев оценки эффективности. В работе определен перечень критериев оценки эффективности системы защиты информации.

Проведен анализ актуальных публикаций по тематике аудита, выявлено несколько направлений и методов. Выделены основные этапы проведения аудита кибербезопасности:

- проведение оценки рисков, связанных с вероятностью реализации угроз безопасности в отношении активов;
- оценка соответствия информационной системы существующим стандартам в области ИБ;
- выработка рекомендаций по внедрению новых и повышению эффективности существующих механизмов безопасности ИС;
- получение максимальной отдачи от средств, инвестируемых в создание (совершенствование, модернизацию) комплексной системы ИБ.

В результате проведенного анализа был разработан алгоритм действий, включающий множество компонентов, в том числе регулирующие нормативно-правовые требования и схему проведения аудита.

Научный руководитель — В. В. Селифанов

Проблемы выбора показателей эффективности защиты информации в техническом задании

В. Е. Антипов, Е. О. Самчук
Сибирский государственный университет
геосистем и технологий, Новосибирск

Воздействие различных факторов на информационную систему может повлиять на обеспечение защиты информации. Требования к системе защиты информации информационной системы включаются в техническое задание, которое должно быть разработано в соответствии с национальными стандартами, нормативными и методическими документами. Требования к мерам и средствам защиты информации (СЗИ), которые применяются в информационной системе, достаточно высоки, так как внедряемые средства должны иметь лицензии, сертификации ФСТЭК и ФСБ и т. д. Но нет ответа на один из важнейших вопросов: насколько то или иное выбранное решение подходит для конкретного спектра задач и какова его эффективность в реалиях заданной информационной системы.

Эффективность рассчитывается для решения следующих задач: приемлемость практического использования СЗИ в конкретной ситуации; сравнение альтернативных вариантов систем; диагностирование прочих факторов и их вкладов в достижение цели; выявление путей повышения эффективности СЗИ.

Проектирование, организация и применение СЗИ всегда содержат элементы неопределенности, так как напрямую связаны с будущими событиями, которые нельзя предугадать. К сожалению, сертификации, лицензирования и испытания полностью не устраняют неопределенность свойств СЗИ и ее отдельных элементов, как и не учитывают случайные атаки. Поэтому только вероятность может являться объективной характеристикой качества СЗИ и должна быть положена в основу комплекса показателей и критериев оценки эффективности СЗИ.

Необходимость оценивать эффективность СЗИ показателями, носящими вероятностный характер, должным образом обеспечит требуемый уровень безопасности информационной системы и ее составляющих.

Научный руководитель — В. В. Селифанов

**Реализация и анализ гибридной атаки на криптографическую систему
NTRU при малых значениях параметров
с использованием алгоритма квантового поиска**

А. О. Бахарев

Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

Квантовые вычисления — это быстроразвивающаяся область компьютерных исследований, которая ставит под угрозу криптографическую стойкость стандартов асимметричного шифрования, используемых в настоящее время. В 2016 г. *National Institute of Standards and Technology* (NIST) объявил конкурс *Post-Quantum Cryptography Competition* (PQCC), по завершении которого будет принят новый — квантово-устойчивый — стандарт асимметричного шифрования. Претендентами являются подходы на основе решеток, кодов, хэш-функций, изогений и многочленов от многих переменных.

Одним из финалистов 3-го раунда конкурса PQCC является криптографическая система с открытым ключом NTRU, основанная на решетках. Основной метод криптоанализа системы NTRU сводится к решению задачи поиска кратчайшего вектора решетки (SVP), в общем случае являющейся NP-трудной. Перспективными являются разработка и анализ квантовых алгоритмов, которые позволяют ускорить решение данной задачи. В [1] был представлен гибридный подход к поиску кратчайшего вектора решетки, в рамках которого используется квантовый алгоритм поиска (*Quantum search*). Целью настоящей работы является анализ эффективности указанного выше подхода для атаки на криптосистему NTRU.

Получена реализация гибридного алгоритма поиска кратчайшего вектора решетки с использованием симулятора квантовых вычислений IBM *Quantum Experience*. Сделан вывод о параметрах системы NTRU, против которой можно эффективно применять гибридную атаку с использованием существующих квантовых симуляторов.

*Работа выполнена при поддержке лаборатории криптографии
JetBrains Research.*

I. Laarhoven T., Mosca M., van de Pol J. Finding shortest lattice vectors faster using quantum search. // *Des. Codes Cryptogr.* 2015. Vol. 77. P. 375–400.

Научные руководители —
канд. физ.-мат. наук Н. Н. Токарева, А. В. Куценко

**Генератор псевдослучайных чисел
с доказанными статистическими свойствами**

В. А. Журавлев

Новосибирский государственный университет

Генераторы псевдослучайных чисел (ГПСЧ) широко используются в системах защиты информации и непрерывно исследуются в современной криптографии. В данном докладе будет представлен класс ГПСЧ, который, во-первых, успешно проходит самые сильные современные наборы статистических тестов, а во-вторых, как уже доказано, генерирует нормальные последовательности, т. е. для всякой сгенерированной последовательности $x_1 x_2 \dots$ и всякого двоичного слова w

$$\lim_{t \rightarrow 0} v_t(w) / (t - |w|) = 2^{-|w|},$$

где $v_t(w)$ — количество вхождений w в последовательности $x_1 \dots x_{|w|}$, $x_2 \dots x_{|w|+1}$, \dots , $x_{t-|w|+1} \dots x_t$.

Научный руководитель — д-р техн. наук, проф. Б. Я. Рябко

**Об эквивалентных ключах. Анализ криптосистемы
с открытым ключом, основанной на сложности решения системы
полиномиальных уравнений в целых числах**

Е. В. Завалишина

Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

В 2016 г. *The National Institute of Standards and Technology* представил доклад под названием *Report on Post-Quantum Cryptography* [1], в котором сообщил, что некоторые задачи, лежащие в основе классической криптографии, могут быть решены квантовыми компьютерами, и предположил начало эры постквантовой криптографии.

В связи с этим в настоящей работе была предпринята попытка создать новый алгоритм шифрования данных с открытым ключом, основанный на решении системы однородных полиномиальных уравнений в целых числах [2]. Далее для описанной криптосистемы было выявлено наличие эквивалентных ключей, а также способ их нахождения через решение системы сравнений для случая с шифрованием двух чисел [3].

В данной работе описывается способ нахождения эквивалентных ключей для случая с произвольным количеством шифруемых чисел, показан общий вид системы сравнений, используемой для нахождения ключей, найдено количество таких ключей для произвольного числа шифруемых значений и модулей, а также выявлены некоторые свойства таких ключей, которые могут быть использованы для оптимизации алгоритма их поиска.

*Работа выполнена при поддержке лаборатории криптографии
JetBrains Research.*

1. National Institute of Standards and Technology NIST Internal or Interagency Reports (IR) 8105 Report on Post-Quantum Cryptography. Gaithersburg, Maryland, April 2016. 15 p.

2. Волков Е., Баранов А., Завалишина Е. Криптографическая система с открытым ключом // Second Conf. Software Eng. Inf. Manage. (SEIM-2017). 2017. С. 41–44.

3. Завалишина Е. Криптоанализ базовой версии криптосистемы с открытым ключом, основанной на сложности решения системы уравнений в целых числах // Материалы 58-й Междунар. науч. студ. конф. / Новосиб. гос. ун-т. Новосибирск: ИПЦ НГУ, 2020. С. 78.

Научный руководитель — канд. физ.-мат. наук, доц. Н. Н. Токарева

S-блоки с высокой компонентной алгебраической иммунитетом

Д. А. Зюбина

Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

Одной из основных разновидностей блочных шифров является подстановочно-перестановочная сеть, которая является комбинацией подстановок (S-блоков) и перестановок. S-блок представляет собой векторную булеву функцию, построенную на основе n булевых функций от n переменных. Для того чтобы шифр был достаточно устойчив к алгебраическим атакам, необходимо, чтобы компонентная алгебраическая иммунитет S-блока принимала максимально возможное значение.

Алгебраической иммунитетом $AI(f)$ булевой функции f называется минимальное число d такое, что существует булева функция g степени d , нетождественно равная нулю, для которой выполняется равенство $fg = 0$ или $(f \oplus 1)g = 0$ [1]. Для функции f от n переменных максимально возможное

$AI(f) = \left\lfloor \frac{n}{2} \right\rfloor$. Компонентной алгебраической иммунитетом $AI_{comp}(F)$ векторной булевой функции называется минимальная алгебраическая иммунитет ее компонентных функций, т. е. функций $f_b(x) = \langle b, F(x) \rangle$, где $b \in \mathbb{Z}_2^n, b \neq 0$ и $\langle a, b \rangle = a_1 b_1 \oplus \dots \oplus a_n b_n$ — скалярное произведение векторов по модулю 2.

В данной работе для построения S-блока с максимальной компонентной алгебраической иммунитетом был реализован метод нахождения линейного подпространства размерности n в множестве булевых функций от n переменных с максимальной алгебраической иммунитетом и нулевой функции.

Гипотеза. В множестве, состоящем из булевых функций от n переменных с максимальной алгебраической иммунитетом и нулевой функции, существует линейное подпространство размерности n .

В частности, было получено, что существует в точности 1888 таких линейных подпространств размерности 3 для булевых функций от 3 переменных. В результате на полученных данных возможно построение S-блока, устойчивого к алгебраическим атакам.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

1. Meier W., Pasalic E., Carlet C. Algebraic attacks and decomposition of Boolean functions // Eurocrypt 2004. LNCS. 2004. Vol. 3027. P. 474–491.

Научный руководитель — канд. физ.-мат. наук Н. Н. Токарева

Разработка программы для запутывания кода на уровне промежуточного представления LLVM

И. А. Корякин

Новосибирский государственный университет

На сегодняшний день защита программного обеспечения с помощью обфускации кода — это один из эффективнейших способов воспрепятствовать модификациям и анализу программных продуктов.

Обфускация или запутывание кода — это приведение исходного текста или исполняемого кода программы к виду, сохраняющему ее функциональность, но затрудняющему анализ, понимание деталей ее работы, декомпиляцию и модификацию.

Целью данной работы является выбор подхода для реализации, а также реализация обфусцирующего компилятора.

В данной работе рассматриваются популярные обфускаторы для языков C/C++. Были рассмотрены следующие обфускаторы:

- Obfuscator-LLVM, который работает на уровне промежуточного представления LLVM. Соответственно, может работать с языками, использующими это промежуточное представление;
- Tigress, который работает на уровне исходного кода. Поддерживает только язык программирования C. Тяжел в интеграции в существующие системы сборки.

В ходе работы был выбран подход преобразования промежуточного представления, а также выбрано конкретное промежуточное представление и библиотеки для работы с ним.

В результате работы была реализована программа, производящая запутывание исходного кода программ посредством модификации промежуточного представления.

Научный руководитель — д-р техн. наук, доц. К. В. Павский

Применение расширений процессорной архитектуры x86 для затруднения анализа программного кода

Р. К. Лебедев

Новосибирский государственный университет

Защита программного кода является трудноразрешимой задачей: если программа не требует интернет-подключения для работы, вся ее логика должна быть реализована в доступном пользователю исполняемом файле, а значит, сложность восстановления кода обеспечена только трудоемкостью процесса обратной разработки.

Для обратной разработки исследователям доступны такие инструменты, как дизассемблеры, декомпиляторы и средства символьного исполнения. Они значительно ускоряют работу: декомпиляторы позволяют читать исполняемый код в виде Си-подобного псевдокода, а средства символьного исполнения — автоматически составлять условия достижимости различных ветвей кода, в том числе и нежелательных для автора программы.

Целью данной работы является исследование возможностей по противодействию этим инструментам при помощи редко используемых инструкций процессорной архитектуры x86, наличие которых в программе может приводить к их частичной или полной неработоспособности.

В качестве исследуемого расширения был выбран набор инструкций AES-NI, реализующий раунды шифра AES в качестве одиночных процессорных инструкций. Это расширение широко поддерживается современными процессорами, однако нечасто используется в программах, не связанных с криптографией.

В ходе работы был предложен метод обфускации констант, основанный на инструкции AESENC из состава AES-NI. Прототип этого метода был реализован при помощи преобразований байткода LLVM для языков C и C++ (компилятор Clang).

В результате работы было установлено, что предложенный метод действительно затрудняет использование широко распространенных инструментов обратной разработки. Декомпиляторы IDA и Ghidra оказались не способны упростить выражения для запутанных методом констант, а инструмент символьного исполнения angr вовсе прекращает работу с ошибкой, встретив инструкцию AESENC.

Научный руководитель — д-р техн. наук, доц. К. В. Павский

Разработка и экспериментальное исследование адаптивного статистического теста

В. В. Малханов

Новосибирский государственный университет

В современном мире большое значение имеют генераторы случайных чисел. Они широко используются в многих задачах, включая моделирование, программирование и криптографию. Так, например, в криптографии генерируемые последовательности имеют огромную роль: от их качества зависит секретность информации. Поэтому очень важно гарантировать определенный уровень близости используемой последовательности к истинно случайной, что приводит к необходимости их тестирования. Такие тесты называются статистическими.

Существует множество статистических тестов, некоторые из которых объединяются в так называемые батареи. Все эти тесты можно классифицировать по скорости используемых алгоритмов и, конечно, эффективности в плане проверки случайности. На текущий момент не существует идеального теста, поэтому задача создания новых более эффективных и быстрых тестов представляет большой теоретический и практический интерес.

В работе планируется разработка нового статистического теста *The order test* на C++ и экспериментальная оценка его мощности относительно других известных тестов. Данный тест основан на идеях универсального кодирования, предлагающих простым переупорядочиванием эффективно сжать искоемое сообщение. Для оценки его мощности используются различные генераторы случайных чисел, взятые из библиотек TestU01 и PractRand. Помимо этого, планируется сравнить полученные результаты с тестом из того же класса «Стопка книг», а также с такими батареями статистических тестов, как Alphabit, Rabbit и Crush.

В результате удалось реализовать *The order test* таким образом, что одна итерация алгоритма выполняется за константное время независимо от размера алфавита. Что касается проведенного исследования, то *The order test* ничуть не уступает тесту «Стопка книг» по эффективности. Однако главным преимуществом исследуемого теста все же является его небольшое время выполнения относительно других известных батарей: Alphabit, Rabbit и Crush.

Научный руководитель — д-р техн. наук, проф. Б. Я. Рябко

Экспериментальное исследование эффективности тестов для проверки генераторов случайных чисел

Р. А. Сафаров

Новосибирский государственный университет

В данном исследовании проведено исследование эффективности тестов «Стопка книг», TestU01 и тестов PractRand для проверки генераторов случайных чисел. На основе проведенного исследования был сделан вывод о том, какой из тестов находит отклонения на большем количестве генераторов и при меньшей длине последовательности.

В данной работе приведены результаты тестирования генераторов случайных чисел и представлены длины последовательностей в битах, выдаваемые генератором, с которых начинаются первые отклонения от случайности, определяемые данным тестом. Если отклонения обнаружены, то с увеличением длины входной последовательности отклонения возрастают. Протестированы 14 различных генераторов случайных чисел, среди которых можно выделить линейные конгруэнтные генераторы, потоковый криптографический генератор RC4, функцию rand в ОС Linux, Mersenne Twister и AES+Feistel+Reverie.

В результате было показано, что найдены отклонения от случайности:

1. **PractRand** (стандартный, 1 терабайт) — на 10 генераторах;
2. **TestU01** (SmallCrush, Crush, BigCrush, Alphabit, Rabbit) — на 9 генераторах;
3. «**Стопка книг**» — на 9 генераторах.

Прочие преимущества

1. С помощью **PractRand** и **TestU01**, как правило, легче всего интерпретировать результат.
2. С помощью **PractRand** и «**Стопка книг**», как правило, проще всего автоматизировать тестирование через интерфейс командной строки.
3. Стопка книг работает быстрее, чем другие наборы тестов.

Различные недостатки

1. **PractRand** требовал больше битов ввода для тестирования, чем другие наборы тестов. Это может быть проблемой, если ваш RNG очень медленный или иным образом ограничен по объему производимых данных.
2. «**Стопка книг**» требовал только двоичных битов ввода для тестирования.

Научный руководитель — д-р техн. наук, проф. Б. Я. Рябко

ИНСТРУМЕНТАЛЬНЫЕ И ПРИКЛАДНЫЕ ПРОГРАММНЫЕ СИСТЕМЫ

УДК 004.4'2

Исследование способа расширения Web-IDE процесс-ориентированного языка Reflex

В. А. Витченко

Институт автоматики и электрометрии СО РАН, Новосибирск

Разработка промышленных процесс-ориентированных алгоритмов управления киберфизическими системами сопряжена с рядом сложностей и рисками, связанными с высокой стоимостью ошибок. Для оптимизации процесса разработки вынуждены использовать нестандартные средства разработки.

В ИАиЭ СО РАН разрабатывается интерактивная среда разработки RIDE 2.0 на базе платформы Eclipse Theia. RIDE 2.0 позволяет разрабатывать промышленные процесс-ориентированные алгоритмы управления на языке Reflex.

Одна из особенностей проекта — необходимость предоставить сторонним разработчикам возможности расширения функциональности. Решение должно предполагать изолированность основной функциональности и должно быть направлено на упрощение создания расширений для среды разработки.

В ходе работы был проведен анализ специфики задачи, платформы Eclipse Theia, кодовой базы RIDE 2.0, существующих аналогов расширяемых сред разработки. В качестве решения был разработан генератор репозитория, позволяющий инициализировать настроенный проект расширения.

Генератор создан на базе библиотеки Yeoman, опубликован в реестре www.npmjs.com.

Практическая ценность полученных результатов заключается в сокращении трудоемкости создания расширений для RIDE 2.0, унификации способа расширения.

Полученные результаты позволяют упростить поддержку RIDE 2.0 и сократить трудозатраты разработчиков.

Научный руководитель — д-р техн. наук В. Е. Зюбин

Автоматизация поиска режима синхронизации мод волоконного лазера на эффекте нелинейного вращения поляризации

Н. А. Воробьев

Новосибирский государственный университет
Институт автоматики и электрометрии СО РАН, Новосибирск

Важным преимуществом волоконных лазеров является способность генерации ультракоротких импульсов высокой интенсивности, применяющихся как в научных исследованиях, так и в производственных сферах. Такие сверхкороткие импульсы возникают только тогда, когда созданы необходимые условия для синхронизации мод. Одним из наиболее простых и эффективных способов достижения синхронизации мод является использование эффекта нелинейного вращения поляризации (НВП). В этом случае требуется подобрать определенное состояние поляризации в оптическом волокне. Однако режимы, получающиеся при использовании эффекта НВП, неустойчивы к температуре, механическим воздействиям и любым изменениям волоконной схемы, а поиск режима синхронизации мод возможно провести лишь вручную.

Эффективным решением данной проблемы является автоматизация поиска режима синхронизации мод волоконного лазера путем управления состоянием поляризации. Для этого в волоконный лазерный резонатор помещается электронно-управляемый контроллер поляризации (КП), меняющий состояние поляризации в волокне в зависимости от установленных значений тока. Затем, изменяя параметры КП по определенному алгоритму и анализируя выходной сигнал лазера с помощью измерительных приборов, можно найти набор параметров, при которых реализуются условия для синхронизации мод.

Таким образом, целью работы является разработка аппаратно-программного модуля для автоматизированного поиска режима синхронизации мод волоконного лазера. Модуль включает управление параметрами КП, сбор данных с измерительных приборов и идентификацию режима синхронизации мод по этим данным, а также эволюционный алгоритм, находящий нужные параметры для КП. К используемой в работе волоконной лазерной схеме такой подход к поиску режимов генерации применяется впервые.

На конференции будут представлены результаты применения эволюционного алгоритма к поиску режима синхронизации мод, а также методы идентификации этого режима, используемые алгоритмом для оценки качества синхронизации мод.

Научный руководитель — Н. Н. Смолянинов

Создание облачной среды разработки для процесс-ориентированного языка Reflex

И. А. Горнев

Институт автоматики и электротехники СО РАН, Новосибирск

Процесс-ориентированный язык Reflex предназначен для разработки алгоритмов управления киберфизическими системами. Киберфизические системы подразумевают взаимодействие программируемого оборудования с физической окружающей средой. Они используются повсеместно, например при промышленных производствах или в промышленном интернете вещей. Киберфизические системы обладают рядом свойств, таких как открытость, событийность, синхронизм, логический параллелизм и др. Построение эффективных и безопасных алгоритмов управления такими системами на языках общего назначения затруднено. Процесс-ориентированная парадигма программирования, в частности язык Reflex, учитывает особенности киберфизических систем, поэтому подходит для данной задачи. Синтаксис языка Reflex основан на синтаксисе языка C, что облегчает порог вхождения. Язык Reflex доказал свою эффективность и неоднократно применялся на промышленных объектах.

Несмотря на доказанную эффективность языка Reflex, его использование затруднено отсутствием специализированной интегрированной среды разработки. Было разработано расширение для редактора Notepad++, предоставляющее часть функциональности, но оно не включает сторонние инструменты, такие как верификаторы, а также не предоставляет возможности использования в качестве веб-приложения.

Цель работы — создание интегрированной среды разработки, специализированной для языка Reflex. Архитектурными требованиями к разрабатываемому продукту стали модульность архитектуры, возможность расширять функциональность среды разработки, возможность разворачивать продукт как в виде приложения для рабочего стола, так и в виде веб-приложения. В ходе работы был проведен анализ средств реализации, разработана расширяемая архитектура, а также был реализован прототип среды разработки под названием RIDE. Для реализации среды разработки использовался фреймворк Eclipse Theia, в который интегрировались процесс-ориентированные языки, описанные с помощью фреймворка Eclipse XText.

Научный руководитель — Т. В. Лях

Автоматическая генерация верифицирующего кода для динамической верификации процесс-ориентированных Reflex-программ

А. Л. Гривцова

Институт автоматики и электротехники СО РАН, Новосибирск

В современном мире киберфизические системы (КФС) нашли широкое применение в науке и промышленности. Одной из важнейших задач при разработке программного обеспечения (ПО) для таких систем является верификация управляющей программы, наличие ошибок в которой может повлечь катастрофические последствия. Наиболее распространенным подходом является динамическая верификация. Но верификация усложняется из-за специфики КФС, заключающейся в интеграции цифровой и физической компонент.

В ИАиЭ СО РАН был предложен гибридный метод динамической верификации ПО КФС, объединяющий тестирование и мониторинг свойств на имитаторе объекта управления. В методе используется четырехкомпонентная модель динамической верификации процесс-ориентированных Reflex-программ [1], которая предполагает организацию программы в виде четырех алгоритмических блоков: алгоритма управления (*Controller*), блока управления сценариями (*Dispatcher*), блока верификации (*Verifier*), виртуального объекта управления (*Plant*). Метод был реализован в комплексе автоматической верификации управляющих Reflex-программ. Однако на данный момент код программы на языке Reflex для каждого блока необходимо реализовывать вручную, что существенно замедляет процесс верификации.

Целью настоящей работы являлась разработка ПО для автоматической генерации кода на языке Reflex. Для описания требований к управляющей программе была выбрана темпоральная логика. Разработаны правила трансляции для автоматического преобразования требований на языке темпоральной логики в код на языке Си, который затем встраивается в блоки *Verifier* и *Dispatcher*. Таким образом, результаты настоящей работы позволяют ускорить динамическую верификацию ПО КФС.

1. Зюбин В. Е. Процесс-ориентированное программирование: учеб.-метод. пособие. Новосибирск: НГУ, 2011. 192 с.

Научный руководитель — Т. В. Лях

Разработка модуля отладки Reflex-программ на цифровых моделях объекта управления

А. А. Двинянин

Институт автоматики и электрометрии СО РАН, Новосибирск

Разработка управляющих программ киберфизическими системами (КФС) сопряжена с рядом трудностей. ПО КФС имеют особенности, которые усложняют их разработку (открытость, событийность, логический параллелизм, цикличность). Это приводит к тому, что разработчики ПО КФС используют специализированные языковые средства (МЭК-61131, G LabVIEW и т. д.). Одним из таких языковых средств является Reflex [1] — процесс-ориентированный язык, предназначенный для программирования устройств логического управления.

В Институте автоматики и электрометрии СО РАН разрабатывается RIDE — интегрированная среда разработки для языка Reflex. RIDE предполагает базовую функциональность сред разработки, статическую и динамическую верификацию на виртуальном объекте управления, а также возможность динамической отладки. Также был разработан транслятор Reflex-кода в код для микроконтроллера Atmega16, но отсутствуют средства отладки для Reflex-программ.

Целью данной работы являлось создание модуля динамической отладки микроконтроллера Atmega16 для среды разработки языка Reflex.

В работе были проанализированы существующие средства отладки, разработаны архитектура и алгоритм отладки. Был модифицирован транслятор Reflex-кода в С-код посредством фреймворка Xtext, реализована пошаговая отладка и отладка в режиме точек останова с возможностью мониторинга и модификации состояния системы и формирования пошагового отчета, отладка с ручной имитацией объекта управления и на цифровой модели.

1. Зюбин В. Е. Процесс-ориентированное программирование: учеб.-метод. пособие. Новосибирск: НГУ, 2011. 192 с.

Научный руководитель — Т. В. Лях

Построение веб-средств спектрального анализа на основе онлайн-баз данных эталонных спектров атомных систем

В. Е. Жилицкий

Новосибирский государственный университет

Изучение спектральных данных, разложение спектра и определение составляющих его элементов является актуальной задачей, часто встречающейся в науке и технике. Зачастую задача решается путем визуального сравнения пиков со всеми спектрами ожидаемых элементов или простым поиском самых явных пиков и сравнением его с ограниченными встроеными табличными данными.

Однако сейчас существуют информационные системы, хранящие наиболее полные и регулярно обновляющиеся базы данных с подробными спектрами, доступные онлайн (NIST, SPECTR-W3, VAMDC и др.), применение которых существенно расширило бы возможности и упростило анализ спектральных данных для пользователя. Более того, перенос алгоритмов анализа на сервер позволяет реализовать принцип «тонкого» клиента, что может ускорить вычисления, повысить их точность и снизить требования к измерительным устройствам.

По итогу работы был спроектирован и реализован набор php-скриптов с JavaScript интерфейсом, использующий данные информационной системы «Электронная структура атомов».

На первом этапе был реализован интерфейс, простые средства визуализации на chart.js и средства загрузки и хранения пользовательских данных в формате .json. Далее была добавлена отправка SQL запросов в базу данных ИС ЭСА, потом был реализован функционал обработки данных для удаления систематической ошибки и определения уровня шума, после чего были реализованы алгоритмы последовательного и приоритезированного анализа пользовательского спектра с маркировкой детектированных линий, выделения составляющих его элементов и определения их относительной интенсивности.

Алгоритмы протестированы и отлажены, показано преимущество от использования серверных вычислений и онлайн-баз данных, планируется интеграция в систему ИСА.

Научный руководитель — канд. физ.-мат. наук В. Г. Казаков

Комплекс программных средств для решения логистических задач компании «Мир мебели»

М. В. Латышев

Иркутский государственный университет

Современный бизнес трудно представить без применения информационных технологий. Они используются на этапах планирования, производства, доставки, хранения и т. д. Для обеспечения этих процессов разработано множество программных продуктов, которые помогают максимально эффективно решать задачи для бизнеса.

К наиболее значимым задачам для компаний, занимающихся доставкой, относится проблема транспортной логистики. Многие заинтересованы в эффективном решении данной проблемы. Одной из таких компаний, работающих в Иркутской области, является «Мир мебели». Она осуществляет продажу мебели с применением технологических решений. В основные ее задачи входит оптимизация доставки заказов.

В настоящее время существует множество программных продуктов, решающих задачи логистики. Они актуальны и хорошо применимы в своей области. Но не существует такого программного продукта, который совмещал бы в себе и транспортную логистику, и оптимальную упаковку товаров в контейнеры.

Реализованное приложение имеет такие входные параметры, как список заказов, которые необходимо доставить, и список автомобилей, осуществляющих доставку. Каждый заказ включает данные о товарах и их уникальных особенностях, адрес и параметры доставки и местонахождения товара. Далее программа осуществляет автоматизированное планирование доставки с учетом проверки вместимости груза в автомобиле и ограничений на перевозку, контроля планового времени доставки, расположения товаров на складах, затрат на ГСМ. Таким образом, решаются важные задачи минимизации количества задействованных в доставке автомобилей и повышения количества доставок одним автомобилем в течение дня.

В результате проделанной работы был создан программный продукт с использованием языков программирования 1С и Python. На данный момент разработанный продукт проходит внедрение в компании «Мир мебели».

Научный руководитель — канд. физ.-мат. наук, доц. О. В. Зубков

Интеграция алгоритмов доказательства с нулевым разглашением в смарт-контракты Ethereum

А. А. Лаханский, Д. А. Сафенрейтер
Новосибирский государственный университет
Лаборатория криптографии JetBrains Research

Ethereum — платформа, расширяющая возможности сети блокчейна биткойна. Она позволила применить концепцию блокчейна не только для децентрализованных валют, но и для любых задач, в которых необходимо гарантировать выполнение действий каждой из сторон.

Ethereum выполняет вычисления на всех узлах сети, что приводит к высоким затратам, ограничениям по сложности и низкой конфиденциальности. zk-SNARK позволяет проверять вычисления в цепочке: части алгоритма zk-SNARK добавляются в сеть в виде заранее скомпилированных смарт-контрактов. После того как генератор создал ключи доказательства и верификации, любой доказывающий может использовать этот ключ для создания доказательства вне сети. А далее, алгоритм проверки может выполняться внутри смарт-контракта.

Существенной проблемой при использовании zk-SNARK является высокая сложность понимания выполняемых протоколом вычислений и, как следствие, сложная работа с данными вычислениями.

ZoKrates, набор инструментов для работы с zk-SNARK в Ethereum, устраняет этот пробел. Он помогает создавать автономные программы и связывать их с блокчейном Ethereum, таким образом расширяя возможности децентрализованного приложения.

ZoKrates имеет ряд недостатков:

- невозможна реализация сложных схем;
- все алгоритмы необходимо реализовывать вручную, а потому при изменении смарт-контракта также необходимо будет сгенерировать новые параметры.

Расширение функционала языка ZoKrates осуществляется посредством добавления модулей в виде файла формата .zok и их последующего добавления в стандартную библиотеку проекта stdlib. Затем при сборке весь функционал из stdlib доступен для использования в коде. Таким образом, ZoKrates открыт к добавлению нового функционала и редактированию уже существующего.

В рамках работы были имплементированы новые криптографические примитивы, расширяющие возможности ZoKrates.

Работа выполнена при поддержке лаборатории криптографии JetBrains Research.

Построение адаптивного архиватора

В. В. Лусников

Новосибирский государственный университет

Сейчас все больше компаний и государственных учреждений для хранения данных используют облачные решения — онлайн-хранилища, в которых данные хранятся на многочисленных распределенных в сети серверах. Перед непосредственным хранением данные сжимают без потерь некоторым архиватором, вследствие чего уменьшаются накладные расходы на хранение.

Для сжатия данных без потерь используют архиваторы с различными алгоритмами сжатия, такими как глубокие нейронные сети (СМ) и цепи Маркова (ДМС). В связи с этим разные архиваторы показывают лучшие результаты сжатия на файлах разного формата, иными словами, при большой выборке файлов для оптимального сжатия потребуется больше одного архиватора. Поэтому существует потребность в создании стратегии выбора архиватора из существующих для оптимального сжатия конкретного файла, а также в реализации архиватора на ее основе. Такой архиватор называется адаптивным, а процесс архивации данным архиватором — адаптивным сжатием.

Целью данной работы являлась разработка архиватора, сжимающего в среднем на большой выборке файлов лучше существующих. Такой архиватор будет выбирать для каждого файла лучший архиватор из предложенных, тем самым осуществляя адаптивное сжатие. За основу схемы выбора архиватора было решено взять раундовую схему, при которой процесс выбора архиватора происходит итеративно: на следующую итерацию проходят только те архиваторы, которые оказались лучшими при сжатии небольшого фрагмента от исходного файла. Для исследования были взяты лучшие архиваторы в свободном доступе. Проверка качества сжатия проводилась на коллекциях файлов Calgary и Canterbury corpus.

В результате исследования была найдена лучшая стратегия выбора — раундовая статическая схема, на ее основе было создано программное обеспечение — адаптивный архиватор, который в среднем дает 20 % улучшение качества сжатия по сравнению с архиваторами, участвующими в исследовании.

Научный руководитель — д-р техн. наук, проф. Б. Я. Рябко

Метод расширения Web-IDE языка Reflex проблемно-ориентированными модулями

Г. Д. Манаков

Институт автоматики и электрометрии СО РАН, Новосибирск
Новосибирский государственный университет

Процесс-ориентированный язык Reflex был разработан в Институте автоматики и электрометрии СО РАН [1]. Для повышения удобства и скорости коллективной разработки была спроектирована и реализована удаленная среда разработки Web-IDE.

Возникла необходимость организации подключения/отключения проблемно-ориентированных модулей к ядру Web-IDE с целью обеспечения независимости модулей от ядра.

На основе анализа существующих решений был сформирован список требований к методам расширения ядра Web-IDE процесс-ориентированного языка Reflex проблемно-ориентированными модулями, основные из которых: а) обеспечение инкапсуляции ядра Web-IDE от модулей расширения, б) унифицированный интерфейс взаимодействия, в) возможность разработки модулей расширения с собственным графическим интерфейсом.

В качестве базовой гипотезы для разрабатываемого механизма расширения Web-IDE предложено оформлять отдельный модуль расширения в виде пары веб-сервер на Spring и файла-конфигурации. Модули расширения размещаются в выделенной директории на сервере. При запуске ядра IDE выделенная директория сканируется бэкэнд-компонентом, конфигурационная информация передается фронтэнд-компоненту ядра, который формирует пользовательский интерфейс, включающий элементы управления модулями.

Использование новой архитектуры дает преимущество во время разработки в связи с облегчением расширения функционала среды разработки и облегчением формирования и добавления новых модулей в Web-IDE за счет отсутствия необходимости создавать отдельные расширения.

1. Liakh T. V., Rozov A. S., Zyubin V. E. Reflex language: A practical notation for cyber-physical systems // System Informatics. 2018. Vol. 2. No. 12. P. 85–104.

Научный руководитель — В. Е. Зюбин

**Разработка информационно-вычислительной системы
по прогнозированию токсикологических свойств химических веществ**

Д. А. Маслич

Новосибирский государственный университет экономики и управления

Огромное количество новых химических веществ, появляющихся на рынке, требует разработки более совершенных методов и систем предсказания химико-биологических и токсикологических свойств. Заменой традиционным методам становится использование информационных технологий, баз данных и компьютерных методов предсказания токсикологических свойств на основе уже изученных соединений, которые гораздо дешевле по стоимости и затраченному времени.

В работе разработана интеллектуальная система, которая позволяет на основе дескрипторов графов структурных формул и топологических индексов структурного графа на основе искусственных нейронных сетей прогнозировать химико-биологические и токсикологические свойства химического вещества. Предсказание параметров токсичности происходит в два этапа. На первом этапе используется байесовское решающее правило, которое вычисляет класс опасности. Классификация химических веществ производится на четыре класса опасности.

Затем в каждом классе опасности строились математические модели, основанные на технологиях нейронных сетей, и по ним вычислялся количественный показатель токсичности. Эксперименты проводились на созданной базе данных по параметрам токсичности объемом 4474 химических веществ для обучения и 800 химических веществ для экзамена.

Процент правильных решений для алгоритма классификации на четыре класса опасности составил от 94 до 96 %. Среднеквадратическая ошибка для предсказания количественного показателя токсичности составила не более 0,2, что соответствует экспериментальному определению этого показателя.

Для удобства ввода графической информации о молекулах разработан графический редактор, который позволяет рисовать структурную формулу молекул для дальнейшего использования в предсказании токсикологических свойств химических веществ и создании баз биологических и структурных данных.

Научный руководитель — канд. техн. наук, доц. А. Л. Осипов

Циркулярная укладка атрибутированного иерархического графа с портами

А. М. Меркулов

Новосибирский государственный университет

Графы широко применяются для описания и структурирования информации в предметных областях, где необходимо промоделировать связи между объектами. Для визуального анализа данных и связей между ними требуется построить изображение графа, учитывающее заданные заранее критерии. Существует программное обеспечение для работы с графами, позволяющее получить данное изображение, получая на вход различные форматы хранения графов. Созданная в Институте систем информатики им. А. П. Ершова СО РАН система Visual Graph позволяет на основе атрибутированных иерархических графовых моделей исследовать сложные структурированные данные большого размера через их визуальные представления.

Читаемость получаемого изображения графа сильно зависит от способа укладки его элементов (вершин и дуг) на плоскости. Циркулярная укладка представляет собой стиль визуализации графов, при котором вершины равномерно распределены по окружности, и находят свое применение в областях, где объекты, представленные вершинами графа, равны по приоритету и ни один из них не занимает привилегированного положения. Циркулярные укладки графов применяются для визуализации кольцевых и звездных сетевых топологий, биологических и социальных сетей, более мелких кластеров в большом графе.

В докладе будет представлен алгоритм циркулярной укладки атрибутированных иерархических графов и расширение для системы Visual Graph, реализующее этот алгоритм. Была придумана эвристика для минимизации пересечений «дуга — дуга». Для решения проблемы пересечений «дуга — вершина» используются округлые вставки в дуги. В расширении присутствует возможность укладывать графы с портами.

Научный руководитель — д-р физ.-мат. наук, проф. В. Н. Касьянов

**Разработка информационной системы
для автоматизации предоставления прав доступа**

Р. А. Нещади́мов, М. В. Дро́гина, Д. Л. Белоу́сов
Томский государственный университет
систем управления и радиоэлектроники

В настоящее время существуют предприятия с большим государственным участием в уставном капитале. Они могут затрагивать сферы добычи природных ресурсов, банковской деятельности, транспорта. Масштабы этих предприятий позволяют обеспечивать процессы жизнедеятельности целых регионов, однако отсюда возникает вопрос промышленной безопасности и необходимости их бесперебойного функционирования. Для их стабильной работы должны соблюдаться условия информационной защиты и независимости от зарубежных решений.

В связи с необходимостью независимости от зарубежных решений осуществляется переход с иностранного программного обеспечения (ПО) на отечественное. При большом количестве сотрудников для обеспечения безопасности необходимо иметь четкое разграничение прав доступа при работе с информацией. Одним из путей увеличения эффективности и автоматизации данного процесса является применение к нему информационной системы.

Цель данной работы — разработка кроссплатформенной информационной системы (ИС) управления доступом, совместимой с отечественной операционной системой Astra Linux.

ИС является клиент-серверным веб-приложением, клиентская часть которого выполнена посредством библиотеки React для языка Javascript, позволяющей разрабатывать гибкие, адаптирующиеся под устройство пользователя пользовательские интерфейсы, а серверная часть — посредством фреймворка Spring для языка Java, имеющего ряд готовых инструментов для работы по клиент-серверной схеме.

Данное веб-приложение позволяет сотрудникам посредством создания и отправки заявок запрашивать и получать доступ к информационным системам, а владельцам и администраторам ИС — предоставлять доступ либо отказывать в предоставлении доступа к курируемым ими системам, в зависимости от содержания заявки. Заявки хранятся в архиве, откуда можно получить к ним доступ.

Научный руководитель — А. А. Матолыгин

Разработка Android-приложения для системы «Деканат»

Р. Е. Полевач, М. В. Черняев

Сибирский федеральный университет, Красноярск

В настоящее время практически каждый человек имеет доступ к интернету через свой смартфон для какого-либо взаимодействия с внешним миром. В основном люди пользуются специальными приложениями, чтобы взаимодействовать с определенными сервисами, так как они предоставляют большее удобство в отличие от интернет-браузера.

В Сибирском федеральном университете разработан и активно используется сервис для взаимодействия студентов, преподавателей и деканата. Однако до настоящего времени официальная разработка имела только веб-версию и крайне плохо отображала свой интерфейс на мобильных устройствах. Поэтому решено было создать Android-приложение (а в перспективе — и его версию под iOS), соединяющееся с базой данных веб-сервиса и отображающее информацию в удобном и современном виде.

В мобильную версию приложения «Деканат» заложена поддержка работы с двумя типами пользователей: студентами университета и его преподавателями. С разработанным приложением студенты смогут легко отслеживать свои успехи, посещаемость, получать оповещения, собирать статистику о своей посещаемости. Если студент является старостой, то он сможет легко проставлять через данное приложение посещаемость студентов своей группы. Преподаватели смогут отслеживать статусы ведомостей, пересдач и комиссионных ведомостей, просматривать список приказов по конкретному студенту. Приложение может отличать студента от преподавателя и наоборот, т. е. при авторизации автоматически осуществляется переход в соответствующий каждой роли личный кабинет. Поэтому приложение является единым как для студентов, так и для преподавателей.

Мобильное приложение реализовано с применением таких технологий, как язык программирования Java; локальная база данных Room для оптимизации некоторых модулей приложения; нативные компоненты от Google, которые соответствуют их философии дизайна мобильных приложений. Также для осуществления запросов к API сервера была использована библиотека Retrofit.

Основными пользователями Android-приложения являются студенты, которые могут проверять статус своих дисциплин, а также преподаватели, которые будут иметь доступ ко всему функционалу веб-версии.

Научный руководитель — А. К. Погребников

Создание ПО для отображения многомерных медицинских данных

А. А. Попова

Новосибирский государственный университет
Институт химической кинетики и горения СО РАН, Новосибирск

В последнее время идет тенденция к мобилизации и автоматизированию обработки и отображения данных. Популярность набирают системы, способные предоставить в понятной форме сложную информацию. Одними из таких данных являются медицинские параметры крови человека. Создание системы, представляющей пользователю удобный интерфейс для получения сведений об уровне своего здоровья, является важным элементом в области медицины.

Целью работы являлось создание мобильного приложения для платформы ОС Android (> 4.1 версии), предоставляющего удобный интерфейс для отслеживания медицинских параметров. Для реализации проекта была выбрана среда разработки Android Studio и язык Java. Для считывания больших данных в приложении создана функция загрузки данных и таблицы формата XLSX, а также встроен модуль оптического распознавания символов (OCR). Приложение оснащено функциями сохранения, просмотра, удаления из истории анализов. Также была создана база медицинских данных, включающая названия анализов, используемые в лабораториях и больницах, и их референсные значения.

В дальнейшем данное приложение будет протестировано в клинике Мешалкина и будет доработан функционал с учетом пожеланий пользователей. Планируется расширить базу данных медицинских анализов, включив в нее связи между медицинским параметром и заболеваниями, которые могут вызываться отклонениями этого параметра.

Научный руководитель — д-р физ.-мат. наук В. П. Мальцев

Преобразование конфигурации 1С:Предприятия в UML-модель

В. А. Попова

Иркутский государственный университет

Современные программные системы функционируют продолжительное время и претерпевают значительные изменения в своей структуре из-за внедрения новой функциональности и доработки уже существующей. Поэтому разработчики часто используют механизмы, позволяющие визуально отследить взаимосвязь объектов системы на диаграммах языка объектного моделирования UML [1].

Иерархическая структура конфигураций, разработанных на платформе 1С:Предприятие [2], напоминает иерархическую структуру UML-моделей. Но, в отличие от редакторов UML, 1С:Предприятие не содержит механизмов для визуализации структуры. Ввиду этого возникла идея: разработать программу для конвертации конфигурации 1С в UML с целью дальнейшей визуализации объектной структуры.

Стандартом для обмена UML-моделями является формат XMI, поддерживающийся графическими редакторами UML. Поскольку имеется возможность выгрузить программу 1С в формат XML, можно выполнить преобразование из XML в XMI, а затем в UML-редакторе получить визуальное представление структуры конфигурации.

Формат XMI оказался не слишком читабельным, поэтому было решено сначала представить конфигурацию в упрощенном формате, а затем выполнить преобразование в XMI.

Для решения задачи написана программа на встроенном языке программирования 1С:Предприятия для конвертации конфигурации в промежуточный формат, а также вторая программа на языке XSLT для преобразования промежуточного формата данных в XMI.

Разработанный программный комплекс позволяет легко получать диаграммы классов UML для любой конфигурации 1С:Предприятия.

1. About the Unified Modeling Language Specification Version 2.5.1 [Электронный ресурс]. URL: <https://www.omg.org/spec/UML/About-UML/> (дата обращения: 18.02.2020).

2. 1С:Предприятие 8.3.17. Документация [Электронный ресурс]. URL: <https://its.1c.ru/db/v8317doc> (дата обращения: 19.02.2020).

Информационная система учета достижений студентов

Е. А. Резинкина, А. Д. Астахова
Кемеровский государственный университет

Привлечение обучающихся к участию в жизни вуза является одной из задач, реализуемых высшими учебными заведениями как социальными институтами в рамках социализации индивидов и их дальнейшей мотивации к участию в научной, спортивной, творческой и общественной деятельности. Существующая в нашем учебном заведении система поощрений и учета достижений студентов в данных областях не в полной мере решает поставленную задачу. Целью данной работы является разработка информационной системы (ИС) для учета достижений студентов для последующего поощрения.

Проведенный обзор показал, что существует ряд систем для учета достижений и поощрения пользователей. Однако, как правило, данные системы являются составными частями более сложных систем и предназначены исключительно для решения проблем в своей предметной области.

Основные функции ИС:

- получение достижения студентом из имеющегося списка достижений;
- добавление нового достижения в систему;
- получение награды за присужденные студенту достижения;
- просмотр рейтинга студентов по различным фильтрам;
- просмотр профилей других студентов;
- просмотр всех достижений системы;
- просмотр полученных достижений в профиле;
- просмотр статуса заявок на получение и добавление достижения.

ИС учета достижений студентов включает две подсистемы: мобильное приложение для студента и веб-приложение для администратора и модератора. В основу реализации положен сервисно-ориентированный подход. Мобильное и веб-приложения получают информацию из базы данных, используя специально написанные rest-сервисы. В качестве средств реализации для ИС был выбран контейнер сервлетов Tomcat, СУБД PostgreSQL, Spring Boot, Java for Android Studio, JavaScript-библиотеки React.

На данный момент ведется разработка информационной системы.

Научный руководитель — канд. техн. наук, доц. С. Ю. Завозкин

Алгоритм выбора фильтра для мониторинга зрительных вызванных потенциалов

А. В. Рябчикова

Новосибирский государственный университет

Для выполнения нейрохирургических операций последние годы стал активно применяться интраоперационный нейрофизиологический мониторинг (ИОНМ). В случае операций в затылочной доле головного мозга или в зонах, прилегающих к зрительным нервам и трактам, для целей ИОНМ используется метод зрительных вызванных потенциалов (ЗВП). Для улучшения соотношения сигнал/шум с целью получения четкого ответа при наименьшем усреднении применяется частотная фильтрация сигнала, отсекающая наиболее высокоамплитудные компоненты фоновой биоэлектрической активности мозга. Однако и частота пика мощности фоновой активности, и период волн ЗВП могут существенно различаться у разных пациентов. В результате возникает необходимость индивидуального подбора параметров частотной фильтрации сигнала ЗВП на каждой операции. Выполнение этого подбора нейрофизиологом «вручную» отнимает достаточно много времени, отвлекая его от проведения ИОНМ; при этом результаты такого подбора могут оказаться неоптимальными.

В связи с этим представлялось целесообразным разработать приложение, позволяющего анализировать необработанные сигналы, получаемые при регистрации ЗВП, и подбирать индивидуальные параметры их частотной фильтрации. Для решения этой задачи была разработана описанная ниже программа.

На вход программа получает экспортируемые данные из ПО, встроенного в медицинское оборудование, применяет различные значения высоких и низких частот с помощью фильтра Батерворта, определяя наилучшие диапазоны по следующим критериям: четко выделяемым пикам с большой амплитудой на требуемом интервале; отсутствию высокоамплитудных пиков на интервалах без реакции на вспышку; хорошо повторяющимся пикам при последовательных записях ЗВП в постоянных условиях (перед началом операции). На выходе программа выдает рекомендованные границы фильтров высокой и низкой частоты и график с применением этих параметров. Дальше настройки выставляются в программной оболочке системы ИОНМ, и мониторинг проводится в ней.

Научный руководитель — канд. биол. наук. Е. А. Левин

Разработка системы сопровождения научных конференций

В. С. Чефонов, Н. С. Чефонов
Сибирский федеральный университет, Красноярск

При проведении различного рода научных конференций нередко возникает необходимость автоматизации процессов регистрации участников, а также последующей обработки поступивших материалов. Зачастую научные конференции предполагают строго определенный формат, который подкрепляется порядком работы и регламентом, а значит, есть необходимость реализации отдельного программного продукта для организации подобных мероприятий. Каждая конференция — это частное событие, однако сам процесс принятия заявок, проверки материалов, отбора статей и т. д. не является уникальным. Для успешного решения подобных задач было решено реализовать информационную систему, которая могла бы включать функционал, позволяющий минимизировать очное взаимодействие организаторов конференции и авторов публикаций, а также ускорить процесс принятия (отклонения), редактуры и утверждения публикаций.

При реализации такого рода системы необходимо использовать наиболее универсальный подход с возможностью индивидуальной настройки. Реализованная система предполагает наличие клиентской и серверной частей, которые взаимодействуют между собой посредством API. Серверная часть включает общий функционал, в то время как клиентская часть является уникальным приложением, представляющим пользовательский интерфейс.

Реализованная система включает 3 различных модуля, каждый из которых предполагает наличие определенной роли у пользователя. Автор имеет возможность подать заявку на участие в конференции, ознакомиться с требованиями, а также отправить все необходимые материалы. Организатору доступен функционал по созданию конференции, определению формы и управлению статусом заявки. Рецензент получает отправленные организатором заявки на рассмотрение и также может принять или отклонить публикацию с возможностью оставить комментарий.

Внедрение разработанной системы сопровождения научных конференций позволит организаторам автоматизировать рутинный труд, что, несомненно, положительно скажется на качестве их работы.

Научный руководитель – канд. техн. наук А. В. Хныкин

Выявление родственных связей в базе данных репрессированных

Д. А. Шарин

Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ

Информационный ресурс «Книга памяти жертв политических репрессий» предоставляет возможность производить сбор, хранение и поиск по базе данных. Важной задачей является анализ данных в базе, в том числе определение родственных связей.

Большинство имеющихся записей о репрессированных не содержат явных данных о родственных связях, коллективом авторов был проведен анализ данных в базе и предложены признаки для определения родственной связи [1]. Анализ показал, что даже полное совпадение по какому-либо признаку не является гарантией родственной связи между людьми. Таким образом, для подтверждения родственной связи необходим вердикт эксперта. Была поставлена задача: разработать подсистему для выявления родственных связей в базе данных репрессированных.

Разработан алгоритм для определения родственной связи «родители — дети». Для реализации данного алгоритма был выбран паттерн проектирования «Декоратор». Разработанная подсистема позволяет формировать выборки возможных родственных связей, состав выборки зависит от параметров, заданных экспертов в специальной форме. Состав выборки отображается в виде списка возможных родственников для человека [2].

1. Хабитуев Б. В. и др. Статистический анализ данных по репрессированным гражданам Республики Бурятия // Вестн. Бурят. гос. ун-та. Гуманитарные исследования внутренней Азии. 2020. Вып. 2. С. 37–42.

2. Шарин Д. А. Поиск родственных связей в базе данных репрессированных // Сб. науч. тр. XII Всерос. науч.-техн. конф. аспирантов, студентов и молодых ученых. Информатика и вычислительная техника (ИВТ-2020). Ульяновск, 2020. С. 301–306.

Научный руководитель — Б. В. Хабитуев

**Разработка подхода к генерации мультиагентной системы
для поддержки процессов реализации транспортных мегапроектов**

Ш. Н. Юлдашов

Новосибирский государственный университет

Мультиагентные системы — это довольно удобный инструментарий моделирования процессов принятия решений развития транспортной инфраструктуры. Судьба этих проектов нередко зависит не столько от их содержания, сколько от компетенций, поведения агентов самого разного типа (отдельных стран, их объединений, региональных, лоббистов, подрядчиков, магистралей-конкурентов и т. д.) и доступных им механизмов влияния.

Разработка мультиагентной системы — сложный и длительный процесс, отдельные стадии которого можно облегчить за счет генерации некоторых блоков. Анализ опыта разработки мультиагентных систем (ИЭОПП СО РАН) позволил нам выделить следующие этапы: 1) формирование перечня альтернативных вариантов трансконтинентальной магистрали (ТКМ); 2) анализ особенностей и содержания процедур принятия решений на этапах подготовки; 3) формализация основных компонентов окружающей среды, имеющих критическое влияние на процесс принятия решения; 4) выявление и оценка влияния акторов, участвующих в реализации проекта, обеспечении поддержки проекта на разных уровнях (информационной, законодательной и др.); 5) формирование агрегированных типов агентов на базе вышеприведенного анализа (спецификация их свойств и атрибутов, составление логических схем взаимодействия агентов друг с другом и др.); 6) формирование и формализация правил поведения агентов, их атрибутов при взаимодействии как с окружающей средой, так и друг с другом и т. д.; 7) формирование инфопотоков для проведения отладочных процедур.

Наиболее продуктивным для упрощения и ускорения процесса разработки и отладки мультиагентной системы представляется этап генерации прототипов агентов стран-участниц ТКМ. Подход основывается на использовании шаблона, который в дальнейшем дорабатывается в ручном режиме.

Научный руководитель — канд. экон. наук Т. Н. Есикова

Исследование механизмов конфигурирования для процесс-ориентированных языков

О. Д. Янголь

Институт автоматики и электрометрии СО РАН, Новосибирск
Новосибирский государственный университет

Промышленная автоматизация является востребованной в современном мире. Из-за специфики задачи промышленной автоматизации языки программирования общего назначения не подходят для ее решения. Один из специализированных методов — процесс-ориентированный подход, в рамках которого разработано несколько языков. Другой подход, доминирующий в настоящее время на рынке, — так называемые языки для программируемых логических контроллеров, описанные в стандарте IEC 61131-3, фиксирующие решения 1950-х, 1960-х гг. по программированию систем управления.

При создании управляющих программ один из вопросов — обеспечение межплатформенной переносимости создаваемых программ. В стандарте IEC 61131-3 для решения этого вопроса используется механизм конфигурирования через файл конфигурации. Для ПО-языков в настоящее время такие механизмы отсутствуют. При этом в силу специфики ПО-подхода прямое копирование IEC 61131-3 решения невозможно.

Целью работы является изучение возможностей конфигурирования ПО-языков. В ходе работы был проведен анализ специфики задачи, стандарта IEC 61131-3, ПО-подхода и сформирован список требований к механизму конфигурации языка роST, который является ПО-расширением одного из языков стандарта — Structured Text. Затем были описаны предлагаемые расширения синтаксиса и семантики языка роST. В дальнейшем предлагается реализация и практическая апробация разработанного механизма на тестовой задаче.

Разработанный механизм конфигурирования снижает трудоемкость портирования созданных роST-программ на различных типах ПЛК и может быть использован для других ПО-языков.

Научный руководитель — д-р техн. наук В. Е. Зюбин

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ, ИНТЕРНЕТ ВЕЩЕЙ

УДК 004

Применение каналов коммуникации в аппаратно-программном комплексе

Е. С. Бощенко, М. А. Пикуль, Д. Е. Павликов, А. В. Тищенко
Томский университет систем управления и радиоэлектроники

Управление какими-либо техническими устройствами обеспечивается различными каналами коммуникации. Коммуникационный канал — это реальная или воображаемая линия связи, по которой сообщения движутся от коммуниканта к реципиенту [1]. Существует проводной и беспроводной каналы связи. В обоих видах сигнал передается через колебания, но в случае с проводным способом передачи данных радиус действия ограничивается физическим объектом, к примеру проводом, также возможны поломки на линии. В случае с беспроводным каналом связи ограничения возможны лишь со стороны собственной силы и зависят от технических характеристик модулей управления.

Для увеличения надежности коммуникационных каналов предлагается использовать беспроводные каналы связи. Это уменьшит риски коммуникации при поломке на физической линии и денежные расходы на приобретение дополнительных физических каналов для расширения коммуникационной связи.

Разработана система удаленного контроля линий электросети бытового помещения. Данная система реализована на основе микроконтроллера, содержащего WiFi-модуль, через который осуществляется удаленное управление системой посредством веб-сайта, разработанного специально под данную систему, электромагнитное реле, которое замыкает либо размыкает цепь электросети при подаче сигнала посредством канала связи. В случае внешних аварийных ситуаций, препятствующих управлению системой при помощи сети Интернет в бытовом помещении, система прошла модернизацию. Добавлен дополнительный модуль с возможностью отправки SMS-сообщений и обмена данными по GPRS, а также подключено резервное питание к самой системе.

1. Соколов А. В. Общая теория социальной коммуникации. СПб., 2002. 51 с.

Математическое моделирование многокоординатного задающего органа дельта-типа для управления промышленным манипулятором

А. С. Бояршинова

Южно-Уральский государственный университет, филиал в г. Миассе

В настоящее время в различных областях промышленности применяются промышленные манипуляционные роботы, используемые для выполнения широкого спектра технологических задач. В большинстве случаев роботы действуют согласно определенной программе и не требуют человеческого вмешательства, но для решения ряда задач необходимо непосредственное управление со стороны оператора, при этом не исключается влияние на него вредных факторов (климатических факторов, взрывоопасной среды, радиационного поля), что для снижения влияния данных воздействий требует построения систем дистанционного наблюдения и дистанционного управления. Разработка систем данного типа предполагает создание эффективных человеко-машинных интерфейсов по каналам управления и наблюдения. С учетом того что манипуляторы и манипуляционные системы могут иметь сложные кинематические схемы, а объект манипулирования в схвате должен иметь возможность линейного позиционирования и углового ориентирования в пространстве R^3 , эффективным является применение задающих органов (джойстиков) дельта-типа, состоящих из основания и подвижной платформы. Платформа соединена с основанием посредством рычажного механизма параллельной структуры с датчиками углового положения промежуточных тяг. На подвижной платформе закреплена управляющая ручка с тремя вращательными степенями свободы и возможностью ее охвата кистью руки оператора. В исследовании решены следующие задачи: разработка и исследование математического алгоритма функционирования и точности определения линейного положения подвижной платформы и угловой ориентации рукоятки джойстика, а также его практическая реализация на встраиваемом вычислительном устройстве с использованием следующих алгоритмических методов и методов моделирования: методов трилатерации (вариации решения задачи трех сфер), метода Монте-Карло, построения имитационной модели ошибок датчика углового положения. Решение поставленных задач позволяет при заданных кинематических параметрах джойстика обеспечить измерение линейного положения в пространстве с погрешностью не более $\pm 0,4$ мм на диапазоне ± 80 мм, измерение угловой ориентации в пространстве — с погрешностью не более $\pm 0,3^\circ$ на диапазоне ± 45 .

Научный руководитель — М. В. Носиков

Применение алгоритмов сегментации для обнаружения опухолей головного мозга

М. А. Кочанов

Новосибирский государственный университет

По данным Министерства здравоохранения РФ [1], показатель смертности от онкологических заболеваний в России в настоящее время составляет более чем 200 человек на 100 тыс. населения. При этом у 65–80 % умерших от онкологии обнаруживаются опухоли в головном мозге. Вопрос диагностики и контроля динамики нейроонкологических заболеваний у пациентов с первичными и вторичными опухолями головного мозга крайне актуален. Раннее обнаружение новообразований значительно улучшает результативность целевой терапии (оперативной, лучевой, химиотерапевтической и иммунологической).

В работе рассматривается задача семантической сегментации опухолей головного мозга на МРТ-изображениях. Из множества подходов для решения данной задачи было выбрано использование методов глубокого обучения. Основным источником для исследования являются медицинские данные пациентов, взятые из [2]. Были проведены их анализ и первичная обработка с целью создания тренировочной выборки для дальнейшего обучения алгоритмов. Проанализированы существующие алгоритмы и архитектуры глубокого обучения как для задачи классической сегментации, так и для задачи *Image-To-Image Translation* [3], решение которой в рамках данного проекта будет эквивалентно решению задачи сегментации. Результаты сегментации было решено оценивать с помощью следующих метрик: precision, recall, dice, IoU. По их показателям были выбраны наиболее эффективные алгоритмы. С целью улучшения точности сегментации на выбранных алгоритмах были проведены эксперименты, в результате которых найдены оптимальные гиперпараметры и функции ошибок.

1. Министерство здравоохранения Российской Федерации. URL: <http://minzdrav.gov.ru>

2. Brain Tumor Segmentation Challenge 2020. URL: <http://braintumorsegmentation.org>

3. Image-to-Image Translation with Conditional Adversarial Nets. URL: <https://phillipi.github.io/pix2pix/>

Научный руководитель — Б. Н. Тучинов

Применение сигналов транспондера для реализации механизма интеллектуального управления БПЛА

А. А. Лапицкий

Вятский государственный университет, Киров

Создание дистанционно управляемых и беспилотных летательных аппаратов (БПЛА) различных типов и назначения является крайне важным направлением развития современной авиации. С ростом численности подобных летательных аппаратов все более актуальным становится вопрос о безопасности их использования с точки зрения столкновения с такими же БПЛА. С учетом того что скорость некоторых моделей может достигать 200 км/ч и более, также встает вопрос и о быстродействии систем управления данными летательными аппаратами. Кроме того, необходимо учитывать ограниченный объем аккумуляторных батарей.

В работе рассмотрен механизм использования для взаимодействия группы БПЛА системы радиосвязи, подобно той, что применяется в транспондерах самолетов гражданской авиации. Предлагается в рамках блока управления БПЛА реализовать модуль, позволяющий при обнаружении опасности столкновения с другими летательными аппаратами корректировать курс БПЛА. Данный модуль позволит дрону иметь «картинку» их окружения и с учетом складывающейся обстановки выполнять необходимые маневры. Главным компонентом модуля является база знаний с хранящимися в ней продукционными правилами, которые сформированы экспертами в области безопасности полетов. Таким образом, блок интеллектуального управления БПЛА будет отслеживать приближающиеся объекты и на основании набора продукционных правил выполнять корректировку курса. Для первоначальных экспериментов была реализована программная модель данного блока, формирующая процедуру взаимодействия узлов для уклонения от столкновения.

Проведенные эксперименты на специальных наборах сгенерированных сигналов, имитирующих связь с другими БПЛА, показали корректность предлагаемых алгоритмов и возможность реализации процедур уклонения от приближающегося объекта в любом из доступных направлений. Данный подход обеспечивает защиту от столкновения не только с другими летательными аппаратами, но и с окружающими предметами. Реализация интеллектуального блока управления возможна на базе любой современной ПЛИС. Это позволит не только повысить скорость реагирования БПЛА, но и значительно снизить потребляемую мощность.

Разработка средств анализа и обработки данных мониторинга технического состояния зданий и сооружений

А. А. Ластович

Новосибирский государственный технический университет

Мониторинг состояния конструкций (SHM) является одним из важных приложений интернета вещей. Исторически системы SHM разрабатывались с использованием проводных сенсорных сетей. Однако низкие затраты на установку и обслуживание беспроводных сенсорных сетей (WSN) сделали их привлекательной альтернативной платформой. Датчики (сенсоры), распределенные по всему объему сооружения или расположенные в наиболее критических местах, собирают и передают данные о состоянии объекта. Эти датчики собирают информацию об окружающей среде, такую как влажность, окружающая вибрация, нагрузка, напряжение и т. д. Сенсорные узлы передают собранные данные базовой станции либо напрямую, либо путем пересылки пакетов друг друга. Агрегация и обработка данных необходимы для обнаружения и локализации структурных повреждений, могут выполняться на разных уровнях (например, непосредственно на периферийных узлах и головных узлах кластеров (*gateway*) и/или на центральном сервере) в зависимости от конкретных целей приложения. При передаче данных возможно их искажение и, следовательно, неверная интерпретация. Необходимо средства, позволяющие анализировать вероятность такой ошибки.

В докладе рассмотрены вопросы разработки средств анализа и обработки данных мониторинга технического состояния зданий и сооружений. Вместо отправки большого объема собранных данных сенсор их обрабатывает, формирует и отправляет вывод о состоянии объекта. Это позволит не только снизить зависимость результатов анализа от ошибок в радиоканалах, но и значительно снизить энергопотребление системы. Для обработки данных используются предварительно полученные правила или методы машинного обучения (в данном случае удобно использовать метод *k*-Nearest Neighbors). Пересылка итогового сообщения выполняется с использованием помехоустойчивого кодирования. Для вычисления вероятности фатального искажения сообщения предложен алгоритм временной сложности $O(n)$. Проведен анализ полученных результатов.

Научный руководитель — канд. физ.-мат. наук В. В. Шахов

Система управления совместными действиями беспилотных аппаратов

Е. Е. Пришляк

Институт автоматики и электрометрии СО РАН, Новосибирск
Новосибирский государственный университет

По мере развития технологий в областях робототехники и машиностроения перед машинами ставятся все более сложные задачи. В основном они подразумевают частичное или полное исключение участия человека из процесса выполнения работы. Это требуется, например, в сложных условиях, когда человек может быть подвергнут опасности или если выполнение миссии невозможно без применения этих технологий. Изучение существующих примеров использования беспилотных аппаратов показало, что сейчас одной из основных задач является разработка алгоритмов управления совместными действиями группы разных устройств, их движения, перемещения груза или слежения за объектом.

Целью работы является создание системы управления совместными действиями беспилотных аппаратов, включающей набор алгоритмов, задающих взаимодействие между устройствами, средств моделирования и специального программного обеспечения. В качестве аппаратов рассматриваются квадрокоптер и колесный робот. В процессе выполнения задания квадрокоптер отслеживает движение колесного робота посредством обработки кадров с бортовой видеокамеры с использованием технологии компьютерного зрения на базе `opencv` и AR-маркеров на роботе. На основе полученных данных об обстановке и местоположении робота на наземном пункте управления формируются траектория движения и управляющие команды. Для исследования особенностей работы алгоритмов и отработки компонентов системы применяется численное моделирование.

В работе представлена конструкция собственного колесного робота, структура программного обеспечения, предложен алгоритм совместного действия нескольких роботов на основе принципа «лидер-ведомые». Алгоритм включает непрерывное вычисление требуемых значений линейной и угловой скорости каждым роботом с целью поддержания заданного положения относительно лидера и вычисление вектора смещения требуемого положения для обхода препятствий.

Результаты работы могут использоваться в сельском, коммунальном хозяйствах, на промышленных объектах.

Научный руководитель — канд. техн. наук А. С. Мальцев

Автоматизация станции синхротронного излучения с использованием средств TANGO и Sardana

А. Н. Скляров

Новосибирский государственный университет
Институт ядерной физики им. Г. И. Будкера, Новосибирск

Современная экспериментальная станция на источнике синхротронного излучения, по сути, представляет собой научную лабораторию, которая при необходимости может быть сконфигурирована под различные исследовательские задачи. В связи с этим для эффективной работы станции важно обеспечить быстрое подключения нового оборудования и быструю переконфигурацию.

Для решения этой задачи вполне подходят программные пакеты TANGO и Sardana, опробованные на многих источниках синхротронного излучения (например, Max IV, Elettra, Alba, ESRF). В данных пакетах устройства определены как абстрактные классы, конкретная реализация которых остается за пользователем. Такой подход позволяет свести модернизацию программного обеспечения на станции к описанию алгоритма проведения эксперимента в новой конфигурации или плагина, определяющего новое устройство в системе.

Представленная работа была направлена на реализацию данного подхода на технологической станции источника ВЭПП-4 ЦКП «СЦСТИ» ИЯФ СО РАН с описанием в системе всех необходимых устройств и демонстрацию нескольких экспериментальных режимов.

Научный руководитель — канд. физ.-мат. наук Я. В. Ракшун

Разработка средств анализа надежности и живучести беспроводных сенсорных сетей, применяемых в нефтегазовой отрасли

В. В. Хагай

Новосибирский государственный технический университет

В НГТУ проводятся исследования технологий интернета вещей для мониторинга протяженных трубопроводов. В качестве основы указанных сетей мониторинга выбраны беспроводные сенсорные сети (БСС). Большая часть обслуживающего полевого оборудования и активов нефтегазовой отрасли установлены в труднодоступных местах, используются в суровых условиях, что накладывает ограничения на использование проводных сетей. БСС имеют ряд преимуществ: отсутствие необходимости прокладки кабеля, легкость развертывания и модификации, низкие эксплуатационные расходы. Однако, для того чтобы в полной мере задействовать потенциал указанных сетей, необходимо разработать средства для анализа их надежности и живучести.

Информация, передаваемая сенсором по беспроводному каналу, может подвергаться искажениям из-за непреднамеренных помех, создаваемых внешними источниками или другими узлами сети. Также актуальна угроза умышленного подавления радиоканала в результате атаки с использованием специального генератора электромагнитных импульсов. Кроме того, на надежность системы мониторинга влияют отказы локальных стоков (контроллеров), управляющих работой кластера сенсоров и выступающих в качестве шлюза для интеграции с опорной сетью. Для преодоления сбоев могут использоваться как штатные методы обеспечения надежности сетевых протоколов, так и специальные средства диагностики и восстановления системы.

В докладе предложены методы оценки работоспособности специализированных БСС. Представлена соответствующая модель, основанная на аппарате Марковских цепей, для случая, когда известны только причины отказов и их интенсивность. Если известна топология сети, оценка отказоустойчивости вычисляется с использованием полиномов надежности графов. Для частных случаев полиномы надежности получены в аналитическом виде. В общем случае значение полинома надежности вычисляется с использованием метода Монте-Карло. Обсуждаются перспективы дальнейших исследований.

Научный руководитель — канд. физ.-мат. наук В. В. Шахов

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОБУЧЕНИЯ

УДК 378.14

Цифровая трансформация вузов: платформа ранней профориентации и дистанционного обучения

А. И. Алексеев

Северо-Восточный федеральный университет им. М. К. Аммосова, Якутск

В работе рассмотрены массовые открытые образовательные курсы по математике и информатике с профессиональной ориентацией учащихся 9–11-х классов на профессиях цифровой экономики. Актуальность такой MOOK имеет две основные причины: незнание большинством родителей, детей и педагогов новых цифровых профессий в цифровой экономике и нехватка учителей информатики в Республике Саха (Якутия).

Образовательные материалы о новых профессиях могут содержать обновляемые новости. Традиционное профориентационное тестирование может быть заменено рекомендательной системой. Такой подход позволит школьнику чувствовать вовлеченность в профессию и образование.

Образовательный контент и задачи должны балансироваться, как это реализовано на популярной платформе stepik.org. Анализ статистики неверных ответов на тесты и ухода учащегося из MOOK позволяет выявить нежелательный резкий скачок сложности в образовательном материале. Исходя из свойств конкурирующих развлекательных цифровых платформ, важнейшими характеристиками MOOK должны стать: открытость, сетевой характер, геймификация и сопричастность [1].

Общие компетенции, так называемые *soft skills*, не удается эффективно преподавать через MOOK. Тем не менее данная платформа может быть интегрирована с выездными мероприятиями СВФУ для обращения внимания учителей, школьников и их родителей к проблеме обучения общим компетенциям. Общие компетенции как умение обучаться самостоятельно, знание языков, умение работать в команде, коммуникабельность и лидерские навыки могут прививаться в школе. Важен фокус внимания, понимание актуальности формирования общих компетенций, поскольку большинство профессиональных компетенций устаревают за пятилетку.

1. Celaya I., Ramírez-Montoya M. S., Naval C., Arbués E. The educational potential of the podcast: an emerging communications medium educating outside the classroom // Proc. 7th Int. Conf. Technol. Ecosystems Enhancing Multiculturality. 2019. P. 1040–1045.

Научный руководитель — канд. физ.-мат. наук Э. И. Шамаев

**Информационное обеспечение лабораторной работы
«Измерение композиционных материалов на приборе серии P2M»**

Р. А. Василенко, О. С. Жабин
Томский государственный университет

В настоящее время композиционные материалы широко распространены во всех сферах жизнедеятельности человека. Композиты используют как в медицине, так и в военно-промышленном комплексе, так как они имеют ряд преимуществ по сравнению с традиционными материалами: легкость, надежность, высокую износостойкость и др. Для получения электромагнитных характеристик данных материалов широко применяются скалярные и векторные анализаторы цепей. Одним из приборов, используемых на радиофизическом факультете ТГУ, является скалярный анализатор цепей P2M, предназначенный для измерения модулей коэффициентов отражения, передачи, а также коэффициента стоячей волны по напряжению (КСВН) в диапазоне от 10 МГц до 40 ГГц. Данный прибор разработан НПФ «Микран» (Томск).

Для проведения измерения электрофизических параметров композиционных материалов была разработана программа на графическом языке LabVIEW. Данная программа позволяет автоматизировать сбор и обработку данных при исследовании композитов: отображать коэффициенты прохождения, отражения и поглощения; получать данные о мнимой части магнитной проницаемости в зависимости от частоты и толщины опытного образца; сохранять результаты измерений в текстовый и графический типы файлов.

Разработана лабораторная работа в системе Moodle по курсу «Технологии организации приборных интерфейсов», которая знакомит студентов с программированием измерительного прибора, процессом измерения электромагнитных характеристик композиционных материалов.

Данная лабораторная работа заключается в создании виртуального прибора на основе предложенной программы. В эту программу студент должен внести изменения для выполнения конкретного измерения. После написания программы студент должен провести тестирование программы и измерение для заданного набора образцов.

Данная работа также является примером использования LabVIEW для автоматизации эксперимента.

Научный руководитель — доц. А. А. Жуков

Разработка системы онлайн-тестирования средствами языка программирования PHP

С. В. Волченко

Вторая Новосибирская гимназия

На сегодняшний день в мире разработано множество систем, предназначенных для проведения онлайн-тестирования. В условиях дистанционного обучения создание подобного рода систем стало актуальным направлением.

Целью данной исследовательской работы является разработка высокопроизводительной системы тестирования с поддержкой устаревших устройств и сохранением всех функций, имеющихся в современных системах тестирования.

Для достижения поставленной цели были решены следующие задачи:

- выполнен анализ существующих систем онлайн-тестирования и технологий разработки подобного рода систем;
- разработан «личный кабинет учителя», модуль вывода теста и обработки результатов;
- реализована защита от спама, кражи и повреждения данных;
- выполнено тестирование системы на устройствах с различными характеристиками.

В ходе анализа существующих систем онлайн-тестирования было выявлено, что ни одна современная система не обеспечивает полного функционала без JavaScript, а некоторые системы вообще не работают без данной технологии.

В качестве тестовой версии были разработаны:

- серверная часть системы тестирования средствами PHP с применением `php-xml`, `php-gd`, `php-mbstring`, `php-curl`;
- клиентская часть, разработанная на основе языка гипертекстовой разметки HTML, без применения языка программирования JavaScript.

На данный момент пользователю доступно создание, редактирование, прохождение тестов, просмотр и скачивание результатов. Предусмотрена защита от отправки вредоносного программного обеспечения на сервер. Система тестирования на текущий момент размещена по адресу: <https://online-test.qwerty281projects.tk/>.

В рамках работы была выполнена разработка системы онлайн-тестирования и успешно апробирована учителями и учениками Второй Новосибирской гимназии.

Научный руководитель — А. А. Шарапов

Программный комплекс для проведение оптических экспериментов в образовательных целях

Н. Ю. Долгов

Институт вычислительной математики
и математической геофизики СО РАН, Новосибирск
Новосибирский государственный университет

Целью работы является создание программного средства, позволяющего самостоятельно программировать формирование виртуальных 3D-сцен и проведение оптических экспериментов, например, в образовательных целях. В качестве основы для разработки взята работа [1], в которой приводится необходимый математический аппарат для обработки взаимодействия лучей линейно поляризованного света (ЛПС) с оптически изотропными прозрачными объектами и требования к программной реализации. В данной работе мы рассмотрели конкретное решение данной задачи.

Разработан программный комплекс, позволяющий рассчитывать по падающему на линзу лучу ЛПС выходные лучи ЛПС. На основании этого комплекса пользователи смогут создавать различные образовательные программы.

Основные объекты виртуальной сцены — сферические линзы. Это обусловлено конструкцией большинства микроскопов. Используемые линзы построены (например, при помощи склеивания) на основе простых сферических линз (двояковыпуклых, плоско-выпуклых, вогнуто-выпуклых, двояковогнутых, плоско-вогнутых и выпукло-вогнутых).

3D-модель сферической линзы строится с использованием теоретико-множественных операций пересечения и объединения над набором примитивов: полупространство, внутренность или внешность сферы, конуса, цилиндра. Программный интерфейс включает набор конструкторов линз, который переводит описание линзы во внутреннее представление.

В докладе приводятся принципы работы библиотеки: модель линейно-поляризованного луча света, геометрическая модель 3D-сцены, состоящей из объектов (линз), и алгоритмы расчета взаимодействия луча с объектами, а также пример сцены по демонстрации колец Ньютона.

1. Debelov V. A., Kushner K. G., Vasilyeva L. F. Lens for a Computer Model of a Polarizing Microscope // *Mathematica Montisnigri*. 2018. Vol. 41. P. 151–165.

Научный руководитель — д-р техн. наук В. А. Дебелов

Создание электронного самоучителя по сборке системного блока

В. К. Иванов
МАОУ Гимназия № 33, Улан-Удэ

Компьютеры и информационные технологии широко используются во многих сферах деятельности, в связи с этим важным вопросом является поднятие общей компьютерной грамотности. По результатам исследования ученых из Института статистических исследований и экономики знаний НИУ ВШЭ только у 12 % россиян компьютерная грамотность находится выше базового уровня [1].

Чаще всего мы имеем дело с персональными компьютерами (ПК). Понимание того, как работает и из чего состоит ПК, крайне важно для любого современного пользователя. Одной из популярных тем сейчас является самостоятельная сборка ПК.

Собрать персональный компьютер — сложная задача. Не разбирающийся в компьютерах человек может допустить ряд серьезных ошибок, которые могут повлиять как на работоспособность компьютера, так и на бюджет сборки.

В русскоязычном сегменте сети Интернет существует огромное количество ресурсов, посвященных сборке, эксплуатации и обслуживанию компьютеров. Недостатком этих ресурсов является то, что информация на них зачастую не структурирована и сложна для понимания неопытного пользователя. Таким образом, задача создания курса для начинающих пользователей, посвященного устройству ПК, его эксплуатации и сборке, является актуальной.

В данной работе проведен анализ материалов, разработана структура курса. Курс включает теоретический материал с фото- и видеонаполнением и контрольно-измерительные материалы (в виде различных тестов). Планируется размещение курса в сети Интернет на платформе Moodle.

1. Уровень владения цифровыми навыками в России и странах ЕС [Электронный ресурс] // Нац. исслед. ун-т «Высшая школа экономики». URL: <https://issek.hse.ru/news/377859466.html> (дата обращения: 25.02.2021).

Научный руководитель — Б. В. Хабитуев

Выбор инструментов для разработки обучающего симулятора производства

И. О. Кислицын

Новосибирский государственный технический университет

В НГТУ разрабатывается система для имитационного моделирования производственных процессов, призванная упростить прототипирование различных конфигураций производства за счет интуитивного интерфейса и высокой производительности.

При создании подобных крупных программных систем большое значение уделяется выбору инструментов. Ошибка при данном выборе может привести к труднорешаемым архитектурным проблемам, а также к низкой производительности программы.

К разрабатываемому симулятору предъявляются следующие нефункциональные требования:

- **быстродействие:** имитационное моделирование требует больших объемов вычислений, должна быть возможность выполнять их за разумное время;
- **кроссплатформенность:** так как система будет использоваться в обучении студентов, должна быть возможность запуска программы на широком спектре операционных систем.

Чтобы удовлетворить обоим требованиям, для разработки была выбрана архитектура «клиент — сервер». Она предоставляет компромиссное решение, позволяя перенести ресурсоемкую часть программы на сервер. Для обеспечения кроссплатформенности было решено реализовать программу в виде веб-приложения. Это значит, что для языка программирования должен быть доступен веб-фреймворк.

Используемый язык программирования также должен иметь механизмы для обеспечения быстродействия. Для разработки системы имитационного моделирования среди прочего были рассмотрены следующие языки и веб-фреймворки:

- Typescript + Express,
- Elixir + Phoenix,
- Python + Django.

В ходе сравнения инструментов по различным критериям, таким как возможность распараллеливания, производительность при выполнении одинаковых задач и совместимость основных принципов дизайна данных инструментов с поставленными задачами, было принято решение разрабатывать симулятор на языке Elixir с веб-фреймворком Phoenix.

Научный руководитель — канд. техн. наук А. В. Фаддеенков

Информационная система для обучения программированию

А. Н. Турусов

Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ

Информационные технологии являются неотъемлемой частью жизни любого человека. В связи с этим возникает потребность в специалистах, которые будут заниматься созданием и обслуживанием программных комплексов и информационных систем. Таким образом, освоение базовых навыков программирования становится обязательным требованием для многих специальностей [1, с. 373].

Обучение программированию происходит в процессе решения задач с помощью базовых понятий: математических операций, управляющих конструкций и структур данных. Одной из особенностей обучения программированию является возможность применения автоматических проверяющих систем. На данный момент в русскоязычном сегменте сети Интернет существует несколько ресурсов, реализующих возможности проверяющих систем: acmp.ru, informatics.msk.ru, acm.timus.ru. Однако обучающиеся могут столкнуться с проблемой нехватки задач начального уровня сложности и отсутствия обучающих материалов. Недостатком существующих систем является то, что в них зачастую затруднительно проводить контроль обучающихся, а также добавлять собственные материалы. В связи с этим была поставлена задача: разработать платформу для обучения программированию с использованием проверяющей системы [ejudge](http://ejudge.ru) [2].

На данный момент был проведен анализ предметной области, разработана ER-model и структура базы данных. Прототип информационной системы находится на этапе разработки.

1. Универсальные компетентности и новая грамотность: от лозунгов к реальности / ред. М. С. Добрякова, Д. А. Фрумин; при участии К. А. Баранникова, Н. Зиила, Дж. Мосс, И. М. Реморенко, Я. Хаутамяки; Нац. исслед. ун-т «Высшая школа экономики». М.: Изд. дом Высшей школы экономики, 2020. 472 с.

2. [Ejudge contest management system](http://ejudge.ru) [Электронный ресурс]. URL: <https://ejudge.ru> (дата обращения: 18.02.2021).

Научные руководители — Д. Ф. Дерюгин, Б. В. Хабитуев

КОМПЬЮТЕРНАЯ ГРАФИКА, МУЛЬТИМЕДИАТЕХНОЛОГИИ И СИСТЕМЫ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

УДК 004.023

Библиотека чтения/записи виртуализированных битовых потоков данных

В. В. Агулов

Новосибирский государственный университет
Института автоматики и электрометрии СО РАН, Новосибирск

Работа посвящена решению задачи частичной замены контента в потоке данных, например при вставке местных передач в цифровом ТВ. Современные потоки данных используют большое число протоколов передачи данных, вложенных один в другой наподобие матрешки. Каждый протокол использует собственные заголовки, разбивающие «полезные данные» на отдельные фрагменты, поэтому для парсинга вложенного потока данных необходимо все «полезные данные» текущего протокола скопировать в отдельный буфер. В современном цифровом ТВ используется пятикратная вложенность (TS — PLP — MPTS — ES — Video), что подразумевает четырехкратное копирование данных. Очевидно, что для большинства задач это очень ресурсоемкое решение.

Целью данной работы является написание библиотеки на C++, объекты которой имитировали бы работу обыкновенного указателя, но в то же время позволяли бы организовать доступ к произвольным фрагментам памяти так, будто она является последовательной памятью. У предлагаемого класса «виртуального указателя» определены привычные методы перемещения указателя, чтения и записи стандартных типов данных, включая чтение/запись нескольких бит. Это позволяет легко переписать код для парсинга любого протокола через работу с «виртуальным указателем». В работе рассматриваются преимущества и недостатки использования виртуальных указателей, приводится сравнение производительности для характерных задач.

При написании данной библиотеки использовался язык C++, стандарт std C++ 17, компилятор MSVC, система сборки стандартная из Visual Studio 2017.

Научный руководитель — доц. И. Г. Таранцев

Роль метрики в обработке изображений

Э. Н. Дайырбаева

Институт информационных и вычислительных технологий, Алматы

Объективная метрика качества изображения является важной частью систем обработки изображений. Одним из значительных и важных применений объективных метрик оценки качества изображений является оценка эффективности алгоритмов и систем обработки изображений. Несмотря на большое количество публикаций по данной тематике, задача оценки качества восстановленных изображений рассматривается лишь в немногих. Но в последнее время задача восстановления утраченных областей изображения получила значительное внимание.

Возможны два подхода к оценке качества изображений: количественная оценка с помощью использования математических методов и субъективная оценка на основе экспертных оценок. Под метрикой (в обработке изображений) понимается функция, определяющая расстояние от оцениваемого изображения до эталонного или идеального изображения в пространстве изображений.

В современной литературе имеется множество описаний объективных метрик, которые можно разделить на три класса: эталонные, неэталонные, псевдоэталонные.

Для проведения сравнений были выбраны следующие ортогональные преобразования: Адамара, Хаара, наклонное, DCT. Для сравнения воздействия шума на финальный результат в работе были выбраны такие показатели, как средний модуль отклонения, среднеквадратичный модуль отклонения, максимум модуля отклонения и отношение сигнала к шуму PSNR. Примененные для сравнения программные эксперименты проведены в приложении Matlab. Функциональные возможности данного приложения позволили произвести ортогональные преобразования с цифровым изображением, а также вычислить желаемые метрики.

Данное исследование финансируется Комитетом науки Министерства образования и науки Республики Казахстан (грант № AP08857179).

Научный руководитель — PhD, доц. А. С. Еримбетова

Компьютерное моделирование опыта по получению интерференционных полос равной толщины

Н. Ю. Долгов

Институт вычислительной математики
и математической геофизики СО РАН, Новосибирск
Новосибирский государственный университет

Явлению интерференции в компьютерной графике посвящено не так много работ. Это связано с тем, что расчет изображений с учетом интерференции света требует больших вычислительных затрат, при этом получаемый результат зачастую отличается незначительными деталями от изображений без учета интерференции. Известны физические опыты, например по получению полос равной толщины, колец Ньютона и т. п., которые невозможно рассматривать без учета интерференции. Они приводятся во многих учебниках [1], пособиях по лабораторным работам [2] и др.

Для реальной сцены, состоящей из источника параллельного монохроматического света, падающего на клин из прозрачного оптически изотропного материала, линзы с апертурой и экрана, разработана соответствующая компьютерная модель в виде 3D-сцены. В докладе рассматриваются особенности реализации на всех этапах *прямого физически корректного моделирования* на основе трассировки линейно-поляризованных лучей.

Были рассчитаны изображения на основе: 1) приближенной формулы из [1, 2], 2) трассировки лучей от источника до наблюдателя (экрана). Установлено, что упрощения, сделанные в [1, 2] при выводе формул, приводят к сдвигу изображения полос.

Судя по доступной литературе, впервые выполнено физически корректное компьютерное моделирование, демонстрирующее явление интерференции для клина в виде полос равной толщины.

1. Ландсберг Г. С. Оптика: учеб. пособие для вузов. 6-е изд. М.: Физматлит, 2003. С. 110–115.

2. Определение длины волны света и оценка параметров когерентности излучения при помощи колец Ньютона: метод. указания к лаб. работе № 26 / сост. Т. И. Папушина, А. В. Михельсон. Екатеринбург: УрФУ, 2015. 20 с.

Научный руководитель — д-р техн. наук В. А. Дебелов

Разработка учебного симулятора вождения автомобиля с использованием технологии виртуальной реальности

Я. П. Лукьянов, М. А. Бекиев

Высший колледж информатики при НГУ, Новосибирск

Виртуальная реальность (англ. *virtual reality*, VR, искусственная реальность) — созданный техническими средствами мир, передаваемый человеку через его ощущения: зрение, слух, осязание и др. Одно из перспективных направлений VR является моделирование реальных явлений, где эксплуатация устройств и механизмов связана с повышенным риском либо большими затратами. Виртуальная реальность позволяет произвести наработку навыков действий в условиях чрезвычайных и иных непредвиденных ситуаций, подготовить высококвалифицированных специалистов в выбранной области.

В рамках данной работы была исследована возможность управления автотранспортом в очках виртуальной реальности с использованием шлема HTC Vive Pro. На первом этапе была проработана концепция работы учебного симулятора вождения автомобиля в виртуальной реальности, сформулированы основные аспекты приложения, проработан игровой процесс, графика, дизайн, интерфейс и взаимодействие с пользователем. На втором этапе были изучены функции Unity, синтаксис языка C# и основы 3D-моделирования в программе Blender. В заключительном этапе была написана и отлажена программа по взаимодействию пользователя с виртуальным миром. Разработка может быть использована как для реального обучения управлению транспортным средством, так и в развлекательных целях.

Данная работа предназначена для получения практических навыков управления вождением автомобиля и сдачи правил дорожного движения. С помощью симулятора вождения можно понять основные принципы управления автомобилем, подготовиться к теоретическому экзамену в ГИБДД. Приложение имеет прикладной характер и предназначено для самостоятельной подготовки начинающих водителей с использованием технологии виртуальной реальности.

Научный руководитель — канд. физ.-мат. наук И. А. Козулин

Система для визуализации трехмерных сферического и эллиптического пространств

Д. И. Мигранов

Новосибирский государственный университет
Институт автоматики и электрометрии СО РАН, Новосибирск

Неевклидовы пространства и геометрии находят применение в различных областях знаний, включая математику и физику. Возможности компьютерной графики позволяют не только представить свойства неевклидовых пространств в наглядной форме, но и предоставить пользователю возможность взаимодействовать с ними.

Один из методов визуализации таких пространств основывается на их рассмотрении с точки зрения групп пространственных преобразований, например их групп движений. С помощью элементов этих групп, которые могут быть представлены в виде матриц, возможно задавать положения объектов в пространстве. Далее (также посредством применения матрицы) необходимо осуществить проецирование на экран.

Примером неевклидова пространства является трехмерное сферическое пространство, в качестве модели которого может рассматриваться гиперсфера. Группа его движений — $O(4)$, группа ортогональных матриц 4×4 , которые могут использоваться для позиционирования объектов. Еще одно неевклидово пространство — эллиптическое, получаемое из сферического отождествлением антиподальных точек гиперсферы. Для позиционирования объектов в нем могут использоваться те же самые матрицы.

Целью данной работы является создание системы для визуализации трехмерных сферического и эллиптического пространств с видом от первого лица. Для упрощения ориентирования в неевклидовых пространствах был реализован эффект дымки, упрощающий восприятие наблюдателем расстояния до объекта. Также был предложен и реализован интерфейс управления перемещением наблюдателя в таких пространствах с видом от первого лица.

При написании системы использовался язык программирования C++ и графический API Direct3D 11, что дало возможность использовать возможности графических ускорителей для визуализации таких пространств. Возможным является текстурирование объектов сцены, а также задание их поведения в зависимости от времени или пользовательского ввода.

Научный руководитель — д-р техн. наук, доц. В. Е. Зюбин

Конструктор интерактивных художественных игр на основе веб-технологий

А. М. Сенотрусова, М. А. Нейкина, А. Е. Будаев
Иркутский государственный университет

В последние несколько лет компьютерное сообщество вновь обратило свое внимание на интерактивные художественные игры (визуальные новеллы, текстовые квесты, *Interactive fiction* или IF). Этому есть несколько причин, среди которых наиболее важной является возвращение интереса к чтению через создание интерактивной литературы со свободным сюжетом, возможностью выбора читателем линии главного героя и непредсказуемым окончанием повествования. Кроме того, современные технологии позволяют создавать такого рода продукты уже на совсем другом уровне, нежели в 80-х гг. XX в. Тут и красочная графика с элементами анимации, и эргономичный интерфейс с возможностью выбора сюжетной линии одним кликом или вовсе с использованием голосового управления, десктопные и мобильные реализации.

Нашей задачей было создание конструктора для разработки интерактивных художественных игр, освоение которого было бы вполне доступно обычным пользователям компьютера, не владеющим навыками программирования, зато умеющим создавать иллюстративный материал и «лихо» закрученные сюжеты. В тоже время функциональные возможности такого редактора должны позволять внедрять в контент повествования максимум мультимедийных компонентов, гибко подстраиваться под сценарный граф автора и обладать нетривиальными решениями визуализации сюжетных сцен.

Итогом работы стал вариант простого конструктора для редактирования и публикации визуальной новеллы. В основу технологического решения легли веб-технологии: язык разметки HTML 5, CSS, JS. Функциональные возможности первой версии конструктора охватывают большинство основных потребностей для редактирования визуальной новеллы приключенческого стиля, дополнительные функции могут быть добавлены с помощью ряда пользовательских тегов. Сейчас осуществляется попытка внедрения возможностей включения в интерфейс голосового управления и получения версии интерактивной художественной игры для мобильного устройства.

Научный руководитель — канд. физ.-мат. наук А. Г. Балахчи

НАУКОЕМКОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

УДК 004.048

Application of neural network methods for automatic taxonomy enrichment for the Russian language

D. A. Pirozhkova
Novosibirsk State University

The solutions of many NLP tasks are based on the use of various knowledge bases, such as Wikidata or the well-known thesaurus WordNet. As a result of the above, the taxonomy enrichment of the knowledge base is an important and relevant issue. In general, taxonomy is a science of classification of things or concepts, including the principles underlying such classification. In particular, the definition for the language is as follows: taxonomy is a hierarchy whose tree nodes represent named entities related to other tree nodes with “is-a-type-of” relations. An “Is-A” relation is a domain independent semantic relation that is a strict partial order relation (antisymmetric, irreflexive, transitive) between a subclass concept and a superclass concept. In other words, hypernym relation (“Is-A”) is a semantic relation between two words, where the hypernym is a word with a broader meaning that expresses a general concept.

The main idea is to create methods that automatically enrich any knowledge base with new terms, and at the same time connect them with existing words using “Is-A” relations. The main task is to develop a method for prediction the hypernym relations between terms in texts of the Computer Science category. An analysis of the fullness of terms and the relations between them showed that Russian knowledge bases do not contain enough information about terms in this category.

The RBERT model is used to solve the binary classification problem because it shows good results in solving the multiclass problem in comparison with other methods. Training data is a manually labeled text corpus containing terms from the information technology area. We predict the type of relations between entities and get some local knowledge base. It can be linked to any knowledge base in order to determine its content with the “Is-A” relation and replenish the missing hypernym relations between entities.

Scientific advisor — Cand. Phys.-Math. Sci. T. V. Batura

Сравнение языковых моделей в задаче классификации мнений

С. А. Березин, Я. Н. Антонова, Д. А. Пирожкова, М. Г. Молчанова
Новосибирский государственный университет

Интернет-площадки все чаще становятся платформой для обсуждений различных явлений и событий. Одновременно с этим построение моделей динамики мнений является актуальной и нетривиальной задачей на стыке математики, компьютерных наук и социологии. Однако математические модели работают с числами, а мнения обычно выражаются словесно.

Традиционно ученые решали эту проблему, используя различные формальные оценки, например социологические опросы. В данной работе предлагается новый подход к проблеме числового представления мнений, основанный на нейросетевых языковых моделях.

Для апробации предложенного подхода были выбраны сообщества сторонников основных кандидатов президентской гонки в США в 2020 г. Это политическое противостояние отражалось в ощутимой поляризации общества, и, следовательно, выражаемые мнения должны являться хорошо отличимыми друг от друга [1]. Поэтому в качестве данных были взяты посты и комментарии с площадки Reddit за начало 2020 г. из двух сообществ: /r/The_Donald (сторонники Дональда Трампа) и /r/JoeBiden (сторонники Джо Байдена).

Текстовые комментарии были переведены в многомерные числовые вектора с помощью трех NLP-алгоритмов. В исследовании использовались и сравнивались алгоритмы векторизации текстов: Doc2Vec, Sentence-BERT и *Universal Sentence Encoder* (USE). После этого результаты векторизации каждым алгоритмом оценивались с помощью классификатора (логистическая регрессия).

Для интерпретации и визуализации результатов размерность полученных векторов снижалась до двух и трех измерений с помощью алгоритмов t-SNE и UMAP. В полученных визуализациях отчетливо наблюдались кластеры схожих мнений и разделимость подмножеств сторонников двух политиков.

В результате работы был получен инструмент, позволяющий определять полярность новых высказываний, изучать дисперсию мнений в сообществах и проводить кластерный анализ текстов.

1. Böttcher L., Gersbach H. The great divide: drivers of polarization in the US public // EPJ Data Sci. 2020. Vol. 9. No. 1. P. 1–13.

Реконструкция геометрии аневризмы брюшной аорты

К. А. Борисова

Новосибирский государственный университет

Аневризма брюшной аорты (ААА) — это аномальное локализованное расширение стенки аорты в нижнем отделе сосуда. Ее разрыв является наиболее частой причиной летального исхода среди людей, страдающих заболеваниями сердечно-сосудистой системы. Сейчас аневризма диагностируется с помощью визуального осмотра снимков КТ, но для более точной оценки риска разрыва аорты необходимо использовать количественный подход, основанный на оценках геометрических характеристик.

Целью данной работы является разработка и реализация алгоритма автоматической 3D-реконструкции геометрии аневризмы брюшной аорты по изображениям компьютерной томографии. В качестве метода интерполяции дискретно заданной поверхности брюшной аорты используется метод BQFE (*biquintic finite element*) [1]. Получаемая поверхность принадлежит классу C^2 . Это позволяет оценить локальные тензоры кривизны [2], которые используются при нахождении механических напряжений в ААА. Высокая точность 3D-реконструкции брюшной аорты необходима для успешного эндопротезирования брюшного отдела аорты, 3D-печати стент-графта и в перспективе построения трехмерной расчетной сетки для моделирования гемодинамики брюшной аорты. Также результаты работы могут быть напрямую использованы при прогнозировании риска разрыва ААА.

1. David B. S. Surface Geometric Analysis of Anatomic Structures Using Biquintic Finite Element Interpolation // Ann. Biomed. Eng. 2000. No. 28. P. 598–611.

2. Bradley C. P. Geometric Modeling of the Human Torso Using Cubic Hermite Elements // Ann. Biomed. Eng. 1997. No. 25. P. 96–111.

Научный руководитель — Я. В. Федотова

Ускорение численного моделирования волновых форм первых вступлений в сейсмических данных

М. В. Галушкин

Новосибирский государственный университет

Для решения прямой задачи сейсмологии в неоднородных моделях применяется численное моделирование сейсмического волнового поля. Наиболее широко используются конечно-разностные схемы решения волнового уравнения или системы уравнений упругости. Эти методы являются вычислительно трудоемкими, особенно для трехмерных моделей. При решении некоторых сейсмических задач требуется моделировать не все волновое поле, а только волновые формы первых вступлений. Это потенциально позволяет уменьшить количество требуемых расчетов, так как для каждого шага по времени требуется проводить расчеты не во всей области, а только в полосе за фронтом первых вступлений. Целью данной работы является разработка алгоритма волнового моделирования сейсмических волн в полосе первых вступлений, а также его оптимизация с учетом особенностей современных центральных процессоров.

На начальном этапе алгоритма выполняется расчет времен первых вступлений с помощью численного решения уравнения эйконала. При проведении расчетов только в полосе за фронтом первых вступлений возникает сложность перебора узлов: порядок обхода узлов становится неупорядоченным, тогда как при расчетах во всей области обход узлов идет последовательно. Неупорядоченный обход по узлам сетки приводит к потере производительности, так как происходит большое количество промахов по кэшу. В рамках работы были предложены и протестированы три подхода к упорядочиванию обхода по узлам сетки при расчете в полосе: прямой обход с проверкой в необходимости вычислений в узле; упорядоченный обход по индексу модели; упорядоченный обход по индексу кривой Гильберта для данной модели.

Численные результаты показали, что упорядоченный обход по индексу модели является наиболее эффективным подходом. Тестирование на системе с общей памятью на 6 ядрах для сетки размером $512 \times 512 \times 512$ (3500 шагов по времени) показало время выполнения моделирования 21 291 с для всей области и 1420 с для моделирования в полосе, что соответствует ускорению в 15 раз. Была создана вычислительная библиотека, которая рассчитывает и формирует данные для дальнейшего использования в сейсмологии.

Научный руководитель — канд. физ.-мат. наук, доц. А. А. Дучков

**Оптимизация численных моделей отдельных процессов
для моделирования разрушения тепловыделяющих элементов**

А. Р. Гафиятуллин

Новосибирский государственный университет

Институт проблем безопасного развития атомной энергетики РАН, Москва

При разработке новых типов реакторных установок особое внимание уделяется безопасности их эксплуатации. Один из подходов изучения аварийных ситуаций предполагает создание экспериментальных реакторных установок. Однако такой способ является дорогим и плохо масштабируемым, вследствие чего все большую популярность приобретает моделирование процессов, происходящих в реакторе, с помощью высокопроизводительных вычислительных систем и методов математического моделирования.

При таком подходе особое внимание должно уделяться рациональному использованию возможностей компьютера, учету особенностей его архитектуры. Это позволяет ускорить расчеты задач, которые необходимы для расчетного моделирования процессов в реакторных установках. В ИБРАЭ РАН для анализа процессов с разрушением элементов активной зоны перспективных реакторов разрабатываются расчетные коды, учитывающие эти особенности вычислительных систем.

Представленная работа посвящена оптимизации модуля, предназначенного для моделирования физических процессов, которые сопровождают разрушение твэлов перспективных реакторных установок. Приведено описание методов оптимизации, обоснование их применения с учетом особенностей архитектуры современных компьютеров, а также результаты расчетов тестовых задач с учетом и без оптимизаций.

Исследование выполнено за счет гранта РФФИ № 18-79-10013.

Научный руководитель — канд. техн. наук Э. В. Усов

О перспективах исследования структуры многомерного пространства, порождаемого множеством двумерных изображений

А. П. Девятериков

Новосибирский государственный университет

Считается, что посредством зрения человек воспринимает около 85 % информации о внешнем мире: фото- и видеозаписи (как документальные, так и художественные), трехмерную компьютерную графику и анимацию, микроскопические и астрономические объекты. Изображения трехмерных объектов передаются в мозг посредством 2D-проекции на сетчатку глаза. Сетчатка содержит примерно 130 млн клеток-палочек, обеспечивающих зрение при слабом освещении, и 7 млн колбочек (трех типов, каждый способен различать до 100 оттенков), ответственных за цветное зрение (с возможностью различать до миллиона цветов). Исключительный интерес представляет задача о том, как в мозге реализованы механизмы, способные столь быстро обрабатывать зрительную информацию и на основе 2D-стереоизображений воссоздавать исходные трехмерные объекты.

Рассмотрим значительно более простую систему — 2D-изображение в градациях серого цвета, в котором яркость пикселей задается числами 0 до 15. Используя цвет каждого пикселя в качестве координаты, для изображения 2×2 получим 4-мерное пространство (4^{16} различных вариантов таких изображений), а для 5×5 — уже $25^{16} (\approx 2,3 \cdot 10^{22})$ вариантов, что сравнимо с объемом всей информации, хранящейся в интернете. Таким образом, для изображения 50×50 пикселей ($\approx 2,3 \cdot 10^{54}$ вариантов) число реализованных вариантов ничтожно мало по сравнению с теми, которые никто никогда не видел. Было бы интересно исследовать это многомерное пространство и определить, как в нем распределены точки, представляющие 2D-изображения объектов нашего мира. Цель работы — исследование свойств таких пространств для изображений, получаемых при сканировании «окном» размером 2×2 и 3×3 (а в будущем и больших размеров) набора различных изображений (животные, пейзажи).

В результате работы для выборки из 2000 изображений (с разрешением $\geq 1920 \times 1200$) были определены частоты всех встреченных комбинаций и построено их распределение, изучено влияние сжатия и зашумления изображений, рассчитаны числовые значения, характеризующие цветовые маски изображений и отношения между ними, а также оценена вычислительная производительность разработанной программы.

Научный руководитель — д-р физ.-мат. наук А. Ю. Пальянов

**Проектирование и разработка программного средства контроля
производственных операций и технического обслуживания
телеметрической системы «Луч-М-2014»**

Е. Н. Казырицкий
Новосибирский государственный университет

Телеметрическая система «Луч-М-2014», выпускаемая в ООО НПП ГА «Луч», используется для геофизического исследования скважин в процессе бурения. Аппаратура эксплуатируется при высокой температуре и в экстремальных условиях, встречаются усложненные геологические условия. В связи с этим необходимо проводить техническое обслуживание модулей, производить ремонт и замену комплектующих. Для своевременного обслуживания телеметрической системы важно вовремя узнавать оперативную информацию о текущем состоянии изделий в процессе эксплуатации и ремонта, степени готовности на производстве.

Целью работы является оптимизация и обеспечение доступности информации о контроле производственных операций и техническом обслуживании модулей телеметрической системы «Луч-М-2014».

В рамках работы были изучены и описаны производственные процессы в ООО НПП ГА «Луч», проанализирована специфика технического обслуживания телеметрической системы «Луч-М-2014», спроектировано и реализовано программное средство контроля производственных операций в виде веб-приложения.

Научные руководители —
канд. техн. наук, доц. А. А. Власов, Д. В. Тейтельбаум

**Разработка программного обеспечения потокового измерителя
удельного электрического сопротивления
попутной воды газовых скважин**

Я. В. Кожевникова

Новосибирский государственный университет
Институт нефтегазовой геологии и геофизики СО РАН, Новосибирск

На сегодняшний день значительная часть сеноманских залежей, разрабатываемых ПАО «Газпром», вступила в завершающую стадию разработки из-за обводнения продуктивного пласта. Накопление пластовой жидкости на забое скважины серьезно осложняет технологический процесс добычи газа. Поскольку контроль степени обводненности в настоящее время ведется геохимическими методами с ручным отбором проб, разработка косвенного метода, опирающегося на свойство электропроводности минерализованной воды, позволяющего диагностировать степень обводнения скважины дистанционно с помощью онлайн-мониторинга посредством разработанного в ИНГГ СО РАН аппаратурно-программного комплекса, имеет большое научное и практическое значение.

Целью данной работы является создание приложения, именуемого RegistratorTDS, для аппаратурно-программного комплекса, позволяющего дискретно получать данные о температуре и сопротивлении жидкости с поточного резистивного датчика, обрабатывать их и доводить информацию об общей минерализации жидкости пользователю с помощью графического, консольного и веб-интерфейсов. Принцип построения приложения RegistratorTDS состоит в написании программного ядра, к которому обращаются два независимых пользовательских интерфейса: консольный и графический. Первый применяется для фиксирования длительных измерений в полевых условиях с возможным отключением электропитания от установки, второй — для мониторинга данных в реальном времени. Оформление приложения в виде ядра позволяет решить проблему дублирования кода и изменять внутреннюю структуру программы, не затрагивая ее внешнего поведения. Веб-интерфейс предоставляет возможность пользователю через сайт с регистрацией наблюдать физико-химические данные с различных скважин. Аппаратурно-программный комплекс прошел ряд лабораторных и полевых испытаний на специальном стенде в ИНГГ СО РАН и нефтегазо-конденсатном месторождении Медвежье ООО «Газпром добыча Надым».

Научный руководитель — канд. техн. наук А. А. Власов

**Разработка прототипа имплантируемого датчика
на основе акселерометра для долговременного измерения активности
мелких животных**

А. С. Кутергина

Новосибирский государственный университет

В экспериментах, нацеленных на исследование поведения животных, активно используются акселерометры. Но большинство существующих устройств имеют ряд ограничений при применении на мелких животных.

Целью данной работы было разработать прототип имплантируемого датчика на основе акселерометра и проанализировать возможность его применения для долговременного измерения активности мелких животных.

За основу датчика был взят датчик температуры [1], к которому был добавлен акселерометр. С точки зрения потребления энергии оптимально проводить измерения 1 раз в минуту. Была выдвинута гипотеза, что активность будет удовлетворительно описываться линейной комбинацией двух параметров: разностью векторов ускорения, измеренных с разницей 10 мс, и длиной вектора ускорения. Для проверки этой гипотезы было проведено испытание на 8 лабораторных мышах (96 ч, 1 изм./мин). Измерения проводились одновременно на двух системах, измеряющих активность: PhenoMaster и разработанном прототипе. На основании оценки корреляции с PhenoMaster было установлено, что активность адекватно оценивается начиная с 20-минутного усреднения (корреляция 0,775). Благодаря введению третьего параметра — разности векторов ускорения, измеренных с разницей 1 мин, была получена корреляция 0,798.

Оценка энергопотребления показала, что разработанное устройство массой менее 1 г способно измерять активность и температуру более 6 месяцев, при частоте записи 1 раз в 20 мин (усреднение поминутного считывания) и показывать корректные данные.

1. Пат. 2014153945/28 Российская Федерация. Датчик температурный накопительный // Федеральная служба по интеллектуальной собственности / Д. В. Петровский; ООО «Экспериментальные модели для биологических испытаний». 29.12.2014.

Научный руководитель — канд. биол. наук Д. В. Петровский

**Система для предсказания динамики структуры
молекулярно-генетических сетей на основе филогенетического
анализа узлов: разработка функций для визуализации
элементарных эволюционных изменений**

Е. В. Маловичко

Новосибирский государственный университет

Генные сети, отражающие совокупность координированно функционирующих генов, их белковых продуктов и взаимосвязей между ними, в настоящее время широко используются в системной биологии наряду с филогенетическими деревьями, выражающими эволюционные взаимосвязи между различными видами или другими сущностями, имеющими общего предка. Существуют программные инструменты для визуализации и анализа различных аспектов биологических сетей, однако удобных пакетов, объединяющих эти обе концепции, не известно.

Целью работы является создание системы для предсказания и визуализации топологии генных сетей для немодельных организмов на основе филогенетического анализа узлов сетей их родственных видов, призванное облегчить эволюционный анализ и визуализацию результатов.

Разрабатываемая система позволит на основе графа генной сети и филогенетических деревьев, отражающих эволюционную динамику каждого узла сети — гена, построить новую топологию, тем самым предсказав, как может выглядеть генная сеть. В рамках данной работы была реализована функциональность, позволяющая строить и визуализировать молекулярно-генетические сети и совершать элементарные изменения в них: 1) точечные мутации, 2) дупликации генов, 3) делеции генов. Для построения и визуализации был использован наиболее распространенный подход к исследованию качественных закономерностей организации генных сетей, основанный на теории графов.

Новизна данной работы заключается в объединении качественной информации о топологии генных регуляторных сетей, участвующих в развитии немодельных организмов, с детальным филогенетическим анализом ее компонентов.

Практической ценностью данной работы является выявление различных тенденций развития немодельных организмов, обнаружение дупликаций и потерь генов в различных эволюционных линиях, отслеживание изменений структуры и функций элементов генных регуляторных сетей, участвующих в развитии организмов.

Работа выполнена при поддержке лаборатории кримографии JetBrains Research.

Научный руководитель — канд. биол. наук У. С. Зубаирова

Разработка библиотеки лучевого трассирования для моделирования сейсмических волн

И. Н. Морозов

Новосибирский государственный университет

Лучевое трассирование — распространенный метод моделирования распространения волн не только в сейсморазведке, но также в оптике, акустике, радиофизике и других областях науки. Поэтому важной задачей является разработка эффективных вычислительных библиотек для лучевого трассирования, так как массовое проведение лучей требуется в сейсморазведке или, например, для реалистичной графики в компьютерных играх. Целью работы является доработка существующего прототипа программы лучевого трассирования на языке Python для ускорения расчетов и обеспечения массового трассирования лучей.

Вычислительная библиотека реализуется на языке программирования C++. Реализуется итерационный алгоритм лучевого трассирования [1], в котором траектория луча находится путем минимизации целевой функции времени пробега волны вдоль траектории (принцип Ферма). Важным фактором для ускорения библиотеки является реализация явных формул зависимости времени пробега от смещения точек излома луча [1], что позволяет использовать градиентные методы оптимизации.

Для ускорения библиотеки требуется определить наиболее эффективные и быстрые библиотеки. Был проведен поиск тестирования и отбор библиотек: для интерполяции криволинейных поверхностей (границ) используется бикубическая интерполяция из библиотеки `libInterpolate`, для оптимизации целевой функции — библиотека `nlort`, а библиотека `Eigen` предоставляет математический аппарат для различных задач, которые возникают в ходе работы программы. Разработана структура классов, сериализация и десериализация в формате JSON.

В ходе тестирования показана вычислительная эффективность библиотеки: удалось достичь ускорения в 15 раз относительно первой версии библиотеки. Расчеты проводилось для скоростной модели из 4 слоев (8 сегментов луча) и 100 000 лучей (10 с против первоначальных 152 с). С использованием параллельных вычислений исследование пока что не проводилось.

1. Оболонцева И. Р., Гречка В. Ю. Лучевой метод в анизотропной среде (алгоритмы, программы). Новосибирск, 1989. С. 9–18.

Научный руководитель — канд. физ.-мат. наук, доц. А. А. Дучков

**Визуализация данных по экспрессии генов
для web-интерфейса базы данных OORT**

О. В. Никулина

Новосибирский государственный университет
Институт цитологии и генетики СО РАН, Новосибирск

Эксперимент по секвенированию транскриптома (RNA-seq) у растений стал практически рутинной процедурой для изучения как модельных организмов, так и сельскохозяйственных культур. В результате биоинформатической обработки таких экспериментов получают объемные разнородные данные, представленные нуклеотидными последовательностями транскриптов, аминокислотными последовательностями и их структурно-функциональной аннотацией. Полученные данные важно представить широкому кругу исследователей в виде баз данных (БД).

Для хранения и анализа транскриптомных экспериментов у сельскохозяйственных растений ранее была разработана база данных OORT (*Out Of Reference Transcripts*), которая позволяет пользователям с помощью поисковых запросов извлекать информацию о структуре и функциях ранее неаннотированной части транскриптомов сельскохозяйственных растений, в частности, идентифицировать новые гены устойчивости к заболеваниям и абиотическому стрессу, длинные некодирующие РНК, последовательности мРНК, получать оценки уровня экспрессии этих транскриптов.

Одной из задач визуализации данных этой базы является интерактивное графическое представление результатов оценки уровней экспрессии транскриптов.

Целью данной работы является разработка графических компонент для интерфейса БД OORT. Созданная компонента позволяет отображать и экспортировать данные по экспрессии транскриптов для удобного анализа пользователя.

Научный руководитель — канд. биол. наук Д. А. Афонников

Исследование применимости пакета SPIN для целей верификация Reflex-программ

А. А. Пономаренко

Институт автоматики и электротметрии СО РАН, Новосибирск

В современном мире к программам автоматизации киберфизических систем в науке и промышленности выдвигаются повышенные требования к надежности. В Институте автоматики и электротметрии СО РАН был разработан процесс-ориентированный язык Reflex. Уже разработаны трансляторы, среды разработки, но остается открытой проблема верификации разрабатываемых алгоритмов. Верификация осложняется логическим и физическим параллелизмом процессов, неопределенным временем функционирования, активным взаимодействием с окружающей средой. Один из подходов к верификации — метод проверки моделей, он достаточно хорошо справляется с задачами проверки управляющих программ. Однако вопрос применимости данного подхода к верификации процесс-ориентированных программ остается открытым.

В результате анализа существующих средств, реализующих метод проверки моделей, был выбран пакет SPIN [1]. В качестве входных данных SPIN используются описания моделей системы на языке Promela. Предложенный способ верификации подразумевает преобразование Reflex-программ в Promela-модели и их последующую проверку.

В качестве демонстрационной была выбрана задача автоматизации «Подъемной платформы для маломобильных пользователей». Были составлены программы управления и модели объекта управления на языке Reflex, сформулирован список требований, в том числе в виде LTL-формул. Reflex-процессы программы управления объекта управления были переведены в последовательно активируемые процессы языка Promela. Представлены результаты проверки при использовании SPIN. Эффективность выбранного метода верификации дала обнадеживающие результаты.

1. Gerard J. Holzmann The Model Checker SPIN // IEEE transactions on software engineering. 1997. Vol. 23. No. 5. P. 279–295. URL: <http://spinroot.com/spin/Doc/ieec97.pdf>

Научные руководители —
д-р техн. наук В. Е. Зюбин, канд. физ.-мат. наук Н. О. Гаранина

Разработка модуля для обеспечения интероперабельности атомных систем

А. А. Потехина

Новосибирский государственный университет

В настоящее время информация об анализе спектральных данных используется во многих исследованиях и областях науки, например, таких как астрофизики и археология. Но из-за большого количества и сложной структуры таких данных работа с ними очень трудоемка даже для специалистов. Поэтому стали разрабатывать атомные информационные системы (ИС), которые значительно облегчают работу со спектральными данными, предоставляя функции для работы с ними.

Одной из важнейших задач ИС в области спектроскопии является предоставление актуальных данных. Для осуществления этой задачи требуется обеспечить регулярное обновление данных, т. е. обмен данными (получение и передачу) в системе, что является проблемой интероперабельности таких систем. Однако многие ИС не имеют инструментов для решения данной задачи.

Целью данной работы является проектирование модуля для обеспечения интероперабельности атомных систем.

Перед проектированием модуля было определено следующее: данные, которые могут быть запрошены; формат представления спектральных данных; способ взаимодействия пользователя с системой при обмене данными (API).

По запросу модуль отдает/принимает спектральные данные для одной атомной системы в формате json. В роли API выступает веб-страница, предоставляющая функции получения или отправки файла с данными. При запросе на получение данных на веб-странице указывается имя атомной системы, посылается URL-строка с параметром name, содержащим имя атомной системы. При осуществлении передачи данных URL-строка не содержит параметров. Загрузка и скачивание файла с данными осуществляется с помощью HTTP-метода POST.

Модуль реализован на языке php и JavaScript с использованием библиотек php_mssql, cURL, libjson в среде разработки PhpStorm.

Разработанный модуль успешно протестирован на информационной системе «Электронная структура атомов» при отображении соответствующих данных на квантограмме. Использование данного модуля не вызывает задержек при обработке запроса пользователя.

Научный руководитель — канд. физ.-мат. наук В. Г. Казаков

Измерение геометрических характеристик брюшной аорты

И. А. Прокофьев

Новосибирский государственный университет

Аневризма брюшной аорты является серьезным заболеванием, несвоевременные диагностика и хирургическое лечение которого зачастую заканчивается летальным исходом пациента. Раннее выявление и последующее наблюдение динамики развития аневризматических трансформаций аорты позволяют производить заблаговременное планирование хирургического вмешательства, что значительно снижает риск послеоперационных осложнений и смертности у пациентов. Важным этапом диагностики аневризмы брюшной аорты является измерение ее геометрических характеристик.

Целью данной работы является реализация алгоритма для измерения геометрических характеристик брюшной аорты, в том числе с аневризматическими трансформациями, по данным компьютерной томографии пациента.

Клинически значимые характеристики геометрии аорты описаны в работе [1]. Также после проведения консультации с представителями медицины список характеристик был дополнен такими характеристиками, как углы бифуркации, длина подвздошных артерий и т. д.

Таким образом, в результате проделанной работы реализован алгоритм для измерения геометрических характеристик брюшной аорты. Алгоритм реализован в виде пакета на языке Python 3 с использованием библиотеки `numpy`. Качество работы алгоритма проверено на реальных данных пациентов. В дальнейшем планируется интеграция реализованного алгоритма в разрабатываемый программный комплекс и использование в клинической практике.

1. da Silva E. S. et all. Morphology and diameter of infrarenal aortic aneurysms: a prospective autopsy study // *Cardiovasc. Surgery*. 2000. Vol. 8. No. 7. P. 526–532.

Научный руководитель — Р. Ю. Епифанов

Когнитивная графика в атомной спектроскопии: новая диаграммная техника и многомерное представление в визуализации спектров информационной системы «Электронная структура атомов»

Е. А. Темникова

Новосибирский государственный университет

Многим ученым требуются спектральные данные атомных систем. Такие данные сегодня получают из открытых интернет-ресурсов, информационных систем. В целом анализ спектров — сложная задача, для которой люди используют табличные и визуальные представления данных. Существуют ведущие ИС, такие как NIST, SPECTR-W3, VAMDC. Они позволяют получать данные и в табличном, и в визуальном виде, в том числе спектрограммы, диаграммы Гротриана и др.

Однако на настоящий момент эти представления не связаны между собой. Специалист может работать только с одной визуализацией или с одной таблицей в один моменты времени. Если реализовать возможность связи данных представлений так, чтобы изменение фильтров сортировки давало реакцию на других представлениях, то исследователям было бы проще изучать электронную структуру атомов.

По итогу работы был спроектирован программный модуль и многооконный интерфейс, который будет работать в составе имеющейся системы «Электронная структура атомов».

Модуль включает общий фрейм для представлений, пользователь может установить параметры, которые применяются ко всем отображениям. Также в модуле программно реализован обмен данными и параметрами между другими модулями, включенными в систему.

На первом этапе были объединены таблица переходов и модуль спектрограммы. Выделение строки в таблице переходов сопровождается фильтрацией спектральных данных и отображением выбранного перехода на спектрограмме.

Далее были собраны пользовательские требования к общим настройкам на главном фрейме. И вначале реализована одна настройка — длина волны, которая применяется к спектральному модулю и диаграмме Гротриана.

Данный прототип был протестирован и отлажен. Планируется внедрение в систему ИС ИСА.

Научный руководитель — канд. физ.-мат наук В. Г. Казаков

НЕЙРОННЫЕ СЕТИ И МАШИННОЕ ОБУЧЕНИЕ

УДК 004.032.26

Named entity recognition in scientific texts

S. A. Berezin, A. E. Pauls
Novosibirsk State University

Named entity recognition is no longer a new task in the field of natural language processing. However, most of the research is devoted to solving this problem on the texts of colloquial, artistic, or formal speech styles. In this work, we considered the problem of extracting entities from scientific texts.

We used the SciERC corpus [1], which describes such entity types as “task”, “method”, “material”, “evaluation metric” and so on. This dataset includes annotations for scientific entities, their relations, and coreference clusters for 500 scientific abstracts. These abstracts are taken from 12 AI conference/workshop proceedings in four AI communities, from the Semantic Scholar Corpus.

In our experiments, we used the language models RoBERTa [2] and ELECTRA [3]. The best result we obtained is macro F1 of 0.58, which indicates the complexity of the problem since the same architectures were able to achieve macro F1 score of 0.97 on the general domain CoNLL 2003 dataset. In further research, we plan to improve the target metric by augmentation of training text data, after which the resulting solution will become the basis for software that allows indexing scientific articles and extract information from them.

1. Luan Y., He L., Ostendorf M., Hajishirzi H. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction // EMNLP. 2018. P. 3219–3232.

2. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. RoBERTa: A Robustly Optimized BERT Pretraining Approach // ArXiv, 2019. abs/1907.11692.

3. Clark, Kevin, Minh-Thang Luong, Quoc V. Le and Christopher D. Manning. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators // ArXiv, 2020. abs/2003.10555.

Scientific Supervisor — Cand. Phys.-Math. T. V. Batura

Zero-shot learning approach to the problem of short text classification

N.A. Nikolaev
Novosibirsk State University

A text classification problem is an extremely common data mining task. It consists of automatic extraction of the most significant features from the text, based on which it is possible to determine whether the text belongs to one of the known classes. But for today, interest in this problem does not fall, since this problem takes on new shapes. In many situations, the data can be very unbalanced, up to the complete absence of training sample elements for some known classes. Thus, researchers create models that are trained in conditions that can be termed as zero-shot learning.

Zero-shot learning (ZSL) is a specific type of technique: to train a model on one set of data classes and to evaluate on another that the model has not seen before. ZSL is a promising avenue, however very few works are presented in this area. Moreover, works of various researchers are based on miscellaneous methods, and the quality of models is evaluated on different datasets, which makes a comparison of results unreasonable. In this work, we present a summary of existing solutions, as well as our own method for solving the stated problem. We also consider the problem associated with the lack of unified data for evaluating the quality of classification models and propose our own variant of a dataset to train a robust classifier.

The idea behind this work is to solve the stated text classification problem in a zero-shot setting using well-established transfer learning approach, since definitions of both approaches are similar, but have distinctive features. The main aim is to build a latent feature multidimensional space that can map the same features of the selected object to different modalities. Generally, there are two modalities, one of which is the main one, and the second is an additional, descriptive one. Therefore, in this work, we provide our own way of building such a complex space with the usage of different modern neural networks and loss functions. Results obtained during series of experiments provide important insights into the complexity of the chosen task and show the quality ratings of the proposed models.

Scientific advisor — Cand. Phys.-Math. Sci. T. V. Batura

Facemask detection using MMDetection toolbox

Mukul Kumar Vishwas ¹, Petr Menshanov ¹, Aleksey Okunev ²

¹Novosibirsk State University

²NSU Higher College of Computer Science, Novosibirsk

In this project we created a face mask detection model using MMDetection an object detection toolbox, our model can distinguish if the person wearing the face mask correctly/incorrectly or not, in the current situation this model is extremely useful as this simple precaution will help to stop the spreading of deadly Coronavirus. The output image should satisfy these four points for a successful recognition system: 1) identify the location of the face; 2) identify face with correct mask, no mask or incorrect mask; 3) a boundary box with the probability of the Model's prediction (varies from 0-1); 4) a mask on the confident region (face).

To train our model we collected more than 8 thousand images from various security cameras and manually annotated it using labelme image annotation tool. We choose MMDetection because it's faster, scalable, support variety of dataset's. Our model is divided mainly into four parts *a) Backbone:* we used *Resnet50* to extract features from the images, *b) Neck:* feature pyramid network (FPN) is used to get more information from the extracted features, *c) Dense head:* As name indicated it works on the region where features are very compact, *d) RoI extractor:* Region of interest is the final part of the model here model actually draw the mask and boundary box on the recognized face.

While training the model the images were resized, flipped, normalized and padded it helped the model to have a diversity of training samples. Because we are training our model on security camera images we faced these problems: 1) low resolution images; 2) difficulty to detect distance people; 3) illumination condition in the images; 4) occlusion/overlapping in images.

To measure the accuracy of our model we used mAP (mean Average Precision), it is a popular metric for object detection models. It gives a calculated value between 0 and 1. The initial mAP value for our model was .215 after adjusting hyper parameter we are able to achieve mAP value .436. The model will be integrated with the NSU Higher College of Computer Science, Novosibirsk security cameras.

Future work: currently we are trying to integrate the current model with unsupervised person re-identification model. It will enable us to track the people if they move from one camera's vision to another.

Research adviser — Dr. A. Okunev

Методы искусственного интеллекта в автоматизации техподдержки

А. А. Аборнев, Л. С. Комаров, В. А. Эйхвальд
Томский государственный университет
систем управления и радиоэлектроники

Роль информационных систем в бизнес-процессах предприятий становится все больше с каждым годом. При взаимодействии пользователей с информационными системами могут встречаться случаи нештатного поведения системы. Для решения этих вопросов существуют специальные службы, называемые службами технической поддержки. Пользователи взаимодействуют с технической поддержкой посредством обращений, содержащих в себе текстовое описание вопроса или проблемы.

Основой обращений является текстовая информация, которую необходимо классифицировать для направления верному исполнителю, эту задачу предлагается автоматизировать.

Классификация текста является классической проблемой в обработке естественного языка. Задача состоит в том, чтобы назначить предопределенные категории для данной текстовой последовательности. Для данной задачи подходит нейронная сеть, которая после обучения сможет корректно классифицировать обращения клиентов, независимо от их изложения.

Важным этапом в классификации текста является представление текста в удобный для классификатора вид. В проекте для представления текста используется предварительно обученная модель BERT.

Перед обучением модели осуществляется подготовка выборки данных. Для системы данными служат набор текстовых запросов, который необходимо обработать для эффективной работы модели.

К выборке данных применена лемматизация и чистка от стоп-слов, а для дообучения данной модели на конкретную задачу выборка разделена на токены, соответствующие тем, что встречаются в словаре предварительно обученной модели.

Перед обучением проводится деление токенизированного текста на тренировочную выборку и валидационную, на которой будет измеряться эффективность обученной на тренировочной выборке модели.

Результатом работы является дообученная модель, предназначенная для назначения предопределенных категорий обращениям, поступившим службе технической поддержки.

Научный руководитель — А. А. Матолыгин

Увеличение разрешения спектрограмм для алгоритмов синтеза речи

М. А. Баушенко

Новосибирский государственный университет

Воссоздание речевого сигнала по печатному тексту прочно вошло в нашу жизнь. Синтез речи может понадобиться везде, где получателем информации служит человек. Например, он может использоваться в голосовых помощниках, при прослушивании аудиокниг, объявлениях в аэропортах и ж/д вокзалах, для людей с ограниченными возможностями и в других отраслях. Важными параметрами качества синтезируемой речи являются ее сходство с человеческим голосом, а также способность быть понятной, что является предметом пристального внимания исследователей в последнее время.

Целью работы является улучшение алгоритмов синтеза речи посредством суперразрешения. Для достижения этой цели было предложено использовать алгоритмы сверхразрешения Pix2PixHD [1] и Neural Supersampling [2] в части обработки спектрограмм.

Был обучен суперразрешающий препроцессинг спектрограмм речи. Реализован суперразрешающий препроцессинг для вокодеров WaveNet и Griffin-Lim. После применения полученных синтезаторов для речи проведена оценка натуральности речи. Проведен анализ значений метрики до и после применения улучшения спектрограмм.

1. Sheng L., Pavlovskiy E. N. Reducing over-smoothness in speech synthesis using Generative Adversarial Networks // Int. Multi-Conf. Eng., Comput. Inf. Sci. (SIBIRCON). Novosibirsk, 2019. P. 0972–0974.

2. Xiao L. et al. Neural supersampling for real-time rendering // ACM Trans. Graphics (TOG). 2020. Vol. 39. No. 4. P. 142: 1–142: 12.

Научный руководитель — канд. физ.-мат. наук Е. Н. Павловский

Разработка алгоритма распознавания звуков с использованием ансамбля сверточных нейронных сетей

С. К. Вербицкий

Новосибирский государственный университет

Распознавание звуковых образов является актуальной и быстроразвивающейся исследовательской задачей. Существует достаточно много примеров практического применения систем распознавания звуковых образов: классификация и детекция звуков окружающей среды; классификация музыкальных жанров; распознавание эмоциональной окраски голоса; классификация заболеваний дыхательной системы и их симптомов с помощью записей дыхания на электронный стетоскоп, а также распознавание дикторов и контроль популяции животных с помощью детекции и классификация их звуков (например, пение различных птиц).

Набор данных Audio Set [1] содержит более 2 млн аудиозаписей. Точность систем для маркировки звуков на данный момент на порядок ниже, чем точность современных систем для задачи распознавания изображений на аналогичных огромных датасетах.

Основной целью работы является исследование и улучшение систем, основанных на остаточных сверточных нейронных сетях, а также реализация и сравнение частотно-временных представлений аудиосигнала и других различных подходов с целью повышения эффективности и производительности систем для задачи распознавания звуковых образов.

В результате получена наивысшая точность классификации и маркировки звуков на трех различных наборах данных. На самом крупном датасете для задачи распознавания звуковых образов Audio Set [1] удалось достигнуть 0.462 mAP, что выше результатов предыдущей лучшей системы с 0.442 mAP [2] на 0,02. Одна из предложенных систем быстрее более чем в два раза и имеет в два раза меньше параметров, чем предыдущая лучшая система, являясь при этом более точной с 0.445 mAP.

1. Gemmeke J. F. et all. Audio Set: An ontology and human-labeled dataset for audio events // IEEE ICASSP. 2017.

2. Kong Q. et all. PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition // IEEE/ACM Trans. ASLP. 2020.

Научный руководитель — д-р техн. наук В. Б. Бериков

Сравнение качества работы 2-мерной и 3-мерной архитектур Unet на примере распознавания аорты на снимках КТ

Е. А. Волкова

Новосибирский государственный университет

Успешное применение методов решения биологических задач сегментации на основе нейронных сетей архитектуры Unet [1] повлекло к адаптации таких подходов к медицинскому домену. Характерной особенностью части медицинских данных является не пиксельное, а воксельное представление. К примеру, воксельное представление имеют снимки компьютерной или магнитно-резонансной томографии. Наличие третьей пространственной координаты поднимает вопрос о необходимости применения 3-мерного варианта архитектуры Unet в медицинском домене.

Целью данной работы является сравнение качества работы нейросетевого алгоритма на основе 2-мерного и 3-мерного варианта архитектуры Unet в задаче сегментации аорты на снимках компьютерной томографии.

В рамках данной работы было проведено сравнение качества работы 2-мерного и 3-мерного варианта архитектуры Unet. Показано, что применение 3-мерного варианта архитектуры позволяет добиться лучшего качества решения задачи. Сравнение проводилось на языке Python, для реализации нейронных сетей была использована библиотека PyTorch.

1. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // Int. Conf. Medical Image Computing Computer-Assisted Intervention. Springer, Cham, 2015. P. 234–241.

Научный руководитель — Р. Ю. Епифанов

Решение задачи распознавания образов при помощи алгоритма гибридной СММ-нейросети

С. В. Гилин

Сибирский федеральный университет, Красноярск

В данной работе решается задача распознавания слов в аудиозаписях. Разработка систем распознавания и преобразования речи является одним из самых актуальных и востребованных направлений в области искусственного интеллекта, машинного обучения и анализа данных. Существующие технологии распознавания речи позволили упростить поиск информации в базах данных и поисковых системах с помощью голосовых запросов. Также они способствовали появлению различных устройств, управляемых с помощью голосового набора, систем интерактивного речевого самообслуживания, команд приложений для диктовки текстов, искусственных помощников на мобильных устройствах, онлайн-консультантов на веб-сайтах и многих других направлений IT-технологий.

В работе предлагается собственный алгоритм гибридной СММ-нейросети распознавания речи, представляющий собой синтез алгоритмов скрытой марковской модели и нейронной сети. Также был разработан программный модуль, реализующий работу предложенного алгоритма гибридной СММ-нейросети.

ИНС представляют собой систему соединенных и взаимодействующих между собой простых процессоров (искусственных нейронов). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам.

При оценке распознанных слов звукового файла в работе используется метрика WER (*World Error Rate*).

Алгоритм гибридной СММ-нейросети для решения задачи распознавания слов в аудиозаписях:

1. Через устройство записи речи приложение получает объект распознавания.
2. Производится вычисление уровня помех и искажений в аудиофайле.
3. Выполняется выделение участка, содержащего речь, разбивка на отдельные слова при помощи промежутков тишины.
4. Начинается оценивание параметров полученных промежутков звукозаписи.
5. Выполняется преобразование звуковых промежутков в символьные схемы с помощью СММ.

6. Производится декодинг полученных схем с помощью обученного алгоритма нейросети.

7. Для проверки правильности распознавания используется метрика WER для оценки правильности распознавания слов.

8. Если пункт 7 не дает желаемого результата, то промежутки возвращаются к пункту 5.

При оценке объемности задачи потребовалось собрать соответствующую обучающую выборку. Данные были взяты из популярной видеоплатформы YouTube. При отборе спикеров использовались новостные и образовательные видео, которые затем были преобразованы в формат аудио. В качестве исходного формата для хранения аудиоданных используются аудиозаписи одного из наиболее распространенных форматов для хранения музыки и звука: WAV.

1. Местецкий Л. В. Математические методы распознавания образов. М.: МГУ, 2004.

2. Фланаган Дж. Л. Анализ, синтез и восприятие речи / пер. с англ. А. А. Пирогова. М.; New Jersey: Bell Telephone Laboratories, 1965.

Научный руководитель — канд. физ.-мат. наук, доц. И. В. Баранова

**Анализ методов машинного обучения
для сопоставления разноязычных графов знаний**

В. А. Гнездилова

Новосибирский государственный университет

Граф знаний — это граф, узлами которого являются объекты реального мира, а ребрами — отношения между объектами. Графы знаний используются для получения новых фактов о предметной области. Концепция связанных данных была представлена еще в конце XX в., но большое внимание к себе привлекла только в 2012 г., когда компания Google представила свой граф знаний.

Межязыковые связи между графами позволяют дополнить граф на одном языке информацией из графа на другом языке. Задача построения таких связей решается различными способами: ручным построением связей человеком, логическими алгоритмами и алгоритмами машинного обучения. В данной работе анализировались именно алгоритмы машинного обучения для решения подобного рода задач.

В последние годы исследователями были предложены различные варианты решений задачи сопоставления графов знаний. Однако эти алгоритмы не могут быть использованы при построении недостающих межязыковых связей, так как только в 50 % связь будет построена верно. Кроме того, такие алгоритмы не использовались для сопоставления англоязычного и русскоязычного графов.

В работе рассматриваются особенности сбора данных для обучающей выборки из английской и русской версии графа знаний DBpedia и анализируются результаты алгоритмов сопоставления графов знаний, основанные на отображении графа в унифицированное векторное пространство.

Полученные результаты помогут при разработке собственного алгоритма сопоставления графов знаний.

Научный руководитель — канд. физ.-мат. наук, доц. З. В. Апанович

Разработка автоматизированной системы прокторинга для платформы «Яндекс.Контест»

А. Л. Голованов

Новосибирский государственный технический университет

Тесты и соревнования по программированию зачастую помогают понять навыки разработчика, скорость его работы, а также качество и эффективность его кода. Онлайн-формат позволяет проводить их в любом месте и времени, однако в этом случае нельзя гарантировать его честность. Тестируемый может использовать сторонние материалы, искать похожие решения в интернете или вовсе выдавать чужое решение за свое.

Целью работы является разработка программного комплекса, позволяющего собирать действия пользователя во время прохождения теста на платформе «Яндекс.Контест» и анализировать их на предмет того, насколько честно выполнена работа.

Задачу было решено разделить на три части: сбор данных об активности пользователя, получение и хранение их, обработка и анализ данных для определения нечестной работы.

В начальной версии данными об активности будет являться то, какие вкладки пользователь открывал во время тестирования, покидал ли курсор мыши веб-страницу с тестом, на какие тестовые задания пользователь переключался и когда отправил решение. В дальнейшем список данных будет расширяться.

Сбор данных был реализован с помощью написанного на языке JavaScript расширения для браузеров на основе Chrome. Это расширение собирает данные о действиях пользователя во время прохождения теста на платформе «Яндекс.Контест» и отправляет их на сервер для хранения и обработки.

Для получения и хранения данных было разработано серверное приложение на языке Java с использованием Spring Framework. Сервер получает данные об активности пользователей во время написания теста и сохраняет их в базе данных.

Разработанная рекуррентная нейросеть по сохраненным данным определяет, являлась ли активность пользователя подозрительной и проходил ли участник тестирование самостоятельно или нет.

Проект курируют сотрудники CS-центра новосибирского отделения «Яндекс».

Научный руководитель — канд. техн. наук, доц. В. М. Стасышин

Разработка логико-вероятностных нейронных сетей

В. Д. Дегтярев

Новосибирский государственный университет

Началом изучения искусственных нейронных сетей можно считать 1943 г., когда У. Макколак и У. Питтс формализовали понятие нейронной сети. С тех пор интерес к этой теме только возрастал, что подтверждается появлением большого количества различных моделей нейронных сетей, которым неизменно находили применение в робототехнике, медицине, экономике и т. д.

Практически все модели, которые используются в настоящее время, имеют существенный недостаток — полное или частичное отсутствие интерпретируемости, из-за чего нейронные сети получили название «черный ящик». Они аппроксимируют данные, но не обнаруживают логико-вероятностную связь между входом и выходом.

Целью нашей работы является разработка логико-вероятностной модели искусственной нейронной сети, основанной на вероятностных формальных понятиях. Эта модель обнаруживает вероятностные законы на признаках объектов. По этой причине она не только предсказывает целевой признак, но и строит семантический вероятностный вывод этих предсказаний. Также использование вероятностных формальных понятий позволяет решить проблему в пропусках в данных [1, 2].

Модель может быть применима во многих областях, где предсказание не может быть использовано без его объяснения. Например, такими областями являются медицина, банковская сфера и военное дело.

1. Vityaev E., Degtiarev V., Pak B., Meister Y. Formalization of “natural” classification and “natural” concepts by probabilistic generalization of formal concepts analysis // CEUR Workshop Proc. 2020. P. 59–73.

2. Vityaev E., Demin A., Ponomariov D. Probabilistic generalization of formal concepts // Programming. 2012. Vol. 38. P. 18–34.

Научный руководитель — д-р физ.-мат. наук, проф. Е. Е. Витяев

Мобильное приложение для диагностирования нарушения органов зрения

Э. Р. Добров

Новосибирский государственный университет экономики и управления

В данной работе дается оценка алгоритмов для создания компьютерных программ в сфере прикладного глубокого обучения в техническом зрении. Алгоритмы базируются на применении искусственных нейронных сетей.

Поставлена задача реализовать информационную систему обнаружения нарушения зрения на основе глубоких нейронных сетей.

Основная особенность заключается в том, что для решения поставленной задачи в разработке информационной системы используются методы машинного обучения в офтальмоскопии.

По показателям состязаний Kaggle, проходивших в 2015, 2019 г., полученное успешное решение определения диагноза на основании осмотра глазного яблока достигло 93 %. На сегодняшний день статистика врачебных ошибок говорит, что существует проблема в постановке степени тяжести диагноза при обследовании пациентов. В контексте подготовки НКР ставится научная задача: усовершенствование рекорда распознавания до 99 %.

По завершении разбора и оценки современных алгоритмов в состязании «Обнаружение диагноза глазного дна» на соревновательной площадке Kaggle будет подобрано оптимальное решение выполненного исследования. Измерить качество работы модели запланировано на публичных датасетах Kaggle.

Развитие рынка мобильного здравоохранения с возможностью удаленного мониторинга и консультаций направлено на повышение качества и доступности медицинского сервиса. Телемедицина испытывает большую потребность в том, чтобы иметь возможность автоматически распознавать изображения на мобильных устройствах в диагностических системах заболеваний глаз по снимкам.

Глубокое обучение требует большой вычислительной мощности. В разработке сложных мобильных приложений существует ряд ограничений, связанных с энергопотреблением и мощностью обработки. Для разрешения этих ограничений предложена легковесная архитектура нейронной сети для достижения компромисса между сложностью математической модели и точностью.

В будущем разработанное мобильное приложение планируется внедрить в офтальмологическом отделении медучреждения.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

Выбор оптимальной архитектуры нейронной сети для распознавания жестов

Г. А. Долгов

Томский государственный университет

По статистике Всемирной федерации глухих, в мире насчитывается 70 млн неслышащих людей. Жестовый язык играет важную роль во взаимодействии слабослышащих друг с другом и с окружающими. Поэтому актуальны исследования в области интерпретации жестового языка.

На сегодняшний день в ситуациях, когда необходимо передать какое-либо сообщение устно, слабослышащие обращаются в специализированные кол-центры. В этом случае требуется посредник, что является недостатком в некоторых ситуациях. Для непосредственного общения существуют системы визуального и невизуального распознавания жестов. Визуальный метод имеет недостатки, главным из которых является зависимость изображения от окружающей обстановки для распознавания жеста. Решения невизуального перевода основываются на использовании датчиков, расположенных на пользователе.

Чтобы достаточно точно определить введенный жест, необходимо провести его классификацию по определенным признакам. Но выделение признаков для большого числа жестов является сложной задачей. Для ее решения возможно использование нейронной сети. В процессе обучения она будет выделять необходимые для классификации признаки по набору данных.

Был проведен анализ следующих архитектур нейронных сетей: нейронной сети прямого распространения, глубокой нейронной сети прямого распространения, трехуровневой сверточной нейронной сети, ансамбля из двух трехуровневых сверточных нейронных сетей.

Разработка и тестирование нейронных сетей проводились в интерактивной среде Google Colab, с последующей проверкой на микроконтроллере STM32 с использованием пакета расширения STM32Cube.AI в среде разработки STM32CubeIDE.

По результатам тестирования наибольшую точность распознавания показал ансамбль из двух трехуровневых сверточных нейронных сетей, но результаты зависят от таких факторов, как количество данных для обучения, количество эпох обучения, выбор оптимизатора и количество нейронов в слое.

Научный руководитель — канд. физ.-мат. наук, доц. О. А. Доценко

Ускорение алгоритма генерации русскоязычной речи

И. Е. Ильин

Новосибирский государственный университет

Синтез речи является одной из важных задач в области обработки звука. Сейчас появляется все больше голосовых ассистентов, для которых качество и скорость синтеза речи — определяющие факторы. Также есть востребованность синтеза речи в реальном времени для кол-центров, которые зачастую не могут справиться с большим наплывом клиентов.

На сегодняшний день в области синтеза речи доминирует подход генерации, основанный на нормализованных потоках (*Normalizing Flow-based Synthesis*), поскольку подобные нейронные сети способны генерировать речь, практически не отличающуюся от человеческой. Однако типичные размеры таких моделей зачастую превышают тысячи мегабайт, что свидетельствует о высокой вычислительной сложности. Это, в свою очередь, затрудняет их работу на вычислительных устройствах, не обладающих таковой. Таким образом, ускорение нейронных сетей для генерации речи — актуальная научная задача, что и является целью данной работы.

В качестве базового решения была выбрана современная генеративная архитектура, состоящая из энкодера Tacotron-2 и декодера WaveGlow. В результате профилирования архитектуры было выяснено, что самой медленной частью является декодер WaveGlow. При этом для достижения наилучшего результата было принято решение модифицировать нейронную сеть на уровне архитектуры. В качестве методов оптимизации скорости и размеров сети были выбраны методы дистилляции и структурированного прунинга, которые были применены к уже обученной модели с последующей тонкой настройкой.

В результате работы удалось ускорить работу базовой нейронной сети более чем в два раза на CPU и в полтора раза на GPU, при сохранении качества синтеза на высоком уровне.

Научный руководитель — А. И. Гончаренко

Система компьютерного зрения для анализа урожайности посевов пшеницы

М. В. Кожекин

Новосибирский государственный университет

Обнаружение колосьев пшеницы на цифровых изображениях является важной задачей для оценки различных характеристик пшеницы, таких как плотность популяции, состояние здоровья, размер и стадия зрелости, наличие остей и т. д. Были разработаны методы обнаружения колосьев пшеницы по RGB-изображениям на основе алгоритмов машинного обучения [1, 2]. Высокая вариативность условий наблюдения, генотипических различий и стадий развития делает обнаружение колосьев пшеницы непростой задачей для систем компьютерного зрения в полевых условиях. Для улучшения качества распознавания колосьев используются большие, разнообразные и хорошо аннотированные наборы данных, такие как GWHD, состоящий из 4700 изображений и 190 000 колосьев. На основании этого набора данных была обучена нейронная сеть архитектуры Mask-RCNN. В ИЦиГ СО РАН был предложен новый протокол автоматического получения изображений посевов пшеницы с помощью БПЛА DJI Phantom 4 Pro, оборудованного цифровой камерой видимого спектра. Целью данной работы является разработка системы компьютерного зрения для распознавания колосьев на изображениях (снятых согласно данному протоколу) и их визуализация на интерактивной карте.

Для решения поставленной задачи использовалась нейросеть Mask-RCNN, инструмент для выделения метаданных exiftool и библиотека для построения карты folium. Полученная карта хранится как веб-страница, что повышает гибкость и легкость разворачивания системы. Исходный код системы выложен в открытый доступ [3].

1. Alkhudaydi T. et all. An Exploration of Deep-Learning Based Phenotypic Analysis to Detect Spike Regions in Field Conditions for UK Bread Wheat // Plant Phenomics. 2019. Vol. 2019.

2. Genaev M. A. et all. Morphometry of the wheat spike by analyzing 2D images // Agronomy. 2019. Vol. 9. No. 7. С. 390.

3. Кожекин М. В. Система компьютерного зрения для анализа урожайности посевов пшеницы [Электронный ресурс] // GitHub. URL: https://github.com/SI07h/wheat_detection

Научный руководитель — канд. биол. наук М. А. Генаев

Применение сверточных нейронных сетей для сегментации томографических изображений гидратосодержащих образцов

Т. О. Колесник

Новосибирский государственный университет

В настоящее время проявляется устойчивый интерес к природным газогидратам как к альтернативному источнику природного газа. Синхротронная компьютерная томография позволяет получать трехмерные изображения внутренней структуры образцов с высоким временным и пространственным разрешением, что порождает большие объемы данных. В связи с этим актуальной является задача разработки алгоритмов автоматической сегментации больших объемов томографических данных.

Сегментация 2D-проекций гидратосодержащих образцов представляет собой задачу многоклассовой классификации изображений в оттенках серого; необходимо выделять классы: «Гранула», «Гидрат» и «Флюид». Стандартным подходом является выделение классов по пороговым значениям интенсивности. Такой подход плохо работает для гидратосодержащих образцов, так как классы гранулы и гидрата-флюида являются слабоконтрастными. Поэтому был рассмотрен подход к сегментации гранул на основе сверточных нейронных сетей, в основе которых лежит идея автоэнкодера.

В данной работе было обучено и протестировано несколько моделей на основе архитектур нейронных сетей FCN-VGG16, SegNet и UNet, а также протестированы разные функции потерь *Binary Cross Entropy*, *Jaccard distance* и *Dice distance*. Были рассмотрены различные реализации слоев кодировки и декодировки карт-признаков в выбранных архитектурах, а также подход пакетной нормализации для моделей UNet. В качестве метрики качества сегментации было выбрано усредненное значение бинарной меры сходства *Intersection over Union*, посчитанной отдельно для класса гранулы и гидрат-флюида.

Наилучший результат сегментации по двум классам был достигнут с помощью нормализованной модели UNet, минимизирующей функцию потерь *Jaccard distance*. Было достигнуто качество сегментации гранул 95,1 % на тестовой выборке, что является заметным улучшением по сравнению с 77,9 % для сегментации по пороговому значению интенсивности.

Научный руководитель — канд. физ.-мат. наук, доц. А. А. Дучков

***Zero-shot learning* в задаче классификации намерений
в диалоговых системах**

А. Р. Королев

Новосибирский государственный университет

Zero-shot learning — класс задач машинного обучения, когда в тестовой выборке встречаются классы объектов, ранее не встречаемые в ходе тренировки, например при обучении классификации изображений животных на одних классах и распознавании новых классов в тестовой выборке с использованием только описания новых классов. Существующие методы основаны на сетях глубокого обучения, таких как генеративно-сопоставительные сети, вариационные автоэнкодеры, байесовские сети и др.

Эти же методы могут быть применены в задачах распознавания естественного языка, а более конкретно — в распознавании намерений в универсальных диалоговых системах. Намерение — задание, выполнение которого пользователь хочет от системы. Система распознает намерение и через подключенный сервис выполняет его. Однако если мы подключим новый сервис, то должны перетренировывать диалоговую систему. Но с помощью методов *zero-shot learning* можно добавить новые намерения без переобучения, предоставив только текстовое описание новых намерений, что делает диалоговые системы более универсальными помощниками.

В данной работе проверяется гипотеза об использовании архитектур *zero-shot learning* для задач компьютерного зрения в обработке естественного языка. Проверялись следующие архитектуры:

- CADA-VAE — архитектура на основе вариационного автокодировщика;
- LisGAN — архитектура на основе генеративно-сопоставительной сети.

Проверка проводилась на известном датасете соревнования для диалоговых систем DSTC-8. Для кодирования текстовых данных использовались различные кодировщики предложений (RoBERT, XLM-RoBERTa и др.). Сравнивалась точность предсказания отдельно видимых и невидимых намерений, а также их среднее гармоническое.

Научный руководитель — И. Ю. Бондаренко

Автоматическое выделение разломов на сейсмических изображениях

Д. С. Коростелев

Новосибирский государственный университет

Разломы являются важнейшими компонентами геологических разрезов, влияющими на формирование ловушек нефти и газа, поэтому их интерпретация — важная задача, которая требует автоматизации.

Интерпретация разломов в 3D-кубах сейсмических данных представляет собой задачу бинарной сегментации; требуется определять наличие или отсутствие разлома в каждом вокселе куба. Для интерпретации применяют как детерминистические методы, так и решения на основе сверточных нейронных сетей. Оба подхода имеют свои недостатки: в первом случае велико число ложноположительных ошибок; во втором случае необходимо размечать данные для обучения. Поэтому был рассмотрен метод обучения сверточных нейронных сетей на основе синтетических данных [1].

В данной работе был использован метод адаптации синтетических данных за счет зашумления на основе метода *Fourier Domain Adaptation* (FDA). Были обучены и протестированы модели, построенные на модификации архитектуры UNet для 3D-сегментации. Для проверки качества сегментации были выбраны модификации метрик *Precision* и *Recall*, учитывающие специфичность масок сегментации.

Наилучшее решение было получено при комбинации результата работы архитектуры Unet с ResNet энкодером (обученной на синтетических данных с зашумлением методом FDA) и результата работы детерминистического метода *Thinned Fault Likelihood*. Применение синтетических примеров позволило отказаться от предварительной разметки данных и получить увеличение качества сегментации на 25–30 % по сравнению с детерминистическим методом.

1. Wu X. et all. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation // *Geophysics*. 2019. Vol. 84. No. 3. P. IM35–IM45.

Научный руководитель — канд. физ.-мат. наук А. А. Дучков

Распознавание рукописных слов на кириллице

Д. С. Кулаковский

Специализированный учебно-научный центр
Новосибирского государственного университета

Большинство работ в различных сферах общества поддерживается оборотом документов, которые могут быть как написанными от руки или распечатанными, так и электронными. Часто возникает необходимость перевода физических документов в цифровой формат, что помогает увеличить темп и качество работы соответствующих предприятий.

Основной задачей работы является создание системы распознавания рукописных слов с использованием нейронной сети. Данная система предназначена для перевода изображения со словами в текст с использованием модели нейронной сети, обученной на распознавание прописного текста на кириллице.

Модель нейронной сети сделана на основе архитектуры сверточной нейронной сети, рекуррентной нейронной сети. Сама же структура сверточной нейронной сети состоит из трех слоев свертки, которые распознают небольшое количество параметров; слоя активации, который использует нелинейное преобразование принимаемого значения, и слоя пулинга, который еще больше уплотняет матрицу признаков принимаемого в модель изображения, используя нелинейное преобразование, и выводит ряд значений признаков изображения. Для обучения был выбран датасет, который представляет собой около 63 000 рукописных предложений на русском языке. Он был разделен на тренировочную выборку — 15 % изображений, тестовую выборку — 15 % и контрольную выборку — 70 %. Нейронная сеть будет написана на языке Python с использованием фреймворков TensorFlow, OpenCV и др.

В ходе работы были получены следующие результаты: средняя точность распознавания букв в слове — 75 %. В дальнейшем планируется добавление интерфейса для взаимодействия с моделью.

Научные руководители — А. А. Мезенцева, С. П. Одяк

Обучение нейронной сети для построения критерия ассоциаций

С. И. Кушмуратов¹, Д. В. Поверин²

¹ Новосибирский государственный университет

² Новосибирский государственный технический университет

В полногеномном анализе ассоциаций изучают взаимосвязь между однонуклеотидными полиморфизмами и любыми фенотипическими признаками, как правило, заболеваниями, вызванными наследственными причинами. Для эксперимента берут две группы людей: больных и здоровых. В результате секвенирования ДНК для каждого однонуклеотидного полиморфизма получается набор данных, по которому можно проверить статистическую гипотезу о независимости группы, к которой относится человек, и его генотипа.

Для проверки гипотезы независимости используются критерии хи-квадрат, критерий отношения правдоподобий с ограничением (CLRT), критерии максимизации минимальной эффективности (MERT), критерий тренда Кокрена — Армитеджа (CATT), критерий MAX3, критерий выбора и исключения генетической модели (GMS и GME). Актуальной задачей является поиск критерия, обладающего более высокой мощностью в случае, когда истинная модель наследования неизвестна [1].

Одним из возможных путей построения более мощного критерия является использование методов машинного обучения, в частности нейронных сетей. В рамках данной работы было разработано программное обеспечение на языке программирования Python с использованием библиотек TensorFlow и Keras, проведено обучение нейронной сети на смоделированных данных, исследована мощность нового критерия, обоснована его состоятельность.

1. Postovalov S. Metge R. W. A power comparison of the association tests for genome-wide association studies // 11th Int. Forum Strategic Technol. (IFOST 2016). June, 1–3, 2016. Novosibirsk: NSTU, 2016. Pt. 1. P. 461–464.

Научный руководитель — д-р техн. наук С. Н. Постовалов

Семантическая сегментация изображений для тренажера Duckietown

К. С. Ланчуковская, Д. Е. Шабалина
Новосибирский государственный университет

Семантическая сегментация — это задача поиска групп пикселей, каждая из которых характеризует один смысловой объект, где требуется не только обнаружить и классифицировать объекты, но и определить их границы. Она применяется во многих сферах, в том числе и для алгоритмов управления беспилотным транспортом, которые активно развиваются и применяются в различных практических сферах, таких как индустриальная робототехника, автомобильная промышленность, медицина. В таких системах семантическая сегментация позволяет человеку точнее проанализировать спутниковые изображения и карты, рентгенографические и электронно-микроскопические снимки.

Активно используются различные тренажеры, что позволяет снизить риски при разработке беспилотных управляющих систем. Одним из таких тренажеров является Duckietown — открытый исследовательский проект в области автономно управляемых роботов, который обладает следующими особенностями: у роботов есть камера, но нет других датчиков, низкая мощность процессора, присутствуют искажения камеры. Все это накладывает ограничения на задачу семантической сегментации, которая на данный момент не решена в Duckietown.

Целью работы является разработка модуля, который будет использовать адаптированные алгоритмы семантической сегментации для тренажера Duckietown.

В работе были исследованы классические алгоритмы компьютерного зрения и алгоритмы, основанные на нейросетях, которые предназначены для семантической сегментации изображений. Из всех рассмотренных алгоритмов для дальнейших испытаний были отобраны нейросетевые алгоритмы, основанные на нейросетях SegNet, UNet и DeepLab. Проведены эксперименты, позволившие оценить применимость этих алгоритмов к задаче сегментации изображений на тренажере Duckietown.

Научный руководитель — Т. В. Лях

**Исследование свойств моделей, основанных на ошибке GE2E,
в задачах верификации говорящего**

А. Е. Легченко

Новосибирский государственный университет

Задача распознавания говорящего важна для многих систем, связанных с анализом речи. Например, голосовое управление, фильтрация речи, дикризация, биометрическая аутентификация и др. В задачах распознавания образов, и в частности распознавания речи, широкое применение нашел нейросетевой подход.

В данной работе рассматривается LSTM (*Long-short term memory network*) — модель для верификации говорящего по примеру голоса, среди известного набора дикторов. Для обучения использовалась ошибка GE2E [1] и общедоступный корпус речи VoxCeleb2 [2]. В качестве тестовых данных были выбраны крупные речевые корпуса LibriSpeech и VoxCeleb1 [2], содержащие около 1000 дикторов. Выбранные речевые корпуса имеют различную природу и качество речевых образцов, что помогает получить более полную информацию при тестировании.

Основной целью исследования было установить основные свойства модели в различных условиях и границы ее применимости. В рамках работы была исследована зависимость точности верификации от количества спикеров, объема и качества данных о дикторах, найдена необходимая длительность речевого образца для возможности верификации.

Полученные результаты дают информацию о возможностях данного подхода и требованиях к входным данным для достижения высокой точности верификации.

1. Wan L. et al. Generalized end-to-end loss for speaker verification // Int. Conf. Acoustics, Speech Signal Proc. (ICASSP). IEEE, 2018. P. 4879–4883.

2. Nagrani A., Chung J. S., Zisserman A. VoxCeleb: A Large-Scale Speaker Identification Dataset // Proc. Interspeech. 2017. P. 2616–2620.

Научный руководитель — канд. физ.-мат. наук, доц. Е. Н. Павловский

Поиск областей брюшной аорты, подверженных риску развития аневризматических трансформаций

А. Е. Лукьянов

Новосибирский государственный университет

Аневризма аорты является серьезным заболеванием сердечно-сосудистой системы, течение которого приводит к множеству неблагоприятных последствий для больного. Среди наиболее серьезных осложнений можно выделить развитие тромбоза, приводящего к ампутации органов или конечностей, и внутреннего кровотечения, приводящего к смерти. Ранняя диагностика заболевания позволяет избежать вышеобозначенных последствий [1]. В настоящее время диагностика заболевания осуществляется врачом с помощью анализа снимков компьютерной томографии. Актуальной задачей остается разработка автоматических и полуавтоматических средств диагностики.

Одной из подзадач разработки таких средств является выделение области интереса, т. е. поиск диапазона снимков компьютерной томографии, содержащего клинически значимые признаки для последующей автоматической диагностики заболевания.

Целью работы является построение алгоритма поиска диапазона снимков компьютерной томографии, содержащего клинически значимые признаки для диагностики аневризматических трансформаций брюшной аорты.

В рамках данной работы был обучен классификатор на основе методов машинного обучения, способный выделять требуемую область. Для реализации алгоритма использован язык Python 3, для работы с методами машинного обучения — библиотека PyTorch. В дальнейшем планируется интеграция реализованного алгоритма в разрабатываемый программный комплекс и использование в клинической практике.

1. Tomee S. M. et al. The consequences of real life practice of early abdominal aortic aneurysm repair: a cost-benefit analysis // Eur. J. Vasc. Endovasc. Surgery. 2017. Vol. 54. No. 1. P. 28–33.

Научный руководитель — Р. Ю. Епифанов

**Построение и исследование векторного пространства признаков
коммерческой недвижимости на основе открытых данных**

В. Ю. Мамедов

Новосибирский государственный университет

За 2019 г. в Москве открылось больше 500 новых кафе, баров и ресторанов. Их число увеличивается ежегодно: аналитики отмечают стабильный рост на 5–7 %. В связи со стабильным ростом спроса на аренду недвижимости в таких крупных городах, как Москва, поиск помещения для аренды становится все более трудной задачей. Место расположения напрямую влияет на прибыльность торговых точек бизнеса, являясь одним из основополагающих факторов для достижения рентабельности. Важно найти золотую середину между уровнем проходимости выбранного места, стоимостью аренды, соотношением спроса и предложения, наличием необходимой инфраструктуры, доступностью недвижимости с точки зрения имеющихся ресурсов и др.

В настоящей работе для исследования рынка коммерческой недвижимости выбран г. Москва. На основе открытых данных (OSM, ФНС, открытые данные Москвы) построен и исследован набор векторных пространств. Для построения были использованы алгоритмы машинного обучения и методы линейной алгебры.

Итогом данной работы является веб-приложение (интерактивная карта), с помощью которого можно найти соответствующие запросам заказчика потенциальные места для размещения бизнеса, а также оценить уровень спроса в конкретной сфере. Интерактивная карта реализована с помощью тайловой графики. В работе исследована и реализована предиктивная модель по оценке стоимости аренды в заданной точке.

Работа выполнена при поддержке компании ООО «Промсофт».

Научный руководитель — И. С. Хомич

Использование классических алгоритмов машинного обучения в предсказании цены бриллианта на основании его характеристик

В. Ю. Моховикова, А. М. Фомченко, М. Т. Шустик
Специализированный учебный научный центр
Новосибирского государственного университета

Машинное обучение — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Имеется множество объектов (ситуаций) и множество возможных ответов (откликов, реакций). Существует некоторая зависимость между ответами и объектами, но она неизвестна. Известна только конечная совокупность прецедентов — пар «объект, ответ», называемая обучающей выборкой. На основе этих данных требуется восстановить зависимость, т. е. построить алгоритм, способный для любого объекта выдать достаточно точный ответ. Для измерения точности модели используются определенные метрики.

Такой тип работы с данными очень востребован уже продолжительное время и подходит для достижения поставленной цели. Основная задача нашей работы — создать программу для предсказания цены бриллианта по его характеристикам с высокой точностью.

Для работы была выбрана платформа Google Colaboratory. Для обучения модели найден датасет на сайте [kaggle.com](https://www.kaggle.com), который содержит данные о 57 000 бриллиантов в виде таблицы с характеристиками бриллиантов и целевой переменной: ценой. Характеристиками являются: чистота, ширина, длина, высота, карат, огранка.

Первый алгоритм, реализованный для данной задачи, — линейная регрессия: классический алгоритм обучения с учителем. Для достижения максимально высокого результата были применены и другие алгоритмы, такие как методы *k*-ближайших соседей, лассо-регрессия. У каждого из них имеются свои отличительные особенности. Программирование моделей производилось на языке Python с использованием библиотек *scikit-learn*, *Pandas*, *matplotlib* и др.

В данный момент ведется работа над предоставлением возможности ввода данных конкретного бриллианта с клавиатуры и вывода предсказанной цены. Модель показывает точность около 90 %, что доказывает эффективность методов классического машинного обучения в решении подобных задач.

Научные руководители — А. А. Мезенцева, С. П. Одияк

Оценка скорости роста проростков пшеницы с помощью методов компьютерного зрения

А. С. Наздрюхин

Новосибирский государственный университет

Целью данной работы является разработка системы компьютерного зрения для отслеживания скорости роста и конечного размера проростков пшеницы.

В ИЦиГ СО РАН был поставлен эксперимент, в котором в прозрачном контейнере высаживались 4 группы растений сорта «Новосибирская 67»: 1) контрольная группа; 2) группа, обработанная гормоном — гиббереллиновой кислотой; 3) изогенная линия: мутант по признаку короткостебельности. Имеет один доминантный ген Rth-D1a. Произведена обработка гиббереллином; 4) изогенная линия: мутант по признаку короткостебельности. Имеет два доминантных гена: Rth-D1 и Rth-D1a. Произведена обработка гиббереллином. Процесс съемки фиксировался при помощи фотокамеры, которая осуществляет съемку контейнера каждые 5 мин на протяжении трех дней.

В результате выполнения данного протокола был получен набор данных, состоящий из 1000 изображений. Сто изображений из этого набора данных были размечены вручную с использованием программы LabelMe. На основании полученной разметки обучена нейронная сеть архитектуры Mask RCNN.

Трекинг проростков выполнялся с помощью алгоритмов классического компьютерного зрения. Каждый сегментированный проросток можно описать в виде маски. Тогда необходимо для каждой маски на i -м изображении найти наиболее похожую маску на $i + 1$ изображении. В качестве меры сходства масок была предложена эвристика, основывающаяся на симметрической разнице изображений. Ее можно описать следующим образом: каждая маска описывается множеством (p, x, y) , где p — значение пикселя: 0 или 1, а x, y — расположение пикселя на изображении. Тогда меру сходства масок A и B можно рассчитать по следующей формуле:

$$score = \frac{1}{\sum_{p \in A \Delta B} p}$$

Исходный код выложен в открытый доступ: <https://github.com/niobeus/seedings>

Научный руководитель — канд. биол. наук М. А. Генаев

Автоматическая оптимизация гиперпараметров рекомендательных систем с учетом их характерных особенностей

А. А. Нечаев

Вятский государственный университет, Киров

Рекомендательными системами (РС) называют программные средства, решающие задачу предсказания интереса пользователей к различным объектам в определенной предметной области. Большинство РС основаны на моделях, обученных на истории взаимодействий пользователей и объектов, а также их характеристиках. Проводимые научные исследования свидетельствуют, что качество рекомендаций сильно зависит от значений гиперпараметров используемых моделей. Поиск их оптимальных величин крайне ресурсоемок, а его результаты требуется актуализировать с учетом временного интервала и роста объема обрабатываемых данных.

Цель работы — повышение эффективности метода байесовской оптимизации гиперпараметров для рекомендательных систем на основе коллаборативной фильтрации. В классическом варианте байесовского метода целевая функция рассматривается как «черный ящик», что может оказаться неэффективным в условиях наличия дополнительной информации о ее характере. В рамках данной работы предлагается проанализировать влияние статистических выбросов и взаимозависимости гиперпараметров на результат обучения.

Модификации подвергаются следующие компоненты метода байесовской оптимизации. В функцию обнаружения точек вводятся специфичные для рекомендательных систем штрафы, снижающие приоритет слабо совместимых значений гиперпараметров. Используемая суррогатная модель целевой функции демонстрирует более узкие значения неопределенности при меньшем количестве статистических выбросов. Для расчета значений целевой функции используются урезанные наборы данных, исключаящие взаимодействия по наиболее ожидаемым паттернам.

Апробация перечисленных модификаций выполнялась на широко известных наборах данных для РС с использованием эталонных реализаций моделей и методов. Проведенные эксперименты показывают корректность разработанных алгоритмов и повышение эффективности автоматического поиска гиперпараметров рекомендательной системы в условиях ограниченных ресурсов.

Научный руководитель — канд. техн. наук, доц. В. Ю. Мельцов

Применение сверточных нейронных сетей в задаче сегментации медицинских изображений

А. А. Нечахина

Новосибирский государственный университет

В данной работе рассматривается задача автоматического распознавания острого инсульта на компьютерных томограммах мозга (КТ-снимках). Выявление нарушения кровоснабжения головного мозга на ранних этапах осложнено влиянием технического и человеческого факторов. Использование методов обработки данных, основанных на нейронных сетях, позволяет повысить качество диагностики. Целью работы являлось построение компьютерной модели, которая эффективно идентифицирует очаги поражения. В задаче исходные данные представлены набором КТ-снимков и сегментированных масок для каждого очага поражения.

Решение задачи состояло из нескольких этапов. Сначала был искусственно расширен набор данных с помощью процедуры *augmentation* [2, 3] для повышения обобщающей способности модели. После чего для построения модели была использована сверточная нейронная сеть с архитектурой U-net [1]. Данная архитектура используется для сегментации биомедицинских изображений и позволяет достигать хороших результатов на небольшом наборе данных. Также был использован метод перекрестной проверки (*cross-validation*), который помог получить более достоверную оценку производительности модели на тестовых данных.

1. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // Int. Conf. Medical Image Computing Computer-Assisted Intervention. October, 2015. P. 234–241.

2. Емельянов С. О., Иванова А. А., Швец Е. А., Николаев Д. П. Методы аугментации обучающих выборок в задачах классификации изображений // Сенсорные системы. 2018. № 3.

3. Андриянов Н. А., Андриянов Д. А. О важности аугментации данных при машинном обучении в задачах обработки изображений в условиях дефицита данных // Мясников В. В. Обработка изображений и дистанционное зондирование Земли. Самара, 2020. Т. 2.

Научный руководитель — канд. физ.-мат. наук, доц. В. М. Неделько

**Исследование по управлению биомедицинскими отходами
в больницах с использованием алгоритма когортного интеллекта**

Е. В. Першиков

Новосибирский государственный университет экономики и управления

Неправильное обращение в Управлении биомедицинскими отходами (BMW) не только опасно для окружающей среды, но также может повлиять на людей, работающих в больнице.

Это существенно важно для правильного обращения с отходами, образующимися в больницах, чтобы избежать проблем для здоровья и экологических рисков. BMW включает загрязнение почвы, воды и воздуха, поэтому необходимо правильно управлять биомедицинскими отходами, поскольку люди борются с инфекционными заболеваниями, такими как ВИЧ, гепатит и т. д. Управление больничными отходами включает управление многими видами деятельности, такими как сбор, транспортировка, обработка и удаление отходов.

В последние несколько лет многие исследователи работали над вопросами, связанными с биомедицинскими отходами, и они сосредоточились на сборе, разделении и обработке биомедицинских отходов в больницах с использованием контейнеров. Но также необходимо сосредоточиться на безопасной транспортировке биомедицинских отходов. В последние годы многие авторы работали над оптимизацией сети в сфере BMW.

В данном исследовании проблема управления BMW решается с помощью недавно появившегося когортного алгоритма. Алгоритм когортного интеллекта — это эволюционный алгоритм, который показывает положительные результаты в различных областях. Основная проблема руководства BMW — обеспечение безопасности людей при транспортировке отходов на свалку. Таким образом, необходимо сосредоточиться на проблеме обращения с отходами, используя методы оптимизации маршрута. Эти методы оптимизации используются для оптимизации расстояния между точкой сбора и площадкой для захоронения с помощью задач маршрутизации транспортных средств и методов оптимизации. Основными задачами, рассматриваемыми для этой проблемы, являются: определение возможных маршрутов, изучение риска, связанного со сбором и транспортировкой захоронения, и расчет оптимальной длины маршрута.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

Испытания эффективности работы нейросетевых архитектур TabNet и NODE на данных, представленных таблицами

Н. А. Радеев, А. Р. Королев
Новосибирский государственный университет

Нейронные сети в последние годы достигли впечатляющих результатов в обработке изображений, видеоданных, звуков и текстов. Но с данными, представленными в виде таблиц, дело обстоит иначе. Нейросетевые модели до сих пор не достигли на табличных данных качества работы, которое достигается классическими методами машинного обучения по типу деревьев решений или машины опорных векторов. Тем не менее в последнее время стали появляться разработки в сфере применения нейронных сетей к табличным данным. Одними из таких архитектур являются TabNet и NODE.

Целью нашей работы было проведение испытаний реализации архитектур TabNet и NODE на языке программирования Python 3 с использованием библиотеки PyTorch в решении задач бинарной и многоклассовой классификации на 39 наборах размеченных табличных данных. А также сравнение полученных результатов с результатами лучших на сегодняшний день AutoML-моделей для работы с данными, представленными в виде таблиц.

Измерения проводились с использованием открытого фреймворка для испытания моделей, решающих задачи бинарной и многоклассовой классификации, OpenML AutoML benchmark. Для бинарной классификации показателем качества была метрика roc-auc , а для многоклассовой — log loss .

По полученным результатам был проведен анализ, который показал эффективность нейросетевой архитектуры TabNet в решении задач бинарной и многоклассовой классификации и позволил выдвинуть гипотезы о том, в каких условиях использование архитектуры TabNet является оправданным на тех или иных табличных данных. NODE, по сравнению с TabNet, на большинстве заданий показал результаты лучше, однако NODE требует больше времени тренировки, чем TabNet. Примечательно, что NODE на некоторых датасетах показал лучший результат среди всех внедренных в AutoML benchmark моделей, что демонстрирует высокий потенциал этой модели.

Работа была начата в рамках зимней IT-школы CompTech@Nsk 2021.

Научные руководители — С. О. Деревянченко, А. Е. Морозов

Исследование модифицированной архитектуры LSTM для рекуррентных нейронных сетей

С. В. Савельев, К. Е. Первалов

Новосибирский государственный технический университет

Основное отличие рекуррентных нейронных сетей (RNN) — наличие связей, идущих от нейронов к собственным предыдущим значениям. Это позволяет создать эффект памяти в сети, благодаря которому RNN справляется с задачами, связанными с обработкой последовательностей. Однако обычная RNN плохо выявляет долговременные зависимости из-за проблемы «исчезающего градиента». Для решения этой проблемы, как и для повышения эффективности работы рекуррентных сетей в целом, было создано множество архитектур. Самая популярная из них — LSTM. Основная идея данной архитектуры — создание ячейки памяти, которая сможет долго хранить информацию и грамотно ей распоряжаться. LSTM имеет множество аналогов, разработанных, как правило, для упрощения обучения, но принципиальных отличий в архитектуре они не имеют, как и серьезного превосходства в эффективности в целом. Поэтому в работе многие структурные идеи будут взяты именно с LSTM.

Также среди наиболее удачных решений проблемы долгосрочной памяти есть архитектура SCRN. Она имеет более простое устройство в сравнении с LSTM, но по эффективности ничем не уступает, а иногда даже превосходит благодаря наличию дополнительной ячейки памяти в нейроне.

Основная цель нашей работы — разработка архитектуры рекуррентной ячейки на основе этих двух наиболее удачных решений. Комбинация из двух разных ячеек памяти позволит решать задачи с более сложными временными зависимостями, а модульность элементов архитектуры позволит настраивать нейроны на решение определенного типа задач.

Научный руководитель — канд. техн. наук И. М. Ступаков

**Концепция совместного использования
нейронных сетей и методов на основе опорных точек
при идентификации и распознавании изображений**

Т. В. Толстов, Н. А. Куприянов, Д. Ж. Утепов
Томский государственный университет
систем управления и радиоэлектроники

При осуществлении идентификации и распознавания изображений необходима высокая точность результата поиска необходимого изображения. Достичь высокой точности идентификации и распознавания можно с помощью одновременного использования методов на основе опорных точек и сверточных нейронных сетей.

Выделяют следующие достоинства использования нейронных сетей при идентификации изображений: так как нейронные сети относятся к задаче многомерной оптимизации, они способны самостоятельно подстраивать весовые коэффициенты на основе от входных параметров. Также нейронные сети обладают инвариантностью к изменениям масштаба и сдвигам изображения. Благодаря своей внутренней архитектуре нейронные сети способны воспринимать изображения панорамно, углубляясь в мелкие детали с каждым последующим слоем, следовательно, они способны осуществлять обобщение элементов изображения. Однако при использовании нейронных сетей выделяют и недостаток: сеть может допускать ошибки в случаях, если в изображении появится даже незначительный шум. На основе этого нейронной сети требуется более мягкая подстройка топологии либо передача на входной слой не исходного изображения, а производных данных, к примеру результата вычисления опорных точек.

Для поиска «особых точек» изображения используют методы на основе опорных точек, например масштабно-инвариантную трансформацию признаков. Среди преимуществ выделяют инвариантность к трансформации и смещению изображения. Появление незначительных шумов на изображении лишь незначительно влияет на качество поиска исходного изображения.

С учетом вышеуказанных преимуществ и недостатков алгоритмов для повышения качества реализации идентификации и распознавания изображений необходимо совместное использование нейронных сетей и методов на основе опорных точек.

Научный руководитель — И. К. Харченко

Применение методов машинного обучения для выявления аномалий с банковскими картами

В. В. Чернова
Московский авиационный институт

Решение задачи выявления мошенничества с кредитными картами на сегодняшний день является актуальным. Обычно при решении задачи обнаружения такого мошенничества применяют следующие типы методов: простые статистические методы, при которых фильтр нижних частот позволяет выявлять аномалии в простых случаях использования и подходы на основе машинного обучения, включая обнаружение аномалий на основе плотности, обнаружение аномалий с использованием методов кластеризации, а также алгоритм обнаружения аномалий в изолированном лесу и алгоритм обнаружения аномалий методом опорных векторов.

В интеллектуальном анализе данных обнаружение аномалий (или обнаружение выбросов) — это процедура идентификации редких элементов, событий или наблюдений, вызывающих подозрения, поскольку они значительно отличаются от большинства данных. Аномалии можно разделить на следующие общие категории: точечные, контекстные и коллективные.

В точечных аномалиях один экземпляр данных является аномальным, если он слишком далек от остальных. В контекстных аномалиях все зависит от контекста. Этот тип аномалии часто встречается в данных временных рядов. Коллективные аномалии представляют собой набор экземпляров данных, которые помогают выявлять аномалии в совокупности. Пример использования в бизнесе: кто-то неожиданно пытается скопировать данные с удаленного компьютера на локальный хост, что может быть отмечено как потенциальная кибератака.

Задача обнаружения мошенничества с кредитными картами включает сбор сведений о прошлых транзакциях по кредитным картам с предоставленными признаками мошеннических транзакций, предобработку набора данных и затем построение и оценку модели выявления мошеннических транзакций. Методы, используемые для решения поставленной задачи: алгоритм обнаружения аномалий в изолированном лесу, алгоритм обнаружения аномалий на основе плотности с использованием локального уровня выброса и метод опорных векторов.

Научный руководитель – доц. В. И. Виноградов

Аугментация данных для распознавания игральные кости маджонг

Е. А. Черногор

Иркутский государственный университет

Одна партия в ричи-маджонг состоит минимум из 8 игр для 4 игроков, где у каждого игрока 10 или 11 костей. Это значит, что в течение игры очки считаются минимум 32 раза. Вручную это делать сложно: учитывается большое количество костей и их комбинаций в соответствии с различными игровыми условиями, поэтому возникла задача автоматизации данного процесса.

В данный момент разрабатывается сервис для автоматического подсчета очков, где одной из его частей является обученная модель. Модель с помощью алгоритмов машинного обучения определяет состав игровой «руки» и сопоставляет его с правилами подсчета очков.

Задача сводится к обработке большого количества изображений с целью определения вида кости. Однако в свободном доступе отсутствуют сформированные датасеты. Разметка датасета является трудоемкой задачей, требующей как большого количества разновидностей костей, так и условий, при которых они сфотографированы. Для расширения обучающей выборки использовался подход с аугментацией данных на основе естественных изображений. Формировалась выборка фотографий, сделанных с минимальным количеством искажений, затем к изображениям применялись методы аугментации данных.

Аугментация данных — это методы, используемые для увеличения объема выборки на основе уже имеющейся, с помощью добавления измененных копий исходных данных или создания синтетических данных из исходных.

Для решения данной задачи использовался язык Python и библиотека OpenCV. Реализованы геометрические преобразования, изменения размера и поворот, блики, изменение яркости (интенсивности) изображений, добавлен шум и размытие.

В результате был сформирован датасет, позволяющий обучить модель. На данном этапе развития точность модели удовлетворяет задаче и не предполагает улучшения в ближайшем будущем.

Научный руководитель — А. В. Киселев

Итеративная адаптация к шуму квантования в нейронных сетях

Д. С. Чудаков

Новосибирский государственный университет

В настоящий момент нейронные сети повсеместно применяются в современных продуктовых решениях, начиная от простых классификаторов изображений и заканчивая комплексными системами, включающими классификацию, детекцию и сегментацию изображений.

Нейронные сети крайне требовательны к вычислительным ресурсам не только на этапе обучения, но и на этапе исполнения (*Inference*). В современных решениях необходимо использовать новые архитектуры, новые подходы к обучению и применять последующие оптимизации. Это позволяет использовать в реальных продуктах точные и легкие сети, для которых нет необходимости в дорогостоящем оборудовании, и даже запускать их на мобильных процессорах. Одним из подходов к оптимизации скорости и размера сети после обучения является квантование (*quantization*).

Существенная часть методов квантования нейронных сетей после обучения нацелена на уменьшение шума от квантования благодаря подбору оптимальных порогов. Данные методы страдают от большой просадки точности итоговой сети, особенно это заметно при использовании малого числа бит (6 и 4 бита). В первую очередь это связано с тем, что нет возможности достаточно уменьшить шум после дискретизации в случаях с низким числом бит.

Был предложен метод итеративной адаптации к шуму от квантования, который последовательно уменьшает шум от дискретизации, а затем обучает последующие слои сети работать с возникшим шумом, минимизируя среднеквадратичное отклонение между квантованной и исходной сетью. Данный метод позволил значительно улучшить точности при квантовании нейронных сетей в малое число бит, по сравнению с методами квантования после обучения, при этом метод затрачивает меньше времени, чем полное обучение квантованной сети с нуля (метод обучения с учетом квантования).

Научный руководитель — канд. физ.-мат. наук С. А. Алямкин

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ

УДК 004.422.81

Разработка прототипа облачной IDE для системы фрагментированного программирования LuNA

Е. А. Александров

Новосибирский государственный университет

В наше время существует много систем программирования, которые сложны в использовании. Каждый, кто желает познакомиться с новой технологией программирования, должен потратить достаточное количество времени, чтобы разобраться с ее использованием. Когда пользователь видит, какой объем работы ему нужно сделать, чтобы запустить самый простой тестовый пример, то это может стать препятствием к ее использованию. Высокий порог вхождения сужает круг пользователей, для которых доступна технология.

Облачная IDE снижает порог вхождения, берет на себя проблемы установки необходимых версий языка и компилятора. Создание облачной IDE — это комплексная проблема, включающая разработку архитектуры системы, составление требований, которые удовлетворяют потребностям пользователей, а также разработку удобного интерфейса. Одной из систем с такой проблемой является LuNA — система фрагментированного программирования, предназначенная для поддержки параллельной реализации больших численных моделей на суперкомпьютерах. Облачная IDE предоставляет удобный веб-интерфейс для LuNA, а ее архитектура может быть использована для проработки облачных IDE для аналогичных систем.

В результате работы была разработана облачная IDE для системы LuNA, которая состоит из клиентского веб-интерфейса на ReactJS, веб-сервера для запуска LuNA-программ.

В облачной IDE можно запускать, просматривать вывод и график времени исполнения программ.

В дальнейшем планируется внедрение существующего инструментария для работы с LuNA-программами.

Научный руководитель — В. А. Перепелкин

**Ускоренный алгоритм поиска соседей в системе атомов
для моделирования диффузии методом молекулярной динамики**

А. А. Баранов

Новосибирский государственный университет

Объектом нашего исследования являются алгоритмы и программы для моделирования процессов роста пространственно упорядоченных массивов квантовых точек на структурированной подложке. Структурированными называют подложки, поверхность которых содержит систему упорядоченных ямок, сформированных с помощью литографии.

Такие массивы перспективны для применения в различных электронных приборах: светодиодах матрицах, лазерах на квантовых точках, квантовых компьютерах и т. д. Однако процесс роста на структурированных подложках изучен не полностью. Так, до сих пор остаются открытыми вопросы о механизмах атомной диффузии, механизме зарождения квантовых точек и распределении потенциального рельефа.

Для моделирования этих процессов используется метод молекулярной динамики (МД), позволяющий детально изучать физику микроскопических явлений и рассчитывать различные параметры, которые задают их энергетику. При этом метод МД связан с машинной обработкой большого объема данных и требует использования эффективных алгоритмов.

Наиболее трудоемкой операций при моделировании является процесс поиска ближайших соседей для каждого атома системы. Эта процедура может занимать до 95 % времени работы всей программы. Поэтому именно оптимизация алгоритма поиска ближайших соседей даст максимальный прирост скорости работы всей программы.

Таким образом, цель работы заключается в разработке и реализации ускоренного алгоритма поиска соседних атомов, который будет использоваться в системе для моделирования процесса формирования пространственно упорядоченных квантовых точек.

В ходе работы был разработан ускоренный алгоритм поиска соседей со списком Верле. Также была предложена его параллельная реализация. Было проведено экспериментальное исследование его эффективности и масштабируемости на кластерных вычислительных системах СибГУТИ.

Научный руководитель — д-р техн. наук, доц. К. В. Павский

Разработка платформы для организации параллельных вычислений на множестве мобильных устройств

А. С. Борзов

Новосибирский государственный университет

Производительность смартфонов растет, а скорость передачи данных по сети увеличивается. Большую часть времени вычислительные ресурсы смартфонов простаивают. Имеется множество публикаций, посвященных организации распределенных вычислений на мобильных устройствах. Актуальной задачей является создание программной платформы для исследования алгоритмов организации распределенного решения задач в разных предметных областях и различных условиях применения. Цель работы — создать прототип такой платформы со специализацией на решении задач линейной алгебры. Следует учитывать особенности мобильных вычислителей: возможность внезапного выхода из участия в решении задач, заряд батареи, возможность перегрева устройства и др.

Предлагаемая программная платформа состоит из сервера и Андроид-приложения, управляющего организацией вычислений на смартфонах — вычислительных узлах распределенной системы. От пользователей системы через веб-интерфейс сервера поступают задачи и требования к виду и количеству устройств для их выполнения. Задача — список команд, включающий операции линейной алгебры, условные ветвления, циклы. Узлы подключаются к серверу, уведомляя его о готовности участвовать в расчетах. Когда набирается необходимое количество узлов, то им выдается задача и они приступают к ее совместному решению. При отказе узлов вычисления не прекращаются, а распределяются на оставшиеся устройства, за исключением случаев, когда продолжить вычисление без добавления нового устройства невозможно.

Сервер реализован на языке Java, фреймворк Spring Boot, паттерн MVC и СУБД PostgreSQL. Алгоритмы организации вычислений на узлах реализованы на языке C++, с использованием протоколов TCP, HTTP, а также библиотек BLAS, LAPACK. Для спецификации задач была написана библиотека на Kotlin.

Научный руководитель — М. А. Городничев

**Эффективная реализация алгоритма когерентного суммирования
для обработки данных сейсморазведки
и микросейсмического мониторинга**

М. О. Вершинин

Новосибирский государственный университет

Сейсморазведка — один из ведущих методов исследования земной коры для построения изображений геологической среды. Микросейсмический мониторинг позволяет оценить рост трещины гидроразрыва пласта — популярного метода стимулирования низкопроницаемых пород. Оба метода порождают большие объемы данных (исчисляются терабайтами). При обработке данных используют схожую процедуру когерентного суммирования — сложение записей разных приемников вдоль рассчитанных времен ожидаемого прихода сейсмических волн. Данная процедура обработки является вычислительно сложной, так что актуальной задачей является создание вычислительно эффективной реализации алгоритма когерентного суммирования и написание на его основе библиотеки для использования в приложениях.

Алгоритм когерентного суммирования ограничен скоростью передачи данных из памяти (*memory bound*). При обработке микросейсмических данных качество кэширования может сильно зависеть от входных параметров процедуры. При обработке сейсморазведочных данных требуется изначально непоследовательный доступ к памяти. Данная работа посвящена исследованию разных способов оптимизации алгоритма когерентного суммирования для этих процедур: распараллеливания, векторизации и эффективного доступа в память.

Для обработки микросейсмических данных блочное считывание данных оказалось наиболее эффективным из-за повышения их локальности: количество L1 промахов сократилось с 28 до 2,2 %, получилось расположить циклы так, что компилятор смог автоматически векторизовать программу. Для обработки сейсморазведочных данных наилучшим оказалась векторизация с использованием инструкций *gather*. В обеих процедурах эффективность распараллеливания программы была около 90 % на 6 ядрах. На основе лучших вариантов оптимизации была написана библиотека на языке C++ с распараллеливанием алгоритма с использованием OpenMP, с векторизацией программы для процессоров архитектуры x86 и x86-64 и возможностью импорта ее в Python.

Научный руководитель — канд. физ.-мат. наук, доц. А. А. Дучков

Разработка и реализация алгоритма распределения ресурсов фрагментированных программ на основе профилирования

А. С. Лямин

Новосибирский государственный университет

Методы автоматизации конструирования параллельных программ с развитием технологий постепенно совершенствуются. Создаются системы фрагментированного программирования, автоматизирующие данный процесс, уменьшающие затраты программистов на разработку соответствующих программ. Важной составляющей таких систем является модуль, ответственный за распределение фрагментов данных и вычислений по узлам вычислительной системы. От его эффективности напрямую зависит эффективность всей системы в целом. Существуют как статические алгоритмы, которые анализируют программу на этапе компиляции, так и динамические, которые анализируют нагрузку на узел в момент исполнения и балансируют ее.

Целью данной работы была разработка алгоритма управления распределенными данными фрагментированных программ и встраивание его в систему фрагментированного программирования LuNA. Основные требования к алгоритму:

- отсутствие глобальных коммуникаций между узлами;
- повышение производительности системы.

Был разработан и реализован алгоритм, который исполняет программу в профилированном режиме, после чего использует лог-файлы, полученные на выходе исполняемой программы. Лог-файлы содержат информацию о создании и перемещении фрагментов данных и вычислений. Все время исполнения программы делится на равные промежутки времени — кванты, размер которых задается параметром. Далее нагрузка рассчитывается так: для каждого узла подсчитывается количество исполняемых на нем фрагментов вычислений в каждый квант времени. С использованием этих данных рассчитывается дисбаланс и принимается решение о перемещении конкретных фрагментов вычислений с перегруженных узлов на недогруженные.

В ходе дальнейшей работы планируется улучшение алгоритма распределения нагрузки. Возможными направлениями для развития являются учет топологии кластера при создании распределения алгоритмом, а также создание и поддержка аннотаций в коде LuNA-программы, позволяющих вручную управлять распределением конкретных фрагментов вычислений.

Научный руководитель — В. А. Перепелкин

Разработка алгоритма распределения фрагментов данных в системе конструирования параллельных программ LuNA

Д. Е. Макаренко

Новосибирский государственный университет

Написание параллельных программ является трудоемким процессом, для уменьшения временных затрат создаются системы, автоматизирующие конструирование параллельных программ. Основной задачей для развития данных систем является повышение качества конструированных программ. Одним из основных показателей качества является распределение данных и вычислений по вычислительным узлам. Эффективность исполнения программы напрямую зависит от распределения ресурсов. Для распределения ресурсов используются как динамические алгоритмы, которые производят анализ в процессе выполнения программы и делают нагрузку на вычислительные узлы равномерной, так и статические, которые производят анализ в процессе компиляции программы.

Целью работы являлась разработка алгоритма распределения данных фрагментированных программ и внедрение его в систему конструирования параллельных программ LuNA.

Алгоритм должен удовлетворять основным требованиям, таким как отсутствие глобальных коммуникаций и обеспечение одинаковой загрузки вычислительных узлов.

На основании требований был предложен алгоритм распределения данных на основе анализа исходного кода на наличие рекомендаций для распределения данных по вычислительным узлам и анализа циклических операторов управления на межитерационную зависимость по данным. Данный алгоритм является статическим и выполняется на этапе компиляции программы. Предложенное решение было реализовано на языке Python.

Новый алгоритм успешно протестирован в системе LuNA. Получено увеличение эффективности по времени по сравнению с непараллельной версией программы.

Научный руководитель — В. А. Перепелкин

Разработка и реализация частных алгоритмов управления распределенными данными в системе LuNA

М. А. Масыч

Новосибирский государственный университет
Институт вычислительной математики
и математической геофизики СО РАН, Новосибирск

Важной и трудоемкой частью разработки параллельных программ является управление данными. В распределенных вычислительных системах обеспечение доступности и целостности данных, поддержка сбалансированного распределения данных по процессорным элементам и приемлемой скорости доступа к ним — нетривиальные задачи.

В общем случае для задачи управления распределенными данными не существует универсального алгоритма. Поэтому есть множество различных решений, которые показывают себя лучше или хуже на тех или иных классах задач. Система фрагментированного программирования LuNA также сталкивается с этой проблемой.

В системе LuNA реализован только один алгоритм управления распределенными данными, который не способен обеспечить высокую эффективность в многочисленных задачах численного моделирования.

Таким образом, использование одного алгоритма управления распределенными данными внутри системы LuNA не является целесообразным. Требуется разработать набор частных алгоритмов, которые были эффективными на широком круге задач.

Целью работы является разработка и реализация набора частных алгоритмов управления распределенными данными, использование которых повысит эффективность генерируемых LuNA-программ.

Для достижения данной цели было проведено исследование приложений, решающих распространенные задачи численного моделирования. На основании проведенного исследования, а также специфики системы LuNA были сформированы требования к алгоритмам и их реализации. В соответствии с требованиями были предложены несколько частных алгоритмов, повышающих эффективность генерируемых программ на практически значимом круге задач. Алгоритмы были реализованы в компиляторе системы LuNA. Было проведено тестирование эффективности программ, использующих разработанные алгоритмы.

Разработанные алгоритмы в составе компилятора системы LuNA повышают качество конструируемых программ.

Научный руководитель — В. А. Перепелкин

**Программные средства доступа к ассоциативной базе данных
параллельных программ в диспетчере ресурсов
с открытым исходным кодом SLURM в стандарте PMIx**

Д. Д. Панфилов

Новосибирский государственный университет
Институт физики полупроводников
им. А. В. Ржанова СО РАН, Новосибирск

Одна из реализаций стека программного обеспечения для выполнения параллельных заданий в стандарте MPI включает библиотеки стандарта PMI и менеджер ресурсов SLURM.

SLURM является одним из популярных менеджеров ресурсов для высокопроизводительных систем. Около 60 % суперкомпьютеров из списка Top500 используют SLURM.

SLURM предоставляет доступ к ресурсам, распределяет их между задачами, обеспечивает структуру для запуска, выполнения и мониторинга параллельных программ на вычислительных узлах.

Существуют различные способы запуска параллельных программ с помощью SLURM, например: SLURM выделяет ресурсы для задачи, а затем ее запускает `mpirun`. В работе рассматривается другой способ: SLURM запускает задачи и выполняет инициализацию коммуникаций через интерфейс управления процессами (*Process Management Interface*, PMI). Одна из версий стандарта PMI называется PMIx (*PMI Exascale*).

PMIx — это стандарт интерфейса прикладного программирования (API), обеспечивающий библиотеки и модели программирования высокопроизводительных вычислений с четко определенным доступом к общедоступным сервисам распределенных вычислительных систем.

Потенциальные возможности SLURM в стандарте PMIx позволяют запускать параллельные программы без использования библиотек MPI. Одно из условий для этого является создание функций `publish` и `lookup` доступа к памяти, содержащей информацию пользовательских процессов. В настоящий момент SLURM не поддерживает вызовы этих функций.

В работе рассматриваются и предлагаются решения для доступа к ассоциативной базе данных в стандарте PMIx при использовании менеджера ресурсов SLURM.

Научный руководитель — д-р техн. наук, доц. К. В. Павский

Разработка алгоритмов распределенной сборки мусора в системе фрагментированного программирования LuNA

А. В. Плешков

Новосибирский государственный университет

Программные системы, в которых используется динамически выделяемая память, нуждаются в ее освобождении. Ручное управление памятью, зачастую приводящее к ошибкам в системах без автоматической сборки мусора, существенно усложняется при переходе к распределенной вычислительной системе, поскольку в таких условиях один объект может использоваться совместно несколькими вычислительными узлами. Таким образом, создание автоматического сборщика мусора позволит упростить процесс написания программ, при этом повышая их надежность и эффективность.

Одной из особенностей системы LuNA является отсутствие ссылок между фрагментами данных, что позволяет частично решить проблему во время компиляции программы. Динамическая распределенная сборка мусора, в свою очередь, невозможна без накладных расходов на коммуникации между узлами и runtime-систему. Поэтому было решено поставить целью данной работы реализацию модуля для компилятора, основной задачей которого будет анализ кода на языке фрагментированного программирования LuNA и автоматическое проставление директив к удалению объектов для системы исполнения с помощью механизма отслеживания жизненного цикла объекта на этапе компиляции.

Был реализован модуль статистического анализа, подключающийся в компилятор и преобразующий код программы в древовидную структуру. Такая организация данных позволяет хранить информацию о количестве потреблений каждого фрагмента данных и отслеживать зависимости по данным. Обход итогового дерева дает возможность построить формулу, по которой во время выполнения программы можно будет определить момент или число обращений, после которого можно корректно удалить фрагмент данных.

В дальнейшем существует два основных направления развития системы сборки мусора: создание динамического сборщика мусора в качестве резервного механизма для более эффективной работы программ и добавление новых директив компилятора для тех случаев, когда во время компиляции невозможно определить жизненный цикл фрагмента данных.

Научный руководитель — В. А. Перепелкин

Вычисление надежности случайного графа с использованием библиотеки MPI

К. С. Сергеев

Новосибирский государственный университет

В процессе эксплуатации сети элементы сетей могут подвергаться отказам по причине износа, поломки или же некоторого внешнего воздействия. Такие сети моделируют случайными графами.

Случайный граф как математическая модель сети с ненадежными элементами представляет собой следующее: G — неориентированный граф, описывающий топологию сети; V — множество вершин, соответствующее узлам сети; E — множество ребер, соответствующее множеству каналов связи сети, которые подвержены случайным независимым отказам. Примерами таких сетей могут служить транспортные, телекоммуникационные, телефонные сети.

Под надежностью сети в работе будем понимать вероятность связности соответствующего графа, описывающего сеть. Расчет вероятности связности такой сети является NP-трудной задачей. Расчет вероятности по определению приводит к перебору 2^m вариантов, где m — это число ребер графа. Для вычисления вероятности связности $R(G)$ используется подход, называемый факторизацией: для каждого ребра e графа G вычисляем $R(G)$ как сумму вероятностей $R(G_1)$ и $R(G_2)$, где граф G_1 получен удалением ребра e , а G_2 — стягиванием ребра e .

В работе предложены методы параллельной реализации факторизации на высокопроизводительных кластерах с распределенной памятью, различными схемами общения процессов и оптимизациями для ускорения вычисления с использованием библиотеки MPI.

Научный руководитель — канд. физ.-мат. наук Д. А. Мигов

**Параллельный генетический алгоритм размещения стоков
для максимизации надежности сети с ненадежными узлами**

А. А. Тарасов

Новосибирский государственный университет

Одной из главных задач проектирования и оптимизации сетей, элементы которых подвержены случайным отказам, является получение наиболее надежных вариантов топологий. В качестве надежности в данной работе рассматривается показатель MENC, отказам при этом подвержены узлы сети. Для поиска оптимального решения предлагается метод размещения стоков сети, основанный на использовании генетических алгоритмов. Под оптимальным решением понимается такое размещение стоков, которое максимизирует надежность функционирования сети.

Как правило, структура таких сетей представлена в виде неориентированного графа $G = (V, E)$, где V — множество вершин (узлов сети), E — множество ребер (каналов связи). Для каждой вершины задана вероятность ее присутствия в графе, что соответствует надежности узла. K — выделенное множество вершин, которые соответствуют стокам сети. Предполагается, что K содержит хотя бы один элемент.

В классическом генетическом алгоритме задача формализуется таким образом, чтобы ее решение могло быть закодировано в виде вектора генов, в данной работе будем использовать битовые строки, каждый бит кодирует наличие стока в узле. Случайным образом создается множество генотипов исходной популяции и рассчитывается значение показателя надежности каждой особи, что соответствует функции пригодности. Из полученного множества решений с помощью оператора «селекции» выбираются наиболее подходящие особи, к которым применяются операторы скрещивания и мутации, результатом чего является получение новых особей. Для них вычисляется значение показателя надежности, после производится отбор наилучших решений в следующее поколение. Процесс происходит итеративно, пока не будет выполнен один из критериев остановки.

В работе предложен параллельный генетический алгоритм оптимизации сети, реализованный на C++ с использованием библиотеки MPI.

Научный руководитель — канд. физ.-мат. наук Д. А. Мигов

Программные решения доступа к информации о среде исполнения в стандарте PMIx

Г. Н. Третьяков

Новосибирский государственный университет
Институт физики полупроводников
им. А. В. Ржанова СО РАН, Новосибирск

В настоящее время при построении суперкомпьютеров популярен SLURM (*Simple Linux Utility for Resource Management*) — планировщик заданий для больших кластеров вычислительных узлов с открытым исходным кодом, который поддерживает очередь ожидающих заданий и управляет общей загрузкой ресурсов в процессе выполнения работы. По данным разработчиков, этот менеджер ресурсов используется примерно в 60 % производительных систем из списка Top500.

Архитектура управления SLURM включает нижеследующее. Верхний уровень управления — это резервированная пара контроллеров кластера. Контроллеры управляют вычислительным кластером и содержат демон управления *slurmctld*, который следит за вычислительными ресурсами, а также демон *slurmd*, который управляет запущенным узлом. Внутри архитектуры также существуют и другие демоны, например для безопасной аутентификации.

Slurm запускает задачи и выполняет инициализацию обмена данными через интерфейс управления процессами (*Process Management Interface*, PMI), PMI Exascale (PMIx) [1]. PMIx является стандартизированной версией PMI, разработанной для высокопроизводительных систем.

PMIx обеспечивает общения между различными участниками, ретрансляции запросов и возврате ответов. Приложение построено на основе клиентской библиотеки PMIx, которая содержит клиентские API, определения атрибутов и поддержку связи для взаимодействия с локальным сервером PMIx. Внутрипроцессное межбиблиотечное взаимодействие поддерживается на уровне клиента во избежание ненужной нагрузки на сервер. Текущая реализация PMIx с использованием Slurm не поддерживает доступ к базе данных о процессах, хранящейся на главном сервере.

В работе предлагаются и рассматриваются решения для доступа к информации о среде исполнения в стандарте PMIx.

1. PMIx Consortium. 2017–2019. PMIx Reference Library. [Электронный ресурс]. URL: <https://github.com/pmix/pmix> (дата обращения: 18.02.2021).

ПРОГРАММНАЯ АРХИТЕКТУРА И СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ

УДК 004.4'42

Разработка межверсионного конвертера для программ на языке Reflex 1.0

А. А. Абраменко

Институт автоматики и электрометрии СО РАН, Новосибирск

Использование языков программирования общего назначения для выполнения задач цифровых систем управления приводит к росту сложности алгоритмов и программной архитектуры, увеличению стоимости разработки и поддержки подобных систем. Для решения этой проблемы в Институте автоматики и электрометрии СО РАН был разработан процесс-ориентированный язык программирования Reflex. Базовые цели, которые ставились при разработке языка, — это его адекватность задачам управления, легкое изучение пользователем, комфортное программирование и сопровождение уже созданных программ.

С момента создания язык Reflex претерпел ряд серьезных изменений синтаксиса, что привело к появлению его новой версии, несовместимой с предыдущей. По этой причине возникла задача перевода уже существующих программ на актуальную версию Reflex 2.0. Вследствие большого объема программ ручное переписывание кода, во-первых, трудоемко, а во-вторых, сопряжено с вероятностью внесения ошибок.

Цель работы — разработка транспайлера для программ, написанных на языке Reflex 1.0 в программы на Reflex 2.0.

Для достижения цели были проанализированы специфика языка Reflex и средства автоматической парсер-генерации, сформирован список требований к разрабатываемой системе, спроектирована архитектура транспайлера и реализованы необходимые модули. Работоспособность создаваемого транспайлера планируется проверить на тестовой задаче.

В качестве вспомогательного средства парсер-генерации был выбран Xtext — среда разработки DSL на основе платформы Eclipse. Использование инструментов автоматической парсер-генерации при разработке системы облегчает изменение транспайлера в случае обновления или расширения синтаксиса языка.

Научный руководитель — д-р техн. наук В. Е. Зюбин

**Актуальность технологий цифровой экономики
для развития умных городов в Республике Узбекистан**

Г. Б. Бахтиярова
Ургенчский филиал Ташкентского университета
информационных технологий

В настоящее время наблюдается активное развитие многочисленных проектов, направленных на развитие цифровой экономики в Республике Узбекистан. Это, в свою очередь, открывает широкие возможности для проведения исследований, поиска самых оптимальных возможностей реализации цифровых проектов.

Одним из наиболее популярных и востребованных проектов цифрового развития в стране являются умные города (*Smart Cities*), умные машины (*Smart Cars*) и даже умные дома (*Smart Houses*). Каждый из вышеупомянутых умных проектов подразумевает цифровизацию услуг и переход на софтверную площадку, где каждый гражданин или житель может воспользоваться государственными (и не только) услугами с помощью комплексных программных платформ, спектр задач которых расширяется с каждым днем.

В работе исследуются подходы в разработке сервисов для умных проектов, в частности на примере умного города. В рамках работы было проведено исследование, в котором изучены следующие [1, 2] научные труды, в которых очень четко отражены проблемы применения сервис-ориентированных архитектур для развития умных городов, а также технологий цифровой экономики в целом. Вместе с тем в рамках данной научной работы проведен сравнительный анализ актуальных концепций по цифровизации городов «под ключ» с применением моделей разработки программных продуктов, используемых на сегодняшний день.

1. Dilshodbek Kuryazov, Bekmurod Khujamuratov, Khursand Sherkhanov. Sustainable Service-Oriented Architecture for Smart City Development // Int. Conf. Inf. Sci. Commun. Technol. (ICISCT), 4–6 Nov. 2019.

2. Кононова О. В., Павловская М. А. Технологии цифровой экономики в проектах умный город: участники и перспективы // Совр. информ. технологии и ИТ-образование. 2018. Т. 14, № 3.

Научный руководитель — канд. техн. наук, доц. Т. В. Ярославцева

Бесшовная интеграция языка роST в среду CoDeSys

В. И. Башев

Новосибирский государственный университет
Институт автоматике и электрометрии СО РАН, Новосибирск

В Институте автоматике и электрометрии СО РАН ведется активное исследование, разработка и апробация процесс-ориентированной парадигмы программирования для систем промышленной автоматизации. В рамках парадигмы создан язык роST (*process-oriented Structured Text*) [1], процесс-ориентированное расширение языка ST (*Structured Text*) из семейства языков IEC 61131-3. Для языка роST уже создан парсер, плагин для Eclipse IDE и проблемно-ориентированные модули кодогенерации в язык C и язык ST, что, в принципе, позволяет использовать роST в специализированных инструментальных средствах на базе IEC 61131-3. Однако это предполагает ручное создание проекта, копирование кода и описание конфигурации, т. е. множество рутинных операций.

Целью работы ставилась разработка метода автоматизированной интеграции роST-программ в средства IEC 61131-3.

Было предложено осуществлять интеграцию через формат XML Exchange, разработанный членами ассоциации PLCopen для обеспечения переносимости IEC 61131-3 программ. Программа на языке роST преобразуется в программируемый организационный элемент (тег <rou>), в котором используемые переменные декларируются посредством тега <interface>, а собственно сгенерированный ST-код оформляется через тег <body>.

Предложенный подход был опробован в среде CoDeSys на тестовых задачах «Сушилка для рук» и «Лифт». В ходе работы было выявлено, что использование XML Exchange все же предполагает создание проекта и ручные операции по привязке программы к элементам визуализации в случае ее использования, но при этом значительно упрощает интеграцию кода в существующий проект.

Использование разработанного проблемно-ориентированного модуля для языка роST упрощает взаимодействие данного языка со специализированными средствами IEC 61131-3.

1. Bashev V., Anureev I., Zyubin V. The Post Language: Process-Oriented Extension for IEC 61131-3 Structured Text // Int. Russ. Autom. Conf. (RusAutoCon). Sochi, 2020. P. 994–999.

Научный руководитель — д-р техн. наук В. Е. Зюбин

Файловый проводник

Е. К. Маковецкий

Городской Дворец детского (юношеского) творчества, Омск

Каждому пользователю компьютера приходится работать с файлами и папками, это можно осуществить с помощью файлового проводника.

Цель работы — создание файлового проводника с более расширенным функционалом по сравнению со стандартным проводником Windows.

В проекте реализованы стандартные действия над файлами и папками. Кроме того, есть возможность пометить файл как основной. В таком случае он всегда будет первым в списке (если помеченных файлов много, то они сортируются по выбранной пользователем сортировке).

Для удобства просмотра файлов реализована «Группировка», создающая группу из файлов одного типа, если их количество будет велико. Если количество файлов одного типа превышает установленное в настройках (по умолчанию это 8), то, например, 100 кнопок с файлами типа png превращаются в 1 кнопку-группировку, которую можно раскрыть и закрыть. Эта функция делает поиск нужного файла более быстрым (если забыто точное имя файла).

Имеются функции «Назад», «Вперед», «Вверх», «Обновить». Пользователь может скопировать путь, посмотреть скрытые файлы, воспользоваться вкладками и др.

Программа написана на языке Java.

В целях увеличения быстродействия программы и уменьшения длительности операций была добавлена многопоточность, которая реализована созданием классов, расширяющих Thread. В каких-то частях кода создаются объекты этих потоков, запускаются и выполняют какую-то функцию. Например, в отдельный поток был занесен подсчет количества файлов и папок в какой-либо папке. Получившиеся значения вносятся на панель информации о папке. Также в отдельный поток занесено копирование файлов и папок, движение текста у надписей, если текст слишком большой, и некоторые другие действия в программе.

В результате работы был создан файловый проводник. Отличия данного приложения от стандартного проводника Windows:

- совершенно другой интерфейс и дизайн;
- возможность пометить файл как основной;
- группировка из файлов одного типа;
- более гибкие настройки.

Исследование алгоритмов планирования задач для целей процесс-ориентированного программирования

Д. А. Пермяшкин

Институт автоматики и электрометрии СО РАН, Новосибирск

В современном мире продолжается давно начавшийся процесс автоматизации промышленного производства. Все большее число технологических процессов переводится на исполнение конвейерам, появляется все больше процессов, которые выполнимы только при помощи автоматических станков и других устройств. Один из подходов к созданию управляющих программ в области промышленной автоматизации — процесс-ориентированное программирование. Подход предполагает описание алгоритма управления в виде множества параллельно исполняемых процессов. Парадигма показывает очень хорошие результаты на практике и может рассматриваться как наиболее перспективный подход для программирования киберфизических систем широкого класса.

К сожалению, использование процесс-ориентированного подхода оставляет открытым вопрос о том, каким методом можно создать корректное расписание задач (и вообще проверить, существует ли оно). Чаще всего в данной предметной области используется либо гипотеза о идеальном синхронизме, либо созданное как-то в ручном порядке расписание.

Работы, исследующие автоматическое составление расписаний, фокусируются в основном на планировании задач в операционных системах. И поскольку в операционных системах набор текущих задач постоянно и непредсказуемо изменяется, то эти алгоритмы являются динамическими, т. е. требуют дополнительных вычислительных ресурсов.

При процесс-ориентированном программировании ситуация другая: набор исполняемых процессов задается статически и заранее известен. Таким образом, в контексте процесс-ориентированного программирования становится возможным использовать менее ресурсоемкие алгоритмы статического планирования.

В данной работе рассматриваются статические модификации алгоритмов планирования *earliest deadline first* и *rate monotonic* с учетом особенностей процесс-ориентированной парадигмы программирования, а также обсуждается алгоритм планирования на основе концепции «сцепленных окон».

Научный руководитель — д-р техн. наук, доц. В. Е. Зюбин

ПРОГРАММНАЯ ИНЖЕНЕРИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

УДК 004.032.26

Voice input/output system for an automatic technical support system

S. A. Berezin
Novosibirsk State University

Chatbots are already quite firmly established in our life: we meet them on the websites of online stores, in the applications of banks and mobile operators. However, communication using text is not always convenient and acceptable for the user: for example, at the legislative level, the driver of a car is prohibited from being distracted by typing while driving.

The solution to this problem can be the use of a system of voice interaction with the program: instead of being distracted by typing and reading text, the user can make a request to the program and hear the answer. To implement such a solution, we applied the Jarvis ASR and Jarvis STT speech recognition systems from Nvidia, based on the NeMo toolkit [1]. The choice of these particular architectures was due to the relatively high quality of their work and the uniformity of interfaces — working out the interaction with solutions from one manufacturer turned out to be much easier than setting up the joint work of disparate parts.

The result of the work was a prototype of a software product, which is currently at the implementation stage.

I. Oleksii Kuchaiev, Jason Li, Huyen Nguyen, Oleksii Hrinchuk, Ryan Leary, Boris Ginsburg, Samuel Krizan, Stanislav Beliaev, Vitaly Lavrukhin, Jack Cook, Patrice Castonguay, Mariya Popova, Jocelyn Huang, and Jonathan M. Cohen. NeMo: a toolkit for building AI applications using Neural Modules. 2019.

Scientific Supervisor — Cand. Phys.-Math. T. V. Batura

Анализ тональности и распознавание эмоций для создания интеллектуального помощника

А. Н. Акатов

Новосибирский государственный университет

Интеллектуальные виртуальные помощники обретают все большую популярность. Важным направлением в этой сфере является определение потребностей пользователя. Распознавание эмоций и настроения человека позволяет понять, насколько точно была определена его потребность и насколько он удовлетворен работой бота.

Данная работа является частью проекта по созданию интеллектуального помощника, который может быть особенно полезен слабовидящим людям, и посвящена поиску решений распознавания эмоций. Объектом исследований являются методы анализа тональности текста. В работе рассмотрены классические подходы на основе правил, словарей и методов машинного обучения с использованием предобученных языковых моделей. Последние, согласно недавним исследованиям, позволяют эффективно решать задачи классификации эмоций, уверенно добиваясь хороших результатов.

Целью данной работы является разработка модуля для интеллектуального помощника, способного анализировать тональность и распознавать эмоции пользователя.

Для достижения этой цели были решены следующие задачи: изучены существующие алгоритмы распознавания эмоций, проведены сравнения этих алгоритмов, выбран и реализован наиболее подходящий и рассмотрены способы улучшения результатов.

В дальнейшей работе планируется усовершенствовать выбранный подход, а также рассмотреть возможность распознавания эмоций в голосовой записи запроса пользователя.

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Долгушева Е. В., Пальчунов Д. Е. Теоретико-модельные методы порождения знаний о предпочтениях абонентов мобильных сетей // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 2. С. 5–16.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Разработка автоматизированного анализа шифров на алгебраическую криптоустойчивость

Н. Д. Атутова^{1,2}, Д. А. Зюбина^{1,2}, С. Д. Филиппов³

¹ Новосибирский государственный университет

² Лаборатория криптографии JetBrains Research

³ Санкт-Петербургский государственный университет

В настоящее время защита информации обеспечивается с помощью шифрования данных. Для достаточной надежности необходима высокая стойкость к статистическим и аналитическим методам криптоанализа шифра. Основная идея алгебраического криптоанализа состоит в составлении сложной системы булевых уравнений, описывающих преобразование шифра, и нахождении решений данной системы, соответствующих секретному ключу.

Целью работы является реализация автоматического построения систем уравнений для анализа и получение оценок стойкости шифров к алгебраическим атакам. В рамках данной работы булевы уравнения представляются в форме АНФ: полиномы, в которых используются только операции сложения и умножения по модулю 2, а также константы 0 и 1 [1].

Рассмотрены Simon и Speck — шифры, имеющие структуры LRX- и ARX-шифров, представленные АНБ США в 2013 г. Реализованы алгоритмы построения систем уравнений в форме АНФ для рассматриваемых шифров. Полученные уравнения использованы для реализации методов решений систем и получения оценок стойкости: SAT-решатель, XL-метод, ElimLin, алгоритм Бухбергера. В результате было выяснено, что наибольшую эффективность показали атаки с помощью SAT-решателя.

*Работа выполнена при поддержке
лаборатории криптографии JetBrains Research.*

1. Токарева Н. Н. Симметричная криптография. Краткий курс: учеб. пособие. Новосибирск: НГУ, 2012.

Научные руководители —
канд. физ.-мат. наук Н. Н. Токарева, А. В. Куценко

Разработка веб-сервиса тематического структурирования базы прецедентов компьютерных атак

Ж. С. Бадмажапов

Новосибирский государственный университет

Для противодействия прецедентам компьютерной безопасности разрабатываются различные ПО, одним из главных компонентов таких ПО являются их собственные базы прецедентов, которые не находятся в общем доступе. Но зачастую не каждую проблему безопасности могут решить специализированные ПО.

При таких обстоятельствах возникает необходимость в информационных ресурсах, которые могут предоставить базу прецедентов для различных пользователей [1–3]. Специалист по информационной безопасности может самостоятельно попытаться решить проблемы на основе описания определенного прецедента, который может быть схож по симптоматике или другим атрибутам.

Целью данной работы является разработка веб-сервиса, предоставляющего доступ к прецедентам компьютерных атак. В ходе выполнения данной работы был предложен метод структуризации данных на основе иерархической кластеризации по темам с собственной функцией метрики. В качестве метода поиска ключевых слов был предложен алгоритм латентного размещения Дирихле.

Предполагаемый функционал веб-сервиса включает следующие пункты: доступ к прецедентам компьютерных атак; поиск прецедентов по атрибутам и ключевым словам; возможность добавления новых прецедентов; вывод иерархии прецедентов, включающих искомые ключевые слова.

1. Яхьяева Г. Э., Ясинская О. В. Методы согласования знаний по компьютерной безопасности, извлеченных из различных документов // Вестн. НГУ. Сер. информ. технологии. 2013. Т. 11, Вып. 3. С. 63–73.

2. Яхьяева Г. Э. и др. Вопросно-ответная система для управления информационными рисками на основе теоретико-модельной формализации предметных областей // Информ. технологии. 2017. Т. 23, № 2. С. 97–106.

3. Головин Н. А. и др. Применение методов машинного обучения для структурирования базы прецедентов компьютерных атак // Материалы Междунар. конф. «Знания — Онтологии — Теории» (ЗОНТ-2019), Новосибирск, 2019. С. 122–128.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Моделирование цифровых двойников ролей на основе d0sl

А. И. Ваганова

Новосибирский государственный университет

Цифровой двойник (ЦД) — это динамическое виртуальное представление физического объекта или системы в течение всего жизненного цикла с использованием данных в режиме реального времени.

Цифровые двойники бизнес-процессов используются для разработки методик управления и производства, тестирования процессов при изменении условий. Для создания таких ЦД необходимо описать знания о предметной области, произвести моделирование регламентов и ролей, участвующих в процессах, используя семантические предметно-ориентированные языки (sDSL, *semantic domain-specific languages*) [1].

sDSL позволяют декларативно описывать знания о предметной области и имплементировать эти знания в исполняемый код. При их использовании упрощаются методы создания спецификаций, процесс заполнения баз знаний, создание систем для извлечения и обработки знаний о предметных областях [2, 3].

Язык d0sl реализует идеи sDSL на практике и позволяет создать DSL (*domain-specific languages*) для моделирования бизнес-процессов и разработки цифровых двойников ролей [1].

Целью настоящей работы является создание цифрового двойника роли в бизнес-процессе. Для этого необходимо определить знания и компетенции оригинала, права доступа к данным из информационных систем и права на принятие решений для управления бизнес-процессами. Исходя из этого, строится цифровая модель роли, которая описывается на предметно-ориентированном языке d0sl. Получившаяся спецификация на языке d0sl далее транслируется в исполняемый код.

1. Gumirov V. S., Matyukov P. Y., Palchunov D. E. *Semantic Domain-specific Languages // Int. Multi-Conf. Eng., Comput. Inf. Sci. (SIBIRCON)*. Novosibirsk: IEEE Press, 2019. P. 0955–0960.

2. Долгушева Е. В., Пальчунов Д. Е. Теоретико-модельные методы порождения знаний о предпочтениях абонентов мобильных сетей // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 2. С. 5–16.

3. Palchunov D., Yakhyaeva G., Yasinskaya O. *Software system for the diagnosis of the spine diseases using case-based reasoning // Proc. SIBIRCON/SibMedInfo-2015*. Novosibirsk, 2015. P. 205–210.

Научный руководитель — д-р физ.-мат. наук, доц. Д. Е. Пальчунов

Анализ и обработка базы данных репрессированных

А. Д. Васильев

Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ

Советская антирелигиозная политика 1920–1930-х гг. была неотъемлемой частью создания социалистического общества. За период «большого террора» в БМАССР было подвергнуто репрессиям около 20 тыс. граждан республики. С 1991 г. в России идет процесс реабилитации жертв политических репрессий.

Сотрудниками Бурятского государственного университета разработан ресурс «Книга памяти жертв политических репрессий в Республике Бурятия». База данных ресурса содержит более 14 тыс. записей о реабилитированных гражданах [1].

Данные в базе являются полной копией дел из архива Федеральной службы безопасности по Республике Бурятия. Вследствие общей малограмотности в 20–30-е гг. дела содержат большое число ошибок, связанных с неверным написанием фамилий, имен, мест ареста, районов и т. д. В 20–30-е гг. административные границы ряда областей республики менялись, в связи с чем в некоторых делах возникает проблема определения места проживания и/или ареста.

Наличие подобных аномалий в базе достаточно сильно влияет на возможности статистического анализа базы. В связи с чем было принято решение провести первичный анализ базы на предмет поиска некорректных данных.

Проведен анализ записей, выявлены основные типы аномалий и приняты меры по их исправлению. На полученном варианте базы проведен статистический анализ базы данных ресурса, получены половозрастные характеристики пострадавших граждан и принадлежность их к районам республики.

1. Хабитуев Б. В., Фартусов Д. Б., Овечкин Н. Д., Александрова Е. И. Создание Web-сайта Книги памяти жертв политических репрессий по Республике Бурятия. Опыт совместной работы программистов и историков // Сиб. ссылка: сб. науч. ст. Иркутск, 2017. Вып. 8(20). С. 81–89.

Научные руководители — Д. Ф. Дерюгин, Б. В. Хабитуев

Разработка и реализация программного приложения для анализа фонетических характеристик русских поэтических текстов

М. В. Вахрушев

Новосибирский государственный университет

Одной из актуальных задач в области использования информационных технологий является создание программного приложения для автоматизации анализа поэтических текстов, призванного облегчить труд филологов.

Комплекс таких разработок ведется в Федеральном исследовательском центре информационных и вычислительных технологий. Целью данной работы являлась разработка модуля транскрипции русских слов с использованием русского фонетического алфавита.

В рамках данной работы была предложена архитектура приложения и реализован алгоритм транскрипции русских слов на языке Python. Представлен процесс интеграции модуля акцентуации, описаны проблемы, которые были обнаружены во время внедрения модуля, а также пути их решения.

В работе представлены результаты тестирования разработанного приложения на примере поэтических текстов русских писателей.

Научный руководитель — д-р техн. наук, проф. В. Б. Барахнин

Научный консультант — канд. филол. наук О. Ю. Кожемякина

**Разработка методов анализа и классификации
экономических новостных текстов, представленных в интернете
для прогнозирования движения фондового рынка**

А. В. Гавриленко

Новосибирский государственный университет

Прогнозирование трендов на фондовых рынках является предельно сложной задачей. В связи с колоссальным ростом количества информации СМИ в сети Интернет и ее влиянием на принимаемые людьми решения, а также наличием активно развивающихся методов для ее обработки и анализа появляется возможность проанализировать информационный фон вокруг конкретной компании и сделать прогноз о ее дальнейшей капитализации.

Данная работа имеет высокую практическую значимость: инвесторы получают возможность эффективно управлять своим портфелем, а компании — повысить эффективность менеджмента.

Цель данной работы — реализация метода для прогнозирования тренда фондового рынка на основе анализа информационного пространства. Под информационным пространством понимаются статьи из различных интернет-изданий по экономической тематике, представленные как на русском, так и на английском языках.

Учитывая формат документов, представленных в Сети, перед классификацией их необходимо привести к нормализованному виду — выполнить первичную обработку [1] исходя из специфики естественного языка документа. В работе решаются следующие задачи:

- анализ существующих методов классификации текстов и построение прогнозов на их основе для фондовых рынков;
- выполнение коррекции, нормализации, анализа и классификации корпуса текстов;
- выявление признаков, полученных в результате п. 2, для построения модели прогнозирования;
- анализ и выводы из полученных результатов.

1. Manning C. D., Raghavan P., Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press, 2008.

Научный руководитель — д-р физ.-мат. наук, доц. Д. Е. Пальчунов

**Разработка системы создания и валидации смарт-контрактов,
позволяющей генерировать тексты договоров на естественном языке**

А. Г. Галиева

Новосибирский государственный университет

Цифровизация экономики неизбежно ведет к глобальной компьютеризации экономических процессов, в том числе описывающих формализованные бизнес-процессы. Для описания формализованных бизнес-процессов ранее были разработаны различные методы, основанные на семантических моделях [1–3]. Однако в целях полной автоматизации бизнес-процессов необходимо решить ключевую задачу упрощения аудита смарт-контрактов для экспертов предметных областей [4].

Целью работы является разработка системы создания и валидации смарт-контрактов, позволяющей генерировать тексты договоров на естественном языке.

Результатом данной работы стала система, позволяющая на основе семантического описания смарт-контракта создавать его исполняемый код и текст договора на естественном языке. Во время генерации системой договора реализуется следующий процесс аудита: упрощенный текст естественного языка преобразуется в набор фрагментов атомарных диаграмм, по которым происходит поиск подходящих в той или иной мере шаблонов с их последующей сборкой. Далее по полученным шаблонам порождается текст естественного языка, который валидируется экспертом предметной области. В результате итеративного редактирования шаблоны доводятся до абсолютного соответствия исходному контракту.

1. Gumirov V. S. et. all. Semantic Domain-specific Languages // Int. Multi-Conf. Eng., Computer Inf. Sci. (SIBIRCON). Novosibirsk: IEEE Press, 2019. P. 0955–0960.

2. Galieva A. G., Palchunov D. E. Logical Methods for Smart Contract Development // Int. Multi-Conf. Eng., Computer Inf. Sci. (SIBIRCON). Novosibirsk: IEEE Press, 2019. P. 0881–0885.

3. Галиева А. Г., Пальчунов Д. Е. Актуальные подходы в автоматизации формализации бизнес-процессов // Материалы VII Междунар. конф. «Знания — Онтологии — Теории» (ЗОНТ-2019). Новосибирск, 2019. С. 415.

4. Галиева А. Г. Разработка методов автоматизированного создания смарт-контрактов на основе формализованных бизнес-процессов // Материалы 58-й Междунар. науч. студ. конф. МНСК-2020; Информационные технологии. Новосибирск, 2020. С. 169.

Научный руководитель — д-р физ.-мат. наук, проф. Д. Е. Пальчунов

Разработка смарт-контрактов для платформы ScientificCoin

Е. Б. Ждаркин

Институт автоматики и электрометрии СО РАН, Новосибирск
Новосибирский государственный университет

ScientificCoin — краудфандинговая интернет-платформа, которая ориентирована на венчурные проекты в области наукоемких технологий. Платформа обеспечивает возможности анонсирования новых инновационных проектов, оценку их потенциала со стороны экспертов, а также венчурное финансирование через заключение смарт-контрактов между инвесторами и компанией или частным лицом, реализующим проект.

Смарт-контракты представляют собой цифровую запись об условиях заключенного соглашения в виде программы (описания операций, в том числе условных, над некоторыми активами, представленными их заместителями — токенами), которые хранятся в блокчейне платформы Ethereum.

В качестве языка разработки был выбран язык Solidity, программы которого транслируются в байткод для виртуальной машины Ethereum. Средой разработки контрактов выступил фреймворк truffle, который дает удобный способ написания и тестирования смарт-контрактов для децентрализованных приложений.

Для предупреждения ошибок при исполнении кода программ в реальном блокчейне, что может стать причиной потери денежных средств пользователями, были разработаны тесты и использованы анализаторы кода. Тесты писались на языке Javascript и проверяли корректность действий функций, связанных с переводом криптовалют и токенов. Анализаторы, которые использовались для аудита кода: HoneyBadger, Manticore, Mythril, Slither и Smartcheck. Все это позволяет предотвратить большой процент уязвимостей контрактов.

При создании интерфейса, позволяющего пользователям запускать свои смарт-контракты в блокчейне и иметь доступ к их функциям, использовалась веб-платформа React. А связь интерфейса с блокчейном обеспечивает провайдер в виде браузерного расширения Metamask, который предоставляет данные с узла Ethereum на машинном языке, а javascript-библиотека web3.js переводит их на удобный язык.

Научный руководитель — д-р техн. наук, доц. В. Е. Зюбин

**Разработка модуля поиска оптимальной последовательности действий
для создания интеллектуального помощника**

А. О. Зайцев

Новосибирский государственный университет

Со временем начали появляться возможности передавать голосовые сообщения, что послужило одной из причин появления умных помощников и чат-ботов. Данная работа является частью большого проекта интеллектуального помощника, создаваемого для помощи слабовидящим людям, и направлена на поиск оптимальной последовательности действий.

Использование одного из существующих алгоритмов поиска не является достаточным для поставленной задачи, так как он либо не учитывает ограничения, либо учитывает их часть. Но для оптимальной работы модуля поиска необходим учет всевозможных ограничений.

В связи с этим целью данной работы является разработка модуля для интеллектуального помощника, выполняющего поиск оптимальной последовательности пользовательских действий с учетом всевозможных ограничений [1].

В рамках работы была изучена предметная область — университет, разработана совместно с коллегами онтологическая модель решаемой задачи [2, 3], рассмотрены и проанализированы существующие технологии решения подобных задач, выбран существующий пригодный алгоритм, способный приближенно выполнять решение поставленной цели, составлен список все возможных ограничений, ведется разработка и тестирование алгоритма на выбранном языке программирования, анализируются результаты его работы, подготавливается основа для будущего расширения алгоритма и дальнейшего внедрения модуля.

1. Филипов И. И., Пальчунов Д. Е., Степанов П. А., Разработка методов семантического поиска в Интернете, основанных на древоводных лингвистических шаблонах // Вестн. НГУ. Сер. информ. технологии. 2019. Т. 17, № 3. С. 111–122.

2. Пальчунов Д. Е., Целищев В. В. Проблема извлечения знаний в системе взаимодействия человека и компьютера: онтологии и пресуппозиции // Философия науки. 2012. № 4(55). С. 20–35.

3. Капустина А. И., Пальчунов Д. Е. Разработка онтологической модели тарифов и услуг сотовой связи, основанной на логически полных определениях понятий // Вестн. НГУ. Сер. информ. технологии. 2017. С. 34–46.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Анализ обращений пользователей юридических форумов средствами интеллектуального анализа данных

Н. О. Зиануров

Новосибирский государственный университет

Не секрет, что в настоящее время на территории Российской Федерации действует множество законов и порой среди них можно запутаться. По данной причине многие люди обращаются за правовой помощью на такие веб-сервисы, как юридические форумы. Однако на юридических форумах большое количество вопросов не получают ответа, поэтому возникает необходимость в автоматическом выделении из данных вопросов отношений к разделам российского права.

Целью работы является разработка обучаемой информационной системы, осуществляющей классификацию юридических обращений на естественном языке [1] относительно правовых отраслей. Для обучения системы используются данные, полученные с русскоязычных юридических форумов [2], имеющих модерацию вопросов. Также в нашей работе выявляются требования к разрабатываемой системе, затрагивающие приведение анализируемых текстов на естественном языке к векторной форме и атрибуты полученных векторов. Так, основным свойством при векторизации слов выступает их относительная частота, веса слов вычисляются при помощи статистической меры TF-IDF, а анализируемые тексты подвергаются нормализации. Для выполнения классификации используются и сравниваются между собой метод Байеса, метод опорных векторов, а также метод k-ближайших соседей.

Как итог, результатом нашей работы являются модели классификации текстов на естественном языке, где тексты представлены в виде вопросов юридического характера.

1. Власов Д. Ю., Пальчунов Д. Е., Степанов П. А. Автоматизация извлечения отношений между понятиями из текстов естественного языка // Вестн. НГУ. Сер. информ. технологии. 2010. Т. 8, № 3. С. 23–33.

2. Степанов П. А., Пальчунов Д. Е., Мирзагитов А. А. Методы анализа диалогов, основанные на теории речевых действий // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 4. С. 102–111.

Научный руководитель — д-р физ.-мат. наук, доц. Д. И. Свириденко

Разработка системы контекстного поиска и извлечения аргументации на естественном языке

Д. К. Зулин

Новосибирский государственный университет

В настоящее время существуют два принципиально различающихся подхода к обработке текстов на естественном языке: методы машинного обучения и синтаксический анализ, основывающийся на знаниях языка. Хотя первый вариант может показать большую производительность, интерпретируемость промежуточных результатов подобных систем крайне мала, а построение декодеров может оказаться слишком трудоемкой задачей. По этой причине в разрабатываемой системе основную роль в интерпретации естественного языка играет система ЭТАП-4 [1], использующая теоретико-модельный подход [2–4].

В докладе будет рассмотрен способ применения синтаксических деревьев, получаемых в результате анализа, рассмотрены стандартные шаблоны аргументации и перенесены на синтаксические деревья, описан формальный язык, позволяющий расширять знания системы по мере извлечения информации из аргументации текста.

1. Дяченко П. В. и др. Современное состояние глубоко аннотированного корпуса текстов русского языка // Тр. Ин-та русского языка им. В. В. Виноградова. 2015. Т. 6. С. 272–300.

2. Найданов Ч. А., Пальчунов Д. Е., Сазонова П. А. Разработка автоматизированных методов предупреждения рисков возникновения критических состояний, основанных на анализе знаний, извлеченных из историй болезней пациентов // Сиб. науч. мед. журн. 2016. Т. 36, № 1. С. 105–113.

3. Долгушева Е. В., Пальчунов Д. Е. Теоретико-модельные методы порождения знаний о предпочтениях абонентов мобильных сетей // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 2. С. 5–16.

4. Ненашева Е. О., Пальчунов Д. Е. Разработка автоматизированных методов преобразования предложений естественного языка в бескванторные формулы логики предикатов // Вестн. НГУ. Сер. информ. технологии. 2017. Т. 15, № 3. С. 49–63.

Научный руководитель — д-р физ.-мат. наук, доц. Д. Е. Пальчунов

Разработка динамической анкеты для рекомендательной системы «Абитуриент — Студент»

А. А. Зыкова

Новосибирский государственный университет

При создании рекомендательной системы «Абитуриент — Студент» [1] является необходимым создание базы знаний о различных университетах и факультетах. Система устроена так, что абитуриенту предоставляются организованные знания, полученные непосредственно от студентов путем заполнения анкет. При проведении анализа заполненных анкет было выявлено нежелание студентов заполнять последние вопросы анкеты по различным причинам, поэтому распределение ответов получается неравномерным. Таким образом, при отображении статистики [2] мы рискуем показать пользователю необъективную информацию по какому-то критерию и ввести его в заблуждение.

Для решения этой проблемы было решено разработать динамическую анкету, которая будет меняться в зависимости от того, насколько много или мало имеется ответов на тот или иной вопрос в базе знаний [3]. Таким образом, пользователи сначала должны видеть вопросы анкеты, заполненные меньшее количество раз.

Анкета реализована в виде сайта с базой данных, хранящей вопросы анкеты и ответы студентов на нее. Это позволяет серверу быстро формировать новую анкету после заполнения предыдущего ее вида и выдавать следующему студенту уже новый вид анкеты.

1. Зыкова А. А., Яхьяева Г. Э. Программная система «Абитуриент — Студент» // Вестн. МГПУ. Сер. Информатика и информатизация образования. 2021. № 1(55).

2. Зыкова А. А. Разработка модуля статистики для рекомендательной системы высших учебных заведений // Материалы Междунар. конф. «Знания — Онтологии — Теории» (ЗОНТ-2019). Новосибирск, 2019. С. 421.

3. Зыкова А. А. Разработка базы знаний для программной системы «Абитуриент — Студент» // Материалы 58-й Междунар. науч. студ. конф. Сер. информ. технологии. Новосибирск, 2020. С. 177.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Разработка электронной торговой площадки фермерских продуктов

А. Е. Кицкалов, Н. Е. Мартынов, Д. А. Черкашин
Сибирский федеральный университет, Красноярск

Сегодня в век технологий бизнес все больше переходит в онлайн-сферу.оборот рынка интернет-торговли в России неуклонно растет. По данным Ассоциации компаний интернет-торговли, с 2010 г. наблюдается его бурный рост с 240 до 1386 млрд руб. в год. Однако некоторые отрасли бизнеса не используют все возможности IT-сферы. Подобную ситуацию сегодня можно наблюдать в агропромышленном комплексе, особенно среди малых и средних предприятий. Главной проблемой небольших фермерских хозяйств является поиск точек сбыта своей продукции.

В настоящее время многим фермерам приходится отдавать товар перекупщикам либо самостоятельно организовывать доведение продукта до конечного потребителя, что приводит к серьезным ресурсным затратам и повышению его итоговой стоимости. Таким образом, мы видим рынок, нуждающийся в инновациях: большому количеству бизнеса требуются площадки для организации своей деятельности.

Решением этой проблемы фермеров может послужить разработка электронной торговой площадки (ЭТП). ЭТП — это программно-аппаратный комплекс организационных, информационных и технических решений, обеспечивающих взаимодействие продавца и покупателя через электронные каналы связи. ЭТП позволяет объединить в одном информационном и торговом пространстве поставщиков и потребителей различных товаров и услуг и предоставляет участникам ЭТП ряд сервисов, повышающих эффективность их бизнеса.

В настоящей работе осуществлена разработка ЭТП для категории B2B. Данный вид ЭТП является наиболее значительным среди посреднических площадок, призванных свести вместе производителей сельскохозяйственной продукции и бизнес, нуждающийся в этой продукции.

Главными функциями площадки является возможность быстрого и простого выставления товара на продажу или его приобретения. Кроме того, важным нововведением площадки по сравнению с аналогами станет интеллектуальная функция организации доставки силами одной из сторон сделки или одной из логистических компаний.

Клиентами нашей площадки станет малый и средний бизнес. Она послужит точкой взаимодействия местных фермеров и их потенциальных клиентов, среди которых можно выделить магазины, рестораны, отели, кафе и заведения быстрого питания.

**Разработка модуля статистики с аналитикой
для стримингового приложения на интегрируемых платформах**

К. Т. Комагина

Сибирский федеральный университет, Красноярск

В настоящее время большое распространение получают стриминговые платформы, которые представляют собой прямое соединение с аудиторией в реальном времени, именуемое прямой трансляцией, и позволяют сохранить его для отложенного просмотра. Для обеспечения доступа к проведению трансляции применяются различные сервисы, использующие разнотипные медиакодеки, а также способы подключения технических устройств для передачи аудио- и видеоданных. Как и при любом роде деятельности, в стриминговой среде имеется потребность в презентации статистических данных и их анализе. Однако такого рода данные зачастую предоставляются интегрируемыми платформами как сырые данные, т. е. неупорядоченные сведения о показателях активности аудитории. Поэтому актуальной задачей является сбор открытых сырых данных с каждой платформы, а также их дальнейшая обработка в более понятный для пользователя вид.

В результате разработанный по сбору данных алгоритм модуля при каждом запуске прямой трансляции собирает информацию о количестве просмотров и комментариях, раз в сутки отправляет запрос на сервера интегрируемых платформ для фиксации изменения числа подписчиков. Далее пользователь в отдельном разделе со статистикой может просмотреть динамику активности на проведенных им трансляциях за выбранный им период: неделю, год или месяц в виде линейной диаграммы с возможностью детализации. Время проведения трансляций также фиксируется алгоритмом и отображается в разделе с помощью столбчатой диаграммы, и пользователь имеет возможность оценить временные показатели, т. е. насколько более длительное проведение трансляции влияет на показатели активности, и на основе этого выбрать для себя релевантный средний показатель. Также на dashboard-панели для пользователя представлена аналитическая сводка, основанная на сравнительном анализе текущего выбранного периода с предыдущими.

Основными потребителями данного сервиса можно считать пользователей интернет-ресурсов, предоставляющих возможности для проведения трансляций, которые заинтересованы в создании и размещении своего контента.

Научный руководитель – канд. техн. наук, доц. Ю. Ю. Якунин

**Определение и декомпозиция задач пользователя
для создания интеллектуального голосового помощника**

А. С. Кондратьев
Новосибирский государственный университет

С развитием компьютерных технологий неуклонно растет число пользователей. Слабовидящие и незрячие люди не могут полноценно пользоваться многими программами с графическим интерфейсом. Данная работа посвящена определению и декомпозиции задач пользователя для создания интеллектуального голосового помощника.

Работа непосредственно связана с набирающей популярность областью — анализом естественного языка [1–3] — и рассматривает проблему определения и декомпозиции пользовательских задач.

Целью данной работы является разработка модуля определения и декомпозиции задач, обеспечивающего достаточную точность и масштабируемость.

В рамках работы была изучена предметная область — университет, существующие методы анализа текста, отобраны наиболее удачные подходы и архитектуры. На доступных наборах данных была протестирована языковая модель BERT, показавшая отличные результаты при небольших затратах на процесс обучения. Используя данную модель, можно легко переобучать модуль при накоплении новых массивов реальных данных от пользователей. Полученные при помощи BERT и простого линейного слоя метки действий можно декомпонировать на подзадачи при помощи модели предметной области, над которой в данный момент работает другой участник проекта. Ведется работа по подготовке внедрения модуля в проект.

1. Пальчунов Д. Е. Поиск и извлечение знаний: порождение новых знаний на основе анализа текстов естественного языка // *Философия науки*. 2009. № 4(43). С. 70–90.

2. Корсун И. А., Пальчунов Д. Е. Теоретико-модельные методы извлечения знаний о смысле понятий из текстов естественного языка // *Вестн. НГУ. Сер. информ. технологии*. 2016. Т. 14, № 3. С. 34–48.

3. Власов Д. Ю., Пальчунов Д. Е., Степанов П. А. Автоматизация извлечения отношений между понятиями из текстов естественного языка // *Вестн. НГУ. Сер. информ. технологии*. 2010. Т. 8, № 3.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Разработка рекомендательной системы досуга в городе

А. В. Коробова

Новосибирский государственный университет

Сфера досуга и развлечений всегда пользовалась популярностью и была востребована у людей. Часто люди пытаются найти места или мероприятия для проведения своего свободного времени. Однако не всегда поиск бывает удобным. Часто пользователям приходится самим разбираться с большим объемом неструктурированной информации, полученной из разных источников [1]. На первых местах всегда крупные организации или рекламные позиции, это уменьшает возможность найти небольшие компании, рассчитанные на меньшее число клиентов или на более необычное проведение досуга. Часто пользователю приходится самому искать места для проведения досуга по своим определенным критериям, и мало где можно получить рекомендацию, что вызывает затруднения, ведь иногда человек не знает конкретной цели или о существовании тех или иных мест.

Целью данной работы является разработка приложения рекомендательной системы, которое обеспечивает сбор, хранение и поддержание актуальной информации, а также выполняет функцию рекомендации мероприятий и мест для проведения досуга.

В ходе выполнения проекта был разработан алгоритм рекомендации, который при наличии выборки данных, зная, какие места выбирают определенные пользователи, может предсказать наиболее вероятный выбор для пользователя, опираясь на его параметры (возраст, пол, предпочтения в отдыхе, бюджет на развлечения). Также было реализовано клиент-серверное приложение, которое обрабатывает запросы пользователя и выдает рекомендации, подобранные для конкретных пользователей с помощью моделей машинного обучения, а также предоставляет пользователю информацию в структурированном и удобном для восприятия формате.

1. Яхьяева Г. Э., Ясинская О. В. Методы согласования знаний по компьютерной безопасности, извлеченных из различных документов // Вестн. НГУ. Сер. информ. технологии. 2013. Т. 11, Вып. 3. С. 63–73.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Рекомендация сценариев реагирования для инцидентов информационной безопасности на основе обучения представлениям

И. А. Краева

Новосибирский государственный университет

Частью стратегии обеспечения информационной безопасности в организациях является использование сценариев реагирования на инциденты. Сценарий реагирования представляют собой набор действий, направленных на минимизацию воздействия на бизнес, восстановление атакованных систем и предотвращение подобных инцидентов в будущем.

В системах автоматизации процессов управления инцидентами — SOAR (*Security Orchestration, Automation and Response*) и IRP (*Incident Response Platform*) — рекомендация сценариев выполняется на основе правил. Также есть ряд исследований, предлагающих использовать прецедентный анализ [1, 2]. Эффективность существующих методов зависит от компетенции экспертов предметной области, задающих правила или описывающих прецедентную модель.

В работе предлагается метод для рекомендации наиболее подходящих сценариев реагирования на основе обучения представлениям. Новый метод не требует привлечения экспертов предметной области и основан на анализе сценариев, по которым ранее выполнялось реагирование на инциденты. Метод предполагает использование нейронной сети для отображения инцидентов в пространство, в котором инциденты с одинаковыми сценариями реагирования находятся ближе друг к другу, чем инциденты с разными сценариями. Полученные представления используются для многозначной классификации на основе коэффициента сходства.

Эксперименты, проведенные на наборе данных с реальными инцидентами, показывают, что предлагаемый метод может быть эффективно использован для рекомендации сценариев реагирования.

1. Яхьяева Г. Э., Ясинская О. В. Применение методологии прецедентных моделей в системе риск-менеджмента, направленного на раннюю диагностику компьютерного нападения // Вестн. НГУ. Сер. информ. технологии. 2012. Т. 10, Вып. 2. С. 106–115.

2. Головин Н. А., Савин Н. П., Яхьяева Г. Э. Применение методов машинного обучения для структурирования базы прецедентов компьютерных атак // Материалы Междунар. конф. «Знания — Онтологии — Теории» (ЗОНТ-2019). Новосибирск, 2019. С. 122–128.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Проектирование чат-бота для диверсификации портфеля ценных бумаг

А. С. Кузнецов

Новосибирский государственный университет

С каждым годом все больше людей стремятся защитить свои сбережения от инфляции, однако банковские ставки по вкладам их не устраивают, поэтому они идут на биржу ценных бумаг и открывают индивидуальный инвестиционный счет. Зачастую эти люди не обладают достаточным количеством знаний для правильного составления портфеля и успешной торговли ценными бумагами, также предполагается, что они не желают тратить много времени на освоение сложных профессиональных инструментов, которые никак не защищают инвесторов от возможности потерять свои деньги, допуская необдуманные действия на рынке ценных бумаг, например составление недиверсифицированного портфеля.

Для решения этой проблемы был спроектирован чат-бот, позволяющий начинающим инвесторам вести простейший учет своих ценных бумаг. Бот хранит и обрабатывает информацию о реальном инвестиционном портфеле, введенную посредством диалога с ботом. Также чат-бот отображает текущее состояние инвестиционного портфеля, по запросу отображает прибыль или убыток от акций, позволяет выставлять собственные оповещения для акций из портфеля. Помимо этого, сформулированы основные принципы, по которым чат-бот должен осуществлять анализ отраслевой направленности портфеля и формулировать варианты его изменения для увеличения диверсификации портфеля.

Взаимодействие с пользователем происходит через кроссплатформенный мессенджер Telegram с использованием Telegram Bot API.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

**Разработка модуля обработки тавтологий в тексте
для автоматизированной обучающей системы**

А. Д. Лебедева

Новосибирский государственный университет

Среди анализаторов текста становится популярно максимально автоматизировать исправление текста. Все чаще для этого используют технологии семантического моделирования [1]. Для упрощения задачи пользователя при использовании анализатора можно встроить рекомендательную систему, которая поможет человеку быстро избавиться от ошибок. В случае анализатора тавтологий необходимо заменять слова на синонимы или личные местоимения.

Цель работы — разработать рекомендательную систему исправления найденных тавтологий в тексте. Тавтология в анализаторе представляется в виде взвешенного графа связанных слов. Система анализирует каждый граф и подбирает место исправления ошибки и слово для замены [2].

В работе рассмотрена система с точки зрения теории графов. Разбиение цепного графа происходит с помощью раскраски вершин. Количество цветов для раскраски определяется количеством синонимов, подходящих для замены. Синонимы подбираются с помощью тезауруса RuWordNet [3]. Также в определенных ситуациях возможна замена часто повторяющихся существительных личными местоимениями, и наоборот: отслеживание чрезмерного употребления личных местоимений.

1. Яхьяева Г. Э., Абсайдульевой А. Р. Семантический подход к моделированию фонда оценочных средств // Вестн. НГУ. Сер. информ. технологии. 2018. Т. 16, Вып. 2. С. 113–121.

2. Лебедева А. Д. Алгоритм анализа тавтологий в тексте // Тез. докл. Междунар. конф. «Мальцевские чтения — 2020». Новосибирск, 2020. С. 77.

3. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. М.: Изд-во МГУ, 2011. 512 с.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Алгоритм поиска ответа на нечетко-вероятностные вопросы

Е. О. Легостаева

Новосибирский государственный университет

Для компаний, работающих с конфиденциальной информацией, важен процесс управления информационными рисками, предсказание компьютерных атак. Справиться с прогнозированием различных рисков может помочь постановка вероятностных вопросов. Возникает потребность в разработке информационной системы и алгоритма, позволяющего дать ответ на нечетко-вероятностные вопросы.

Существует вопросно-ответная система RiskPanel, предоставляющая возможность анализировать риски, связанные с компьютерными атаками [1]. Эта система содержит модуль QA-RiskPanel, который является вопросно-ответной системой и использует базу прецедентов компьютерных атак в качестве источника знаний [2, 3]. Существует веб-интерфейс, позволяющий пользователю порождать нечетко-вероятностные вопросы [4]. Разработан алгоритм, позволяющий найти ответ на вопросы, сформулированные пользователем системы. Алгоритм формирует две(одну) новые базы прецедентов из начальной базы, опираясь на вероятность, указанную в вопросе. Поиск ответа на вопрос по новым базам приводит к получению двух(одного) значений вероятности — вероятностного интервала, являющегося ответом на искомый вопрос.

1. Пальчунов Д. Е. и др. Программная система управления информационными рисками RiskPanel // Программная инженерия. 2011. № 7. С. 29–36.

2. Яхьяева Г. Э. и др. Вопросно-ответная система для управления информационными рисками на основе теоретико-модельной формализации предметных областей // Информ. технологии. 2017. Т. 23, № 2. С. 97–106.

3. Головин Н. А. и др. Применение методов машинного обучения для структурирования базы прецедентов компьютерных атак // Материалы Междунар. конф. «Знания — Онтологии — Теории» (ЗОНТ-2019). Новосибирск, 2019. С. 122–128.

4. Легостаева Е. О. Разработка интерфейса модуля нечетко-вероятностных вопросов для вероятностной вопросно-ответной системы // Тез. докл. Междунар. конф. «Мальцевские чтения — 2020». Новосибирск, 2020. С. 78.

Научный руководитель — О. В. Данилова

Разработка четырехуровневой семантической модели для создания интеллектуального помощника

Р. А. Лешов

Новосибирский государственный университет

По мере развития компьютерных технологий их начинают применять во всех сферах жизни. Вместе с этим становится все больше пользователей компьютерных технологий, среди которых встречаются люди, имеющие проблемы со зрением. Ввиду того что большая часть информации передается визуально, у таких пользователей возникают трудности с использованием программ и сервисов, не имеющих специальных средств управления для слабовидящих людей. Данная работа посвящена вопросу разработки четырехуровневой семантической модели предметной области для создания интеллектуального голосового помощника для данной категории пользователей.

Задача построения семантической модели включает спецификацию смысла ключевых понятий предметной области. В текущее время эта задача решается в основном путем построения ER-диаграмм, которые отражают взаимосвязи между сущностями предметной области, но в то же время упускают такие важные факторы, как знания и прецеденты в предметной области. Разработка модели, учитывающей эти факторы, позволила бы программе лучше понимать нужды и намерения пользователя.

Цель работы — разработать и применить для создания интеллектуального помощника семантическую модель предметной области, включающую онтологическую модель, модель знаний и прецедентов.

В рамках работы будет рассматриваться специфическая предметная область — университет. Тем не менее модель будет реализована с учетом возможности ее использования и в других областях.

Были решены следующие задачи: проведен анализ литературы, изучены методы создания онтологий, описана методология исследования, разработана онтология предметной области университета, реализована реляционная база данных для хранения модели, реализованы миграции базы данных на языке Python, а также настроено взаимодействие базы данных с онтологией при помощи запросов SPARQL.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Использование экстрактивных суммаризаторов для подготовки обучающих данных для абстрактивных суммаризаторов

Н. С. Мамеев

Новосибирский государственный университет

Во многих областях встречается задача кластеризации текстов на естественном языке. Кластеризация является промежуточным шагом, который пользователь делает для удобной последующей обработки, в которой можно будет оперировать группами схожих элементов. Кластеризация является описательной процедурой, которая не делает выводов о содержимом кластера. Чтобы придать кластеру текстов смысловую нагрузку, пользователю необходимо самостоятельно разобрать несколько текстов и выделить из них общую часть, описывающую весь кластер. Для группы текстов на естественном языке данная задача может решаться в автоматическом режиме с помощью суммаризации текста — класса методов анализа текстов, целью которых является построение краткого описания или заголовка.

Основные подходы к суммаризации текста: экстрактивный, когда предложениям исходного текста присваивается вес, описывающий, насколько данное предложение отражает смысловую нагрузку исходного текста. Резюме составляется из предложений с наибольшим весом. Данный метод является обучением без учителя и требует лишь задания функции веса, но структура резюме целиком зависит от структуры исходного текста. Абстрактивный подход: резюме является результатом генерации нового текста, содержательно обобщающего исходный. Данный подход позволяет создавать более гибкое резюме, но реализуется с помощью машинного обучения, из-за чего требует достаточно большое количество данных для обучения. В открытом доступе находятся наборы для ограниченного количества предметных областей, а генерация собственного набора — затратный процесс, поскольку требует разметки большого числа текстов человеком, что ограничивает применение данного подхода.

Целью работы является разработка программной системы, подготавливающей обучающие данные для абстрактивных суммаризаторов с помощью экстрактивных суммаризаторов.

Применение данной программной системы облегчит использование абстрактивных суммаризаторов за счет автоматической подготовки набора обучающих данных. От пользователя будет требоваться только составление набора неразмеченных данных для выбранной предметной области.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Параллельная реализация CI-CLOPE для кластеризации категориальных данных

А. С. Михайлов

Новосибирский государственный университет

В настоящее время наблюдается сильное развитие направления анализа больших массивов данных. Поиск различных закономерностей может дать существенную пользу для дальнейшего прогнозирования.

Особое место в этой области занимает кластеризация, суть которой заключается в объединении объектов в группы так, чтоб внутри группы объекты были похожи, а с представителями других групп — отличались [1].

Каждый объект обладает каким-либо набором признаков, что и выделяет его среди других. Удобно сравнивать объекты по признаку, имеющему числовую характеристику, с другой стороны, признаки могут быть не количественными, а номинальными, что заставляет сравнивать объекты друг с другом попарно. LargeItem, Seed, F-Tree, Rock, Clope относятся к алгоритмам, работающими именно с такими данными [2]. Clope обладает высокой скоростью работы и качеством кластеризации, поэтому и был взят за основу для дальнейшего улучшения.

Clope строит свое разбиение, основываясь на одном глобальном критерии, и не позволяет задать количество кластеров, которое хотелось бы получить, заранее. CI-Clope — модификация, которая решает эту проблему. Суть его работы заключается в том, чтобы не запускать повторно оригинальный алгоритм в надежде, что получится необходимое количество кластеров, а объединить или разбить существующий набор кластера, чтоб довести до необходимого. В рамках данной работы также представлена оптимизация данного алгоритма — параллелизация процесса слияния и разбиение кластеров. Данное изменение позволяет не использовать исходный алгоритм многократно, а работать уже с полученным результатом.

1. Palchunov D., Yakhyaeva G., Dolgusheva E. Conceptual Methods for Identifying Needs of Mobile Network Subscribers // CLA. 2016. С. 147–160.

2. Yang Y., Guan X., You J. CLOPE: a fast and effective clustering algorithm for transactional data // Proc. 8th ACM SIGKDD Int. Conf. Knowledge Discovery Data Min. ACM. 2002. С. 682–687.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Применение рекурсивных нейронных сетей в задаче иерархической сегментации изображения

П. В. Мызников

Новосибирский государственный университет

Сегментация изображения является одной из базовых задач компьютерного зрения. Среди самых популярных практических приложений сегментации изображения можно выделить анализ медицинских снимков, разметку спутниковых снимков, построение визуальных поисковых систем.

Одним из видов сегментации изображения является иерархическая сегментация — такое разделение изображения на сегменты, при котором сегменты могут быть повторно декомпозированы. Иерархическая сегментация — более сложная задача, и в то же время она позволяет извлекать дополнительную информацию. Например, она используется в задачах распознавания сцен.

Если проследить историю развития алгоритмов сегментации, можно заметить переход от детерминированных алгоритмических подходов к подходам машинного обучения и далее к глубокому обучению. Так, самым современным методом сегментации считается архитектура нейронной сети Mask R-CNN [1]. Тем не менее, как показали эксперименты, при применении этого подхода для решения иерархической сегментации остается большое пространство для улучшения [2].

В данной работе предлагается построить рекурсивную архитектуру нейронной сети, реализующую иерархическую сегментацию входного изображения. Предполагается, что естественное сходство требуемого результата и архитектуры нейронной сети позволит увеличить эффективность метода.

Результатом работы является нахождение оптимальных параметров рекурсивной нейронной сети в ходе проведения экспериментов на заданной выборке изображений и при сравнении полученных значений с актуальными аналогами.

1. He K., Gkioxari G., Dollár P. and Girshick R. Mask R-CNN // IEEE Int. Conf. Computer Vision (ICCV). Venice, 2017. P. 2980–2988.

2. Myznikov P., Huang Y. A New Method for Hierarchical Image Segmentation from Visual Designs // 54th Annu. Conf. Inf. Sci. Systems, CISS 2020 [9086192]. Institute of Electrical and Electronics Engineers, 2020.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Разработка автоматизированных методов наполнения онтологии предметной области при помощи виртуального помощника

А. С. Орловский

Новосибирский государственный университет

Виртуальные помощники (*virtual assistants, conversational agents*, чат-боты) применяются для общения с пользователем на естественном языке с целью сокращения издержек на выполнение аналогичных задач усилиями реальных людей. В данной работе представлен виртуальный ассистент (ВА), помогающий пользователю на основе его запроса отыскать наиболее релевантные товары в интернет-магазинах компьютерной техники [1].

На основе микросервисной архитектуры разработаны компоненты ВА.

- **«Менеджер диалога»**, содержащий основную бизнес-логику помощника, принимающий реплики пользователя и отправляющий свои реплики в ответ в зависимости от результатов работы других сервисов. Используется API внешней службы DialogFlow для распознавания интенгов.

- **«Обработчик естественного языка»**, ответственный за извлечение именованных сущностей из строк пользовательского ввода и из содержимого каталогов товаров в интернет-магазинах. Используется библиотека Pullenti SDK для извлечения атрибутов товаров.

- **«Поисковый агент для анализа сайтов»**, предоставляющий релевантную информацию о товарах на сайтах интернет-магазинов по текстовым запросам.

- **«Менеджер онтологий»**, ответственный за хранение собранных с сайтов данных о товарах, логический вывод новых знаний и обработку запросов на извлечение товаров с требуемыми атрибутами [2].

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Капустина А. И., Пальчунов Д. Е. Разработка онтологической модели тарифов и услуг сотовой связи, основанной на логически полных определениях понятий // Вестн. НГУ. Сер. информ. технологии. 2017. Т. 15, № 2. С. 34–46.

Научный руководитель — д-р физ.-мат. наук, проф. Д. Е. Пальчунов

**Описание и алгоритмическая поддержка массовых операций
в вычислительных моделях**

Д. Р. Парфенов

Новосибирский государственный университет

Вычислительные модели являются одним из центральных понятий в теории структурного синтеза параллельных программ на вычислительных моделях. Они служат для описания некоторой предметной области с целью автоматической генерации программ, решающих задачи из этой предметной области. Однако в простых вычислительных моделях отсутствует возможность определять массовые вычисления на уровне самой модели: пользователь вынужден самостоятельно реализовывать модули для эффективной работы с массивами, что делает процесс разработки более сложным, а также потенциально снижает производительность.

Решить данную проблему призваны рекурсивные вычислительные модели и вычислительные модели с массивами. В данной области было проведено достаточно много теоретических исследований, но практически удобные способы представления и их алгоритмическая поддержка требуют дополнительной проработки. Целью исследования было предоставить удобный способ описания частных видов таких вычислительных моделей, а также обеспечить работоспособность алгоритмов планирования на вычислительных моделях и корректность исполнения сгенерированных программ.

В ходе исследования было введено несколько новых операторов, которые призваны упростить работу с массивами в вычислительных моделях. Разработанные операторы позволяют задавать рекурсивные вычислительные модели и вычислительные модели с массивами частного вида, с помощью которых возможно решение задач в практически важных предметных областях, например матрично-векторные операции, возникающие в задачах численных методов.

Для проверки предложенного решения на практике был разработан и реализован прототип интерпретатора вычислительных моделей. Предложенное решение облегчает разработку модулей и позволяет генерировать более эффективные параллельные программы, а также не требует существенных изменений в существующем алгоритме планирования вычислений для вычислительных моделей без массивов.

Научный руководитель — В. А. Перепелкин

Фреймворк AutoML для BigData на базе Apache Spark

А. Е. Паульс, С. О. Деревянченко, Н. А. Николаев, Л. Б. Анисютин
Новосибирский государственный университет

Все больше и больше задач реального мира решаются с помощью методов машинного обучения. В некоторых задачах машины уже превзошли человека, в других — применение человека невозможно в принципе ввиду большого количества данных и необходимости быстрого результата. Но даже несмотря на активное развитие и обширное внедрение машинного обучения во все сферы жизнедеятельности, универсальное решение все еще не найдено. Поэтому для решения конкретной бизнес-задачи нередко приходится привлекать экспертные знания специалиста по машинному обучению или даже целую команду для ручного построения пайплайна предобработки данных, подбора модели и различных параметров. Для полной или частичной автоматизации этих процессов и были предложены методы автоматического машинного обучения (AutoML).

В рамках нашего исследования рассматриваются методы AutoML для табличных данных. Этот формат данных используются во многих реальных задачах, например: в обнаружении мошенничеств (*fraud detection*) и кредитном скоринге. Существуют различные фреймворки и сервисы для AutoML: некоторые некоммерческие и с открытым кодом (*autogluon*, *auto-sklearn*), другие — нет (Google Cloud AutoML, H2O).

Типичный подход к AutoML для табличных данных заключается в стандартном пайплайне — предобработке данных, инженерии признаков, подборе модели и гиперпараметров. Особенность конкретного фреймворка заключается в реализации каждого из шагов и примененных оптимизациях. Например, большинство существующих фреймворков не подходит для работы с так называемыми большими данными в распределенной системе.

Исследованы и протестированы основные подходы к AutoML на табличных данных, предложен бейзлайн для AutoML для больших данных на базе фреймворка для распределенных вычислений Apache Spark, а также исследована возможность применения моделей на основе нейронных сетей для табличных данных.

Научный руководитель — И. Ю. Бондаренко

**Применение теоретико-модельных методов
для автоматизированного извлечения знаний из медицинских текстов**

Р. С. Погодин

Новосибирский государственный университет

Работа посвящена автоматизации извлечения знаний из описаний болезней, представленных текстами на естественном языке, с помощью применения теоретико-модельных методов [1].

Для решения данной задачи знания представляются в виде фрагментов атомарных диаграмм, содержащих только двухместные предикаты [2]. Далее, по двухместным предикатам строится смысловое дерево — промежуточная форма представления знаний, удобная для трансляции в древовидные форматы [3]. На финальном этапе на основе смыслового дерева конструируются инфоресурсы облачной платформы IACPaaS. Также поддерживается экспорт извлеченных знаний в форматы Semantic Web, такие как RDF и OWL, что позволяет применять существующие методы и инструменты для порождения новых знаний [4].

Разработаны методы извлечения знаний из медицинских текстов. Особенности этих текстов являются предложения с сочинительными связями, уточняющими и сравнительными оборотами, а также терминами-словосочетаниями. Данные методы реализованы в программной системе.

1. Найданов Ч. А., Пальчунов Д. Е., Сазонова П. А. Разработка автоматизированных методов предупреждения рисков возникновения критических состояний, основанных на анализе знаний, извлеченных из историй болезней пациентов // Сиб. науч. мед. журн. 2016. Т. 36, № 1. С. 105–113.

2. Ненашева Е. О., Пальчунов Д. Е. Разработка автоматизированных методов преобразования предложений естественного языка в бескванторные формулы логики предикатов // Вестн. НГУ. Сер. информ. технологии. 2017. Т. 15, № 3. С. 49–63.

3. Погодин Р. С. Автоматизированное извлечение знаний из медицинских текстов // Совр. наука: актуальные проблемы теории и практики. Сер. естественные и технические науки. 2020. № 7. С. 96–99.

4. Капустина А. И., Пальчунов Д. Е. Разработка методов интеграции автоматических средств логического вывода для порождения знаний в онтологической модели // Вестн. НГУ. Сер. информ. технологии. 2019. Т. 17, № 3. С. 29–42.

Научный руководитель — д-р физ.-мат. наук, проф. Д. Е. Пальчунов

Разработка рекомендательной системы электронных курсов

Т. М. Подкур, Е. А. Шестакова
Новосибирский государственный университет

В последнее время становится актуальной задача цифровизации образования [1, 2]. В сети Интернет появляются образовательные ресурсы, которые составляют конкуренцию учебным заведениям как по уровню образования [3], так и по востребованности среди студентов. Основной формой контента на таких ресурсах являются видеокорсы.

При поиске интересующих электронных курсов пользователь вынужден просматривать множество различных образовательных платформ. При этом каждая из них имеет свой интерфейс, классификацию тематик и параметры сортировки курсов. Данные различия платформ, а также постоянное изменение и расширение набора курсов существенно усложняют поиск [4].

Для решения этой проблемы было предложено разработать рекомендательную систему, помогающую пользователю в подборе электронных курсов. Такая система будет состоять из двух модулей: модуля сбора и обработки контента из сети Интернет и модуля формирования профиля пользователя и построения персональных рекомендаций.

Была разработана архитектура приложения, сформированы модели курса и профиля пользователя. Разработана мера связи между курсами. Разработан и реализован алгоритм составления единой классификации курсов, а также алгоритм актуализации значений профиля, который осуществляет автоматическое обновление данных на основе действий пользователя, где под действием понимается любое взаимодействие пользователя с курсами или профилем собственных предпочтений.

1. Яхьяева Г. Э., Абсайдульевой А. Р. Семантический подход к моделированию фонда оценочных средств // Вестн. НГУ. Сер. информ. технологии. 2018. Т. 16, вып. 2. С. 113–121.

2. Зыкова А. А., Яхьяева Г. Э. Программная система «Абитуриент — Студент» // Вестн. МГПУ. Сер. Информатика и информатизация образования. 2020. № 1(55).

3. Баталин К. В., Яхьяева Г. Э. Система управления оценочными средствами // Вестн. НГУ. Сер. информ. технологии. 2020. Т. 18, № 2. С. 5–14.

4. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000. 384 с.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

**Разработка гибридных моделей
на основе синтеза логического вывода и нейронных сетей**

Д. В. Протасов

Новосибирский государственный университет

Проблема интерпретации результатов работы нейронных сетей сопряжено с тем, что сети представляют из себя «черный ящик», не позволяющий определить, как был получен результат. Поэтому была предпринята попытка создания системы, позволяющей интерпретировать результат в удобном для человеческого восприятия виде.

В системе предполагается наличие онтологии, из которой на вход нейросети поступают данные, сама нейросеть (в рассматриваемом случае решающая задачу классификации), механизм вывода и связующий модуль.

Механизм вывода основан на технологии Semantic Web и позволяет на основе данных онтологий, описанных на языке OWL DL, получить требуемую интерпретацию с помощью машины логического вывода [1, 2]. Для этих целей подходит фреймворк DL-Learner.

Связующий модуль ответственен за непосредственное управление (опциональным) запуском нейросети, анализом данных и отправкой обработанных данных механизму вывода.

Модуль предоставляет возможность работы в 3 режимах: **полном цикле работы**: работа нейросети, автоматический анализ разбиения на классы; **частичном цикле работы**: выход нейросети уже существует, передаются данные о разбиении на классы; **ручном вводе** конфигурационного файла о разбиении на классы.

Тестирование модели показало, что прототип уже способен решать задачи с малым объемом данных.

1. Корсун И. А., Пальчунов Д. Е. Теоретико-модельные методы извлечения знаний о смысле понятий из текстов естественного языка // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 3. С. 34–48.

2. Капустина А. И., Пальчунов Д. Е. Разработка онтологической модели тарифов и услуг сотовой связи, основанной на логически полных определениях понятий // Вестн. НГУ. Сер. информ. технологии. 2017. Т. 15, № 2. С. 34–46.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Оценка лавинной опасности по метеорологическим измерениям на выборке с несбалансированными классами

Н. А. Радеев

Новосибирский государственный университет

Оценка лавинной опасности является сложной задачей, для решения которой нужно специальное образование и соответствующий опыт. Также данный процесс требует присутствия специалиста в месте, для которого производится оценка. А количество тестов, которые необходимо провести над снежным покровом для получения оценки, приближается к сорока. Все это делает задачу оценки лавинной опасности трудноразрешимой.

Целью данной работы была разработка базового решения, основанного только на метеорологических данных. Для этого использовался набор данных с измерениями метеорологических показателей за 20 лет и данные о лавинной опасности за тот же период времени.

Были решена задача выделения из данных признаков, несущих в себе важную информацию о состоянии погоды на момент оценки лавинной опасности. Была решена проблема несбалансированных классов. Так как дней с максимальной и минимальной лавинной опасностью меньше, чем со средней, то некорректно обучать алгоритм классификации на таких данных без коррекций. Для этого был использован метод взвешивания классов: редкие классы имеют больший вес при обучении алгоритма.

Полученные результаты доказали, что одних метеорологических данных недостаточно для оценки лавинной опасности, чем подтвердили гипотезу о том, что решающим фактором в оценке лавинной опасности является состояние снежного покрова. Анализ результатов также показал, что резкие изменения погодных условий являются фактором повышения лавинной опасности.

1. Pozdnoukhov R. S. Purves M. Kanevski Applying machine learning methods to avalanche forecasting // Ann. Glaciology. 2008. No. 49.

2. Головин Н. А., Савин Н. П., Яхьяева Г. Э. Применение методов машинного обучения для структурирования базы прецедентов компьютерных атак // Материалы Междунар. конф. «Знания — Онтологии — Теории». Новосибирск, 2019. С. 122–128.

3. Радеев Н. А. Сбор и предобработка данных для системы предсказания лавинной опасности // Тез. докл. Междунар. конф. «Мальцевские чтения — 2020», Новосибирск, 2020. С. 91.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Мультимодальное распознавание эмоций при видеосвязи

О. А. Седухин

Новосибирский государственный университет

Система автоматического распознавания эмоций на основе машинного обучения может иметь много разных применений: оценка удовлетворенности клиента или слушателя, придание эмоционального интеллекта электронному собеседнику, создание игр, в которых сюжет может меняться в зависимости от эмоций игрока.

В первую очередь в докладе уделено внимание существующим на данный момент наборам данных для задачи распознавания эмоций. Далее на практике сравниваются несколько методов решения задачи.

1. Построение архитектуры модели по стандартному подходу: экстракторы признаков и «голова». В докладе рассмотрены два основных варианта архитектуры «голова»: *late fusion* и *early fusion*, сравнивается их эффективность. Также очень важно качество работы каждого экстрактора признаков по отдельности. В докладе описывается, каким образом можно его измерить и повысить.

2. Обучение генеративно-состязательной сети с целью получения хорошего экстрактора признаков с использованием большого количества неразмеченных обучающих данных, которые легко доступны в интернете.

3. Использование семантической модели [1, 2] для построения логических выводов на основе полученных нейронными сетями промежуточных данных, а также для улучшения качества псевдоразметки в задаче автоматического расширения набора обучающих данных.

4. Использование сетей с механизмом внимания, получивших распространение в последние годы. Такие сети являются совершенно новым подходом к анализу видео.

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Долгушева Е. В., Пальчунов Д. Е. Теоретико-модельные методы порождения знаний о предпочтениях абонентов мобильных сетей // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 2. С. 5–16.

Научный руководитель — д-р физ.-мат. наук, доц. Д. Е. Пальчунов

Исправление конфликтов во множестве вероятностных событий

В. А. Скокова

Новосибирский государственный университет

В середине прошлого века появилось новое направление в исследовании искусственного интеллекта, которое получило название «экспертные системы». Такие системы эмулируют действия человека-эксперта в определенной области.

Классическая экспертная система содержит базу знаний, которая используется для вывода результата. Зачастую эксперту необходимо дать оценку событиям, которые находятся в базе знаний и могут быть связаны между собой. Информация может иметь значительный объем и содержать противоречивые данные [1, 2], и время, затраченное на ее анализ, может быть очень большим. Кроме того, эксперт может допустить ошибку.

В рамках данной работы предлагается алгоритм поиска и исправления конфликтов во множестве вероятностных событий. События представляются в виде формул логики предикатов, представляющих вероятностные события, и их вероятностных значений, представляющих субъективную меру уверенности эксперта в их действительности [3]. По множеству формул строится система линейных алгебраических уравнений с ограничениями. Несовместность полученной системы указывает на противоречивость данных. В таком случае предлагаются новые решения системы, удовлетворяющие ограничениям. По этим решениям можно скорректировать вероятности, введенные экспертом.

1. Пальчунов Д. Е., Яхьяева Г. Э. Нечеткие алгебраические системы // Вестн. НГУ. Сер. математика, механика, информатика. 2010. Т. 10, вып. 3. С. 75–92.

2. Palchunov D., Yakhyaeva G. Application of Boolean-valued models and FCA for the development of ontological model. CEUR Workshop Proceedings, 1921.

3. Яхьяева Г. Э., Карманова А. А., Ершов А. А., Савин Н. П. Вопросно-ответная система для управления информационными рисками на основе теоретико-модельной формализации предметных областей // Информационные технологии. 2017. Т. 23, № 2. С. 97–106.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

**Разработка итерационного алгоритма
для построения лингвистических шаблонов, находящих схожие
по смыслу грамматические конструкции русского языка**

А. Ю. Смирнов

Новосибирский государственный университет

В современном мире существуют информационные системы, работа которых зависит от входных данных, представленных в виде неструктурированных текстов на естественном языке, например: поисковые системы, голосовые помощники. Чтобы корректно обработать поданные данные, необходимо понять, какой смысл несут входящие в текст языковые конструкции. Для этого существуют специальные алгоритмы, которые в своей работе полагаются на применение лингвистических шаблонов. Лингвистический шаблон — описание языковой конструкции, ее лексического состава и грамматических свойств. Пример лингвистического шаблона: *N работает в NP*, где *N* — имя существительное, *NP* — группа имен существительных. С помощью лингвистического шаблона можно сопоставить абстрактную языковую конструкцию определенной группе конкретных языковых конструкций и обработать их соответствующим образом. На данный момент существуют готовые реализации алгоритмов, позволяющие создавать лингвистические шаблоны для английского языка. Для русского же языка все шаблоны разрабатываются вручную, что является существенным недостатком.

Целью данной работы является разработка алгоритма, который бы позволял искать лингвистические шаблоны для текста на русском языке в автоматическом режиме, причем данный алгоритм должен быть итерационным, т. е. использовать результат предыдущего шага, тем самым постоянно увеличивая количество найденных шаблонов.

В докладе рассматриваются:

- существующие подходы для построения лингвистических шаблонов в автоматическом режиме;
- архитектура и алгоритм приложения;
- практические результаты использования разработанного алгоритма по сравнению с существующими реализациями.

Научный руководитель — д-р техн. наук, доц. В. Б. Барахнин

Разработка интеллектуального голосового интерфейса для помощи людям с проблемами зрения в навигации внутри помещений

А. С. Трегубов

Новосибирский государственный университет

Взаимодействие с людьми, ограниченными в возможностях зрительного восприятия с помощью визуального интерфейса, не представляется возможным. Однако развитие современных технологий и анализа тестов на естественном языке позволяет создавать голосовые интерфейсы, особенно актуальные для слепых и слабовидящих людей.

Цель работы — создание интеллектуального голосового интерфейса в виде электронного ассистента для описанной предметной области.

В рамках работы был создан веб-сервис для интеллектуального голосового помощника. Разработана структура базы данных с учетом дальнейшей интеграции с онтологическими моделями. Реализован функционал для декомпозиции пользовательских задач и формирования последовательности их выполнения с учетом ограничений. Для тестирования был реализован чат-бот на платформе Telegram.

Преимуществом разрабатываемого диалогового интерфейса является возможность формирования уточняющих вопросов для выявления потребностей пользователя за счет использования 4-уровневой онтологической модели предметной области [1, 2].

В дальнейшем планируется развитие работы и интеграция с системой навигации внутри помещений для людей с ограниченными возможностями, состоящей из серверной части, мобильного приложения, трекера, надеваемого на руку пользователя, и системы датчиков, расположенных по периметру помещения.

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Найданов Ч. А., Пальчунов Д. Е., Сазонова П. А. Разработка автоматизированных методов предупреждения рисков возникновения критических состояний, основанных на анализе знаний, извлеченных из историй болезней пациентов // Сиб. науч. мед. журн. 2016. Т. 36, № 1. С. 105–113.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

Разработка методов извлечения и обработки аргументов из текстов естественного языка, основанных на логике предикатов

Д. Усова

Новосибирский государственный университет

Тематика моделирования рассуждений на данный момент актуальна, для нее не существует универсального решения. Возможность предоставлять автоматическую аргументацию была бы полезна в таких областях, как медицина, политика, социология и т. д. Многие существующие решения предлагают лишь анализ синтаксической структуры предложений, при этом теряется семантика утверждений. Многие системы оперируют числами, но при этом теряют множество деталей, которые важны для понимания смысла аргументации. Также существует ряд проблем (например, проблема омонимичных слов), решение которых не является тривиальным.

Предлагается подход на основе шаблонов рассуждений и применения специальной структуры — ситуации, с помощью которой формулируется понятие аргумента. Ситуации строятся при помощи фрагментов атомарных диаграмм и необходимы для отражения смысла утверждений, а не только их синтаксической структуры [1]. Чтобы получать новые знания из извлеченных из текстов аргументов, вводится ряд отношений между ситуациями, при этом часть отношений построена на онтологических связях между понятиями [2]. Данный подход позволяет переходить от одних аргументов к похожим аргументам. При осуществлении этого перехода применяется отношение сходства, что позволяет действительно извлекать новые знания [3]. Разработаны методы перехода от имеющихся аргументов к похожим и порождения аргументов, подобных данным.

Конечной целью является построение теории аргументации, основанной на формализации ситуаций, и реализация разработанных методов в программной системе, предназначенной для извлечения, анализа и поиска аргументации для текстов и утверждений естественного языка.

1. Корсун И. А., Пальчунов Д. Е. Теоретико-модельные методы извлечения знаний о смысле понятий из текстов естественного языка // Вестн. НГУ. Сер. информ. технологии. 2016. Т. 14, № 3. С. 34–48.

2. Власов Д. Ю., Пальчунов Д. Е., Степанов П. А. Автоматизация извлечения отношений между понятиями из текстов естественного языка // Вестн. НГУ. Сер. информ. технологии. 2010. Т. 8, № 3. С. 23–33.

3. Пальчунов Д. Е. Поиск и извлечение знаний: порождение новых знаний на основе анализа текстов естественного языка // Философия науки. 2009. № 4(43). С. 70–90.

Научный руководитель — Д. Е. Пальчунов

Разработка чат-ботов, поддерживающих анализ и выражение эмоций

Г. К. Хлиманкова

Новосибирский государственный университет

Целью данной работы является изучение технологий, поддерживающих анализ и выражение эмоций. Актуальность данной темы состоит в том, что в современном мире компьютерная графика и анимация применяются в огромном количестве областей: науке, промышленности, архитектуре, образовании и др. Анализ эмоций полезен при создании интеллектуальных помощников, общающихся с пользователем на естественном языке [1], для точного определения желаний и потребностей пользователя [2].

Мимику компьютерного персонажа убедительной и выразительной позволяет сделать риг, т. е. система контроллеров для управления деформациями объекта. Риг позволяет аниматору управлять движением тела и лица персонажа. Для классификации эмоций и разработки алгоритма применена технология FACS, созданная Уоллесом и Экманом. Эта технология оперирует так называемыми единицами действия, имеющими свои коды. В ней исследованы и описаны все существующие движения губ, которые могут вытягиваться, раздвигаться, стягиваться, расплываться, сужаться и сжиматься.

Разрабатывается приложение, осуществляющее риггинг лица персонажа. Его основной функцией является рисование эмоций / изменение положения частей лица у персонажа по настройкам пользователя. Строятся основные шаблоны, по которым можно визуальнo менять эмоции, для реализации этого подхода разрабатывается предметно-ориентированный язык [3].

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Palchunov D., Yakhyaeva G., Dolgusheva E. Conceptual methods for identifying needs of mobile network subscribers // CEUR Workshop Proceedings. 2016. 1624. P. 147–160.

3. Gumirov V. S., Matyukov P. Y., Palchunov D. E. Semantic Domain-specific Languages // Int. Multi-Conf. Eng., Computer Inf. Sci. (SIBIRCON). Novosibirsk: IEEE Press., 2019. P. 0955–0960.

Научный руководитель — д-р физ.-мат. наук, доц. Д. Е. Пальчунов

Рекомендательная система поддержки диалога AskMeSmth

М. М. Хлыбова

Новосибирский государственный университет

В настоящий момент из-за ряда факторов растет число людей, имеющих проблемы с коммуникацией. Один из подходов развить коммуникационные навыки — обсуждение различных тем, которые могут формироваться исходя из вопросов. Вопросы могут различаться по структуре и предметной области [1–3]. Так как с их подбором могут возникнуть различные сложности, одним из решений данной проблемы является приложение, которое будет предлагать пользователям вопросы на различные темы.

В ходе обзора похожих приложений одним из параметров оценки был дизайн. Для приложения необходим лаконичный и интуитивно понятный дизайн. Ряд решений был принят на основании сравнения с дизайном многих часто используемых систем: темная и светлая тема, переключение элементов с помощью свайпов.

Индивидуальность вопросов для каждого пользователя — одна из особенностей системы, выделяющая ее среди других аналогов. Данную функциональность осуществляет рекомендательная система, реализованная на основании алгоритма GroupLens, с использованием различных расширений данного алгоритма для обработки большего объема данных. Также рассмотрены подходы к решению проблем «холодного старта» как для новых оцениваемых элементов (вопросов), так и для новых пользователей.

Так как база вопросов пополняется, то возникает вопрос дублирования содержания вопросов и поиска решений, касаемых уникальности вопросов в системе. Вводится метрика схожести между текстами вопросов.

1. Яхьяева Г. Э., Ясинская О. В., Карманова А. А. Вероятностная вопросно-ответная система в области компьютерной безопасности // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, вып. 3. С. 132–145.

2. Яхьяева Г. Э. и др. Вопросно-ответная система для управления информационными рисками на основе теоретико-модельной формализации предметных областей // Инф. технологии. 2017. Т. 23, № 2. С. 97–106.

3. Palchunov D., Yakhyaeva G. Representation of Knowledge Using Different Structures of Concepts // CEUR Workshop Proceedings. 2020. Vol. 2729. С. 69–74.

Научный руководитель — О. В. Данилова

Разработка системы поддержки документооборота для управления экспертизой проектов

Д. А. Худяков

Новосибирский государственный университет

Информационные технологии на сегодняшний день глубоко проникают во все виды деятельности. Также информационные технологии проникают как в образовательную деятельность [1, 2], так и в научную сферу деятельности.

Для того чтобы получить грант научной группе, необходимо пройти строгий конкурсный отбор. Заявки делятся по экспертным секциям и управляются их координаторами. На каждую конкурсную заявку координатор экспертной секции должен найти не менее двух разбирающихся в тематике заявки экспертов.

Кроме того, в рамках взаимодействия координатора секции и эксперта между ними происходит обмен множеством различных документов. С ростом количества отводимых для управления координатору заявок увеличивается его нагрузка, что повышает вероятность ошибки и недосмотра. Основная задача координатора — наиболее правильно выбрать эксперта и в случае получения экспертного заключения убедиться в его качестве. Вместе с этим координатор занимается заполнением и организацией хранения сотен файлов, отправкой и отслеживанием почты, на что тратит время и силы.

В связи с данной проблемой было решено создать веб-приложение, предоставляющее координаторам экспертных секций помощь в управлении экспертизой. Разработанное приложение помогает частично автоматизировать бизнес-процесс взаимодействия координатора и эксперта тем, что берет на себя заботы по генерации документов из имеющейся об экспертах информации, отправке почты и хранению документов на основных шагах бизнес-процесса.

1. Яхьяева Г. Э., Абсайдульевой А. Р. Семантический подход к моделированию фонда оценочных средств // Вестн. НГУ. Сер. информ. технологии. 2018. Т. 16, № 2. С. 113–121.

2. Баталин К. В. и др. Программная система управления образовательным процессом ИТОС // Вестн. НГУ. Сер. информ. технологии. 2018. Т. 16, № 4. С. 20–30.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

**Разработка модуля классификации орфографических ошибок
для автоматизированной обучающей системы**

Е. С. Шевчук

Новосибирский государственный университет

С развитием человечества все большему количеству людей стало необходимо писать множество текстов, что нужно делать без ошибок. Для того чтобы их не допускать, нужно знать правила русского языка либо можно использовать анализаторы текста. Но они только исправляют ошибки, не объясняют, в чем она заключается. Это ведет к заучиванию наизусть правильного написания слов, что малоэффективно в смысле повышения грамотности.

Одним из решений данной проблемы может быть изучение допущенных в тексте ошибок, подкреплённых правилами, что в дальнейшем поможет не тратить время на использование программ для проверки текстов. Для классификации с помощью алгоритма C4.5 был составлен обучающий набор данных [1, 2]. Для его оптимального размера часть классов включает несколько правил русского языка. По этой же причине часть ошибок не входит в обучающее множество. Разработанный модуль работает по следующему алгоритму: после нахождения ошибки в словах каждая из них будет предварительно проанализирована. Затем, если ошибка не определена, заполняется ряд признаков, которые будут использованы при классификации с помощью построенного на обучающем множестве дерева решений. После определения класса ошибки определяется правило, по которым совершена ошибка. Таким образом, в ходе данной работы было сформировано обучающее множество и разработан подробный алгоритм классификации ошибок с последующим анализом для выдачи результата пользователю в удобном для него виде.

1. Шевчук Е.С. Алгоритм классификации орфографических ошибок на основе алгоритма машинного обучения C4.5 // Тез. докл. Междунар. конф. «Мальцевские чтения — 2020». Новосибирск, 2020. С. 102.

2. Breiman L. et all. Classification and Regression Trees. Wadsworth, Belmont, California, 1984.

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

Сравнение алгоритмов активного обучения на задачах классификации изображений

А. С. Щербин

Новосибирский государственный университет

При разработке программных систем, использующих методы машинного обучения, часто возникает необходимость добавить обучающие примеры. Увеличение количества данных при достаточной их информативности значительно улучшает качество алгоритмов машинного обучения. Подход, использующий имеющийся алгоритм для выбора примеров, которые следует добавить в обучающее множество, называется активным обучением [1].

Для оценки алгоритмов активного обучения часто используют кривую обучения, где по оси абсцисс указывается количество примеров в обучающем множестве, а по оси ординат — значение метрики качества алгоритма. Также важным параметром является задача, которую решает алгоритм машинного обучения.

Целью данной работы является сравнение разработанного ранее алгоритма активного обучения на основе метрики информативности [2] с существующими классическими методами активного обучения. Для исследования были использованы два набора данных: Imagenette и Imagewoof [3]. Первый набор данных содержит изображения различных бытовых вещей, в нем присутствуют следующие классы: лещ, английский спрингер-спаниель, кассетный магнитофон, бензопила, церковь, валторна, мусоровоз, бензопилка, мяч для гольфа, парашют. Второй набор содержит изображения собак 10 различных пород. Задача классификации пород собак оказалась более сложной, чем задача классификации бытовых предметов.

1. Settles B. Active Learning Literature Survey // Comput. Sci. Tech. Rep. 1648. University of Wisconsin-Madison, 2010. P. 3–15.

2. Щербин А. Разработка методов активного обучения для улучшения качества модели классификации изображений // Материалы 58-й Междунар. науч. студ. конф. МНСК-2020. Сер. инфор. технологии. Новосибирск: НГУ, 2019. С. 180.

3. Github репозиторий FastAI [Электронный ресурс]. URL: <https://github.com/fastai/imagenette> (дата обращения: 18.02.21).

Научный руководитель — канд. физ.-мат. наук, доц. Г. Э. Яхьяева

**Проектирование общей структуры
интеллектуального помощника в контексте экспертных систем
и систем поддержки принятия решений**

А. А. Якобсон

Институт математики им. С. Л. Соболева СО РАН, Новосибирск
Новосибирский государственный университет

В современном мире виртуальные цифровые помощники стремительно развиваются в разных областях, таких как интернет-маркетинг, автоматизация бизнес-процессов.

Ядро системы состоит из нескольких ключевых компонентов: интерфейса человек-компьютер [1], модуля интерпретации ввода, базы знаний предметной области [2], интеллектуального модуля анализа поведения пользователя [3]. Основной функционал обеспечивает интеллектуальный модуль анализа поступающих в помощника запросов. Он включает интерфейс взаимодействия с базой знаний, контекстные алгоритмы или нейросети, выполняющие обработку данных с опорой на базу знаний. При этом система должна допускать использование нескольких логических блоков анализа данных. Этого можно достигнуть, разработав интерфейс обмена информацией между блоками. Полученная система будет масштабируема в смысле количества аналитических компонентов, единственным ограничением будет совместимость компонентов между собой, а также доступной базой знаний.

В результате разработана архитектура виртуального цифрового помощника, пригодная для использования в различных предметных областях, в зависимости от используемой базы знаний.

1. Деревянко Д. В., Пальчунов Д. Е. Формальные методы разработки вопросно-ответной системы на естественном языке // Вестн. НГУ. Сер. информ. технологии. 2014. Т. 12, № 3. С. 34–47.

2. Palchunov D., Yakhyaeva G., Yasinskaya O. Software system for the diagnosis of the spine diseases using case-based reasoning // Proc. Int. Conf. Biomed. Eng. Comput. Technol. (SIBIRCON/SibMedInfo — 2015). Novosibirsk, 2015. P. 205–210.

3. Palchunov D., Yakhyaeva G., Dolgusheva E. Conceptual methods for identifying needs of mobile network subscribers // CEUR Workshop Proceedings. 2016. 1624. P. 147–160.

Научный руководитель — д-р физ.-мат. наук Д. Е. Пальчунов

СЕТЕВЫЕ ТЕХНОЛОГИИ, ТЕХНОЛОГИИ ИНТЕРНЕТА И ИХ ПРИЛОЖЕНИЯ

УДК 004.043

Анализ и оптимизация радиочастотного покрытия мобильных сетей

С. Т. Афанасьев

Новосибирский государственный университет

Сегодня компании-операторы сотовых сетей уделяют большое внимание вопросам оптимизации радиочастотного покрытия, от которого зависит качество предоставляемых услуг. Сигналы мобильных сетей распространяются базовыми станциями оператора, рациональное расположение которых в городских условиях требует сбора колоссального количества точных данных и их сложного математического анализа. Распространению сигналов в городах препятствуют плотная застройка и эффекты интерференции электромагнитных полей.

Данная работа посвящена разработке программного обеспечения, которое предоставляет функционал автоматизированного сбора анонимных данных для смартфонов абонентов, и последующему анализу собранных данных.

Целями работы являются:

1) исследование уровня и качества сигналов сотовых сетей в Новосибирске для поиска проблемных участков и предоставления рекомендаций по их устранению;

2) расширение функционала программного комплекса, созданного в рамках ВКР бакалавра, для проведения исследования эффекта интерференции мобильных сетей в городских условиях.

В ходе работы были реализованы и протестированы следующие функционалы:

1) сбор и отправка данных о качестве соединения с привязкой к геолокации и временной отметкой с мобильных устройств абонентов;

2) построение и визуализация карт качества покрытия по данным накопленной статистики в виде слоев.

Практической ценностью данной работы является автоматизированный сбор данных непосредственно с мобильных устройств пользователей, что позволит при анализе получить максимально правдоподобную картину уровня качества мобильных сетей, с точки зрения абонентов. Внедрение данного комплекса в работу операторов позволит сократить издержки на измерения показателей их сотовых сетей.

Научный руководитель — канд. техн. наук В. Г. Дроздова

Информационная система «Позвоночные животные Бурятии»

З. Б. Балданов

Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ

В век информационных технологий все актуальнее становится проблема хранения собранных и проанализированных данных с полевых исследований. В связи с этим важной задачей становится разработка информационного ресурса для публикации накопленных данных в открытом доступе в сети Интернет.

В Бурятском государственном университете проведена работа по описанию позвоночных животных Бурятии. База охватывает все виды позвоночных животных, зарегистрированных на территории Бурятии, включает 556 видов из 6 классов. В базе содержатся сведения по распространению видов на территории Бурятии и за ее пределами, информация о регистрациях залетных видов птиц и редких животных, а также приводятся сведения о динамике численности и лимитирующих факторах редких видов и краткие описания морфологии и биологии видов. Список видов в базе составлен в соответствии с последними изменениями таксономии и номенклатуры. На данный момент база данных представлена в таблице Excel.

Была поставлена задача разработать информационный ресурс для публикации описанной базы в сети Интернет.

Первым этапом работы является анализ предметной области и разработка прототипа информационного ресурса. На данный момент разработана ER-модель и структура базы данных. Прототип информационной системы находится на этапе разработки.

Научный руководитель — Б. В. Хабитуев

Реализация онлайн-сервиса по анализу данных локального сейсмического мониторинга

М. А. Бедарева

Новосибирский государственный университет

В связи с развитием и удешевлением сейсмической аппаратуры расширяются возможности установки локальных сетей сейсмического мониторинга. Они помогают осуществлять контроль над техногенной сейсмичностью или состоянием зданий и сооружений: устойчивостью и цельностью конструкции. Также достаточно хорошо отработана технология сбора и хранения данных. Но для эффективного решения прикладных задач возникает потребность в онлайн-сервисе по анализу данных сейсмического мониторинга для доступа разного рода специалистов к различным данным сетей мониторинга.

Целью является создание удобного программного инструмента для работы с данными мониторинга локальной сейсмичности и устойчивости сооружений с возможностью запрашивать данные с сервера и проводить их анализ. В результате работы был реализован алгоритм анализа длинных рядов сейсмических записей с накоплением амплитудных спектров колебаний. Затем на спектрах выделяются резонансные частоты, критерием выделения является наличие четких максимумов для трех кратных частот.

Разработан сервис на основе Django-Dash приложения на языке программирования Python с использованием библиотеки Plotly для визуализации данных, а также языков HTML и CSS. Онлайн-сервис реализует следующие функции:

- 1) запрос и доступ к данным;
- 2) анализ данных (выделение резонансных частот и магнитуды сейсмического события);
- 3) инструменты интерактивной графики;
- 4) отсылка экстренных сообщений при превышении пороговых значений.

Научный руководитель — канд. физ.-мат. наук А. А. Дучков

Построение прототипа системы для быстрой разработки веб-приложений в целях обучения в подходе, альтернативном MVC

О. А. Зенина

Новосибирский государственный университет

В современной информационно-образовательной среде веб-ресурсы являются одним из наиболее важных источников информации.

Общая тенденция развития требований к веб-ресурсам характеризуется более сложной структурой данных, развитым гипертекстовым пространством, динамическим интерфейсом, использованием мультимедиа компонентов, большим объемом содержащейся информации.

Отсюда вытекает ряд проблем, например: поддержка целостности ресурсов, длительный и трудоемкий процесс разработки сложной гипертекстовой структуры, предоставление различных интерфейсов для разных типов пользователей.

Для целей более быстрой разработки в данный момент применяются два подхода: использование фреймворков или CMS. Первый подход более гибкий, но доступен только разработчикам с некоторым опытом, второй предназначен для пользователей, которые не хотят обучаться веб-программированию.

Поэтому построение системы, использование которой было бы возможно новичкам в программировании для создания современных веб-ресурсов, является актуальной темой для современности.

Целью данной работы является построение системы для быстрой разработки веб-ресурсов.

Для достижения данной цели было предложено ввести уровень метамодели данных, основанный на объектно-ориентированных парадигмах. Физическим представлением данного уровня является некоторое количество таблиц в базе данных, которые содержат служебную информацию о данных строящегося ресурса, например: типы зарегистрированных в системе данных, типы разрешенных связей между элементами различных типов данных и т. д.

Данный подход был использован при построении системы, с помощью которой можно обучаться веб-программированию и быстро создавать веб-ресурсы.

Научный руководитель — В. Г. Казаков

Разработка мобильного приложения студента НГТУ НЭТИ под платформы iOS и Android

А. П. Коваленко, Д. К. Ковалева

Новосибирский государственный технический университет

В связи с пандемией коронавируса все больше и больше ощущается необходимость связи преподавателей и деканата со студентом, и старые способы коммуникации не справляются с этой задачей. В последнее время почти у каждого имеется при себе мобильное устройство, а мобильные приложения стали революцией XXI в.

Разработка мобильного приложения является развитием проекта, описанного в [1]. В приложение включено все, что необходимо студенту на ежедневной основе: авторизация в информационной системе университета, просмотр расписания, просмотр успеваемости за контрольные недели и семестры, просмотр сообщений от преподавателей и деканата, получение уведомлений о новых сообщениях, получение уведомлений о начале пары, доступ к ресурсам электронно-библиотечной системы. Так как в вузе также обучаются и иностранные студенты, была реализована система, которая достаточно просто и удобно предоставляет локализацию приложения на иностранные языки.

Для написания приложения под платформу iOS использовался язык Swift, а также библиотеки RxSwift, Moya, SwiftGen и SwiftLint. В качестве основной архитектуры была выбрана архитектура MVVM-C.

Для написания приложения под платформу Android использовался язык Kotlin, а также библиотеки Retrofit, RxJava2, Jetpack. В качестве основной архитектуры была выбрана MVVM.

Для управления процессом разработки использовались Issues в Gitlab, для разработки дизайна выступал онлайн-сервис Figma. Для обеих платформ была создана локализационная таблица, которая позволяет при компиляции приложения генерировать локализационные строки.

1. Шибалова Ю. В., Коваленко А. П. Разработка мобильного приложения обучающегося НГТУ НЭТИ // Материалы 58-й Междунар. науч. студ. конф. МНСК-2020. Сер. информ. технологии. Новосибирск: ИПЦ НГУ, 2020. С. 24.

Научные руководители —
канд. техн. наук, доц. В. М. Стасышин, О. Е. Аврунев, М. А. Городничев

Реализация распределенного кэширования HTTP-запросов для сайтов с динамическим содержимым на основе реляционных баз данных

Т. Л. Манасян

Новосибирский государственный университет

За последнее десятилетие большинство веб-приложений перешло от формата выдачи сервером готовой HTML-страницы клиенту на формат сайтов с динамическим содержимым, когда для построения одной страницы клиент может отправлять до нескольких десятков запросов на сервер. В связи с этим нагрузки на серверные части веб-приложений, предоставляющие нужный контент, резко возросли, поэтому особенно актуальной является проблема увеличения их производительности. Одним из решений данной проблемы является кэширование обрабатываемых сервером HTTP-запросов.

Целью данной работы является разработка библиотеки для распределенного кэширования HTTP-запросов для сайтов с динамическим содержимым на основе БД.

Для достижения данной цели было выбрано HTTP-кэширование на основе сильного валидатора ETag. Была разработана архитектура библиотеки и составлен список требований к итоговому решению. Разработанная библиотека использует состояния таблиц БД серверной части веб-приложения, поэтому далее была выбрана распределенная СУБД для хранения информации о данных состояниях, обновляющихся в автоматическом режиме при каждом обновлении соответствующей таблицы.

В существующих решениях при вычислении значения заголовка ETag, серверной части веб-приложений приходится выполнять бизнес-логику, включая обращения к БД. Это может существенно увеличить время обработки запросов. Новизна работы заключается в том, что при использовании библиотеки при вычислении ETag бизнес-логика приложения не затрагивается. Значение заголовка строится исключительно на основе состояний таблиц, от которых зависит результат конкретного HTTP-запроса.

Практическая ценность работы состоит в уменьшении времени обработки HTTP-запросов, зависящих от данных из БД, серверными частями веб-приложений, уменьшении сетевого трафика и затрат времени на парсинг HTML-страниц клиентскими частями при использовании разрабатываемой библиотеки.

Научный руководитель — д-р техн. наук, доц. В. Б. Баракнин

Система управления задачами

А. О. Мирошниченко, Н. В. Беребердин, Е. А. Литвинов
Городской дворец детского творчества, Омск

Существует большое количество различных ежедневников и планировщиков задач, например Google Календарь или Googlekeep, но у них есть недостаток: при помощи этих систем невозможно отслеживать выполнение задач учениками. Заявка на создание данного приложения от психологической службы Дворца творчества была продиктована востребованностью внедрить в образовательный процесс технологию тайм-менеджмента.

Цель работы — разработать клиент-серверное веб-приложение для системы управления задачами, способное формировать группы исполнителей, создавать задачи, назначать задачи на группы исполнителей и отслеживать их выполнение.

Задачи:

- изучить запрос на разработку системы от психологической службы,
- разработать клиентское веб-приложение для отображения и управления данными,
- разработать серверное приложение для хранения и обработки данных,
- организовать взаимодействие клиентского и серверного приложений.

Средства разработки: язык программирования JavaScript, библиотека Bootstrap, фреймворк веб-приложений Express для Node.js, AJAX-запросы.

Клиентская часть, т. е. то, с чем пользователь непосредственно взаимодействует, позволяет создавать и обрабатывать задачи. Визуальный интерфейс позволяет пользователю управлять списком своих дел, например: создавать, изменять, отмечать выполненными или удалять их.

Серверная часть позволяет хранить и предоставлять доступ к данным пользователя и его делам. Создан механизм регистрации и аутентификации пользователей на сайте. Все данные пользователей хранятся в файле .json в виде массива с объектами. Создан механизм для подтверждения адресов электронной почты. Реализован функционал, позволяющий обрабатывать запросы пользователя на создание, изменение, завершение и удаление дел, а также запросы на регистрацию в сервисе и последующий вход в аккаунт. Пароли хранятся в виде хэш-сумм, которые создаются однонаправленной криптографической функцией SHA-256.

Адрес ресурса клиента: <https://k16-omsk.ru/diary>. Адрес ресурса администратора: <https://k16-omsk.ru/admdiary>.

Научный руководитель — И. А. Циглер

Обновление информации в веб-интерфейсе NSUts при помощи push-уведомлений

А. С. Мокроусов

Новосибирский государственный университет

NSUts — система автоматической оценки заданий по программированию. Она используется для проведения олимпиад и тренировок по спортивному программированию. Система проверяет отправленные участниками программы (решения) на то, что они проходят все тесты и укладываются в ограничения по времени и памяти. Проверка осуществляется на выделенных компьютерах — тестирующих клиентах.

В настоящее время данные в веб-интерфейсе NSUts (например, о статусе тестирования решения или объявлениях жюри) обновляются путем опроса сервера, что имеет ряд недостатков с точки зрения производительности и задержки обновления содержимого. Более логичной и эффективной альтернативой является передача данных от сервера по мере их появления (push-уведомления).

Был реализован набор компонентов, предоставляющих API для уведомлений о событиях, например об изменении статуса тестирования решения или появлении нового вопроса от участника олимпиады, а также для обработки этих уведомлений в веб-интерфейсе системы.

Реализация состоит:

- из серверного компонента, поддерживающего постоянные WebSocket, иницилируемые веб-интерфейсом и тестирующими клиентами (PHP/Ratchet);
- компонента для тестирующего клиента (Python);
- компонента для веб-интерфейса, позволяющего устанавливать обработчики для уведомлений (TypeScript).

Эти компоненты используются для обновления области уведомлений на странице веб-интерфейса (об ответах жюри на вопросы, новостях олимпиады и др.), информации о статусе тестирования решения и номере текущего теста, что показало работоспособность решения. В дальнейшем планируется исследовать его производительность в «боевых» условиях.

Научный руководитель — Д. В. Иртегов

Информационная система для проведения образовательных мероприятий

А. А. Мункоев

Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ

В последнее время в России постоянно растет число образовательных мероприятий для школьников и студентов. Одновременно с этим растет и интерес к этим мероприятиям. Например, среднее число участников первого этапа Всероссийской олимпиады школьников с 2011 по 2015 г. составляет 9 млн человек [1]. Ежегодно Институт математики и информатики проводит серию образовательных мероприятий для школьников. Общее число участников на этих мероприятиях превышает 1000 человек.

При организации образовательных мероприятий с большим количеством участников возникает ряд проблем: сбор и обработка информации, получение согласия на обработку персональных данных, распределение по аудиториям, рассылка результатов и т. д. Часть из этих проблем может быть решена при помощи внедрения информационной системы. В связи с этим была поставлена задача — разработать информационную систему, которая позволит частично автоматизировать процесс организации образовательных мероприятий.

Первым этапом работы является разработка прототипа информационной системы. На данный момент разработана ER-модель, структура базы данных и прототип системы. Прототип системы реализован с использованием РНР фреймворка Yii2 и готовится к внедрению.

1. Абатурова В. В. и др. Всероссийская олимпиада школьников в таблицах и цифрах // Академия повышения квалификации и профессиональной переподготовки работников образования. М., 2015. С. 41.

Научный руководитель — Б. В. Хабитуев

Проектирование и реализация веб-приложения «Расписание ИГУ»

В. А. Попова, Д. А. Гармаева
Иркутский государственный университет

Для организации представления расписания учебного процесса в настоящее время целесообразно применять программные системы, которые позволяют не только визуально отображать информацию, но и обеспечить оперативное внесение изменений.

В данный момент в Иркутском государственном университете отсутствует единая для всех подразделений информационная система отображения расписания учебного процесса.

Обзор существующих систем для представления расписания занятий показал, что многие особенности структуры учебного учреждения не могут быть учтены в таких программах.

Авторами ранее было создано веб-приложение «Расписание ИМИТ ИГУ» [1], которое успешно внедрено в 2019 г. в учебный процесс Института математики и информационных технологий ИГУ. С помощью инструмента «Яндекс.Метрика» установлено, что в среднем в день сайт посещают 220 человек (около трети общего контингента института). По результатам опросов посетителей о юзабилити веб-приложения была проведена доработка интерфейса, а также выполнена оптимизация приложения.

Для создания веб-приложения «Расписание ИГУ» выбраны технологии Node.js, Vue.js, MySQL. На текущий момент разработан пользовательский интерфейс, реализован механизм редактирования и просмотра расписания занятий, добавлена поддержка мобильной версии, настроен обмен данными с корпоративной информационной системой университета через HTTP-сервисы.

В настоящее время проводится тестирование разработанной функциональности, далее планируется приступить к нагрузочному тестированию.

*Исследование проведено при финансовой поддержке гранта
Иркутского государственного университета для молодых ученых
№ 091-20-302 «Разработка информационной системы “Расписание ИГУ”».*

1. Расписание ИМИТ ИГУ [Электронный ресурс]. URL: <http://raspmath.isu.ru/> (дата обращения: 17.02.2020).

Научный руководитель — канд. физ.-мат. наук, доц. А. С. Казимиров

Common Addon management System

В. Прокопенко
СУНЦ НГУ

Развитие, доработка и расширение являются неотъемлемой частью почти каждого проекта. Специально для этого были придуманы системы контроля версий. Они облегчают редактирование проекта, храня его различные версии. Однако этого может не хватать. Не все проекты создаются с нуля. У них появляются зависимости, файлы, взятые откуда-либо. В процессе жизни проекта зависимости меняются, добавляются и удаляются. Так, таких зависимостей может становиться очень много и управлять ими становится очень сложно. Также некоторые проекты создаются, чтобы на их основе создавались новые (например, языки программирования). Для этого пишутся специальные менеджеры пакетов или плагинов. Таких проектов накопилось уже довольно много (npm, pip, cargo и т. д.). И у каждого из них есть свой собственный менеджер пакетов.

Цель проекта — создание веб-сервиса (системы) для управления проектом, его зависимостями и расширениями. Также необходимо максимально упростить добавление зависимостей, создание дополнений и плагинов.

Сервис создан на фреймворке Django (язык программирования — Python). Интерфейс построен на HTML 5 и CSS 3. В качестве СУБД используется PostgreSQL.

Почти каждый элемент имеет свой стиль, чтобы можно было сразу определить, что в нем находится. Все проекты имеют темы — тэги. По ним легче искать нужные зависимости. Вся логика разделена на части, модули.

Сегодня в любом веб-сервисе используется JavaScript. Он применяется везде: в интерфейсе, логике, даже целые страницы. Нередко им злоупотребляют, реализовывая функции, которые уже есть в HTML и CSS. Так почему бы не создать сервис, не используя его вообще?

В результате был создан сервис, содержащий систему контроля версий, управления зависимостями и дополнениями. Любой проект или его часть можно добавить как зависимость, если это возможно. К любому проекту можно добавить расширение, если это возможно. Проекты сгруппированы по темам. Файлы проекта можно редактировать в веб-сервисе.

Научные руководители — И. В. Цимбалист, Г. Н. Цимбалист

Основные принципы разработки программного обеспечения устройства HTTP-фильтрации для промышленного интернета на платформе Linux

Д. А. Сафенрейтер
Новосибирский государственный университет

В области разработки ПО для промышленного интернета существует ряд проблем, которые не позволяют полностью задействовать ресурсы промышленной сети:

- необходимость повышения уровня безопасности промышленных сетей, один из путей которой связан с фильтрацией контента;
- проблема повышения пропускной способности канала с учетом фильтрации.

Осуществляя фильтрацию контента, можно повысить безопасность канала доступа в интернет, а также значительно увеличить уровень безопасности промышленной сети.

Модель передачи данных в сети Интернет базируется на стеке протоколов TCP/IP и прикладных протоколах. Информация проходит через четыре уровня, каждый из которых описывается протоколом передачи.

Основная работа ПО фильтрации будет сосредоточена на протоколе прикладного уровня HTTP, поскольку он используется внутри промышленной сети для доступа к внутренним веб-серверам и к веб-серверам за пределами сети. Выбранный метод фильтрации — отбор по URL-адресу, так как он наиболее широко применяется на данный момент.

Для того чтобы сделать фильтрацию, необходимо осуществить перехват пакетов трафика и выделить из них те, которые относятся к протоколу HTTP, сформировать из них сессии, выделить URL-адрес и проверить его на наличие в списках запрещенных/разрешенных ресурсов и пропустить или отфильтровать пакеты сессии.

Для непосредственного перехвата пакетов используются различные библиотеки. Одна из наиболее популярных — PCAP, которая служит для создания программ анализа сетевых данных, поступающих на сетевую карту компьютера.

В итоге ПО фильтрации имеет многоуровневую структуру от перехвата пакетов до механизма принятия решения о фильтрации.

Используемая платформа — LINUX, как наиболее популярная основа при создании сетевых устройств.

Научный руководитель — канд. техн. наук К. И. Будников

Разработка автоматизированной системы формирования заявок на отправку корреспонденции

В. С. Якимов, В. С. Андрианов
Сибирский федеральный университет, Красноярск

Промышленные предприятия достаточно часто испытывают проблемы в организации своего внутреннего документооборота, что приводит к значительному снижению их эффективности при выполнении основных видов деятельности. В интересах одной такой компании была разработана современная информационная система, позволяющая автоматизировать процесс приема и отправки корреспонденции между его отделами.

Одной из главных задач модуля является упрощение и ускорение рабочего процесса внутри компании. Для этого в системе разработаны справочники, которые позволяют значительно ускорить и упростить заполнение пользователем заявки на отправку корреспонденции. Также для своевременного оповещения пользователя в модуле присутствует система, которая отправляет электронные письма получателям и отправителям корреспонденции.

Особенностью проекта является то, что он построен на «Чистой архитектуре» (*Clean Architecture*). Это делает систему гибкой, независимой от какого-либо фреймворка, пользовательского интерфейса или базы данных. Система становится легко тестируемой.

Разработанная автоматизированная система представляет собой веб-приложение на платформе ASP.NET Core, разделенное на две части. Серверная часть отвечает за работу с базой данных, обработку данных и отправку электронных писем. В ней используются такие технологии, как Clean Architecture, Entity Framework Core, CQRS, Repository, MediatR и JWT, а для отправки электронных писем – SMTP-сервер. Клиентская часть отвечает за взаимодействие с пользователями и серверной частью. В ней используются такие технологии, как MVC, DevExtreme, Windows Authentication, Bootstrap и jQuery. Взаимодействие между серверной и клиентской частями происходит при помощи формата JSON.

Дальнейшее внедрение разработанной автоматизированной системы на предприятии позволит пользователям сократить время на отправку внутренней корреспонденции, так как отпадет необходимость в хождении по кабинетам, распечатке, самостоятельном отслеживании статуса документов.

Научный руководитель — канд. техн. наук А. В. Хныкин

ТЕОРИЯ И ПРАКТИКА ЦИФРОВОЙ ЭКОНОМИКИ

УДК 004

Classification of students with binary decision trees

Denizay Cam

University of Applied Sciences Coburg, Germany

At the University of Applied Sciences Coburg, the performance monitoring of students is currently still carried out manually. The students are evaluated individually by a person with their own years of experience based on their data and classified as “not at risk” or “at risk”. This process has to be carried out regularly and exactly and is very time-consuming. In addition, it can be difficult to order a large amount of data and generate a logical conclusion.

Based on this problem, the following question was asked for this research: Is it possible to classify students of the University of Applied Sciences Coburg based on available data with the binary decision tree algorithm? In order to classify, it was of major importance to collect information from the students. This information was taken from the university’s internal platforms MyCampus, Primus and Moodle. Over 50 different data were collected from which information could be extracted. At the beginning, these data are mostly unorganized, incomplete and partly unimportant. For this reason, it is important to carry out a pre-processing. The data from the students were selected, preprocessed, and therefore prepared for the binary decision tree. A decision tree is a tree-structured classification model. Because of the visual representation, the decision tree gives the advantage that the result is easier to understand. Similarly, the tree also represents the argumentation of the decision at the same time. A decision tree consists of the root which further splits into branch, depending on the outcome of the branches the next branch or the terminal are formed.

To start an evaluation with the decision tree, it is started from the root of the tree. According to the answer, the route of the tree is followed until a suitable branch is reached. Finally, the branch represents the corresponding class membership. The collected information from the students was transformed into new data sets. By aggregating, new data were generated from which new information could be derived. The following questions were converted to attributes and should be answered by the decision tree:

1. The difference between the exams that should have been written and those that have not been written yet.
2. Whether four of the Magic five exams have been passed.
3. The critical exams are passed.
4. Exist deadlines for modules.

5. What is the student's grade point average.

In relation to the research question, split-criteria have to be defined, which the decision tree needs to be able to classify the data. The following split-criteria were considered for the attributes:

- 1) ≤ 2 or > 2 ,
- 2) yes or no,
- 3) yes or no,
- 4) yes or no,
- 5) $\leq 2,5$ or $> 2,5$.

Therefore, after the evaluation of the raw data, it was considered which requirement must be fulfilled in order to categorize the students into the classes "not at risk" or "at risk". With the help of test data for which the class membership is already known different combinations of attributes and split criteria were tried out until a result with a very low error rate was finally generated. After the improved algorithm has almost no errors, it can be applied to new data sets for which class membership is not yet known. The aim of the work was to generate knowledge from a set of data. A model should be developed that is trained with data in order to be able to make prognoses for new data. This is done in a multi-step data analysis process. And so, the above research question could be proven.

Scientific supervisor — Prof. E. Gerhardt

Conducting a sentiment analysis of product reviews in an online store

Dilek Yakut

University of Applied Sciences Coburg, Germany

The amount of digital data is constantly increasing due to Big Data. The manual processing of these in companies is becoming more difficult because of human capacities. As a result, technical solutions are needed to include current, significant data in business processes and to stay competitive through a stable information base. About 80 percent of the available data is text-based. So, linguistic analysis techniques are important to enable the machine extraction of information from unstructured data. The present topic illustrates the relevance of text mining in an online store. Sentiment analysis is performed in two different ways on real product reviews from the online shop of Zalando, a German company that sells fashion and cosmetics products. Product ratings are generated by the intrinsic motivation of customers. Shoppers voluntarily share their opinions about the shopping experience and the product, which provides the company with information free of charge. Accordingly, a valuable source of knowledge is available. The knowledge gained from this enables the realization of possible improvements in the sales process.

The analysis of expressions of opinion on the web with text mining methods makes it possible to break down texts into individual parts and to work out the polarity of positive, neutral or negative sentiment about a product. The focus is on these expressions, which are already created as positive or negative in a sentiment dictionary and thus have information content for the analysis. There is only one sentiment vocabulary for this in German from the University of Leipzig, the SentiWS. The first publication of SentiWS in version v1.8b is from 19.05.2010, which contains 16,406 positive and 16,328 negative words. The individual words in it are given a sentiment value. This is a value from the interval $[-1;1]$. The highest sentiment value in the positive vocabulary according to SentiWS is 1 for the expression “succeeded” and the lowest in the negative range is -1 for the word “danger”.

With an efficient method, information is to be extracted from large quantities of evaluations in a short time, so that insights from them do not have to be lost. To evaluate the concepts, the results are compared with human perception in order to measure their significance. In method 1 “MAX-WH”, a word frequency analysis is carried out in the selected seven evaluations with the data analysis software MAXQDA and the polarity towards the product is determined using the German Sentiment Dictionary. The method 1 is therefore only dictionary-based.

In method 2 “DY”, words that have a meaning for the sentiment are manually created in MAXQDA and checked for correspondence with the ratings, so

compared to method 1, a context-based analysis is targeted here. Subsequently, the sentiment vocabulary SentiWS is matched with these words via Excel to form the polarity based on the terms and to identify the customer attitude.

The analysis delivers correct values in both methods and the sentiment matched human perception. A sentiment analysis in reviews in the online store can therefore be successfully realized in order to receive information about the polarity. Despite this, a critical questioning of the results is important. In the dictionary-based method, words are checked for agreement in both parts of the dictionary — positive and negative vocabulary. In the process, words received erroneous sentiment values because they are evaluated in the lexicon independently of context. Thus, incorrect values influence the calculation. For example, the term “big” is assigned a positive value of 0.3694 because it is associated with an impression, truly describing a point of criticism in an evaluation for a shoe. Furthermore, significant words are not taken into calculation, even though they say crucial things from a business point of view. This results from the fact that there is no match with any vocabulary.

The problem of misinterpretation and the assignment of erroneous values is solved in context-based method through the class definition (positive and negative expressions) and differentiation of search terms. Accordingly, the words were specifically checked in the matching dictionary. The problem here is that not all words that express a sentiment in the present topic are contained in the thesaurus and consequently have no effect on the result, even if they are created as search terms.

The gaps in the methods make it clear that especially text mining tools such as Word Sense Disambiguation and Text Chunking, which are based on the topic context, are highly relevant for obtaining meaningful analysis results. For both methods, using adequate templates and dictionaries is effective for circumventing the problems. For long-term success, it makes sense to create these manually, to expand them permanently and to avoid ready-made templates.

Scientific supervisor — Prof. E. Gerhardt

Базовые требования к методике бизнес-анализа процесса цифровой трансформации предприятия

К. В. Баулина

Новосибирский государственный университет экономики и управления

В настоящее время наблюдается сдвиг парадигмы в области информатизации предприятий, вследствие чего ИТ и бизнес представляют собой единую материю. Традиционные бизнес-модели утратили свою жизнеспособность. Актуальность приобретает проектирование цифровых бизнес-моделей, направленных на предоставление цифровых сервисов через платформы в периметре цифровой экосистемы. Эта тенденция способствует развитию процесса бизнес-анализа на предприятиях, поскольку возникает необходимость переключения с автоматизации бизнес-процессов на реализацию процессов цифровой трансформации бизнеса.

Современные отечественные и зарубежные авторы уделяют большое внимание описанию ключевых элементов цифровой трансформации бизнеса: цифровых бизнес-моделей, цифровых ценностей, цифровой экосистемы, цифровых пространств и цифровых платформ. Но анализ источников показывает, что отсутствует методика бизнес-анализа, описывающая формирование всех перечисленных элементов и их совместное влияние на реализацию процесса цифровой трансформации на предприятии.

Целью исследования является определение базовых требований к методике бизнес-анализа процесса цифровой трансформации предприятия.

Задачи исследования включают: анализ существующих подходов к бизнес-анализу процесса цифровой трансформации предприятия; выявление достоинств и недостатков проанализированных подходов; определение требований к методике бизнес-анализа процесса цифровой трансформации предприятия.

Выделенные требования позволят сформировать методику бизнес-анализа процесса цифровой трансформации предприятия, ориентированную на проектирование цифровой бизнес-модели, формирование цифровых пространств, моделирование цифровой экосистемы, проектирование цифровых платформ и сервисов, анализ модели распределения цифровых ценностей.

Методика бизнес-анализа, соответствующая требованиям цифровой трансформации, ведет к повышению эффективности реализации цифровых стратегий на предприятии.

Научный руководитель — канд. экон. наук, доц. П. М. Пашков

Информационная система управления цифровыми двойниками гидроэнергетических объектов

М. М. Бацагин

Новосибирский государственный университет экономики и управления

При управлении цифровыми двойниками гидроэнергетических объектов важную роль играет соответствующий инструментальный аппарат. Цифровой двойник — это цифровая копия физического объекта или процесса, помогающая оптимизировать эффективность бизнеса. Точное описание инженерного узла (детали) с визуальным 3D-изображением и историей эксплуатации объекта повышает эффективность процессов управления строительством гидроэнергетических объектов.

Информационная система включает:

- 1) подсистему 3D-моделирования для визуального отображения объекта;
- 2) базу данных для хранения информации о самом объекте, а также об истории изменений и эксплуатаций;
- 3) уровень бизнес-логики приложения для обработки всей функциональности информационной системы, соединение с базой данных и 3D-моделями;
- 4) уровень интерфейса, включающий десктоп-версию и мобильное приложение.

Анализ научных публикаций показывает повышение эффективности системы управления организации (предприятия) за счет использования современных информационных технологий [1, 2].

Научная новизна исследования состоит в разработке архитектуры информационной системы для управления цифровыми двойниками гидроэнергетических объектов в рамках концепции цифровой фабрики.

Практическая значимость заключается в том, что разработанная информационная система позволит повысить эффективность управление цифровыми двойниками при строительстве гидроэнергетических объектов.

1. Терещенко С. Н., Тишина Н. В. Процесс автоматизации государственного адресного реестра // Инновации в жизнь. 2019. № 3(30). С. 48–57.

2. Терещенко С. Н., Моисеева Е. Д. Использование чат-бота как элемента цифровой трансформации университета // Наука Красноярья. 2020. Т. 9, № 4-4. С. 239–244.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

Разработка чат-бота для повышения эффективности взаимодействия с клиентами в сфере автострахования

О. А. Бизяев

Новосибирский государственный университет экономики и управления

В данной работе дается оценка эффективности применения чат-бота для информационного взаимодействия с клиентами в сфере автострахования. Алгоритмы реагирования на запросы чат-бота построены на технологиях применения искусственного интеллекта.

При страховании автомобиля по программе ОСАГО или КАСКО у клиента, как правило, возникают типовые вопросы, требующие быстрого реагирования от компании, предоставляющей соответствующие услуги. Большинство запросов можно автоматизировать при помощи технологии чат-ботов. Чат-боты доказали свою эффективность по обработке информационных запросов во многих отраслях [1, 2].

Проведен анализ публикаций в сфере обработки естественного языка и опыте использования чат-ботов в различных секторах экономики.

Спроектирован и разработан чат-бот, позволяющий обрабатывать типовые клиентские запросы в сфере автострахования. Для создания чат-бота использовалась платформа DialogFlow (компания Google). Общий объем интентов (англ. *Intent*, единица обработки запроса) составил 84 единицы. Максимальная глубина сохранения контекста запроса составила 4 уровня. Для взаимодействия с базой данных использовалась технология WebHook, язык обработки PHP. Чат-бот интегрирован с мобильным телефоном.

Внедрение чат-бота позволило снизить нагрузку на сотрудников, обрабатывающих входящие клиентские запросы, на 30 %.

В настоящей работе впервые составлена методика проектирования и внедрения чат-бота для сферы автострахования, позволяющая повысить эффективность обработки информационных запросов. Разработанное решение позволит повысить эффективность работы компании в сфере автострахования.

1. Терещенко С. Н. Возможности технологий машинного обучения для управления бизнес-процессами университета // *Int. J. Adv. Stud.* 2019. Т. 9, № 2. С. 68–71.

2. Терещенко С. Н., Моисеева Е. Д. Использование чат-бота как элемента цифровой трансформации университета // *Наука Красноярья.* 2020. Т. 9, № 4-4. С. 239–244.

Сравнительный анализ методов формализации описаний нормативных документов вуза

Е. А. Гаврилова

Новосибирский государственный университет экономики и управления

В условиях большого объема нормативных документов процессы правоприменения и нормотворчества в вузе не должны отставать от постоянно меняющихся объектов регулирования. Для этого следует унифицировать процессы подготовки нормативно-правовых документов и обеспечить их непротиворечивость, что является сложной теоретической и практической задачей.

Практика показывает наличие значительных трудностей увязать в одну документ, содержащие большое количество правовых норм. Обширность базы нормативной документации, ее постоянное увеличение и модернизация приводят к появлению противоречий и пробелов в нормативном регулировании. Для выявления противоречий и контроля согласованности нормативных документов необходимо формализовать описания правовых норм и обеспечить их создание в единой информационной среде [1].

Вопрос возможности цифровой формализации права рассматривается давно и вызывает острые дискуссии в научной сфере, поэтому целью исследования является проведение сравнительного анализа методов формализации нормативных документов и выбор опорного метода для построения системы управления качеством нормативных документов вуза. В процессе исследования рассмотрены такие методы, как правовая символика, декларативные методы, логика предикатов, модальная логика, продукционные системы.

Деонтическая логика как часть модальной логики, оперирующая такими понятиями, как обязательство, запрет и дозволение, выбрана в качестве наиболее подходящего варианта для формализованного описания нормативных документов, позволяющего вывести процесс разработки и применения норм на качественно новую ступень за счет автоматизированного выявления противоречий и пробелов нормативного регулирования в вузе.

1. Бобров Л. К., Рыжков О. Ю. Формализованная запись нормативных документов для компьютерной обработки // Науч.-техн. информация. Сер. 2: Информ. процессы и системы. 2013. С. 7–14.

Научный руководитель — д-р техн. наук, проф. Л. К. Бобров

Моделирование архитектуры суперсервисов в процессе построения цифровых государственных услуг

О. В. Ермушина

Новосибирский государственный университет экономики и управления

Согласно национальной программе «Цифровая экономика», был разработан проект цифровой трансформации государственного управления, в основе которого была заложена концепция суперсервисов.

В понятие суперсервиса закладывается проактивное предоставление государственных услуг гражданам с момента наступления какой-либо жизненной ситуации. Изначальная задумка заключалась в том, что государство само напоминает гражданину о наступлении жизненной ситуации, предоставляет информацию о том, что нужно сделать, а все документы запрашивает самостоятельно из электронных реестров по единому идентификатору.

Такой подход значительно бы снизил бумажные обороты документов, сделал процесс наиболее прозрачным, простым и понятным для пользователя, сократилась бы дезинформация населения о положенных им выплатах и пособиях. Посещение государственных органов становится необязательным для получения госуслуг, а получить выплату или пособие становится возможным «в один клик».

На данный момент вопрос цифровизации государственного управления актуален не только внутри нашей страны, но и по всему миру. Так, схожую идею суперсервисов пробовали внедрить в Австралии, Индии, США и др. В России концепция суперсервисов не была продумана основательно, поэтому на данный момент вместо 25 основных суперсервисов заработал один «Поступи в вуз онлайн» не в полной мере. Основная проблема заключается в межведомственном взаимодействии и отсутствии единого координатора суперсервисов.

Целью данного исследования является построение модели архитектуры суперсервисов в процессе цифровой трансформации государственных услуг российского сегмента.

В рамках поставленной цели решаются следующие задачи: изучение отечественного и зарубежного опыта организации суперсервисов, проблемный анализ существующей системы, построение перспективной модели архитектуры суперсервисов.

Данная модель будет учитывать особенности межведомственного координирования и поможет в решении выявленных проблем.

Научный руководитель — канд. экон. наук, доц. П. М. Пашков

**Алгоритмы автоматической валидации и оценки стоимости
изготовления деталей на станках с ЧПУ
в клиент-серверной системе цифрового производства**

В. К. Закиров

Новосибирский государственный университет

Анализ 3D-моделей деталей ради определения возможности их изготовления на станках является актуальной задачей современного промышленного производства. Такой анализ необходим и для изделий, обрабатываемых с помощью технологий лазерной резки, гидроабразивной резки и гибки металла. Традиционная система производства не так эффективна, так как требует взаимодействия заказчика и технолога производства, из-за чего процесс от получения 3D-модели до начала изготовления не автоматизирован и времязатрачено. Для решения этой проблемы необходима система автоматической валидации, которая будет анализировать модели и немедленно информировать о возможности изготовления детали и ее стоимости.

В работе планировалось разработать алгоритмы валидации, определяющие, возможно ли изготовить модель из плоского листового металла с помощью операций гибки и резки. Также было необходимо разработать программный модуль валидации 3D-модели и оценки стоимости ее производства. Этот модуль было необходимо встроить в клиент-серверную систему на базе LEDAS Cloud Platform, отвечающую за взаимодействие заказчика и технолога. Для реализации алгоритмов должны быть использованы языки C++ и TypeScript, библиотека Open Cascade.

На основе анализа базы моделей из листового металла было выявлено, что в них чаще всех других встречаются ребра длины, равной толщине листа металла. Был разработан подход, основанный на определении толщины листа металла и валидации изготавливаемости модели на основе определенной толщины. Для деталей из листового металла, изготавливаемых с помощью гибки, был также разработан метод построения развертки на плоскость. Все методы были реализованы, встроены в существующую систему и используются в промышленном производстве.

Научный руководитель — канд. физ.-мат. наук И. А. Рыков

Компьютерная имитационная модель для обеспечения соблюдения правил дорожного движения с использованием Unity Engine

Н. А. Исаев

Новосибирский государственный университет экономики и управления

В наше время дорожное движение становится все более серьезной проблемой. Данное исследование было направлено на разработку компьютерной визуальной имитационной модели для моделирования дорожных ситуаций. Система значительно улучшит управление дорогами, а также предоставит властям «песочницу» для проверки разработанных ими дорожных маршрутов и конфигурации правил дорожного движения без риска негативного воздействия на жизнь людей. Программа также позволяет сравнивать различные конфигурации правил дорожного движения, чтобы выявить лучшую для тех или иных условий. В этом случае система действует как инструмент для прогнозирования потенциальных пробок и лучшего понимания потока трафика, что позволит принимать более обоснованные решения проблем, связанных с дорожным движением.

Для разработки системы использовались функции из библиотеки Unity Game Engine. Плагин Unity Game Engine под названием Playmaker также активно использовался, поскольку он предоставлял интерфейс FSM, который значительно упростил и повысил эффективность разработки функций. Помимо Playmaker, Modular Traffic (еще один плагин Unity Game Engine, дополняющий Playmaker) предоставляет необходимые модели и поведение объектов для симуляции.

Информация, полученная в ходе исследования, может быть применена в различных областях, таких как системы прогнозирования дорожного движения и системы навигации (например, Google Maps, Waze). Самым важным вкладом будут данные, собранные при использовании системы. Эти данные помогут разработать более совершенные системы навигации. Также можно было бы использовать систему для прогнозирования предстоящих условий дорожного движения, например для моделирования того, что произойдет, если крупное шоссе будет закрыто. Развитие системы позволит повысить реалистичность поведения пешеходов и транспортных средств путем доработки физической модели объектов, а оптимизация движка приложения — реализовать имитацию большего масштаба.

Научный руководитель — канд. техн. наук, доц. А. Л. Осипов

Построение модели зрелости управления релизами

С. А. Карапетян

Новосибирский государственный университет экономики и управления

С ростом цифровых технологии возрастает спрос на программное обеспечение, отвечающее новым задачам. Повышенный спрос на обновленное программное обеспечение часто заставляет разработчиков спешить с выпуском новых продуктов на рынок, не дожидаясь их полной отладки. Это приводит к ошибкам и сбоям приобретенного программного продукта, в результате растет недовольство клиентов и разработчикам приходится выпускать patch для устранения ошибок.

В связи с этим появилась задача по проработке «процесса управления релизами», которая прописывает, как своевременно выпускать и переустанавливать программный код на боевой контур, чтобы реализованный функционал при этом продолжал правильно работать.

Управление релизом — это процесс, который касается управления, планирования и контроля создания программного продукта, на всех применимых этапах и в каждой среде: от разработки до тестирования и развертывания. Для правильного управления релизами в компании необходимо построить модель зрелости.

Уровень 0 — общих элементов нет, продукт впервые разрабатывается и выходит в релиз, моделируется база данных ИС. На данном уровне происходит управление релизами баз данных, а не программного кода.

Уровень 1 — общие справочные данные. Общий набор справочных данных создается отдельными системами, на этом уровне проходит процесс отправки разработанного кода с одного релизного контура на другой.

Уровень 2 — синхронизация базы данных и разрабатываемого кода. На данном уровне начинает появляться дизайн, управляемый событиями, чистый программный код, который легко синхронизируется с предыдущим релизом программного кода без всевозможных конфликтов с информационной базой системы.

Уровень 3 — синхронизация бизнес-мероприятий. На этом уровне берется масштабируемая модель, управляемая событиями, из предыдущего уровня и добавляется ограничение: события, которые используются для синхронизации данных и программного кода между релизными контурами и системой.

Модель зрелости управления релизами используется как инструмент для эффективного выбора процесса управления релизами в зависимости от уровня, на котором находится создаваемый программный продукт.

Цифровые технологии предсказания численности населения в РФ

Е. В. Клепикова

Новосибирский государственный университет экономики и управления

В современном мире для эффективности осуществления социально-экономической концепции развития страны нужно владеть информацией о численности населения. Государство проявляет озабоченность в решении проблем, связанных с демографией, что отразилось в Указе Президента Российской Федерации об утверждении Концепции демографической политики Российской Федерации на период до 2025 года № 1351 от 09.10.2007 г.

Исследуется эффективность разработанных математических моделей для осуществления ретроспективного прогноза численности населения для периода 1946–2020 гг. В настоящее время статистические модели предсказания широко применяются в прогнозировании динамики численности населения, среди которых особо выделяются моделирование с помощью регрессии и метод передвижки возрастов, которые обладают высокой областью использования и эффективной степенью достоверности осуществления прогнозов. В практике прогнозирования динамики численности населения наиболее часто встречаются нелинейные модели.

Среди пяти трендовых моделей (линейной, логарифмической, степенной, экспоненциальной и полиномиальной) наилучшими статистическими характеристиками обладает полиномиальная модель. Коэффициент детерминации оказался равным 0,9833, критерий Фишера — 4298,257, объем выборки — 75 в период с 1946 по 2020 гг. Модель оказалась значимой. Показатель MAE = 1,643, показатель MAPE = 1,201. Точность прогноза, вычисленная по формуле $T = 100 - \text{MAPE}$, оказалась равной 98,799 %.

Предложены и исследованы линейные и нелинейные модели регрессионного типа, показывающие высокую предсказательную способность. Проведена проверка эффективности разработанных статистических методик, основанных на линейных и нелинейных моделях. Разработаны инструментальные программные средства для предсказания численности населения РФ по разработанным моделям в рамках цифровой экономики с использованием языка программирования Python.

Научный руководитель — канд. техн. наук, доц. А. Л. Осипов

Прогнозирование спроса для интернет-магазина доставки еды

М. С. Климонов

Иркутский государственный университет

Целью работы является проектирование и создание системы интернет-маркетинга для предприятия доставки еды.

Был проведен анализ рынка, сформировано уникальное торговое предложение, стратегия развития, под требования которой были подобраны наиболее эффективные инструменты интернет-маркетинга.

Для автоматизации всех бизнес-процессов, связанных с заказами через интернет, было принято решение использовать CMS Битрикс и ресторанный систему для учета iiko. Стандартное решение оказалось недостаточно гибким, поэтому был разработан собственный модуль для интеграции.

Далее была разработана система из 3 основных компонентов.

1. Сбор данных — были установлены счетчики метрик сайта, различных рекламных кабинетов, для сбора и подсчета офлайн-данных была использована ERP-система.

2. Аналитика — данные из рекламных кабинетов были переданы в Google Analytics, включая интеграцию с OWOX.

3. Визуализация — реализован механизм передачи собранных данных в Power BI с предобработкой и их визуализацией с помощью интерактивных отчетов.

Полученная информация позволила провести более эффективный анализ деятельности предприятия, были предложены способы для оптимизации и модернизации существующих бизнес-процессов.

В итоге удалось оцифровать деятельность предприятия. Это натолкнуло на мысль о возможности доработать данную систему до цифровой тени. Для тестирования данной гипотезы выбран один параметр для прогнозирования — спрос. В систему добавлены внешние параметры: курс валют, осадки, температура, праздничный, признак рабочего дня, день недели и др.

На основе этих данных были построены эконометрические модели для прогнозирования спроса с использованием нескольких алгоритмов машинного обучения.

Наилучшие результаты по двум показателям точности (коэффициенту детерминации, SKO) показала модель случайного леса.

Научный руководитель — канд. физ.-мат. наук А. С. Казимиров

Модель архитектуры информационной системы контроля соблюдения правил и требований к использованию опасного производственного объекта на основе технологии смарт-контрактов

Е. Д. Коровченко

Новосибирский государственный университет экономики и управления

В эпоху развития цифровой экономики технология смарт-контрактов может применяться не только для соблюдения условий коммерческого договора, но и для более сложных взаимоотношений между сторонами. Смарт-контракты на базе применения технологии блокчейн могут стать связующим звеном между владельцем опасного производственного объекта (ОПО) и органами Ростехнадзора, позволяя отслеживать объект, подлежащий обязательной экспертизе промышленной безопасности, узнавать о его техническом состоянии, а также просматривать выполнение всех предписаний, выданных экспертной организацией владельцу объекта, что позволяет существенно сократить количество аварий на подобных объектах.

Анализ литературных источников показывает, что на текущий момент отсутствуют модели и методы, позволяющие внедрить технологию смарт-контрактов, для создания информационной системы (ИС) контроля за соблюдением требований промышленной безопасности (ТПБ).

Цель научного исследования — определение архитектуры ИС контроля за соблюдением ТПБ на основе технологии смарт-контрактов.

Методика исследования включает: выявление комплексов технических средств, применяемых на данный момент Ростехнадзором за контролем соблюдения ТПБ; анализ проблем в области промышленной безопасности; изучение возможности применения смарт-контрактов; моделирование архитектуры ИС контроля соблюдения ТПБ.

В основе предложенного подхода лежит построение модели архитектуры ИС, которая позволит оперативно реагировать на экстренные ситуации, не подвергая опасности производственный процесс.

Отличительной особенностью ИС является контроль за соблюдением правил и требований к использованию ОПО со стороны всех звеньев, участвующих на основе использования смарт-контрактов.

К результатам исследования относится модель архитектуры ИС контроля соблюдения правил и требований к использованию ОПО. Практический результат исследования является возможностью использования данной архитектуры ИС со стороны всех звеньев, участвующих в процессе обеспечения ТПБ.

Анализ иноязычных версий сайтов российских вузов как инструмент привлечения абитуриентов

А. И. Кочеева

Новосибирский государственный университет экономики и управления

Актуальность проблемы увеличения количества иностранных граждан, обучающихся в России, напрямую зависит от уровня привлечения абитуриентов в вузы. Сегодня все высшие учебные заведения используют для привлечения абитуриентов цифровые технологии, в частности веб-сайты.

В работе проведен анализ иноязычных версий сайтов российских вузов по нескольким параметрам: техническим параметрам, контент-структуре, удобству пользователя и поведению пользователя на сайте. Выборку составили 34 веб-сайта. Анализ поведения пользователей на сайте позволил выявить непродолжительность сеансов и большое количество уходов с сайта.

По итогам исследования установлено, что все иноязычные сайты российских вузов нуждаются в оптимизации по различным показателям. Также даны рекомендации для улучшения сайтов по каждому из параметров.

По техническим параметрам:

- исправить скорость загрузки сайта 9 вузам;
- установить переадресацию с протокола HTTP на HTTPS 18 вузам;
- повысить уровень оптимизации мега-тегов 29 вузам;
- откорректировать правильность файла sitemap.xml 22 вузам.

При организации контент-структуры: обеспечить наличие общей и коммерческой информации, а также разнообразие форматов и улучшить качество контента.

Для удобства пользователей рекомендовано обеспечить сайты поисковой строкой, 19 вузам — блоком «Контакты», а 23 вузам — формой обратной связи. Также адаптировать веб-сайты для мобильных устройств.

Результаты работы могут быть использованы вузами для оптимизации иностранных версий сайтов учебного заведения, а также для дальнейшего исследования по смежным темам.

Научный руководитель — канд. физ.-мат. наук, доц. А. И. Пестунов

Чат-бот как эффективный инструмент для кол-центра вуза

Р. Р. Латышева

Новосибирский государственный университет экономики и управления

В статье дается оценка эффективности применения чат-бота при управлении информационными запросами вуза. Алгоритмы чат-бота базируются на технологиях применения искусственного интеллекта.

Поставлена задача снизить загрузку сотрудников кол-центра по обработке информационных запросов за счет применения искусственного интеллекта.

Основная особенность заключается в том, что для решения поставленной задачи в разработке чат-бота используется алгоритм машинного обучения по обработке естественного языка (англ. NLP — *Neuro-linguistic programming*).

Чат-бот (от англ. *chat* — болтать, *bot* — робот) — это компьютерная программа, которая может общаться с человеком с помощью текстовых или голосовых сообщений (команд), взаимодействие осуществляется через простой и понятный интерфейс.

В ходе анализа научных работ в сфере искусственного интеллекта было выявлено, что использование чат-ботов повышает эффективность процессов обработки внешних запросов [1]. НГУЭиУ — один из крупнейших вузов Сибири, в котором учатся более 10 тыс. студентов. Структурной единицей по обработке входящих информационных запросов выступает информационный центр, выполняющий функции кол-центра.

Научная новизна исследования заключается в разработке методики по внедрению чат-бота для обработки информационных запросов в вузе.

Практическая значимость работы заключается в том, что предложенные инструменты для кол-центра вуза позволят повысить эффективность обработки информационных запросов.

1. Терещенко С. Н. Возможности технологий машинного обучения для управления бизнес-процессами университета // *Int. J. Adv. Stud.* 2019. Т. 9, № 2. С. 68–71.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

Методика построения цифровой бизнес-модели факторинговых сервисов

П. Ю. Мельникова

Новосибирский государственный университет экономики и управления

Цифровая трансформация, в частности переход на цифровые сервисы, является актуальным мировым трендом, в том числе для банковских продуктов и услуг. Приоритетным направлением цифровизации факторинга является переход на электронный документооборот, развитие мультифункциональных цифровых платформ (онлайн-маркетплейсов), развитие аналитических систем со сбором данных (*data mining*) и машинным обучением (*machine learning*) для лучшей оценки сделок.

Цифровизация предоставляет широкие возможности для проникновения факторинга в новые отрасли, а также делает факторинг еще более удобным и привлекательным инструментом финансирования. Одним из ключевых шагов цифровизации является определение цифровой стратегии и построение цифровой бизнес-модели. При этом требуется учитывать всю специфику и сложность процесса факторинга, поэтому стандартные методы и подходы к построению бизнес-модели не подойдут для данной отрасли. Факторинг только начинает набирать популярность на российском рынке банковских продуктов, вопросы его цифровизации пока не проработаны, и в связи с этим возникает актуальная задача построения цифровой бизнес-модели факторингового сервиса [1].

Целью исследования является описание методики построения цифровой бизнес-модели для факторинговых сервисов.

В процессе исследования были решены следующие задачи: изучен бизнес-процесс факторинга, концепция и принципы факторинговых продуктов и сервисов, выявлены ключевые особенности факторинга и его специфика, проанализированы основные подходы к построению цифровых бизнес-моделей, построена цифровая бизнес-модель факторингового сервиса и описана методика построения цифровой бизнес-модели факторинговых сервисов.

Предложенная методика позволит учитывать специфику факторинга при проектировании и построении цифровых факторинговых сервисов, а также создавать новые цифровые продукты на рынке факторинга.

1. Pashkov P., Pelykh V. Digital transformation of financial services on the basis of trust // Book Proc. 50th Int. Sci. Conf. Econ. Social Dev. 2020. P. 375–383.

**Разработка социально ориентированного приложения
MatchAuPair на компонентной основе.
Реализация логики и пользовательского интерфейса (роль «семья»)**

Т. Г. Нестерова
Новосибирский государственный университет

Существующие средства поддержки социального взаимодействия со стороны государства, частных компаний и общества неэффективны из-за недостаточной гибкости используемых инструментов. Агентства по кадровому обеспечению часто ориентированы на оказание только информационной поддержки и не стремятся учитывать специфику задач клиента, что негативно сказывается на качестве подбора кадров при высокой стоимости услуг. Кроме того, они, как правило, не оказывают услуг по сопровождению. Таким образом, возникает потребность в оптимальном решении данной проблемы, учитывающем современные подходы и возможности передовых IT-инструментов.

Одним из таких решений является разрабатываемое приложение MatchAuPair, представляющее собой прозрачную систему выбора кандидатов. Целью данной работы является реализация клиентской части приложения. Для этого проведен сравнительный анализ подходов и технологий создания мобильных приложений, разработана расширяемая модульная архитектура и набор компонентов.

Для реализации выбран гибридный JavaScript фреймворк React Native, позволяющий использовать общий интерфейс для разных реализаций, упрощая портирование на различные системы. На данный момент поддерживаются мобильные устройства с операционными системами Android и iOS, в дальнейшем планируется переход на mobWeb и веб-платформы.

Реализованное приложение предоставляет выбор одной из доступных ролей со следующими функциями: несколько вариантов регистрации (через email, Google, Apple, Facebook); поиск с использованием фильтров, например выбор по возрасту, языкам, продолжительности пребывания; общение в текстовых чатах; добавление профилей в списки «Избранное» и «Заблокированные»; изменение данных профиля; настройка уведомлений и конфиденциальности.

Предполагается дальнейшее развитие проекта, включающее добавление поддержки других платформ, включение дополнительных ролей, расширение функций для существующих ролей, интеграцию со сторонними сервисами, например для верификации профилей.

Разработка программного обеспечения для анализа и оптимизации стратегий на фондовом рынке с использованием высокочастотных данных

М. А. Овсянникова

Новосибирский государственный университет

В современном обществе деятельность многих людей связана с фондовым рынком. Для успешной работы в данной сфере необходимо изучение рынка, построение торговых стратегий с их последующим анализом и оптимизацией.

Стратегия, составленная инвестором с использованием торговых индикаторов, может как помочь получить максимальную прибыль, так и повлечь значительные убытки. Поэтому после составления стратегии надо понять, является ли полученная стратегия оптимальной или ее нужно дорабатывать перед применением в реальных условиях. Для проверки состоятельности стратегии удобно использовать данные рынка прошлых лет. Однако нужный объем внутрисдневных данных прошлых лет не позволяет проводить вычисления вручную, из-за чего необходимо разработать ПО, выполняющее задание, анализ и оптимизацию стратегий с использованием высокочастотных данных российского фондового рынка, а также современных индикаторов, метрик и алгоритмов.

Целью работы является разработка программного обеспечения, позволяющего обрабатывать внутрисдневные котировки акций Московской фондовой биржи с целью построения на их основе торговых стратегий, проведения их анализа и оптимизации.

В ходе работы был реализован следующий набор функций.

1. Реализовать потоковое считывание данных российского фондового рынка с отображением графика выбранного инструмента.
2. Добавить возможность расчета и отображения в виде графиков значений торговых индикаторов с настраиваемыми параметрами.
3. Реализовать задание стратегий, проведение их анализа и оптимизации с предоставлением отчета с графиками и метриками качества.

Практической ценностью данной работы является автоматизация вычисления, анализа и оптимизации торговых стратегий и наглядное представление полученных результатов.

Научный руководитель — канд. экон. наук А. В. Костин

Информационные технологии в прогнозировании материального ущерба от пожаров

А. Д. Пихурова

Новосибирский государственный университет экономики и управления

В мировой практике проблема защиты от пожаров является одной из наиболее актуальных задач, требующих тщательного анализа и прогнозирования ущерба от них. С использованием методов цифровой экономики требуется предвидеть степень ущерба от пожаров и обеспечить защиту от них с упреждением, т. е. нужно разрабатывать методики предсказания пожарной опасности с целью обеспечения наименьшего ущерба.

Так как пожары в большинстве случаев являются своеобразной продукцией социума, то для их оценивания лучше использовать следующие цифровые параметры: численность населения, объем основных фондов и валового внутреннего продукта (ВВП).

Исследовались следующие многофакторные модели:

$$Y_t = AX_1^\alpha X_2^\beta X_3^\gamma e^{\lambda t},$$

где Y_t — ущерб от пожаров в t -м году; X_1 — ВВП; X_2 — численность населения РФ; X_3 — объем основных фондов РФ; λ — параметр, характеризующий экспоненциальный временной тренд.

Ниже представлена многофакторная нелинейная модель с использованием экспоненциального тренда.

$$Y_t = X_1^{1,174171743} X_2^{-0,5403032} X_3^{0,363051976} e^{-0,11850482t}.$$

Модель обладает следующими статистическими свойствами: коэффициент детерминации 0,997; критерий Фишера 1048,438; объем выборки 16 в 2003–2018 гг.; параметр RMSE 2,18; параметр MAPE 11,83.

Представлен комплекс программных средств на языке программирования Python для прогнозирования экономического ущерба от количества пожаров по значимым макроэкономическим параметрам в рамках цифровой экономики.

Научный руководитель — канд. техн. наук, доц. А. Л. Осипов

Разработка чат-бота для управления процессами строительства объектов гидроэнергетики

Д. С. Рехтин

Новосибирский государственный университет экономики и управления

Дана оценка эффективности применения чат-бота при управлении процессами строительства объектов гидроэнергетики. Алгоритмы чат-бота базируются на технологиях применения искусственного интеллекта.

Строительство объектов гидроэнергетики требует постоянного технического и инженерного консультирования по многим вопросам. Использование чат-бота для оперативных ответов инженерно-техническому персоналу является актуальной задачей для повышения эффективности управления процессами строительства объектов гидроэнергетики.

Проведен анализ публикаций в сфере машинного обучения и применении чат-ботов в различных сферах экономики. Чат-боты эффективно применяются во многих отраслях [1].

Спроектирован и разработан чат-бот. Ядро чат-бота основано на алгоритмах машинного обучения по обработке естественного языка (англ. NLP — *Neuro-linguistic programming*). Для создания чат-бота использовалась платформа DialogFlow компании Google. Чат-бот настроен на ответы по инженерно-технической документации. Решение интегрировано с мобильным телефоном и позволяет сотрудникам оперативно получать ответы при помощи смартфона.

Научная новизна работы заключается в разработке методики проектирования и внедрения чат-бота для инженерно-технической консультации сотрудников при управлении строительством гидроэнергетических объектов.

Практическая значимость работы состоит в следующем: разработанное решение позволит повысить эффективность управления процессами строительства гидроэнергетических объектов.

1. Терещенко С. Н., Моисеева Е. Д. Использование чат-бота как элемента цифровой трансформации университета // Наука Красноярья. 2020. Т. 9, № 4-4. С. 239–244.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

Методика проектирования цифровых финансовых сервисов в рамках платформенного подхода

Л. А. Шевченко

Новосибирский государственный университет экономики и управления

В настоящее время наблюдается резкий рост востребованности цифровых платформ в финансовом секторе. Это связано с тем, что компании, занимающиеся ИТ, проектируют и внедряют цифровые платформы для взаимодействия клиентов и стейкхолдеров, предоставляя им различные сервисы: сервисы в области электронной коммерции и т. п. Вокруг платформы формируется круг стейкхолдеров, предоставляющий разнообразные продукты и привлекающий поток пользователей.

Анализ научных публикаций показывает, что переход к цифровым платформам наиболее эффективен по сравнению с использованием отдельных цифровых финансовых сервисов, так как цифровые сервисы проектируются и предоставляются различными, зачастую разрозненными подразделениями организации; сервис создается на том стеке технологий, который доступен компании; сервисы имеют фрагментированные процессы и цифровые каналы с разной динамикой развития, изолированный пользовательский опыт на уровне каждого канала.

На данный момент нет методики, которая позволяет проектировать цифровые финансовые сервисы в рамках платформенного подхода. Целью исследования является разработка методики проектирования цифровых финансовых сервисов на основе платформенного подхода.

Методика исследования включает: анализ понятия «цифровой финансовый сервис», изучение и выбор метода проектирования цифровых финансовых сервисов, описание подхода построения цифровых сервисов с опорой на платформенный подход.

В основе предложенной методики лежит отказ от внедрения разрозненных цифровых финансовых сервисов. Она направлена на построение цифровой платформы, включающей финансовые сервисы, основным драйвером которых будут являться клиенты, и возможность построить взаимодействие с ними по всем возможным каналам коммуникации.

Предложенный подход проходит апробацию на предприятии финансовой отрасли, которое участвует в цифровой трансформации.

1. Pashkov P., Pelykh V. Digital transformation of financial services on the basis of trust // Book Proc. 50th Int. Sci. Conf. Econ. Social Dev. 2020. P. 375–383.

Цифровая трансформация системы управления хозяйственными процессами коммерческого банка

Ю. А. Шигаева

Новосибирский государственный университет экономики и управления

В современном крупном коммерческом банке операции по планированию и учету внутрихозяйственных операций составляют значимую часть. В крупном банке в хозяйственные операции входят: управление договорами аренды недвижимости, договорами с клининговыми компаниями, службами охраны и многое другое. Сохраняется тенденция вывода на аутсорсинг непрофильных служб.

Цифровая трансформация коснулась всех процессов управления коммерческих банков, в том числе и хозяйственных. Анализ научных публикаций показывает, что применение современных информационных технологий [1, 2] позволяет повысить эффективность системы управления организации (предприятия).

Предлагается модель архитектуры информационной системы коммерческого банка в рамках цифровой трансформации банковских процессов, которая включает следующие подсистемы:

- 1) компьютерное зрение для обработки входящих договоров (и других документов);
- 2) блок интеллектуального анализа данных на основе алгоритмов машинного обучения для определения оптимальных условий управления договорами и хозяйственной собственностью;
- 3) голосового помощника для повышения оперативности и мобильности взаимодействия сотрудников банка и подрядчиков;
- 4) интеграционную шину для связи с корпоративной системой банка.

Разработана методика интеграции корпоративной информационной системы коммерческого банка с облачными сервисами технологий искусственного интеллекта. Разработанная модель позволит коммерческим банкам развивать архитектуру корпоративной информационной системы и соответствовать вызовам цифровой трансформации.

1. Терещенко С. Н., Тишина Н. В. Процесс автоматизации государственного адресного реестра // Инновации в жизнь. 2019. № 3(30). С. 48–57.

2. Терещенко С. Н., Моисеева Е. Д. Использование чат-бота как элемента цифровой трансформации университета // Наука Красноярья. 2020. Т. 9, № 4-4. С. 239–244.

Трансформация архитектуры информационной системы высшего учебного заведения во время пандемии COVID-19

Г. С. Ширманова

Новосибирский государственный университет экономики и управления

Пандемия COVID-19 повлекла массовый переход высших учебных заведений на формат обучения с применением дистанционных технологий. Значительные изменения затронули практически все процессы управления учебным заведением [1, 2]. Переход на онлайн-обучение стал причиной ускоренного перехода на новое программное обеспечение как для самого процесса обучения (например, Zoom), так и для управления бизнес-процессами вуза. Данный процесс вносит значительные изменения в архитектуру корпоративной информационной системы университета. Например, использование системы видеоконференций для проведения лекций и практических занятий влечет необходимость интеграции подсистемы расписания, учебных графиков, системы контроля посещений занятий и др.

Поддержание единой интегрированной корпоративной информационной системы становится довольно трудоемкой задачей. Отдельным вызовом стали мессенджеры и социальные сети, через которые во время пандемии шел значительный поток информационного обмена как между студентами и преподавателями, так и между преподавателями и руководством вуза. Фактически являясь составной частью информационного пространства вуза, мессенджеры и социальные сети практически не интегрированы в корпоративную информационную систему.

Научная новизна исследования состоит в разработке архитектуры корпоративной информационной системы вуза, позволяющей интегрировать новые подсистемы, включая внешние сервисы, без нарушения концепции «атомарности ввода данных». Практическая значимость заключается в том, что разработанная архитектура позволит повысить эффективность построения корпоративной информационной системы вуза.

1. Терещенко С. Н. Возможности технологий машинного обучения для управления бизнес-процессами университета // *Int. J. Adv. Stud.* 2019. Т. 9, № 2. С. 68–71.

2. Терещенко С. Н., Моисеева Е. Д. Использование чат-бота как элемента цифровой трансформации университета // *Наука Красноярья.* 2020. Т. 9, № 4-4. С. 239–244.

Научный руководитель — канд. техн. наук, доц. С. Н. Терещенко

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

УДК 004.032.26

Automatic summarization of scientific texts

A. E. Pauls, S. A. Berezin
Novosibirsk State University

Many “intelligent” tasks that were previously considered impossible for automation, for example, image classification, text sentiment analysis, predicting credit default, etc., are now successfully solved using machine learning methods. In particular, recent advances in automatic natural language processing make it possible to solve the problem of constructing a summary with surprising quality. Many people face the need to build a summary — students, scientists, office workers — all those who interact with large amounts of textual information. Since the amount of textual information is dramatically increasing, there is a need for better methods of summarization.

Traditionally, summarization is divided into two types — extractive, when some sentences or parts of them are extracted from the source text without modification, and abstractive, when sentences or words can be changed, or even supplemented with other text. Extractive summarization is considered to be simpler and more reliable, but people usually use reformulation to shorten thoughts to get a better and more compact summary. Therefore, most modern works are devoted precisely to abstractive summarization.

Two current State-of-the-Art approaches deserve the most attention — BART, and PEGASUS. The main idea behind them is to pre-train a Transformer with a special task — noise reduction and restoration of original texts in BART and restoration of masked “important” sentences in PEGASUS. Authors of PEGASUS and BART argue that the design of noise or importance functions is a field for future research.

We propose a task for pretraining Transformer to recover masked sentences, as in PEGASUS, but the “importance” of sentences is calculated based on the number and type of named entities (in original paper ROUGE is used — coincidence of n-grams of randomly selected words). We assume that this will improve the quality of summarization of texts from different domains (e. g. scientific texts), which are rich in named entities.

Scientific adviser — Cand. Phys.-Math. Sci. T. V. Batura

**Разработка системы сбора отзывов
о товарах и услугах и анализа их тональности**

Е. А. Багадаева

Новосибирский государственный университет

Сегодня социальные сети играют огромную роль в жизни современного человека. Бизнес быстро освоил новую возможность для установления контакта как с действующими, так и с потенциальными клиентами, а также для получения обратной связи в виде отзывов на предлагаемые товары и услуги. Продвижением в социальных сетях обычно занимаются специалисты в области маркетинга социальных сетей. Как правило, таким специалистам приходится анализировать огромное количество текстовых отзывов. Для решения подобной задачи актуально применение программного решения с целью автоматизации процесса анализа и экономии времени специалиста.

Разрабатываемая система автоматического анализа отзывов включает графический пользовательский интерфейс, парсер социальных сетей и классификатор тональности. Для парсинга комментариев (отзывов о товарах и услугах) из социальной сети «ВКонтакте» используется предоставляемый ею API. Для оптимизации работы парсера с целью преодоления существующих ограничений на количество запросов к API в секунду используется язык VKScripts. При подготовке текстов отзывов к анализу выполняется их препроцессинг, очистка от «мусора» и стоп-слов, а также лемматизация. Классификация отзывов по тональности выполняется с помощью алгоритмов классического машинного обучения. В качестве данных для обучения использованы открытые датасеты из сети Интернет.

Дальнейшая работа в этой области предполагает улучшение качества классификатора тональности, добавление новых инструментов аналитики.

Научный руководитель — канд. техн. наук Ю. А. Загоруйко

Построение имитационных моделей по текстам на естественном языке

А. А. Гавриленко

Новосибирский государственный университет

В настоящее время возрастает потребность в использовании имитационного моделирования при проектировании, разработке, оптимизации технических систем и технологических процессов, информационных и управляющих систем в различных прикладных областях. Отсутствие у потенциальных пользователей, являющихся специалистами в конкретных прикладных областях, профессиональной подготовки в области имитационного моделирования препятствует его широкому использованию.

В связи с этим представляется интересным построение имитационной модели способом, который бы не требовал много усилий со стороны потенциальных пользователей. Один из таких способов — описание моделей на естественном языке.

Для построения имитационных моделей таким способом была расширена существующая система имитационного моделирования MTSS. С помощью библиотеки Stanford NLP, основанной на машинном обучении, из текста на английском языке извлекаются знания об элементарных моделях, используемых системой MTSS для построения имитационной модели в определенной прикладной области.

Дальнейшая работа в этой области предполагает переход с английского языка описания моделей на русский.

Научный руководитель — канд. техн. наук Ю. А. Загорулько

Формирование выборки для решения задачи кросс-языковой идентификации сущностей

Д. И. Гусев

Новосибирский государственный университет

При решении задачи кросс-языковой идентификации сущностей [1] необходимым является создание соответствующего набора данных о различных объектах реального мира. Существующие наборы основаны на зарубежных языках, таких как английский, китайский, японский и т. д. Русского среди них нет.

Для решения этой проблемы было решено сформировать выборку на основе встраивания к межъязыковому выравниванию сущностей [2], которое не зависит от машинного перевода между кросс-языковыми наборами данных.

Полученный набор данных представляет собой коллекцию триплетов англо- и русскоязычных версий проекта DBpedia с встраиванием структур и дополнительным уточнением вложений, используя тройки атрибутов с встраиванием атрибутов.

Набор данных будет использован для дальнейшей генерации кросс-языковых связей. Методы построения графов знаний, основанных на одноязычной структурной информации, могут не соответствовать объектам, имеющим разные факты в мультязычных графах знаний [3]. Таким образом, полученный набор данных будет использован для построения мультязычных графов знаний и уточнения описания сущностей.

1. Knowledge Graph Embeddings and Explainable AI / Federico Bianchi, Gaetano Rossiello, Luca Costabello et al. // arXiv:2004.14843. 2020.

2. Cross-lingual Entity Alignment via Joint Attribute-Preserving Embedding / Zequn Sun, Wei Hu, Chengkai Li // arXiv:1708.05045. 2017.

3. Cross-lingual Knowledge Graph Alignment via Graph Matching Neural Network / Kun Xu, Liwei Wang, Mo Yu et al. // arXiv:1905.11605. 2019.

Научный руководитель — канд. физ.-мат. наук, доц. З. В. Апанович

Разработка методов извлечения терминологических цепочек из корпуса текстов

М. Ю. Гуськова

Новосибирский государственный университет

Интенсивное развитие современных информационных технологий привело к значительным изменениям в процессах обработки научно-технических и публицистических текстов, обучения и извлечения знаний. Эти процессы напрямую связаны с понятием «термин».

Термин — это языковой знак, являющийся носителем элементарной информации в виде отдельного научного понятия, входящего в систему понятий определенной предметной области. Термин может быть отдельным словом или словосочетанием, в частности терминологической цепочкой. На долю словосочетаний приходится около 80 % всех терминов предметной области, поэтому вопрос их эффективного извлечения из текста достаточно актуален.

Цель работы — разработка методов автоматического извлечения терминологических цепочек из корпуса текстов ограниченной предметной области.

Для обработки текста и создания словаря предметной области пользователь должен описать структуру терминологических цепочек в терминах морфологических, синтаксических и семантических характеристик. Для этого разработана модель, описывающая правила сборки цепочек. Например, по правилу «Глагол (мышления) + Предлог + Существительное (Т. п.)» будет собрана цепочка «Думать о будущем».

Для проведения исследования собран корпус текстов из предметной области «Психология», состоящий из 20 статей подходящей тематики, и на его основе создан словарь, содержащий около 10 тыс. терминов.

Также разработан алгоритм поиска словосочетаний, удовлетворяющих заданным пользователем параметрам. На вход алгоритму поступают результаты работы морфологического анализа текста. Далее выполняется анализ найденного списка и поиск в нем терминологических цепочек, подходящих под выбранную систему правил.

Таким образом, разработан подход к извлечению терминологических цепочек с помощью системы правил, задаваемых пользователем.

Научный руководитель — канд. физ.-мат. наук Е. А. Сидорова

**Исследование и анализ SLAM-алгоритмов и фреймворков
для построения карт и локализации на наборах данных тренажера
Duckietown и имитатора Duckietown**

А. Д. Девятовская, Н. Е. Бирючков
Новосибирский государственный университет

Одной из фундаментальных задач в сфере мобильной робототехники является локализация объекта и построение карты окружающей территории. Для решения этой задачи используются алгоритмы построения карты местности и одновременного позиционирования (*Simultaneous Localization And Mapping* — SLAM). SLAM выполняет минимизацию ошибок определения позиции камеры и вычисление инверсной глубины ключевых точек. За счет этого выравнивается траектория движения робота (камеры) и обозреваемое пространство, на основе которого формируется карта мира.

В настоящее время существует исследовательский проект Duckietown в области автономных роботов. Он состоит из двух частей: роботов (*duckiebots*) и города (*duckietown*), по которому они передвигаются. В городе есть размеченные дороги, светофоры, Arpil-знаки и препятствия. Этот тренажер применяется для отладки и тестирования SLAM-алгоритмов для промышленных роботов и автопилотирования. Тренажер отражает особенности промышленных автопилотируемых систем с ограниченными ресурсами: роботы оснащены только камерами, а также имеют слабую вычислительную мощность. Однако в тренажере не решена задача SLAM. В силу особенностей тренажера накладывается ограничение на использование стандартных SLAM-алгоритмов (большинство алгоритмов требуют LIDAR-сенсоры).

Таким образом, целью данной работы является анализ и адаптация SLAM-алгоритмов и фреймворков, которые возможно использовать для роботов *duckiebots* с учетом всех ограничений.

В ходе исследования был проведен анализ существующих алгоритмов и фреймворков, были разработаны инструкции по настройке фреймворка для работы с тренажером и имитатором и был протестирован на различных датасетах фреймворк OpenVSLAM, который удовлетворяет всем техническим ограничениям роботов. Для тестирования разработанных алгоритмов использовались наборы данных тренажера и имитатора Duckietown.

Научный руководитель — Т. В. Лях

Разработка алгоритма извлечения фактов из текстов биомедицинских статей

С. О. Деревянченко

Новосибирский государственный университет

В современном мире ежедневно появляется огромное количество новой текстовой информации. Ее публикацией занимаются миллиарды различных источников. В такой ситуации одной из важнейших задач обработки информации является извлечение самой важной и полезной ее части из всего объема. Популярным и полезным подходом здесь является извлечение фактов из текстов.

Фактом в тексте назовем структурированную информацию о некоторых объектах, встречающихся в данном тексте, и их отношении друг с другом. Таким образом, задача извлечения фактов из тестов является задачей поиска структурированной информации в неструктурированных текстах и относится к направлению Text Mining.

В данной работе ведется разработка нового алгоритма решения задачи извлечения фактов из текстов, основными характеристиками которого являются:

- отсутствие необходимости учителя (алгоритм обучается на размеченный корпус данных);
- легкая адаптируемость к новым темам (качество и работоспособность алгоритма оцениваются на текстах из биомедицинской тематики, однако подразумевается сохранение качества при использовании алгоритма в новом домене);
- низкое требование к аппаратному обеспечению (возможность обучения и работы алгоритма на компьютерах средней мощности, например обычных ПК).

Основная идея алгоритма заключается во взаимодействии модели тематического моделирования LDA (*Latent Dirichlet allocation*) и разбора дерева зависимостей предложения. Использование этих составных частей позволяет алгоритму удовлетворять заявленным выше требованиям.

Научный руководитель — канд. техн. наук П. С. Деменков

Использование методов машинного обучения для управления манипулятором

А. А. Дунаевский

Сибирский государственный университет
телекоммуникаций и информатики, Новосибирск

Расчет движения роботов-манипуляторов для взаимодействия с движущимися предметами представляет значительную сложность с алгоритмической точки зрения. Использование алгоритмов управления, базирующихся на теории автоматического управления, характеризуется низкой степенью переиспользования кода и вследствие этого высокой трудоемкостью разработки. Вместо них систему предлагается строить на основе обучающегося алгоритма, который обучается, взаимодействуя с некоторой средой. Целью настоящей работы является использование методов машинного обучения для упрощения создания систем управления манипуляторами.

В качестве объекта для системы управления был выбран робот, бесконечно отбивающий мячик для настольного тенниса вверх. Для моделирования робота применялся физический движок с компьютерной средой трехмерного моделирования и фреймворк Tensorflow.

В начале был использован классический подход, основанный на использовании пропорционально-интегрально-дифференциального регулятора, затем использовалось обучение с подкреплением на основе алгоритма DQN. Для обучения некоторому действию необходимо задать награду, которая будет его описывать. После описания условия награждения модели данный алгоритм позволяет генерировать траектории, при которых мяч не будет вылетать за пределы рабочего поля манипулятора. В трехмерной среде моделирования был проведен ряд экспериментов, в ходе которых система управления обучалась отбивать мяч так, чтобы он не падал за пределы поля.

Данный метод примечателен тем, что его можно использовать при разработке систем управления для решения задач в заранее неизвестном окружении, необходимо только задать динамические и кинематические параметры робота и параметры подкрепления. Дальнейшее исследование предполагает применение алгоритма на реальной установке.

Научный руководитель — О. Д. Климин

**Предсказание химического состава шлака
в процессе легирования стали**

Е. А. Заварзин, А. А. Мелехин, А. В. Хафизов,
А. С. Савкина, Д. Д. Городецкий, К. Келлер
Новосибирский государственный университет
Московский технический университет связи и информатики
Московский государственный университет им. М. В. Ломоносова
Высшая школа экономики, Санкт-Петербург
Омский государственный университет им. Ф. М. Достоевского
Грацкий технический университет

Легирование — один из этапов обработки стали. Расплавленная сталь заливается в специальный металлический ковш и греется электричеством для получения нужного состава сплава. Процесс нацелен на вывод серы, регулирование количества легирующих присадок, содержания железа, оксидов алюминия и кремния. Перед началом нагрева берется проба химического состава шлака. Затем химический состав корректируется добавлением легирующих компонентов. Процесс повторяется до тех пор, пока не будет достигнут нужный химический состав стали.

Мы предлагаем ML-модель, обученную на данных о процессе легирования Евраз ЗСМК, предсказывающую химический состав последней пробы шлака на основе исходных данных процесса легирования.

Предобработка позволила после исключения пропусков и выбросов отфильтровать около 50 % исходных данных. В работе рассматривались модели линейной регрессии, градиентного бустинга, LGBM и случайного леса, а также два подхода к решению задачи. Первый способ предполагал избавление от линейно зависимых признаков и удаление признаков методом *backward elimination*, а второй — итеративный процесс обучения моделей, где каждая следующая модель принимала на вход не только исходные признаки, но и предсказания предыдущих моделей.

Результатом работы является модель предсказания химического состава шлака легированной стали с точностью по MAPE от 0,4 до 6,5 %.

Работа была выполнена в рамках зимней школы CompTech'21.

Научные руководители — С. А. Васильев, А. В. Зубков

**Разработка подсистемы синтаксического анализа
системы перевода текстов**

Ю. М. Кислицына
Вятский государственный университет

Методы автоматического анализа естественного языка и компьютерные технологии обработки информации — это наиболее динамично развивающиеся области компьютерной лингвистики в наши дни. Одной из главных проблем компьютерной лингвистики является машинный перевод.

Машинный перевод представляет собой процесс связанного перевода текстов компьютерной программой с одного естественного языка на другой. Системы на основе грамматических правил производят анализ текста, при этом перевод выполняется на основе встроенных словарей выбранной языковой пары, а также грамматик, включающих семантические, морфологические и синтаксические закономерности обоих языков. Одним из этапов реализации подобного является поверхностно-синтаксический разбор.

Для решения проблемы автоматического синтаксического анализа были выбраны некоммерческие лингвистические платформы, среди которых одно из первых мест занимает *Natural Language Toolkit*. Это набор библиотек для языка программирования Python, ориентированный на выполнение основных процедур автоматической обработки текстов.

В ходе разработки была выявлена тенденция, что если готовые решения содержат несколько языков, то все языки принадлежат одной языковой группе. Так как русский язык и английский принадлежат разным языковым группам, следовательно, особенности построения предложений сильно отличаются друг от друга. Это стало причиной, по которой данные языки были проанализированы разными алгоритмами. При разработке данной подсистемы была выявлена проблема распознавания омонимии. Данное явление распространено во многих языках, и одним из способов решения является использование вероятностных методов.

Научный руководитель — канд. техн. наук, доц. М. Л. Долженкова

Каскадирование модулей на основе искусственных нейронных сетей для осуществления спортивных прогнозов

А. К. Крутиков

Вятский государственный университет, Киров

Прогнозирование и планирование в спорте — необходимые и важные аспекты подготовки как отдельных спортсменов, так и работы больших спортивных клубов, центров спортивной подготовки и школ олимпийского резерва. Для обработки и анализа таких значительных объемов данных необходимо использовать средства вычислительной техники и современные информационные технологии. В последнее время для формирования прогнозов успешно применяются различные методы прикладного искусственного интеллекта. Как наиболее эффективное средство в области прогнозирования спортивных событий показал себя аппарат искусственных нейронных сетей (ИНС).

В лаборатории «Интеллектуальные системы» им. В. А. Байкова Вятского государственного университета разработан прототип простейшей системы, предназначенной для прогнозирования результатов в индивидуальных видах спорта. Прототип реализован в среде MATLAB. Система состоит из модулей, в основе каждого из которых лежит определенная модель ИНС.

В данной работе предложена структура системы прогнозирования, состоящая из нескольких модулей и представляющая из себя многоступенчатый конвейер. Каждый модуль при этом обучается отдельно, а выборка строго соответствует модели и функции модуля. Обработанные специальным образом исходные данные поступают на вход первого уровня каскада. Начиная со второго уровня, на вход поступают динамически формируемые предыдущей ступенью данные. На выходе каскада получаем вектор, являющийся результатом прогноза. Детализация прогноза и распределение по модулям специализированных данных (скоростных, силовых, антропометрических и других показателей спортсменов) позволяют особым образом выделить наиболее важные прогнозоформирующие параметры. Эксперименты показали эффективность предлагаемой архитектуры, несмотря на незначительное увеличение времени обучения ИНС. Кроме того, структура самого каскада, как и тип используемой сети, существенно зависит от специфики данных, формирующих обучающие выборки, а также от спортивной дисциплины, в рамках которой осуществляется прогнозирование.

Научный руководитель — канд. техн. наук, доц. В. Ю. Мельцов

Разработка паттерна онтологического проектирования для систем автоматического извлечения параметрической информации из текстов

Д. К. Митюшин

Новосибирский государственный университет

В современном мире количество создаваемой и передаваемой текстовой информации растет невероятно быстро. Обработка такого объема данных вручную требует немало усилий. Именно поэтому актуальной становится проблема автоматизации такой работы. Особый интерес представляют параметрические конструкции — тип информации, которую можно оценить: измерить, охарактеризовать и сравнить.

Целью данной работы является разработка методов и ресурсов, обеспечивающих извлечение и анализ параметрических данных в текстах.

За основу решения был взят подход к извлечению информации при помощи правил и шаблонов. В результате систематизации была разработана онтология различных видов параметрических конструкций, в которых определена их структура и взаимосвязь друг с другом. Она представляет собой онтологический паттерн содержания: пополняя ее понятиями предметной области, можно проектировать онтологии, ориентированные на создание систем извлечения параметрической информации из специализированных текстов.

Примером такой системы является программный модуль, извлекающий из историй болезни данные о состоянии пациентов с последующим упорядочиванием этих состояний по времени. На вход системе поступает онтология классов параметрических конструкций и правила по их извлечению. Библиотека Diglex находит в тексте фразы, соответствующие этим шаблонам, далее на их основе создаются экземпляры классов онтологии, а результат представляется в виде модели хронологии изменений симптоматики пациента.

Разработанный в ходе исследования паттерн онтологического проектирования позволяет извлекать из текстов произвольной предметной области различные параметрические конструкции, например: сравнение с эталоном («на 2 градуса выше нормы»), периодичность («пять раз в год»), динамика («давление повысилось») и др. Кроме того, инструмент представляет извлеченные конструкции в удобном, формализованном виде — как набор экземпляров классов онтологии.

Научный руководитель — канд. физ.-мат. наук Е. А. Сидорова

Обучение суперсетей для сэмплирования оптимальных архитектур для различных программно-аппаратных комплексов

Т. И. Науменко

Новосибирский государственный университет

Развитие нейронных сетей дало возможность решать ряд задач с точностями, недоступными ранее. Однако нейронные сети требовательны к вычислительным ресурсам, и часто лучшая точность достигается за счет увеличения размера нейронной сети и, как следствие, времени ее работы. При этом для различных вычислительных устройств требуются различные архитектуры нейронных сетей, например, GPU эффективно работают с большими матричными умножениями, в то время как для ARM-процессора больше подходят менее ресурсоемкие операции.

Одним из современных подходов к выбору оптимальной архитектуры нейронной сети являются алгоритмы автоматического поиска архитектур нейронных сетей. С их помощью можно определять оптимальные архитектуры из зафиксированного пространства поиска для произвольного программно-аппаратного комплекса. При этом алгоритмы автоматического поиска требуют обучать найденную сеть с нуля, что затрудняет их внедрение в бизнес-процесс.

Данная работа адресует указанную выше проблему с помощью обучения специальной сети — суперсети, которая является суперпозицией большого количества нейронных сетей. Преимуществом подхода, описанного в данной работе, является то, что он позволяет из суперсети извлекать одиночные модели, которые без дообучения показывают хорошие результаты для задач классификации.

Для демонстрации успешной работы алгоритма была обучена суперсеть, составленная из операций *mobile inverted bottleneck* на датасете *Imagenette*. Далее, выбрали 1000 случайных архитектур, извлекли обученные веса из суперсети и измерили качество работы полученных сетей, предварительно накопив статистику слоев *BatchNorm* для каждой сети.

В результате было получено, что средняя точность выбранных сетей составляет 93,4 %, лучшая архитектура имеет точность 94,7 %; при этом лучший опубликованный результат на *Imagenette* на текущий момент составляет 95,9 %. Далее для каждой сети была измерена скорость работы на GPU, CPU и ARM и показано, что среди всех архитектур с точностью выше 94 % наименьшее время работы для различных вычислительных платформ достигается на различных архитектурах.

Разработка метода поиска прототипов «естественных» понятий

Б. Пак

Новосибирский государственный университет

Категоризация — один из фундаментальных когнитивных процессов, заключающийся в отнесении познаваемого объекта к некоторой категории (классу). Теория прототипов представляет собой один из подходов к проблеме категоризации, она возникла в середине 70-х гг. XX в. в работах американского психолога Элеоноры Рош [1], которая ввела «естественные» понятия, отражающие высококоррелированную структуру признаков объектов внешнего мира. Прототипы «естественных» понятий — это наиболее репрезентативные примеры объектов, отражающие эту высококоррелированную структуру.

Поскольку прототипы формируются непосредственно в процессе познания, а познание есть когнитивный процесс, присущий человеку, то разработка метода поиска прототипов является актуальной проблемой современной теории искусственного интеллекта.

Существующие методы поиска прототипов не способны отразить их высококоррелированную структуру. В работе [2] предлагается новый метод поиска прототипов «естественных» понятий, основанный на вероятностном обобщении формальных понятий. На основе данного метода разработан программный комплекс на Python и проведена серия компьютерных экспериментов, иллюстрирующих эффективность метода.

1. Rosch E. Natural categories // Cogn. Psychol. 1973. Vol. 4. P. 328–350.

2. Vityaev E., Pak B. Prototypes of the “natural” concepts discovery // Cognitive Syst. Res. 2021. Vol. 67. P. 1–8.

Научный руководитель — д-р физ.-мат. наук, проф. Е. Е. Витяев

Разработка методов создания информационных чат-ботов для интернет-сайтов

В. Д. Рублев

Новосибирский государственный университет

За последние несколько лет жизненный ритм стал динамичнее, одним из основных факторов, подтолкнувших к этому, стало развитие информационных технологий, ориентированных на упрощение жизни современного человека. Одна из таких технологий — средства для обмена сообщениями, т. е. мессенджеры. И с ростом популярности мессенджеров место чат-ботов в повседневной жизни становится все заметнее. Интеллектуальные интерактивные чат-боты часто используются в мобильных приложениях и меняют модель взаимодействия между организациями и клиентами. Например, большинство типичных вопросов и проблем, которые возникают у клиентов, решаются довольно просто, но для этого организациям приходится писать и публиковать ответы на часто задаваемые вопросы. Чат-боты — это прекрасная персональная альтернатива разделам вопросов и ответов.

Целью работы является разработка технологии создания информационных чат-ботов для поиска по сайтам определенного жанра с возможностью настройки под разные группы сайтов. Для поиска по сайтам используется подход к информационному поиску в контенте веб-сайта на основе его жанровой модели и предварительной тематической кластеризации контента. Также чат-бот позволяет использовать для поиска мета-поисковую систему или может самостоятельно обрабатывать контент сайта.

В ходе выполнения работы был разработан метод организации обратной связи с пользователем, осуществляемой информационным чат-ботом для поиска ответа на конкретный вопрос. Если поисковая система ошиблась с результатами поиска, то пользователю предлагается уточнить запрос и указать, есть ли нужное ему в результатах (но для ответа на вопрос необходимо выделить фрагменты текста) или система полностью ошиблась. В первом случае требуется кластеризовать результаты поиска и выделить ближайшие кластеры, во втором — необходимо искать заново, понизив релевантность уже найденных результатов.

Таким образом, предлагаемый подход позволяет организовать диалог с пользователем и обеспечивает итеративное улучшение результатов поиска.

Научный руководитель — канд. физ.-мат. наук Е. А. Сидорова

**Подход к извлечению отношений между терминами
для автоматического построения лингвистической онтологии
заданной предметной области**

П. М. Чагина

Новосибирский государственный университет

При работе с ограниченными предметными областями необходимы экспертные знания для эффективного выделения терминов данной области и установления их характеристик. Для фиксации таких данных активно применяются лингвистические онтологии. Однако процесс построения онтологии подразумевает обработку большого корпуса текстов.

Целью данной работы является автоматизация процесса пополнения лингвистической онтологии терминами и связями между ними путем применения лексико-синтаксических шаблонов.

В работе разработан специальный язык представления *шаблонов с переменными* для извлечения отношений между терминами в тексте. Каждый шаблон содержит описание результатов его применения.

син1 = [x_1 (“ x_2 “)”) =>**Sin**(x_1 , x_2): $x_1 \neq x_2$

По этому шаблону из предложения “Сначала проверяется необходимость вакуумирования трубы (значение давления)” извлекаются термины x_1 = “необходимость вакуумирования” и x_2 = “значение давления”, между которыми устанавливается отношение синонимии.

Процедура выявления связей при помощи шаблонов состоит из итеративного процесса разбора подходящих фрагментов текста. Алгоритм включает четыре этапа. Входными параметрами являются словарь, содержащий начальное множество терминов и семантических связей, набор шаблонов, корпус текстов. На первом этапе происходит разбор описания шаблонов. На втором этапе на основе каждого шаблона генерируются поисковые шаблоны. На третьем этапе происходит сборка найденных по поисковым шаблонам фрагментов для шаблонов с переменными и их означивание. На четвертом этапе алгоритм пытается построить новые поисковые шаблоны.

При извлечении семантических связей из корпуса технических документов более 75 % обнаруженных связей являлись новыми, а точность извлечения приблизилась к 98 %.

Научный руководитель — канд. физ.-мат. наук Е. А. Сидорова

Применение механизма искусственных нейронных сетей для построения картографических подсистем ГИС в условиях неполноты информации

А. А. Чудинов

Вятский государственный университет, Киров

Построение картографических компонентов геоинформационных систем (ГИС) по десяткам срезам данных и более (начиная от рельефа, типа и высоты растительности, особенности водных ресурсов и заканчивая наличием домостроений, плотности населения, количества и вида домашних животных) сталкивается с проблемой разной детализации слоев и наличием в них пробелов. Недостаток информации может быть связан с неполнотой исходных данных, порчей данных или отсутствием среза как класса. Поиск зависимостей между разными слоями на срезе данных является одной из самых запрашиваемых задач, связанных с геоинформационными системами. Анализ, позволяющий заполнить «белое пятно» содержимым, крайне необходим геологам, газо- и нефтедобытчикам, железнодорожным компаниям, автоперевозчикам и т. д.

Цель данной работы — повышение эффективности построения картографической подсистемы путем применения искусственных нейронных сетей (ИНС) для заполнения пустых участков или доопределения информации в разных слоях ГИС. Применение нейронных сетей совместно с некоторыми другими методами, такими как эвристический анализ, корреляционный анализ, позволяет заполнить «белые пятна» и улучшить детализацию данных на слое, что повышает общую эффективность геоинформационной системы.

В работе рассмотрен механизм обработки слоев информации с привязками к координатам, которые проходят через этап анализа исходных данных с максимально возможным повышением детализации. Этот механизм включает как алгоритмическую обработку (известные методы математической статистики), так и применение ИНС, которые отвечают за обнаружение скрытых взаимосвязей между разными слоями или ошибок в данных, не обнаруженные на предыдущих шагах.

Проведенные эксперименты на программной модели показали, что поиск с помощью нейросетей скрытых или неочевидных зависимостей между слоями превышает возможности других известных методов, а в комбинации с этими методами повышает эффективность решения задачи построения и дополнения картографических компонентов для ГИС.

**Использование методов теоретико-информационной классификации
для решения проблем литературоведения**

А. И. Щеголева

Новосибирский государственный университет

Вопреки существованию большого количества литературоведческих исследований, остается множество литературных произведений, авторство которых неизвестно или ставится под сомнение. Например, не установлено авторство Н. А. Некрасова и А. Я. Панаевой в отношении романов «Три страны света» и «Мертвое озеро». Литературовед Б. Л. Бессонов, изучив существующую информацию об этих романах, пришел к выводу о невозможности их полной атрибуции и высказал аргументированные предположения о том, что большинство глав обоих романов принадлежит Некрасову. Но это предположение сильно противоречит «Воспоминаниям» Панаевой. Поэтому было решено исследовать произведения «Три страны света» и «Мертвое озеро» для определения авторства Некрасова, Панаевой или их современников, так как литературоведческие исследования не смогли прийти к точному ответу.

Цель данной работы — разработать метод атрибуции литературных произведений и реализовать его. С использованием созданного программного решения проведено исследование авторства частей романов «Три страны света» и «Мертвое озеро». Для достижения поставленной цели был выбран теоретико-информационный подход, основанный на сжатии данных, и подобраны его наилучшие параметры для исследования. Были проанализированы возможные авторы выбранных произведений и созданы их обучающие выборки.

С помощью выбранного метода, который не использовался ранее для исследования романов «Три страны света» и «Мертвое озеро», были получены статистически обоснованные результаты относительно авторства исследуемых романов. Также реализованное решение можно использовать для атрибуции других литературных произведений.

В результате исследования авторства произведений «Три страны света» и «Мертвое озеро» показано, что автором большинства частей романа «Три страны света» является А. Я. Панаева. Возможными соавторами этого произведения выявлены Н. А. Некрасов, И. И. Панаев и А. В. Дружинин. Авторами романа «Мертвое озеро» являются А. Я. Панаева и Н. А. Некрасов.

Научный руководитель — д-р техн. наук, проф. Б. Я. Рябко

Указатель авторов

Berezin S. A.	74, 127, 209	Бояршинова А. С.	38
Denizay Cam.....	184	Будаев А. Е.	57
Dilek Yakut	186	Ваганова А. И.	131
Menshanov Petr	76	Василенко Р. А.	46
Mukul Kumar Vishwas	76	Васильев А. Д.	132
Okunev Aleksey	76	Вахрушев М. В.	133
Pauls A. E.....	74, 209	Вербицкий С. К.	79
		Вершинин М. О.	113
Аборнев А. А.	77	Витченко В. А.	15
Абраменко А. А.	122	Волкова Е. А.	80
Агулов В. В.	52	Волченко С. В.	47
Акатов А. Н.	128	Воробьев Н. А.	16
Александров Е. А.	110	Гавриленко А. А.	134, 211
Алексеев А. И.	45	Гаврилова Е. А.	191
Ан В. Р.	5	Галиева А. Г.	135
Андрианов В. С.	183	Галушкин М. В.	61
Анисютин Л. Б.	155	Гармаева Д. А.	180
Антипов В. Е.	6	Гафиятуллин А. Р.	62
Антонова Я. Н.	59	Гилин С. В.	81
Астахова А. Д.	31	Гнездилова В. А.	83
Атутова Н. Д.	129	Голованов А. Л.	84
Афанасьев С. Т.	171	Горнев И. А.	17
Багадаева Е. А.	210	Городецкий Д. Д.	217
Бадмажапов Ж. С.	130	Гривцова А. Л.	18
Балданов З. Б.	172	Гусев Д. И.	212
Баранов А. А.	111	Гуськова М. Ю.	213
Баулина К. В.	188	Дайырбаева Э. Н.	53
Баушенко М. А.	78	Двинянин А. А.	19
Бахарев А. О.	7	Девятериков А. П.	63
Бахтиярова Г. Б.	123	Девятковская А. Д.	214
Бацагин М. М.	189	Дегтярев В. Д.	85
Башев В. И.	124	Деревянченко С. О.	155, 215
Бедарева М. А.	173	Добров Э. Р.	86
Бекиев М. А.	55	Долгов Г. А.	87
Белуосов Д. Л.	27	Долгов Н. Ю.	48, 54
Беребердин Н. В.	177	Дрогина М. В.	27
Березин С. А.	59	Дунаевский А. А.	216
Бизязев О. А.	190	Ермушина О. В.	192
Бирючков Н. Е.	214	Жабин О. С.	46
Борзов А. С.	112	Ждаркин Е. Б.	136
Борисова К. А.	60	Жилицкий В. Е.	20
Бощенко Е. С.	37	Журавлев В. А.	8

Завалишина Е. В.	9	Ластович А. А.	41
Заварзин Е. А.	217	Латышев М. В.	21
Зайцев А. О.	137	Латышева Р. Р.	200
Закиров В. К.	193	Лаханский А. А.	22
Зенина О. А.	174	Лебедева А. Д.	147
Зиануров Н. О.	138	Лебедев Р. К.	12
Зулин Д. К.	139	Легостаева Е. О.	148
Зыкова А. А.	140	Легченко А. Е.	96
Зюбина Д. А.	10, 129	Лешов Р. А.	149
Иванов В. К.	49	Литвинов Е. А.	177
Ильин И. Е.	88	Лукьянов А. Е.	97
Исаев Н. А.	194	Лукьянов Я. П.	55
Казырицкий Е. Н.	64	Лусников В. В.	23
Карапетян С. А.	195	Лямин А. С.	114
Келлер К.	217	Макаренко Д. Е.	115
Кислицына Ю. М.	218	Маковецкий Е. К.	125
Кислицын И. О.	50	Маловичко Е. В.	67
Кицкалов А. Е.	141	Малханов В. В.	13
Клепикова Е. В.	196	Мамедов В. Ю.	98
Климонов М. С.	197	Мамеев Н. С.	150
Ковалева Д. К.	175	Манаков Г. Д.	24
Коваленко А. П.	175	Манасян Т. Л.	176
Кожевникова Я. В.	65	Мартынов Н. Е.	141
Кожекин М. В.	89	Маслич Д. А.	25
Колесник Т. О.	90	Масыч М. А.	116
Комагина К. Т.	142	Мелехин А. А.	217
Комаров Л. С.	77	Мельникова П. Ю.	201
Кондратьев А. С.	143	Меркулов А. М.	26
Коробова А. В.	144	Мигранов Д. И.	56
Коровченко Е. Д.	198	Мирошниченко А. О.	177
Королев А. Р.	91, 104	Митюшин Д. К.	220
Коростелев Д. С.	92	Михайлов А. С.	151
Корякин И. А.	11	Мокроусов А. С.	178
Кочанов М. А.	39	Молчанова М. Г.	59
Кочеева А. И.	199	Морозов И. Н.	68
Краева И. А.	145	Моховикова В. Ю.	99
Крутиков А. К.	219	Мункоев А. А.	179
Кузнецов А. С.	146	Мызников П. В.	152
Кулаковский Д. С.	93	Наздрюхин А. С.	100
Куприянов Н. А.	106	Науменко Т. И.	221
Кутергина А. С.	66	Нейкина М. А.	57
Кушмуратов С. И.	94	Нестерова Т. Г.	202
Ланчуковская К. С.	95	Нечаев А. А.	101
Лапицкий А. А.	40	Нечахина А. А.	102

Нешади́мов Р. А.	27	Скокова В. А.	161
Николаев Н. А.	75, 155	Смирнов А. Ю.	162
Никулина О. В.	69	Табакаева В. А.	5
Овсянникова М. А.	203	Тарасов А. А.	120
Орловский А. С.	153	Темникова Е. А.	73
Павликов Д. Е.	37	Тищенко А. В.	37
Пак Б.	222	Толстов Т. В.	106
Панфилов Д. Д.	117	Трегубов А. С.	163
Парфенов Д. Р.	154	Третьяков Г. Н.	121
Паульс А. Е.	155	Турусов А. Н.	51
Перевалов К. Е.	105	Усова Д.	164
Пермяшкин Д. А.	126	Утепов Д. Ж.	106
Першиков Е. В.	103	Филиппов С. Д.	129
Пикуль М. А.	37	Фомченко А. М.	99
Пирожкова Д. А.	58, 59	Хагай В. В.	44
Пихурова А. Д.	204	Хафизов А. В.	217
Плешков А. В.	118	Хлиманкова Г. К.	165
Поверин Д. В.	94	Хлыбова М. М.	166
Погодин Р. С.	156	Худяков Д. А.	167
Подкур Т. М.	157	Чагина П. М.	224
Полевач Р. Е.	28	Черкашин Д. А.	141
Пономаренко А. А.	70	Чернова В. В.	107
Попова А. А.	29	Черногор Е. А.	108
Попова В. А.	30, 180	Черняев М. В.	28
Потехина А. А.	71	Чефонов В. С.	33
Пришляк Е. Е.	42	Чефонов Н. С.	33
Прокопенко В.	181	Чудаков Д. С.	109
Прокофьев И. А.	72	Чудинов А. А.	225
Протасов Д. В.	158	Шабалина Д. Е.	95
Радеев Н. А.	104, 159	Шарин Д. А.	34
Резинкина Е. А.	31	Шевченко Л. А.	206
Рехтин Д. С.	205	Шевчук Е. С.	168
Рублев В. Д.	223	Шестакова Е. А.	157
Рябчикова А. В.	32	Шигаева Ю. А.	207
Савельев С. В.	105	Ширманова Г. С.	208
Савкина А. С.	217	Шустик М. Т.	99
Самчук Е. О.	6	Щеголева А. И.	226
Сафаров Р. А.	14	Щербин А. С.	169
Сафенрейтер Д. А.	22, 182	Эйхвальд В. А.	77
Седухин О. А.	160	Юлдашов Ш. Н.	35
Сенотрусова А. М.	57	Якимов В. С.	183
Сергеев К. С.	119	Якобсон А. А.	170
Склярлов А. Н.	43	Янголь О. Д.	36

Содержание

ЗАЩИТА ИНФОРМАЦИИ И КИБЕРБЕЗОПАСНОСТЬ

Ан В. Р., Табакаева В. А.	5
Антипов В. Е., Самчук Е. О.	6
Бахарев А. О.	7
Журавлев В. А.	8
Завалишина Е. В.	9
Зюбина Д. А.	10
Корякин И. А.	11
Лебедев Р. К.	12
Малханов В. В.	13
Сафаров Р. А.	14

ИНСТРУМЕНТАЛЬНЫЕ И ПРИКЛАДНЫЕ ПРОГРАММНЫЕ СИСТЕМЫ

Витченко В. А.	15
Воробьев Н. А.	16
Горнев И. А.	17
Гривцова А. Л.	18
Двинянин А. А.	19
Жилицкий В. Е.	20
Латышев М. В.	21
Лаханский А. А., Сафенрейтер Д. А.	22
Лусников В. В.	23
Манакон Г. Д.	24
Маслич Д. А.	25
Меркулов А. М.	26
Нешадимов Р. А., Дрогина М. В., Белоусов Д. Л.	27
Полевач Р. Е., Черняев М. В.	28
Попова А. А.	29
Попова В. А.	30
Резинкина Е. А., Астахова А. Д.	31
Рябчикова А. В.	32
Чефонов В. С., Чефонов Н. С.	33
Шарин Д. А.	34
Юлдашов Ш. Н.	35
Янголь О. Д.	36

ИНФОРМАЦИОННО-УПРАВЛЯЮЩИЕ СИСТЕМЫ, ИНТЕРНЕТ ВЕЩЕЙ

Бощенко Е. С., Пикуль М. А., Павликов Д. Е., Тищенко А. В.	37
Бояршинова А. С.	38
Кочанов М. А.	39
Лапицкий А. А.	40
Ластович А. А.	41
Пришляк Е. Е.	42
Скляр А. Н.	43
Хагай В. В.	44

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОБУЧЕНИЯ

Алексеев А. И.	45
Василенко Р. А., Жабин О. С.	46
Волченко С. В.	47
Долгов Н. Ю.	48
Иванов В. К.	49
Кислицын И. О.	50
Турусов А. Н.	51

КОМПЬЮТЕРНАЯ ГРАФИКА, МУЛЬТИМЕДИАТЕХНОЛОГИИ И СИСТЕМЫ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

Агулов В. В.	52
Дайырбаева Э. Н.	53
Долгов Н. Ю.	54
Лукьянов Я. П., Бекиев М. А.	55
Мигранов Д. И.	56
Сенотрусова А. М., Нейкина М. А., Будаев А. Е.	57

НАУКОЕМКОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Pirozhkova D. A.	58
Березин С. А., Антонова Я. Н., Пирожкова Д. А., Молчанова М. Г.	59
Борисова К. А.	60
Галушкин М. В.	61
Гафиятуллин А. Р.	62

Девятериков А. П.	63
Казырицкий Е. Н.	64
Кожевникова Я. В.	65
Кутергина А. С.	66
Маловичко Е. В.	67
Морозов И. Н.	68
Никулина О. В.	69
Пономаренко А. А.	70
Потехина А. А.	71
Прокофьев И. А.	72
Темникова Е. А.	73

НЕЙРОННЫЕ СЕТИ И МАШИННОЕ ОБУЧЕНИЕ

Berezin S. A., Pauls A. E.	74
Nikolaev N. A.	75
Mukul Kumar Vishwas, Menshanov Petr, Okunев Aleksey	76
Аборнев А. А., Комаров Л. С., Эйхвальд В. А.	77
Баушенко М. А.	78
Вербицкий С. К.	79
Волкова Е. А.	80
Гилин С. В.	81
Гнездилова В. А.	83
Голованов А. Л.	84
Дегтярев В. Д.	85
Добров Э. Р.	86
Долгов Г. А.	87
Ильин И. Е.	88
Кожекин М. В.	89
Колесник Т. О.	90
Королев А. Р.	91
Коростелев Д. С.	92
Кулаковский Д. С.	93
Кушмуратов С. И., Поверин Д. В.	94
Ланчуковская К. С., Шабалина Д. Е.	95
Легченко А. Е.	96
Лукьянов А. Е.	97
Мамедов В. Ю.	98
Моховикова В. Ю., Фомченко А. М., Шустик М. Т.	99
Наздрюхин А. С.	100
Нечаев А. А.	101
Нечахина А. А.	102

Першиков Е. В.	103
Радеев Н. А., Королев А. Р.	104
Савельев С. В., Перевалов К. Е.	105
Толстов Т. В., Куприянов Н. А., Утепов Д. Ж.	106
Чернова В. В.	107
Черногор Е. А.	108
Чудаков Д. С.	109

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ

Александров Е. А.	110
Баранов А. А.	111
Борзов А. С.	112
Вершинин М. О.	113
Лямин А. С.	114
Макаренко Д. Е.	115
Масыч М. А.	116
Панфилов Д. Д.	117
Плешков А. В.	118
Сергеев К. С.	119
Тарасов А. А.	120
Третьяков Г. Н.	121

ПРОГРАММНАЯ АРХИТЕКТУРА И СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ

Абраменко А. А.	122
Бахтиярова Г. Б.	123
Башев В. И.	124
Маковецкий Е. К.	125
Пермяшкин Д. А.	126

ПРОГРАММНАЯ ИНЖЕНЕРИЯ И ИНЖЕНЕРИЯ ЗНАНИЙ

Berezin S. A.	127
Акатов А. Н.	128
Атутова Н. Д., Зюбина Д. А., Филиппов С. Д.	129
Бадмажапов Ж. С.	130
Ваганова А. И.	131
Васильев А. Д.	132

Вахрушев М. В.	133
Гавриленко А. В.	134
Галиева А. Г.	135
Ждаркин Е. Б.	136
Зайцев А. О.	137
Зиануров Н. О.	138
Зулин Д. К.	139
Зыкова А. А.	140
Кицкалов А. Е., Мартынов Н. Е., Черкашин Д. А.	141
Комагина К. Т.	142
Кондратьев А. С.	143
Коробова А. В.	144
Краева И. А.	145
Кузнецов А. С.	146
Лебедева А. Д.	147
Легостаева Е. О.	148
Лешов Р. А.	149
Мамеев Н. С.	150
Михайлов А. С.	151
Мызников П. В.	152
Орловский А. С.	153
Парфенов Д. Р.	154
Паульс А. Е., Деревянченко С. О., Николаев Н. А., Анисютин Л. Б.	155
Погодин Р. С.	156
Подкур Т. М., Шестакова Е. А.	157
Протасов Д. В.	158
Радеев Н. А.	159
Седухин О. А.	160
Скокова В. А.	161
Смирнов А. Ю.	162
Трегубов А. С.	163
Усова Д.	164
Хлиманкова Г. К.	165
Хлыбова М. М.	166
Худяков Д. А.	167
Шевчук Е. С.	168
Щербин А. С.	169
Якобсон А. А.	170

СЕТЕВЫЕ ТЕХНОЛОГИИ, ТЕХНОЛОГИИ ИНТЕРНЕТА И ИХ ПРИЛОЖЕНИЯ

Афанасьев С. Т.	171
Балданов З. Б.	172
Бедарева М. А.	173
Зенина О. А.	174
Коваленко А. П., Ковалева Д. К.	175
Манасян Т. Л.	176
Мирошниченко А. О., Беребердин Н. В., Литвинов Е. А.	177
Мокроусов А. С.	178
Мункоев А. А.	179
Попова В. А., Гармаева Д. А.	180
Прокопенко В.	181
Сафенрейтер Д. А.	182
Якимов В. С., Андрианов В. С.	183

ТЕОРИЯ И ПРАКТИКА ЦИФРОВОЙ ЭКОНОМИКИ

Denizay Sam.	184
Dilek Yakut	186
Баулина К. В.	188
Бацагин М. М.	189
Бизязев О. А.	190
Гаврилова Е. А.	191
Ермушина О. В.	192
Закиров В. К.	193
Исаев Н. А.	194
Карапетян С. А.	195
Клепикова Е. В.	196
Климонов М. С.	197
Коровченко Е. Д.	198
Кочеева А. И.	199
Латышева Р. Р.	200
Мельникова П. Ю.	201
Нестерова Т. Г.	202
Овсянникова М. А.	203
Пихурова А. Д.	204
Рехтин Д. С.	205
Шевченко Л. А.	206
Шигаева Ю. А.	207
Ширманова Г. С.	208

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Pauls A. E., Berezin S. A.	209
Багадаева Е. А.	210
Гавриленко А. А.	211
Гусев Д. И.	212
Гуськова М. Ю.	213
Девятовская А. Д., Бирючков Н. Е.	214
Деревянченко С. О.	215
Дунаевский А. А.	216
Заварзин Е. А., Мелехин А. А., Хафизов А. В., Савкина А. С., Городецкий Д. Д., Келлер К.	217
Кислицына Ю. М.	218
Крутиков А. К.	219
Митюшин Д. К.	220
Науменко Т. И.	221
Пак Б.	222
Рублев В. Д.	223
Чагина П. М.	224
Чудинов А. А.	225
Щеголева А. И.	226
Указатель авторов.	227

Научное издание

МНСК-2021

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Материалы

59-й Международной научной студенческой конференции

12–23 апреля 2021 г.

Корректор *Д. И. Ковалева*

Верстка *А. С. Терешкиной*

Обложка *Е. В. Неклюдовой*

Подписано в печать 05.05.2021 г.

Формат 60 × 84/16. Уч.-изд. л. 14,75. Усл. печ. л. 13,7.

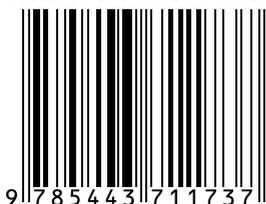
Тираж 49 экз. Заказ № 53.

Издательско-полиграфический центр НГУ

630090, г. Новосибирск, ул. Пирогова, 2

Секция
ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ

ISBN 978-5-4437-1173-7



N* Новосибирский
государственный
университет
***НАСТОЯЩАЯ НАУКА**

