

SIBERIAN

scientific workshop

SibDATA

2022

THE 3rd

17 июня
г. Красноярск

Тезисы докладов



**Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр «Красноярский научный центр
Сибирского отделения Российской академии наук»
(ФИЦ КНЦ СО РАН)**

**Институт вычислительного моделирования
Сибирского отделения Российской академии наук –
обособленное подразделение ФИЦ КНЦ СО РАН (ИВМ СО РАН)**

**Федеральное государственное бюджетное образовательное учреждение высшего
образования «Сибирский государственный университет науки и технологий имени
академика М.Ф. Решетнева» (СибГУ им. М.Ф. Решетнева)**

III Сибирский научный семинар

Data Analysis Technologies with Applications

SibDATA-2022

Сборник трудов

17 июня 2022 года

Красноярск, 2022

III Сибирский научный семинар Data Analysis Technologies with Applications (SibDATA-2022) [Электронный ресурс]: Сборник трудов научного семинара (17 июня 2022 г.). Красноярск: ИВМ СО РАН, 2022. 35 с.
ISBN 978-5-6047078-0-7

Сборник трудов подготовлен по результатам работы III Сибирского научного семинара Data Analysis Technologies with Applications (SibDATA-2022), который проходил 17 июня 2022 г. в г. Красноярске. Сборник содержит аннотации докладов на русском и английском языках. Семинар SibDATA-2022 направлен на обсуждение последних тенденций, инноваций, актуальных практических задач и решений в области технологий анализа данных. Семинар представляет собой научный форум, где ученые, исследователи, практики и преподаватели могут обмениваться и делиться своим опытом и новыми результатами своих исследований, передавать знания молодым ученым и черпать новые идеи. Материалы сборника публикуются в авторской редакции.

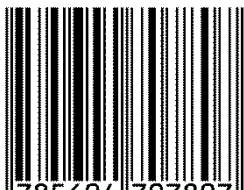
SibDATA-2022 проводится при поддержке
Красноярского математического центра, финансируемого Минобрнауки РФ
в рамках мероприятий по созданию и развитию региональных НОМЦ
(Соглашение 075-02-2022-873)

ПРОГРАММНЫЙ КОМИТЕТ СЕМИНАРА

Басан Е.С.	канд. техн. наук, Южный федеральный университет (Ростов-на-Дону)
Власов А.И.	канд. техн. наук, доцент, Московский государственный технический университет им. Н.Э. Баумана (г. Москва)
Гладков Л.А.	канд. техн. наук, доцент, Южный федеральный университет (г. Ростов-на-Дону)
Глазкова А.В.	канд. техн. наук, Тюменский государственный университет (г. Тюмень)
Дудченко А.В.	научный сотрудник, Институт медицинской информатики университетской больницы Гейдельберга (г. Гейдельберг, Германия)
Карпова Е.Д.	канд. физ.-мат. наук, доцент, Институт вычислительного моделирования СО РАН (г. Красноярск)
Кирик Е.С.	канд. физ.-мат. наук, доцент, Институт вычислительного моделирования СО РАН (г. Красноярск)
Коробко А.В.	канд. техн. наук, Сибирский государственный университет науки и технологий им. М.Ф. Решетнева (г. Красноярск)
Лапина М.А.	канд. физ.-мат. наук, доцент, Северо-Кавказский федеральный университет (г. Ставрополь)
Макарова Е.А.	д-р техн. наук, профессор, Уфимский государственный авиационный технический университет (г. Уфа)
Ноженкова Л.Ф.	д-р техн. наук, профессор, Институт вычислительного моделирования СО РАН (г. Красноярск)

- Пенькова Т.Г.** канд. техн. наук, доцент, Институт вычислительного моделирования СО РАН (г. Красноярск)
- Пустохина И.В.** канд. техн. наук, Российский экономический университет им. Г.В. Плеханова (г. Москва)
- Рыжов А.П.** д-р техн. наук, канд. физ.-мат. наук, профессор, Московский государственный университет им. М.В. Ломоносова (г. Москва)
- Садовский М.Г.** д-р техн. наук, профессор, Институт вычислительного моделирования СО РАН (г. Красноярск)
- Сенашова М.Ю.** канд. физ.-мат. наук, доцент, Институт вычислительного моделирования СО РАН (г. Красноярск)
- Симонов К.В.** д-р техн. наук, профессор, Институт вычислительного моделирования СО РАН (г. Красноярск)
- Тесля Н.Н.** канд. техн. наук, Санкт-Петербургский институт информатики и автоматизации РАН (г. Санкт-Петербург)
- Фаворская М.Н.** д-р техн. наук, профессор, Сибирский государственный университет науки и технологий им. М.Ф. Решетнева (г. Красноярск)
- Ходашинский И.А.** д-р техн. наук, профессор, Томский государственный университет систем управления и радиоэлектроники (г. Томск)
- Шайдуров В.В.** д-р физ.-мат. наук, член.-корр. Федеральный исследовательский центр «Красноярский научный центр» СО РАН, (г. Красноярск)
- Щербаков М.В.** д-р техн. наук, доцент, Волгоградский государственный технический университет (г. Волгоград)

ISBN 978-5-6047078-0-7



9 785604 707807

ISBN 978-5-6047078-0-7

© ИВМ СО РАН, 2022
© СибГУ им. М.Ф. Решетнева, 2022
© Коллектив авторов, 2022

СОДЕРЖАНИЕ

Басан Е.С., Прошкин Н.А., Силин О.И. ТЕХНОЛОГИЯ СОЗДАНИЯ ЛОЖНЫХ ИНФОРМАЦИОННЫХ ПОЛЕЙ ДЛЯ ЗАЩИТЫ БПЛА	5
Исаев С.В., Кононов Д.Д. АНАЛИЗ ДИНАМИКИ И КЛАССИФИКАЦИЯ АТАК НА ВЕБ-СЕРВИСЫ КОРПОРАТИВНОЙ СЕТИ	7
Касьянов В.Н. ВИЗУАЛИЗАЦИЯ СТРУКТУРИРОВАННЫХ ДАННЫХ НА ОСНОВЕ АТРИБУТИРОВАННЫХ ИЕРАРХИЧЕСКИХ ГРАФОВ С ПОРТАМИ	9
Кенц А.С., Зотин А.Г., Хамад Ю.А., Симонов К.В., Кругляков А.С. КОРРЕЛЯЦИЯ ЛАБОРАТОРНЫХ И ЛУЧЕВЫХ МЕТОДОВ ИССЛЕДОВАНИЯ СО СТЕПЕНЬЮ ТЯЖЕСТИ ПАЦИЕНТОВ ПРИ COVID-19	11
Коробко А.В., Шефтель К.С. МОДЕЛЬНО-ОРИЕНТИРОВАННЫЙ ПОДХОД К ФОРМИРОВАНИЮ ОПЕРАТИВНЫХ АНАЛИТИЧЕСКИХ ОТЧЁТОВ	13
Курносенко А.Е., Власов А.И. СОЗДАНИЕ ЦИФРОВОГО ДВОЙНИКА СБОРОЧНОГО ПРОИЗВОДСТВА ЭЛЕКТРОННОЙ АППАРАТУРЫ С ПРИМЕНЕНИЕМ ДВУХУРОВНЕВОЙ ДЕКОМПОЗИЦИИ	15
Медиевский А.В., Зотин А.Г., Симонов К.В., Кругляков А.С. МЕТОДИКА ВИЗУАЛИЗАЦИИ КЛЕТОК ЦЕНТРАЛЬНОЙ НЕРВНОЙ СИСТЕМЫ НА ОСНОВЕ ОБРАБОТКИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ	17
Метус А.М., Морозов Р.В., Ноженков А.И. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ РЕЙТИНГОВОГО ОЦЕНИВАНИЯ ТЕРРИТОРИЙ	19
Мионов В.А., Перетокин С.А. Симонов К.В. К ВОПРОСУ О ПОСТРОЕНИИ РЕГИОНАЛЬНОЙ МОДЕЛИ ЗАТУХАНИЯ СИЛЬНЫХ ДВИЖЕНИЙ ГРУНТА ДЛЯ БАЙКАЛЬСКОЙ РИФТОВОЙ ЗОНЫ	21
Молявко А.А., Кареева Е.Д., Садовский М.Г. ПРИМЕНЕНИЕ ТЕХНОЛОГИИ МРІ ДЛЯ РАСПАРАЛЛЕЛИВАНИЯ В ЗАДАЧЕ СРАВНЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ МЕТОДОМ ШАЙДУРОВА	23
Ничепорчук В.В. АЛГОРИТМИЗАЦИЯ РАЗРАБОТКИ АНАЛИТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ УПРАВЛЕНИЯ БЕЗОПАСНОСТЬЮ ТЕРРИТОРИЙ	25
Пенькова Т.Г., Ноженкова Л.Ф. ИССЛЕДОВАНИЕ ТЕРРИТОРИАЛЬНЫХ КЛАСТЕРОВ МАКРОРЕГИОНА НА ОСНОВЕ МНОГОМЕРНОГО АНАЛИЗА ПОКАЗАТЕЛЕЙ КАЧЕСТВА ЖИЗНИ (НА ПРИМЕРЕ КРАСНОЯСКОГО КРАЯ)	27

Сенашова М.Ю. ОСОБЕННОСТИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ФОТОСИНТЕТИЧЕСКОЙ СИСТЕМЫ ХЛОРОПЛАСТОВ ВОДОРОСЛЕЙ	ГЕНОВ	29
Фаворская М.Н., Пахирка А.И. УСИЛЕНИЕ АРТЕФАКТОВ НА КАРТАХ ГЛУБИНЫ ИЗОБРАЖЕНИЙ ЛИЦ ПРИ ОБНАРУЖЕНИИ ПРЕЗЕНТАЦИОННЫХ АТАК		31
Хамад Ю.А., Медиевский А.В., Зотин А.Г., Симонов К.В., Кругляков А.С. СЕГМЕНТАЦИЯ ОБЪЕМНЫХ ОБРАЗОВАНИЙ НА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ И МЕТОДА Пороговой обработки		33

ТЕХНОЛОГИЯ СОЗДАНИЯ ЛОЖНЫХ ИНФОРМАЦИОННЫХ ПОЛЕЙ ДЛЯ ЗАЩИТЫ БПЛА

Е.С. Басан, Н.А. Прошкин, О.И.Силин

Южный федеральный университет, г. Ростов-на-Дону

В наше время вопросы обеспечения безопасности беспилотных летательных аппаратов стоят достаточно остро. БПЛА можно использовать для реализации критических процессов. Такими процессами могут быть ретрансляция сообщений, поиск пропавших без вести, разведка, охрана объектов и многое другое. При этом БПЛА часто находится за пределами контролируемой зоны и физически незащищен. Защитить БПЛА и каналы связи между БПЛА сложно, поскольку каналы беспроводной связи, используемые для БПЛА, физически доступны. Существует множество вариантов физической защиты, которые включают создание дополнительного шума в канале. В этом случае злоумышленник может попытаться провести информационную атаку или, для начала, просто обнаружить БПЛА. В данном исследовании предлагается способ сокрытия БПЛА путем создания ложных информационных полей. Затем проводится тестирование метода, путем анализа радиоспектра и сравнения поддельных полей с допустимыми. Результаты показали эффективность разработанного программного обеспечения, которое позволяет создавать ложные точки доступа, которые могут быть обнаружены злоумышленником и позволяют скрыть настоящую передачу. В исследовании продемонстрирована работа модуля эмуляции активности в радиоканале связи для нескольких БПЛА, создав несколько поддельных точек доступа, которые передают трафик, состоящий только из пакетов маяков и зондов. Анализ спектров различного, попеременно возрастающего количества точек доступа показал, что с увеличением количества точек доступа изменяется количество несущих частот, что показывает несколько работающих отдельных устройств Wi-Fi, частота появления пиков увеличивается – увеличивается вероятность запутывания противника – имитация активного радиотрафика. Анализ спектра реальной работы устройства показал, что при одновременном включении 13 точек доступа сигнал становится более непрерывным и эффективно запутанным по сравнению с 9 точками доступа. Это связано с тем, что при попеременном увеличении количества точек доступа появляется ошибка во время запуска следующей точки доступа, тем самым можно указать запуск точки доступа в неподходящее время. Когда 13 точек доступа создаются и запускаются одновременно, каждая из них осуществляет широкополосную передачу с равным интервалом времени от соседней, таким образом достигается эффект непрерывного радиотрафика. В заключение отметим, что был проведен эксперимент с простейшим адаптером Wi-Fi и слабой антенной. В реальных условиях следует использовать гораздо более мощный излучатель для увеличения дальности и уровня сигнала, а противник будет использовать более чувствительную антенну. Также для увеличения дальности и уровня сигнала можно использовать узконаправленные антенны, однако необходимо точно понимать местоположение разведки противника.

TECHNOLOGY OF CREATING FALSE INFORMATION FIELDS FOR PROTECTION OF UAV

E.S. Basan, N.A. Proshkin, O.I. Silin

Southern Federal University, Rostov-on-Don

Nowadays, the issues of ensuring the security of unmanned aerial vehicles are (UAV) quite acute. UAVs can be used to implement critical processes. Such processes are messages, searching for missing persons, reconnaissance, guarding objects, and much more. At the same time, the UAV is often outside the controlled area and is physically unprotected. Protecting UAVs and communication channels between UAVs is difficult because the wireless communication channels are not physically secure. There are many options for physical protection, which include the creation of additional noise in the channel. In this case, the attacker may try to carry out an informational attack or detect the UAV. This study proposes a way to hide the UAV by creating false information fields. Then the method is tested by analyzing the radio spectrum and comparing fake fields with valid ones. The results showed the effectiveness of the developed software, which allows to create fake access points that can be detected by an attacker and allow to hide the real transmission. This study demonstrated the operation of the radio frequency emulation module for several UAVs by creating several fake access points that transmit traffic consisting only of beacon and probe packets. An analysis of the spectra of a different, alternately increasing number of access points showed that increasing in the number of access points, the number of carrier frequencies changes, which shows several working separate Wi-Fi devices, the frequency of peaks increases – the probability of adversary confusion increases – imitation of active radio traffic. While 9 simultaneously running access points do not change the effectiveness of obfuscation in any way. An analysis of the spectrum of the actual operation of the device showed that when 13 access points are turned on simultaneously, the signal becomes more continuous and effectively entangled compared to 9 access points. This is due to the fact that when we increase the number of access points alternately, we make an error during the start of the next access point, thus we can specify the start of the access point at the wrong time. When 13 access points are created and launched at the same time, each of them broadcasts at an equal time interval from its neighbor, thus achieving the effect of continuous radio traffic. In conclusion, we note that an experiment was conducted with the simplest Wi-Fi adapter and a weak antenna. In real conditions, a much more powerful radiator should be used to increase the range and signal level, and the enemy will use a more sensitive antenna. Also, to increase the range and signal level, you can use narrowly directed antennas, but it is necessary to accurately understand the location of enemy reconnaissance.

АНАЛИЗ ДИНАМИКИ И КЛАССИФИКАЦИЯ АТАК НА ВЕБ-СЕРВИСЫ КОРПОРАТИВНОЙ СЕТИ

С.В. Исаев, Д.Д. Кононов

Институт вычислительного моделирования СО РАН, г. Красноярск

В настоящее время многие компании используют веб-технологии для организации корпоративных сервисов различного уровня. Веб-сервисы подвержены рискам информационной безопасности, поскольку функционируют в открытой сети Интернет. Кроме того, важной частью функционирования современных информационных систем является задача обеспечения кибербезопасности, в том числе оценка рисков киберактивов. Эта задача является комплексной. Одним из важных компонентов получения информации для анализа являются различные журналы активности, которые генерирует система.

Проведен обзор существующих работ, в которых используются различные методы анализа журналов сервисов. Часто авторы описывают методику анализа и в качестве примера используют тестовые данные, что не позволяет оценить работоспособность подхода на реальных данных. Либо используются реальные данные с короткими временными интервалами, что затрудняет анализ динамики происходящих процессов за различные периоды. В данной работе проводится исследование безопасности корпоративной сети Красноярского научного центра (ФИЦ КИЦ СО РАН) на основе анализа журналов веб-сервисов. Целью работы является анализ безопасности веб-сервисов в динамике за последние 2 года, классификация кибератак по видам, выявление зависимостей между различными параметрами атак. В отличие от существующих работ, анализ выполняется на длительных временных интервалах, что позволяет выявить динамику поведения веб-сервисов по часам, дням, месяцам и годам.

Источниками данных для анализа в работе являются данные веб-сервисов за 2020-2021 годы и неполный 2022 год. Первичная обработка включает агрегацию журналов со всех веб-сервисов и унификацию формата для последующей обработки. Для всех данных выполняется геопривязка источника. Затем выполняется обработка ошибок с агрегацией по различным интервалам времени. Также выполняется обработка атак, которая включает классификацию по типам с последующей агрегацией по геоданным. Классификация атак по типам осуществляется по OWASP с использованием набора правил ModSecurity Core Rule Set, предназначенного для идентификации киберугроз веб-приложений. Обработка и анализ данных осуществляются с помощью, как готовых программных продуктов, так и разработанных самостоятельно.

Результатом проведенного анализа является оценка динамики атак на веб-сервисы по странам, выделение основных групп стран с постоянным профилем атак и высокой их интенсивностью. Выполнено сравнение попарных корреляций различных видов атак, выявлены атаки с высокой корреляцией, которые можно агрегировать при оценке рисков. Предложен эвристический метод сравнения рисков кибербезопасности для различных периодов, использующий классификацию по видам атак. Он не зависит от сравниваемых временных интервалов и объема выборок, т.к. основан на статистических показателях. Метод оценки рисков кибербезопасности может использоваться не только для веб-сервисов, но и в других областях, в которых существует классификация показателей.

ANALYSIS OF DYNAMICS AND CLASSIFICATION OF ATTACKS TO THE WEB SERVICES OF THE CORPORATE NETWORK

S.V. Isaev, D.D. Kononov

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

Currently, many companies use web technologies to organize various corporate services. Web services are subject to information security risks because they operate on the open Internet. In addition, an important part of the functioning of modern information systems is the task of ensuring cybersecurity including risk assessment of cyber assets. This task is complex. One of the important components of obtaining information for analysis is the various activity logs that the system generates.

A review of existing works that use various methods for analyzing service logs is carried out. Often, authors describe the analysis methodology and use test data as an example, which does not allow evaluating the described approach on real data. In other cases, real data are used with short time intervals, which makes it difficult to analyze the dynamics of processes for different periods. In this paper, we study the security of the corporate network of the Krasnoyarsk Scientific Center (FRC KSC SB RAS) based on the analysis of web service logs. The aim of the work is to analyze the security of web services over the past 2 years, classify cyber attacks by type, and identify dependencies between various attack parameters. Unlike existing works, the analysis is performed over long time intervals, which allows you to identify the dynamics of the behavior of web services by hours, days, months, and years.

The data sources for the analysis in the work are web services log data for 2020-2021 and an incomplete 2022. Primary processing includes aggregation of logs from all web services and unification of the format for further processing. Source geolocation is performed for all data. Then, error handling is performed with aggregation over various time intervals. Attacks detection are also processed, which includes classification by type followed by aggregation by geodata. Classification of attacks by type is carried out according to OWASP using the ModSecurity Core Rule Set designed to identify cyber threats to web applications. Data processing and analysis are carried out using both ready-made software products and independently developed ones.

The result of the analysis is an assessment of the dynamics of attacks on web services by country, the identification of the main groups of countries with a constant profile of attacks, and their high intensity. A comparison of pairwise correlations of various types of attacks was carried out, attacks with high correlation were identified, which can be aggregated when assessing risks. A heuristic method of comparing cybersecurity risks for different periods using classification by types of attacks is proposed. It does not depend on the compared time intervals and the volume of samples, because it is based on statistical indicators. The cybersecurity risk assessment method can be used not only for web services but also in other areas in which there is a classification of indicators.

ВИЗУАЛИЗАЦИЯ СТРУКТУРИРОВАННЫХ ДАННЫХ НА ОСНОВЕ АТТРИБУТИРОВАННЫХ ИЕРАРХИЧЕСКИХ ГРАФОВ С ПОРТАМИ

В.Н. Касьянов

Институт систем информатики им. А.П. Ершова СО РАН, г. Новосибирск

Визуализация структурированных или реляционных данных на основе графовых моделей имеет множество сфер применения как в реальных, так и в теоретических областях. Среди них — физические сети связи и электрические сети, с одной стороны, и структуры данных компилятора и диаграммы изменения состояний, с другой. Поэтому в настоящее время на рынке широко представлены наукоемкие программные продукты, использующие методы визуализации информации на основе графовых моделей, такие как Cytoscape, Nigres, Gephi, Graphviz, Tulip, yEd и многие другие.

Поскольку информация, которую желательно визуализировать, постоянно увеличивается и усложняется, возникает все больше ситуаций, в которых классические графовые модели перестают быть адекватными. Требуются и возникают более мощные теоретико-графовые формализмы для представления информационных моделей, обладающих иерархической структурой, поскольку иерархичность является основой многочисленных методов визуальной обработки сложных больших данных в различных областях применения вычислительных систем. Одним из таких формализмов являются так называемые иерархические графы и графовые модели. Этот формализм позволяет выделить в исходном графе множество таких его частей (так называемых фрагментов), что все элементы каждого фрагмента заслуживают отдельного совместного рассмотрения, а все фрагменты выделенного множества образуют иерархию по вложенности. В Институте систем информатики им. А.П. Ершова СО РАН по заказу компании Intel была создана система визуализации Visual Graph, которая основана на иерархических графовых моделях и позволяет исследовать сложные структурированные большие данные, возникающие в компиляторах и в других системах конструирования программ, через их визуальные представления.

Во многих приложениях объекты, моделируемые вершинами графа, являются сложными и содержат непересекающиеся логические части, через которые они (эти объекты) находятся во взаимосвязи, моделируемой ребрами. При представлении графов с вершинами, моделирующими сложные объекты, в существующих форматах описания графов (см., например, GraphML) эти различные части сложных объектов обычно выражаются с помощью так называемых портов вершин, которые при изображении графа могут представляться разными точками изображений вершин, в которых соответствующие вершины соединяются с инцидентными им ребрами.

В докладе введен формализм атрибутированных иерархических графов с портами и рассмотрены новые возможности системы Visual Graph по визуализации структурированных данных большого размера на основе атрибутированных иерархических графов с портами.

VISUALIZATION OF STRUCTURED DATA BASED ON ATTRIBUTED HIERARCHICAL GRAPHS WITH PORTS

V.N. Kasyanov

A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk

Visualization of structured or relational data based on graph models has many applications in both real and theoretical fields. Among them are physical communication networks and electrical networks, on the one hand, and compiler data structures and state transition diagrams, on the other. Therefore, at present, science-intensive software products using information visualization methods based on graph models (such as Cytoscape, Higris, Gephi, Graphviz, Tulip, yEd and many others) are widely represented on the market.

Since the information that it is desirable to visualize is constantly growing and becoming more complex, more and more situations arise in which classical graph models cease to be adequate. More powerful graph-theoretic formalisms have been needed and introduced to represent a hierarchical kind of diagramming objects, since hierarchy is the basis of numerous methods for visual processing of complex big data in various fields of applications. One of these formalisms is the so-called hierarchical graphs and graph models. The hierarchical graph model allows selecting in the given classical graph a set of such its parts (so-called fragments) that all elements of each selected fragment deserve separate joint consideration, and all fragments of the selected set form a nesting hierarchy. At the A.P. Ershov Institute of Informatics Systems by order of the Intel Company the Visual Graph system has been constructed as a visualization system based on hierarchical graph models and aimed to explore complex structured big data that occurs in compilers and other program design systems through their visual representations.

In many applications, objects modeled by graph vertices are complex and contain non-intersecting logical parts through which they (these objects) are in a relationship modeled by edges. When representing graphs with vertices that model complex objects in existing graph description formats (see, for example, GraphML), these different parts of complex objects are usually expressed using the so-called vertex ports, which in a graph drawing can be represented by different points of vertex images, in which the corresponding vertices are connected to the arcs incident to them.

In the paper the formalism of attributed hierarchical graphs with ports is introduced and new possibilities of the Visual Graph system for visualization of large structured data based on attributed hierarchical graphs with ports are considered.

КОРРЕЛЯЦИЯ ЛАБОРАТОРНЫХ И ЛУЧЕВЫХ МЕТОДОВ ИССЛЕДОВАНИЯ СО СТЕПЕНЬЮ ТЯЖЕСТИ ПАЦИЕНТОВ ПРИ COVID-19

А.С. Кенц¹, А.Г. Зотин², Ю.А. Хамад³, К.В. Симонов⁴, А.С. Кругляков⁴

¹*Федеральный Сибирский научно-клинический центр ФМБА России*

²*Сибирский государственный университет науки и технологий им. ак. М.Ф. Решетнева*

³*Сибирский федеральный университет*

⁴*Институт вычислительного моделирования СО РАН*

г. Красноярск

Исследование посвящено поиску корреляции и ретроспективной оценке лабораторных и инструментальных данных со степенью тяжести пациентов при COVID-19. Выполнено усовершенствование методики интерпретации и анализа динамических изменений, связанных с COVID-19, на КТ снимках легких. Методика включает ключевые этапы: предварительная обработка, сегментация с цветовым кодированием, расчет и кумулятивная оценка признаков для выделения областей с вероятной патологией (включая комбинированную оценку признаков). Анализ и интерпретация проводится на формирующейся базе данных пациентов. При этом выделяются следующие критерии: КТ исследование легких в динамике и клинические данные (степень тяжести, температура, сатурация и др.). Результаты лабораторных исследований анализируются с акцентом ключевого показателя – интерлейкин-6. Данный показатель является маркером значимых и серьезных изменений тяжести состояния пациента.

Задачи исследования: сбор общей базы данных пациентов, с подтвержденным COVID-19, с отображением лабораторных и инструментальных показателей; поиск взаимосвязи одного из показателей (интерлейкин-6) и степенью тяжести пациентов по КТ изображениям; поиск специфического КТ паттерна, либо корреляции показателей для поиска причин формирования отсроченных изменений. На данный момент сформированная база данных представлена 2200 пациентами в период с 2020 по 2022 гг. и включает 60,9% женщин и 39,1% мужчин. Кроме степени тяжести по КТ снимкам, у пациентов отслежены изменения в динамике по КТ снимкам и изменения, указанных выше клинических показателей. В ходе исследования изучалось предположение о регрессе остаточных изменений у пациентов, выписанных с грубыми ретикулярными изменениями. Данное предположение подтверждено на «эталонном пациенте» при динамическом исследовании через 6 месяцев с момента болезни. При анализе динамики отмечается регресс достаточно плотных зон с восстановлением их «воздушности». При динамическом наблюдении 20 пациентов отмечаем, что все имеют схожую динамику и выражено положительную МСКТ картину сроком от 4 и более месяцев после перенесенного заболевания. Проведенные исследования позволяют аргументировать предположение о регрессе патологических изменений в легких, но требуют большую выборку пациентов, что также является основой дальнейших исследований. Также, при ретроспективном анализе будет оцениваться корреляция увеличения степени тяжести пациента соизмеримой с ухудшением картины по КТ снимкам и показателей Интерлейкина-6.

CORRELATION OF LABORATORY AND CT METHODS WITH COVID-19 PATIENT SEVERITY

A.S. Kents¹, A.G. Zotin², Yu.A. Hamad³, K.V. Simonov⁴, A.S. Kruglyakov⁴

¹*Federal Siberian Scientific and Clinical Center FMBA of Russia*

²*Reshetnev Siberian State University of Science and Technology*

³*Siberian Federal University*

⁴*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

The study is devoted to the search for a correlation and a retrospective assessment of laboratory and instrumental data with the severity of patients with COVID-19. An improvement was made in the methodology for interpreting and analyzing dynamic changes associated with COVID-19 on CT images of the lungs. The technique includes key steps: pre-processing, segmentation with color-coding, calculation and cumulative feature evaluation to identify areas with probable pathology (including combined feature evaluation). Analysis and interpretation is carried out on the emerging database of patients. At the same time the following criteria are distinguished: CT examination of the lungs in dynamics and clinical data (severity, temperature, saturation, etc.). The results of laboratory tests are analyzed with an emphasis on the key indicator – interleukin-6. This indicator is a marker of significant and serious changes in the severity of the patient's condition.

The aim of the study is a comprehensive assessment of the correlation of laboratory and CT images based on a retrospective analysis of patients with a severe course of the disease and interleukin-6 (IL-6) parameters. Research objectives: collection of a general database of patients with confirmed COVID-19 displaying laboratory and instrumental indicators; search for the relationship of one of the indicators (interleukin-6) and the severity of patients on CT images; search for a specific CT pattern or correlation of indicators to search for the causes of the formation of delayed changes. Presently, the formed database is represented by 2200 patients in the period from 2020 to 2022 and includes 60.9% of women and 39.1% of men. In addition to the severity according to CT images changes in the dynamics according to CT images and changes in the above clinical indicators were monitored in patients. The study examined the assumption of regression of residual changes in patients discharged with gross reticular changes. This assumption was confirmed on the «reference patient» with a dynamic study 6 months after the onset of the disease. When analyzing the dynamics, a regression of rather dense zones with the restoration of their «airiness» is noted. During the dynamic observation of 20 patients, we note that all have similar dynamics and a pronouncedly positive CT images picture for a period of 4 or more months after the disease. The conducted studies allow arguing the assumption about the regression of pathological changes in the lungs, but require a large sample of patients, which is also the basis for further research. Also, in a retrospective analysis the correlation of an increase in the severity of the patient commensurate with deterioration in performance according to CT and Interleukin-6 indicators will be assessed.

МОДЕЛЬНО-ОРИЕНТИРОВАННЫЙ ПОДХОД К ФОРМИРОВАНИЮ ОПЕРАТИВНЫХ АНАЛИТИЧЕСКИХ ОТЧЁТОВ

А.В. Коробко¹, К.С. Шефтель²

¹*Институт вычислительного моделирования СО РАН, г. Красноярск*

²*Сибирский государственный университет науки и технологий
имени ак. М.Ф. Решетнева*

В условиях обилия информационных потоков, всевозрастающих вычислительных мощностей и повышения требований к обоснованности принимаемых решений, сохраняют свою актуальность сводные отчетные формы, особенно в критически важных областях деятельности человека. Несмотря на повсеместное внедрение оперативных информационных панелей и дашбордов, обновляемых автоматически в момент поступления новых данных, важной частью реактивного и проактивного управления являются регламентированные срочные отчетные формы. Печатные отчеты позволяют «фиксировать» состояние отслеживаемой ситуации на момент формирования отчета, отправлять всем заинтересованным лицам одну и ту же информацию, не беспокоясь о возможном изменении информации и манипуляции с данными для формирования не совсем объективной картины.

Несмотря на кажущуюся статичность отчетов, чтобы оставаться актуальными они должны изменяться. На примере оперативной сводки Центра ГО и ЧС видно, что отчет претерпевает постоянные изменения. В зависимости от сезона меняется перечень контролируемых обстановок: подтопления и лесные пожары характерны для весенне-летнего периода, а мониторинг ледовых переправ необходим исключительно зимой. С окончанием отчетного года в некоторые разделы отчета добавляются новые столбцы для сопоставления показателей с нарастающим итогом с аналогичными показателями за предшествующие периоды. Ввиду развития средств инструментального контроля добавляются новые разделы, такие как «космический мониторинг термоточек». С точки зрения программных средств формирования оперативного отчета все эти изменения требуют непосредственного участия человека в адаптации и настройки новой версии отчета. Суть современной концепции «zero code» заключается в тезисе: «если участие человека необходимо, то пусть это будет аналитик, а не программист». Модельно-ориентированный подход позволяет разрабатывать «zero code» системы за счет вынесения бизнес-логики из программного кода в метаданные и разработке алгоритмов формирования системы под управлением этих метаданных. Структура метаданных задает потенциальный функционал системы, а данные, хранящиеся в этой структуре, определяют информационное наполнение системы.

Настоящее исследование посвящено разработке методических и алгоритмических средств поддержки создания и редактирования печатных отчетных форм на базе модельно-ориентированного подхода на примере отчетной формы оперативной сводки Центра ГО и ЧС. В работе рассмотрена структура печатной отчетной формы, выполнена классификация основных элементов отчета с точки зрения информационного наполнения и порядка формирования. Разработан шаблон печатной отчетной формы, включающий статические и динамические элементы. Предложена структура метаданных для поддержки динамического формирования печатных отчетов на основе шаблона. Построена модель процесса формирования отчетов на основе метаданных. Разработан комплекс программных классов, реализующих алгоритмы заполнения шаблона фактическими данными. Выполнен анализ существующих решений и сделан обоснованный выбор программной библиотеки динамического формирования отчетов «Jinja». Предложенные средства апробированы на данных Агентства по гражданской обороне, чрезвычайным ситуациям и пожарной безопасности Красноярского края в рамках задачи формирования печатной формы оперативной сводки.

MODEL-DRIVEN APPROACH TO FORMATION OF OPERATIONAL ANALYTICAL REPORTS

A.V. Korobko¹, K.S. Sheftel²

¹*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,*

²*Reshetnev Siberian State University of Science and Technology, Krasnoyarsk*

Operational reports remain relevant despite an abundance of information flows, ever-increasing computing power and rising requirements to management decisions quality. This is especially true for critical areas of human activity. Short-term operational reports are important part of tactical management. They are actively used along with strategic dashboards that are updated automatically when new data arrives. Printed reports allow you to “fix” the state of the monitored situation at the time of the report generation, send the same information to all interested parties without worrying about possible changes in information and data manipulation.

Printed reporting forms are usually static, but often they must be changed to fit the current agenda. The example is the operational summary of the Center for Civil Defense and Emergencies. The form of the summary is subject to constant changes. The list of controlled situations varies depending on the season: flooding and wildfires are typical for the spring-summer period, and monitoring of ice crossings is necessary only in winter. With the end of the reporting year, new columns are added to some parts of the report form to compare the cumulative totals with those for previous years. The development of instrumental control leads to the addition of new sections, such as "space monitoring of hotspots". Changes to the report form require direct human participation. Bringing in a qualified programmer for each such change is expensive and time consuming. The modern solution to the problem is low code/no code (LC/NC) applications. LC/NC applications can provide a close fit to business requirements, can be implemented quickly, and typically cost much less than full stack systems development. The theoretical basis for such solutions is the model-driven approach (MDA). The use of MDA allows you to save the basic rules of business logic and the available functionality of the application in the form of metadata, giving the user the opportunity to fill the application with their own data and interfaces.

This study is devoted to the development of methodological and algorithmic tools to support the creation and editing of printed report forms based on a MDA in case of the operational summary of the Center for Civil Defense and Emergencies. The paper considers the structure of the printed report form, classifies the main elements of the report in terms of information content and the order of formation. The template for a printed report form has been developed, including static and dynamic elements. The metadata structure is proposed to support the dynamic generation of printed reports based on a template. A complex of software classes has been developed that implements algorithms for filling the template with actual data. The analysis of existing solutions was carried out and a reasonable choice of the Jinja dynamic reporting library was made. The proposed methodological and algorithmic tools are used to solve the tasks of the Agency for Civil Defense, Emergencies and Fire Safety of the Krasnoyarsk Territory in order to generate the printed form of the operational summary.

СОЗДАНИЕ ЦИФРОВОГО ДВОЙНИКА СБОРОЧНОГО ПРОИЗВОДСТВА ЭЛЕКТРОННОЙ АППАРАТУРЫ С ПРИМЕНЕНИЕМ ДВУХУРОВНЕВОЙ ДЕКОМПОЗИЦИИ

А.Е. Курносенко, А.И. Власов

Московский государственный технический университет им. Н.Э. Баумана, г. Москва

Работа посвящена методу создания цифровых двойников на примере сборочного производства электронных модулей на печатных платах. Задача решалась с применением методов дискретного имитационного моделирования, а также связки конструкторской САПР Solid Edge и системы подготовки производства Tecnomatix. Выполнялась двухуровневая декомпозиция производственных объектов. На первом, верхнем уровне декомпозиции технологический процесс сборки разбивался на отдельные операции с последующей классификацией этих операций как выполняемых полностью на автоматизированном оборудовании и вручную с возможным применением средств автоматизации. Далее для каждой классифицированной операции на втором, нижнем уровне осуществлялась декомпозиция на отдельные технологические переходы, сопровождаемая декомпозицией технологического оборудования и оснащения на отдельные функциональные узлы, выполняющие переход или группу переходов. Внимание в работе уделено второму уровню декомпозиции, рассмотренному на примере моделирования операции установки компонентов автоматически – с помощью сборочного автомата и вручную – с участием цифровых моделей антропоморфных манекенов, имитирующих работу операторов-монтажников электронных компонентов на печатные платы. Представлены результаты декомпозиции рабочего места монтажника и полученная в результате комплексная цифровая модель рабочего места установки компонентов. В экспериментальной части работы показаны преимущества подхода с декомпозицией, связанные, в частности, с возможностью получения результатов на различных иерархических уровнях разбиения. На созданной модели проведен ряд экспериментов вида «что, если» с различным составом, компоновкой и количеством автоматизированных и ручных рабочих мест. В первом исследовании представлено накопленное время, которое монтажник тратит на последовательное достижение четырех основных функциональных узлов системы установки компонентов при различной компоновке своего рабочего места с последующим выбором наиболее эффективной конфигурации с точки зрения производительности работы. Во втором исследовании была промоделирована вся технологическая линия сборки в составе автомата трафаретной печати, автомата установки компонентов, печи пайки оплавлением и рабочих мест ручной установки компонентов в двух вариантах компоновки. Продемонстрированы результаты оценки производительности различных вариантов компоновки производственной линии в целях ее балансировки с достижением высокой и равномерной загрузки технологического оборудования. Отмечены достоинства данного подхода к созданию цифрового двойника сборочного производства, среди которых возможность проигрывать различные сценарии вида «что, если» на каждом из иерархических уровней без затратных натурных экспериментов, связанных с внесением изменений в конфигурацию реальной линии, а также без необходимости проведения физического запуска самой линии; получение количественной оценки описываемых производств в виде коэффициента загрузки и производительности технологического оборудования и операций; возможность компоновки рабочего места с учетом эргономики и анатомических особенностей человека-оператора. Среди возможных дополнительных задач, решаемых с применением созданной модели, выделены моделирование запуска в производство новых изделий, оптимизация переналадок технологического оборудования и количества обслуживающего персонала, организация и оптимизация материальных потоков с участием автоматизированного склада. Отмечено отсутствие привязки подхода к конкретной отрасли и возможность его распространения на иные процессы.

CREATING A DIGITAL TWIN ASSEMBLY PRODUCTION OF ELECTRONIC EQUIPMENT USING TWO-LEVEL DECOMPOSITION

A.E. Kurnosenko, A.I. Vlasov

Bauman Moscow State Technical University, Moscow

The paper is devoted to the method of creating digital twins on the example of PCB assembly. The problem was solved using the methods of discrete simulation modeling, as well as a combination of the Solid Edge CAD system and the Tecnomatix preproduction system. A two-level decomposition of production objects was carried out. At the first, upper level of decomposition, the assembly process was divided into separate operations with the subsequent classification of these operations as being performed entirely on automated equipment and manually with the possible use of automation tools. Further, for each classified operation, at the second, lower level, decomposition into separate process steps was carried out, accompanied by the decomposition of technological equipment and tools into separate functional units that perform a step or a group of steps. Attention is paid to the second level of decomposition, considered on the example of simulating the automatic component placement operation – using an assembly equipment and manually – with the participation of digital models of scalable human figures that simulate the electronic component mounting on PCBs. The results of the workplace decomposition and the resulting complex digital model of the component mounting workplace are presented. The experimental part of the work shows the advantages of the decomposition approach, related, in particular, to the possibility of obtaining results at different hierarchical levels of decomposition. On the created model, a number of “what if” experiments were carried out with different composition, layout and number of automated equipment and manual workplaces. The first study presents the accumulated time that an operator spends sequentially reaching the four main functional modules of the component mounting system with various layouts of his workplace, and then choosing the most efficient configuration in terms of work productivity. In the second study, the entire assembly line was simulated, consisting of a screen-printing, a component placing, a reflow soldering equipment, and manual component placement workstations in two layouts. The results of evaluating the performance of various production line layouts are demonstrated in order to balance the line with the achievement of high and uniform loading of process equipment. The advantages of this approach to creating a digital twin of assembly manufacturing are noted, including the ability to play various “what if” scenarios at each of the hierarchical levels without costly full-scale experiments associated with making changes to the configuration of a real line, and also without the need to physically launch the line itself; obtaining a quantitative assessment of the described production floors in the form of a utilization factor and productivity of technological equipment and operations; the possibility of arranging the workplace, taking into account ergonomics and anatomical features of the human operator. Among the possible additional tasks solved using the created model, modeling of the new product introduction, optimization of the technological equipment changeover and the number of service personnel, organization and optimization of material flows with the participation of an automated warehouse are highlighted. It is noted that the approach is not tied to a specific industry and could be applied to other processes.

МЕТОДИКА ВИЗУАЛИЗАЦИИ КЛЕТОК ЦЕНТРАЛЬНОЙ НЕРВНОЙ СИСТЕМЫ НА ОСНОВЕ ОБРАБОТКИ МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЙ

А.В. Медиевский¹, А.Г. Зотин², К.В. Симонов³, А.С. Кругляков³

¹*Красноярский государственный медицинский университет
имени профессора В.Ф. Войно-Ясенецкого,*

²*Сибирский государственный университет науки и технологий им. ак. М.Ф. Решетнева*

³*Институт вычислительного моделирования СО РАН,
г. Красноярск*

Основными задачами исследования являются: разработка методики выделения объектов на нативных микроскопических изображениях и визуализация интересующих клеток с усилением контраста с помощью цветового кодирования. Для работы с клетками, находящимися в трехмерном пространстве, необходимо использовать методы микроскопии, позволяющие визуализировать биологическую ткань послойно. Применяется для этой цели чаще всего флуоресцентная микроскопия, но среди её недостатков можно выделить трудоемкость приготовления препарата и высокую фототоксичность. Поэтому для структурно-функционального исследования объемных тканей лучше использовать метод дифференциально-интерференционного контраста (ДИК). Для трассировки нейронов, помимо ДИК-микроскопа, оптимальным вариантом является фазово-контрастный микроскоп, работающий по аналогичному принципу. Он имеет более простую структуру, что приводит к меньшим затратам и, как следствие, большей популярности. В рамках предлагаемой вычислительной методики для решения задачи визуализации можно выделить три основных этапа обработки и анализа визуальных данных. Этап предварительной обработки, включающий шумоподавление, а также коррекцию яркости и контрастности. Основной этап, на котором происходит формирование контурного представления и цветового кодирования. Заключительный этап – с выделением признаков (маркеров) и интерпретацией результатов.

Предварительная обработка необходима для устранения возможной шумовой составляющей, а также для повышения контрастности и коррекции яркостных характеристик изображений. Коррекция яркостных параметров особенно важна для светлополюсных микроскопических изображений, так как нативные объекты незначительно изменяют амплитуду прошедших световых лучей, из-за чего исследуемые клетки могут сливаться со светлым полем и вызывать трудности в определении их морфологии. Подавление шума реализовано на основе модификации взвешенного медианного фильтра. Было принято решение задавать распределение весовых коэффициентов с учетом расстояния от обрабатываемого пикселя и локальных данных окрестности. Для улучшения яркостных характеристик используется вариация метода на основе технологии Retinex. После подавления шумов и улучшения яркостных характеристик микроскопического изображения формируется контурное представление с использованием шиарлет-преобразования и цветового кодирования. Области отображаются в цвете в соответствии с данными с цветовой кодировкой и образуют сегменты для дальнейшего анализа. Затем вычисляются текстурные характеристики объектов интереса, необходимые для получения оценок для каждого участка нейрона.

Таким образом, для изучения нервных клеток мозга мышей разработана и апробирована методика визуализации, позволяющая выделять границы тел, аксонов и дендритов на нативных микроскопических изображениях. Предлагаемый метод создает условия для сохранения жизнеспособности объектов исследования в течение более длительного времени. Предварительные расчеты на модельных изображениях позволили получить приемлемые результаты по сравнению с методом анализа флуоресцентных изображений.

METHOD OF CENTRAL NERVOUS SYSTEM CELLS VISUALIZATION BASED ON MICROSCOPIC IMAGES PROCESSING

A.V. Medievsky¹, A.G. Zotin², K.V. Simonov³, A.S. Kruglyakov³

¹*Krasnoyarsk State Medical University named after Prof. V.F. Voino-Yasenetsky,*

²*Reshetnev Siberian State University of Science and Technology,*

³*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

The main objectives of the study are: development of a technique for selecting objects on native microscopic images and visualization of cells of interest with contrast enhancement using color coding. In order to work with cells located in three-dimensional space it is necessary to use microscopy methods that allow visualizing biological tissue in layers. Confocal fluorescence microscopy is most often used in such cases, but among its shortcomings one can single out the laboriousness of nerve tissue preparation and high phototoxicity. Therefore, for the dynamic structural and functional study of volumetric tissues it is better to use the method of native microscopy, such as differential interference contrast (DIC). In order to trace neurons, in addition to the DIC microscope, the best option is a phase-contrast microscope, which works on a similar principle. It has a simpler structure which leads to lower costs and as a result more popularity. Within the framework of the proposed computational technique for solving the visualization problem three main stages of processing and analyzing visual data can be distinguished. A pre-processing stage that includes noise reduction as well as brightness and contrast correction. The main stage at which the formation of the contour representation and color coding takes place. The final stage is involves the selection of features (markers) and the interpretation of the results.

Pre-processing is necessary to eliminate the possible noise component as well as to increase the contrast and correct the brightness characteristics of images. Correction of brightness parameters is especially important for bright-field microscopic images, since native objects slightly change the amplitude of the transmitted light rays which is why the cells under study can merge with the bright field and cause difficulties in determining their morphology. Noise suppression is implemented based on a modification of the weighted median filter. It was decided to set the distribution of weight coefficients taking into account the distance from the processed pixel and the local data of the neighborhood. To improve the brightness characteristics a variation of the method based on Retinex technology is used. After noise suppression and improvement of the brightness characteristics of the microscopic image, a contour representation is formed using the Shearlet transform and color coding. Areas are displayed in color according to the color coded data and form segments for further analysis. Then the textural characteristics of the objects of interest are calculated, which are necessary to obtain estimates for each section of the neuron.

Thus, for the study of nerve cells in the brain of mice a visualization technique was developed and tested which makes it possible to distinguish the boundaries of bodies, axons and dendrites in native microscopic images. The proposed approach makes it possible to create favorable conditions that preserve the viability of research objects for a longer time. Preliminary calculations on model images made it possible to obtain acceptable results in comparison with the method of analysis of fluorescent images.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ РЕЙТИНГОВОГО ОЦЕНИВАНИЯ ТЕРРИТОРИЙ*

А.М. Метус, Р.В. Морозов, А.И. Ноженков

Институт вычислительного моделирования СО РАН, г. Красноярск

Представлена аналитическая платформа для оценивания качества жизни муниципальных образований регионов в разрезе реализации национальных проектов. Платформа предназначена для информационно-аналитической поддержки принятия решений в территориальном управлении. Рейтинговое оценивание позволяет определить приоритетные направления развития и территории, требующие повышенного внимания. Реализован оригинальный метод интегрального аналитического оценивания, который обеспечивает комплексное оценивание состояния территорий в два этапа. На первом этапе происходит формирование территориально-ориентированной нормативной модели состояния территории на основе многолетней статистики и экспертных знаний. На втором этапе выполняется расчет интегральной оценки на основе сравнения текущего состояния территории с моделью.

Представлено описание программного обеспечения аналитической платформы. Платформа имеет клиент-серверную архитектуру и web-интерфейс. Серверная сторона обеспечивает консолидацию данных из внешних источников, реализует алгоритмы, предусмотренные методом интегрального аналитического оценивания для расчета отдельных составляющих модели и рейтинговых оценок территорий, поддерживает взаимодействие с пользовательским интерфейсом по технологии REST API. REST API-сервер реализован с использованием PostgREST. В работе представлена схема базы данных. Клиентская сторона обеспечивает создание и редактирование нормативной модели, а также пространственно-временной анализ состояния территорий. Клиентская часть платформы реализована на языке JavaScript с использованием библиотеки jQuery, компонентов DevExtreme и Leaflet для работы с интерактивной картой. В работе представлено описание библиотек и компонентов, используемых для построения пользовательского интерфейса. В текущей реализации платформы backend и frontend серверы физически базируются на одном хосте с СУБД. Основное web-приложение размещается на стороне frontend-сервера. В качестве web-сервера используется Apache. Взаимодействие между backend и frontend серверами осуществляется по HTTP протоколу. В качестве основного формата обмена информацией используется JSON.

Выполнена апробация аналитической платформы для оценивания качества жизни территорий Красноярского края на основе мониторинговых данных Автоматизированной информационной системы мониторинга муниципальных образований («АИС Регион МО»). Результаты оценивания по данным за 2019 год показали, что большинство районов края демонстрируют удовлетворительный и пониженный уровень качества жизни.

* Исследование выполнено при финансовой поддержке РФФИ, Правительства Красноярского края и Красноярского краевого фонда науки в рамках научного проекта № 20-47-242910.

SOFTWARE OF THE ANALYTICAL PLATFORM FOR RATING ASSESSMENT OF TERRITORIES*

A.M. Metus, R.V. Morozov, A.I. Nozhenkov

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

This paper presents the analytical platform for assessing the life quality of municipal territories of the regions in the context of the implementation of national projects. The platform facilitates information and analytical support for decision-making in the territorial management. Rating assessment allows us to identify priority areas for development and areas that require increased attention. The original method of integral analytical assessment has been implemented, which provides a comprehensive assessment of the state of territories in two stages. At the first stage, a geographically-oriented normative model of the state of the territory is formed based on long-term statistics and expert knowledge. At the second stage, an integral assessment is calculated based on a comparison of the current state of the territory with the model.

The description of the analytical platform software is presented. The platform has a client-server architecture and a web interface. The server provides consolidation of data from external sources, implements algorithms of the method of integral analytical evaluation for calculating individual components of the model and estimations of the state of territories, supports interaction with the user interface using REST API technology. The REST API server is implemented using PostgREST. The database schema is presented. The client provides the formation and editing of a normative model, as well as a spatial and time analysis of the state of territories. The client is implemented in JavaScript using the jQuery library, DevExtreme and Leaflet components for working with an interactive map. The description of the library of components used to build the user interface is presented. In the current implementation of the platform, backend and frontend servers are physically deployed on the same host with the DBMS. The main web application is deployed at the frontend server side. Apache is used as a web server. The interaction between backend and frontend servers is carried out via HTTP protocol. JSON is used as the main format for information exchange.

The analytical platform for estimation of the life quality of the territories of the Krasnoyarsk Region was tested on the monitoring data provided by Automated Information System for monitoring municipalities ("AIS Region MO"). The results of the estimation using data for 2019 showed that most of the areas of the region demonstrate satisfactory and low levels of life quality.

* The research was funded by RFBR, Krasnoyarsk Territory and Krasnoyarsk Regional Fund of Science, project number 20-47-242910.

К ВОПРОСУ О ПОСТРОЕНИИ РЕГИОНАЛЬНОЙ МОДЕЛИ ЗАТУХАНИЯ СИЛЬНЫХ ДВИЖЕНИЙ ГРУНТА ДЛЯ БАЙКАЛЬСКОЙ РИФТОВОЙ ЗОНЫ

В.А. Миронов^{1,2}, С.А. Перетокин^{2,3}, К.В. Симонов¹

¹*Институт вычислительного моделирования СО РАН, Красноярск*

²*Красноярский филиал Федерального исследовательского центра информационных и вычислительных технологий, Красноярск*

³*Институт физики Земли им. О.Ю. Шмидта РАН, Москва*

Определение достоверных параметров возможных землетрясений при инженерно-геологических изысканиях необходимы для обеспечения сейсмостойкости объектов с целью избежания разрушений и человеческих жертв. Модель затухания описывает зависимость характеристик колебания грунта от параметров землетрясения, расстояния до источника, локальных условий площадки. Данная модель строится на основе региональных каталогов землетрясений. Данная работа посвящена получению физических характеристик (амплитуды колебаний и спектры реакции грунта) от землетрясений для Байкальской рифтовой зоны. Исходными данными являются 252 записи от 53 землетрясений. Критерием отбора являлись энергетический класс землетрясения ($K \geq 11$) и максимальное эпицентральное расстояние ($R \leq 300$ км). Для получения физических характеристик и обработки записей реализованы вычислительные алгоритмы в среде MATLAB. Применяемые алгоритмы основаны на методологии Тихоокеанского центра инженерных исследований землетрясений и адаптированы к особенностям сейсмического мониторинга Байкальского региона.

Суть обработки заключается в следующем. Определение частот среза и применение 5-полосного акаузального фильтра Баттерворта. Путем интегрирования и дифференцирования в частотной области получение временных рядов ускорений, скоростей, смещений. В конце, получение средних геометрических значений от двух горизонтальных компонент (RotD50). Амплитуды RotD50 не зависят от азимута эпицентра землетрясения относительно регистрирующей станции и используются для построения современных моделей затухания. В ходе обработки для каждого землетрясения вычислены пиковые ускорения грунта, пиковые скорости грунта и амплитуды спектра реакции ускорений для 22 периодов в диапазоне 0.05-10 с. Сравнение полученных данных с четырьмя современными зарубежными моделями затухания показало, что для энергетических классов 12, 13 полученные значения несколько ниже. Однако, пока однозначно сказать о применимости этих моделей затруднительно. Для этого в дальнейшем планируется увеличение набора статистики. Полученные данные позволят сформировать базу данных сильных движений для Байкальского региона. На ее основе можно будет как разработать новую региональную модель затухания, так и обоснованно построить логическое дерево из имеющихся моделей.

Работа выполнена по теме гос. задания № 0144-2019-0010 с использованием данных, полученных на уникальной научной установке «Сейсмоинфразвуковой комплекс мониторинга арктической криолитозоны и комплекс непрерывного сейсмического мониторинга Российской Федерации, сопредельных территорий и мира».

TO THE QUESTION OF THE DEVELOP OF A REGIONAL ATTENUATION MODEL OF GROUND STRONG MOTION FOR THE BAIKAL RIFT

V.A. Mironov^{1,2}, S.A. Peretokin^{2,3}, K.V. Simonov¹

¹*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,*

²*Krasnoyarsk Branch of the Federal Research Center for Information and Computational Technologies, Krasnoyarsk*

³*Schmidt Institute of Physics of the Earth RAS, Moscow*

Determination of reliable parameters of possible earthquakes during geotechnical surveys is necessary to ensure the seismic resistance of objects in order to avoid damage and loss of life. The attenuation model describes the dependence of ground motion characteristics on earthquake parameters, distance to the source, local site conditions. This model is developed on the basis of regional earthquake catalogs. This work is devoted to obtaining physical characteristics (oscillation amplitudes and ground response spectra) from earthquakes for the Baikal rift zone. The initial data are 252 records from 53 earthquakes. The selection criterion was the energy class of the earthquake ($K \geq 11$) and the maximum epicentral distance ($R \leq 300$ km). To obtain physical characteristics and process records, computational algorithms are implemented in the MATLAB environment. The applied algorithms are based on the methodology of the Pacific Earthquake Engineering Research Center and adapted to the features of seismic monitoring of the Baikal region.

The essence of processing is as follows. Determining cutoff frequencies and applying a 5-pole acausal Butterworth filter. By integrating and differentiating in the frequency domain, obtaining time series of accelerations, velocities, displacements. Finally, getting the geometric mean of the two horizontal components (RotD50). Amplitudes RotD50 do not depend on the azimuth of the earthquake epicenter relative to the recording station and are used to develop modern attenuation models. In the course of processing, for each earthquake, peak ground accelerations, peak ground velocities, and amplitudes of the acceleration response spectrum for 22 periods in the range of 0.05–10 s were calculated. Comparison of the data obtained with four modern foreign attenuation models showed that for energy classes 12, 13 the obtained values are somewhat lower. However, it is still difficult to say unambiguously about the applicability of these models. To this end, it is planned to increase the set of statistics in the future. The data obtained will make it possible to form a database of strong movements for the Baikal region. Based on it, it will be possible both to develop a new regional attenuation model and reasonably construct a logical tree from existing models.

The work was carried out on the topic of state assignment №0144-2019-0010. The data used in the work were obtained with large-scale research facilities «Seismic infrasound array for monitoring Arctic cryolitozone and continuous seismic monitoring of the Russian Federation, neighboring territories and the world».

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ МРІ ДЛЯ РАСПАРАЛЛЕЛИВАНИЯ В ЗАДАЧЕ СРАВНЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ МЕТОДОМ ШАЙДУРОВА

А.А. Молявко, Е.Д. Кареева, М.Г. Садовский

Институт вычислительного моделирования СО РАН, г. Красноярск

Сравнение символьных последовательностей является фундаментом во многих науках: комбинаторика, теория вероятностей, вычислительная математика, биоинформатика, лингвистика. Главная проблема в решении этой задачи состоит в том, что в пространстве таких объектов невозможно построить содержательное понятие метрики. Традиционным выходом в этой ситуации является метод, предложенный в 1990 году и называемый выравниванием (alignment). Суть этого метода заключается в возможно более точной подгонке одной последовательности к другой. В случае, если полное точное совпадение невозможно, метод разрешает использовать т. н. элементарные преобразования: замену одного символа, вставку символа либо удаление. За каждую такую операцию назначается штраф; укладка, дающая наименьшее значение суммарного штрафа, считается наилучшей. Значение суммарного штрафа, не являющееся метрикой, является, тем не менее, основанием для сравнения последовательностей. Кроме того, что для сравнительно длинных последовательностей данный метод является высокочувствительным с точки зрения вычислений, у него есть несколько существенных недостатков. К ним относятся расходимость метода и произвол в выборе значений штрафных функций, назначаемых за элементарные операции. Тем не менее, к настоящему времени этот метод фактически стал стандартом.

В настоящей работе предложена альтернатива выравниванию, свободная от расходимости и от свободных параметров (необходимости выбора значений штрафных функций) — метод Шайдунова для сравнения символьных последовательностей, основанный на вычислении свёртки двух символьных последовательностей. Для этого символьную последовательность представляют в виде $K(0,1)$ -последовательностей, где K — мощность исходного алфавита. В каждой $(0,1)$ -последовательности какой-то символ заменяется на единицу (например, a из набора $\{a, c, g, t\}$ для генетических последовательностей), а остальные — на нули. Каждая такая $(0,1)$ -последовательность, соответствующая одинаковой замененной букве, рассматривается как набор коэффициентов полинома степени, на единицу меньшей, чем длина последовательностей. Затем вычисляются свёртки $(0,1)$ -последовательностей, соответствующие одной и той же букве. Набор всех свёрток почленно суммируется; полученная суммарная свёртка и является результатом сравнения. Каждое значение данной свёртки равно числу точных посимвольных совпадений в каждом из всевозможных наложений исходных символьных последовательностей вне зависимости от взаимного расположения этих совпадений.

Реализация метода позволяет использовать два мощных вычислительных инструмента: быстрое преобразование Фурье для вычисления свёртки и возможность крупномодульного распараллеливания для ускорения вычислений. В работе описаны особенности применения БПФ и крупномодульного распараллеливания для реализации метода. Полученные результаты показывают очень высокую эффективность данного метода: так, например, сравнение первой и второй хромосом человека (237 и 222 миллиона символов, соответственно) заняло на персональном компьютере (8 ядер, 32 Gb RAM, 3,2 GHz) 100 минут. Возможны также обобщения метода, вычисляющие число точных совпадений пар, троек и т. д. символов по-прежнему в независимости от текущего положения сравниваемых последовательностей.

APPLICATION OF MPI TECHNOLOGY FOR PARALLELIZATION IN THE PROBLEM OF COMPARISON OF SEQUENCES BY THE SHAIUROV METHOD

A.A. Molyavko, E.D. Karepova, M.G. Sadovsky

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

Comparison of symbol sequences is the core in many sciences ranging from combinatorics, probability theory, computational mathematics to bioinformatics and linguistics. The lack of essential idea for metrics in the space of such objects makes a problem. Alignment proposed in 1990 is a by default technique here. It consists in maximization of fitting two sequence. If an exact match is impossible, the method allows so-called. elementary transformations: mutation, insertion or deletion. Each such operation is assigned with the fine; the lay-out giving the least value of the total penalty corresponds to the best alignment. The score is not a metric; however, it stands behind the comparison of sequences. Besides the computational expense, this method has several significant drawbacks. The divergence and the arbitrariness in the penalty functions choice are among them. However, currently this method is a standard.

Here we propose the alternative method that is free from divergence and parameters arbitrariness; that is Shaidurov method to compare symbol sequences based on calculation of the convolution of a couple of symbol sequences. To do that, a symbol sequence is converted into K (0,1)-sequences, where K is the capacity of the original alphabet. In each (0,1)-sequence, some symbol is replaced with one (for example, a from the set $\{a, c, g, t\}$ for genetic sequences), and all others are replaced by zeros. Each such (0,1)-sequence corresponding to the same replaced letter is considered as a series of coefficients of a polynomial of degree a unit lesser than the length of the sequence. Then the convolutions of (0,1)-sequences corresponding to the same letter are calculated. The set of all convolutions is summed term by term; the resulting total convolution is the result of the comparison. Each value of this convolution is equal to the number of exact symbol matches for every possible overlay of the original character sequences, regardless of the relative position of these matches.

The method implementation allows to use two powerful computational tools: the fast Fourier transform for convolution calculation and the coarse-grained parallelism to speed up the calculations. The paper these issues for the implementation of the method. The results obtained show the very high efficiency of the method: for example, comparison of the first and second human chromosomes (237 and 222 million symbols, respectively) took 100 minutes of PC (8 cores, 32 Gb RAM, 3.2 GHz). The method yields generalizations with calculations of the exactly matches numbers of pairs, triples, etc. of symbols, regardless of the current position of the compared sequences.

АЛГОРИТМИЗАЦИЯ РАЗРАБОТКИ АНАЛИТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ УПРАВЛЕНИЯ БЕЗОПАСНОСТЬЮ ТЕРРИТОРИЙ

В.В. Ничепорчук

Институт вычислительного моделирования СО РАН, г. Красноярск

Цифровизация управления сложными системами неразрывно связана с аналитической обработкой больших объёмов данных. За последнее время появились и активно используются множество технологий, позволяющих извлекать из данных информацию для принятия решений. Это Data Mining, Big и Science Data, Small and Wide Data и другие. Относительно легко такие технологии внедряются в отрасли, изначально ориентированные на обработку данных, например, поисковые сервисы, мобильные операторы связи, ритейл. Флагманами внедрения аналитики являются сферы с результативной автоматизацией: банковская деятельность, налоговая служба. В государственном и муниципальном управлении ситуация сложнее. Количество данных возрастает из-за детализации отчётности. При этом качество информации, а вместе с ней и эффективность управления, падают. Требуют решения проблемы перегруженности специалистов, их переориентация на аналитическую деятельность, формализация процессов принятия решений. Достичь целей цифровизации управления применительно к территориям можно с использованием обобщённых алгоритмов, позволяющих перейти к принятию решений на основе данных. Такие алгоритмы включают технологии Low-code/No-code для роботизированной автоматизация процессов (RPA) и бесшовной межсистемной интеграции. В работе предложен процесс алгоритмизации разработки аналитических моделей на примере управления природно-техногенной безопасностью территорий.

Трансформация управления состоит из следующих этапов: формализация данных, включая их структурирование и систематизацию, построение аналитических моделей разного назначения, формирование представлений результатов обработки в качестве элементов управленческих решений. Кратко опишем особенности каждого этапа. Структурирование данных – перевод их в машиночитаемый вид. Выполняется проектировщиком без погружения в предметную область. На основе анализа отчётных документов, отражающих процессы управления, выделяются системообразующие данные, формируется структура хранилища данных. Одновременно решается задача проектирования сервисов актуализации данных. Слабым местом существующих систем является организация систематического обновления данных без участия разработчика. Систематизация данных – семантический анализ – выполняется с учётом конкретных задач управления. Для этого необходима переработка методических документов в схемы бизнес-процессов. Проблема этого этапа – согласование алгоритмов всеми заинтересованными лицами, от инженеров, пишущих код, до конкретных пользователей разного уровня подготовки. Построение разных аналитических моделей реализует многоэтапную обработку данных. Процессы обработки различны в зависимости от задач управления. Вначале реализуется контроль данных и разведочный анализ с использованием OLAP. Поиск причинно-следственных связей реализован с применением Data Mining и Machine Learning. Степени свободы присущи процессу формирования представлений результатов обработки. Элементы управленческих решений создаются встроенными средствами аналитической технологии или на основе интеграции технологий.

Большое число работ и накопленный опыт применения программных систем в управлении безопасностью территорий показывают возможность максимальной автоматизации бизнес-процессов. Эффективность цифровизации процессов формирования решений заключается в высвобождении ресурсов для выполнения рутинных операций, повышение достоверности информации. Трансформация управления является важнейшим способом противодействия росту рисков чрезвычайных ситуаций.

ALGORITHMS FOR THE DEVELOPMENT OF ANALYTICAL MODELS FOR TERRITORIAL SECURITY MANAGEMENT

V.V. Nicheporchuk

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

Digitalization of the management of complex systems is directly related to the analytical processing of Big Data. Now a lot of technologies for extracting information from data for decision-making are being intensively used. These are Data Mining, Big Data and Scientific Data, Small and Big Data and others. It is relatively easy for such technologies to be implementing in data processing-oriented industries. For example, search services, mobile operators, retail. The flagships of the analytics use are areas with effective automation – banking, tax service, customs. There is a lag in the digitalization of the processes of state and municipal administration. Due to the granularity of reporting, the amount of data increases. The quality of information and the associated management efficiency are reduced. There are a few problems to be solved, including the overload of specialists, their reorientation to analytical activities, and the formalization of decision-making processes. It is possible to achieve the goals of territories management digitalization using generalized algorithms that allow you to move on to data-based decision-making. Such algorithms include Low-code/No-code technologies for robotic process automation (RPA) and seamless cross-system integration. In this the paper proposed the process of algorithmization for the creating of analytical models and this applying for the provision of natural and man-made safety.

Stages of management transformation: data formalization, including their structuring and systematization, construction of analytical models for various purposes, formation of representations of processing results as elements of management decisions. We will briefly describe the features of each stage. Structuring data is translating it into a machine-readable form. Performed by the designer without dive in the subject area. Here the data warehouse structure is formed based on the analysis of reports, reflecting management processes, system-forming data are allocated. The weak point of existing systems is the organization of systematic updating of data without the participation of the developers. Data systematization or semantic analysis is performed for the specific management tasks. Here it is necessary to process methodological documents into business process diagrams of different notations. The problem of this stage is the coordination of algorithms by all interested parties, from engineers writing code to specific users of different levels of training. The construction of various analytical models implements multi-stage data processing. The processing processes vary in complexity depending on the management tasks. At the beginning, data control and exploratory analysis using OLAP are implemented. The search for cause-and-effect relationships is implemented using Data Mining and Machine Learning. Elements of management decisions are created by built-in analytical technology or through technology integration.

A large number of works and accumulated experience in the software systems and services use in the safety management of territories make it possible to maximize the formalization and automation of business processes. The efficiency digitalization of decision-making processes is the release of resources to perform routine operations, increasing the reliability and accuracy of information. Management transformation is the most important way to counteract the growth of emergency risks.

ИССЛЕДОВАНИЕ ТЕРРИТОРИАЛЬНЫХ КЛАСТЕРОВ КРАСНОЯСКОГО КРАЯ НА ОСНОВЕ МНОГОМЕРНОГО АНАЛИЗА ПОКАЗАТЕЛЕЙ КАЧЕСТВА ЖИЗНИ*

Т.Г. Пенькова, Л.Ф. Ноженкова

Институт вычислительного моделирования СО РАН, г. Красноярск

Улучшение качества жизни – главная задача органов государственной власти всех уровней. С целью повышения уровня жизни и обеспечения сбалансированности развития экономики и социальной сферы, в соответствии с Указами Президента Российской Федерации, с 2019 года в нашей стране реализуются национальные проекты. Помимо федеральных министерств, ответственность за достижение плановых показателей и повышение конкурентоспособности регионов возложена на органы власти субъектов Российской Федерации. Однако обширность территорий, неоднородность их природных условий, социально-экономических процессов и инфраструктурного развития требуют кластерного подхода в управлении развитием территорий. Красноярский край – один из крупнейших регионов Российской Федерации, имеющий сложную национальную, климатическую, биологическую, производственную структуры, включающий 17 городских округов, 3 муниципальных округа и 41 муниципальный район. В работе представлены результаты исследования территориальных кластеров макрорегиона на основе применения методов многомерного анализа – анализа главных компонент и кластерного анализа – к данным о качестве жизни муниципальных образований Красноярского края.

В разрезе показателей качества жизни исследованы существующие кластерные структуры. Восемикластерная структура Красноярского края исторически сложилась по географическому принципу и включает в качестве кластеров следующие группы районов: Ангарская группа, Восточная группа, Енисейская группа, Западная группа, Центральная группа, Южная группа, Таймырский автономный округ, Эвенкийский автономный округ. Четырехкластерная структура сформирована экспертами с точки зрения социально-экономических условий развития территорий края, в качестве отдельных кластеров выступают городские округа, территории с крупными промышленными объектами, территории с развитой сельскохозяйственной и лесопромышленной отраслью, отдаленные территории. На основе анализа главных компонент, определены особенности и ключевые характеристики территориальных кластеров. Результаты показали, что восьмикластерная структура преобразуется в четырехкластерную и по существу воспроизводит экспертный принцип деления муниципальных образований. С целью обнаружения структуры в данных и выявления территориальных кластеров на основе близости значений показателей качества жизни, выполнен кластерный анализ с помощью алгоритма DBSCAN. Анализ фактических значений показателей в пространстве главных компонент выявил трехкластерную структуру: первый кластер объединил муниципальные образования, для которых по всем показателям качества жизни характерны средние значения; второй кластер объединил отдаленные и арктические территории, для которых характерны высокие значения показателей социальной сферы и развития обрабатывающих производств; в третий кластер вошли муниципальные образования с развитой сельскохозяйственной и фермерской деятельностью. Анализ интегральных оценок качества жизни в разрезе национальных проектов в пространстве главных компонент позволил выделить структуры, обнаруживающие «отстающие» муниципальные образования края по направлениям «Экология», «Образование» и «Культура», «Жилье и городская среда». Результаты исследования позволили выявить особенности и условия развития муниципальных образований, определить проблемные и перспективные направления экономики и социальной сферы макрорегиона, позволяющие более эффективно планировать и проводить стратегические мероприятия по повышению качества жизни на территориях Красноярского края, используя преимущества кластерного подхода.

* Исследование выполнено при финансовой поддержке РФФИ, Правительства Красноярского края и Красноярского краевого фонда науки в рамках научного проекта № 20-47-242910.

STUDY OF THE TERRITORIAL CLUSTERS OF THE KRASNOYASKY KRAI BASED ON A MULTIDIMENSIONAL ANALYSIS OF LIFE QUALITY INDICATORS*

T.G. Penkova, L.F. Nozhenkova

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

Improving the life quality is the major task of authorities at all levels. In order to advance the standard of living and ensure a balance in the development of the economy and the social sphere, in accordance with the Decrees of the President of the Russian Federation, national projects have been implemented in our country since 2019. In addition to the federal ministries, the authorities of the federal subject of the Russian Federation are responsible for achieving the planned indicators aimed at ensuring economic growth, creating a comfortable environment and increasing the competitiveness of the regions. However, the enormous area of the territories, the heterogeneity of their natural conditions, socio-economic processes and infrastructure development require a cluster approach in the management of the territories development. The Krasnoyarsk krai is one of the largest regions of the Russian Federation, which has a complex national, climatic, biological and production structure; it includes 17 urban districts, 3 municipalities and 41 municipal districts. The paper presents the results of a study of territorial clusters of a macro-region based on the application of multidimensional analysis methods – principal component analysis and cluster analysis – to data on the life quality of municipal territories in the Krasnoyarsk krai.

In the context of life quality indicators, the existing cluster structures were studied. The eight-cluster structure of the Krasnoyarsk krai has historically developed according to the geographical principle and includes the following groups of areas as clusters: Angarsk group, Eastern group, Yenisei group, Western group, Central group, Southern group, Taimyr Autonomous Okrug, Evenki Autonomous Okrug. The four-cluster structure was formed by experts based on the socio-economic conditions; there are the following separate clusters: urban districts, territories with large industrial facilities, territories with developed agriculture and forestry and territories in the northern part of the region. Based on the analysis of principal components, the features and key characteristics of territorial clusters were determined. The results showed that the eight-cluster structure is transformed into a four-cluster structure and essentially reproduces the expert principle of the division of municipal territories. In order to detect the structure in the data and identify territorial clusters based on the proximity of the life quality indicators values, a cluster analysis was performed using the DBSCAN algorithm. An analysis of the actual values of indicators in the space of the principal components revealed a three-cluster structure: the first cluster united territories, for which average values are typical for all life quality indicators; the second cluster united remote and arctic territories, which are characterized by high values of indicators of the social sphere and the development of manufacturing industries; the third cluster included territories with developed agricultural and farming industry. An analysis of the integral assessments of the life quality in the context of national projects in the space of the principal components identifies structures that reveal the "lagging behind" municipal territories of the region in the spheres of "Ecology", "Education" and "Culture", "Housing and the urban environment".

The results of the study made it possible to identify the features and conditions for the development of municipal territories, to identify problematic and promising directions of the economy and social sphere of the macro-region, which allow more efficient planning and implementation of strategic actions to improve the life quality in the territories of the Krasnoyarsk krai using the advantages of the cluster approach.

* The research was funded by RFBR, Krasnoyarsk Territory and Krasnoyarsk Regional Fund of Science, project number 20-47-242910.

ОСОБЕННОСТИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ГЕНОВ ФОТОСИНТЕТИЧЕСКОЙ СИСТЕМЫ ХЛОРОПЛАСТОВ ВОДОРΟΣЛЕЙ

М.Ю. Сенашова

Институт вычислительного моделирования СО РАН, г. Красноярск

Водоросли являются промежуточным звеном эволюции фотосинтетических организмов между цианобактериями и высшими растениями. Множество работ посвящено анализу сходства и различия фотосистем цианобактерий, водорослей и высших растений с точки зрения биологии и генетики. В этой работе делается попытка сравнения генов фотосистем всех этих видов с точки зрения биоинформатики. Отражает ли информация, содержащаяся в символьных последовательностях, соответствующих генам фотосистем, процесс эволюции рассматриваемых организмов? Для ответа на этот вопрос рассмотрена пространственная структура генов фотосистем I и II хлоропластов водорослей. Под пространственной структурой понимается распределение точек, соответствующих частотным словарям генов, в пространстве частот триплетов. Полученное множество точек для визуализации проецируется из пространства частот триплетов в пространство трех первых главных компонент. Для анализа и визуализации результатов используется свободно распространяемая программа ViDaExpert.

Рассматриваются проекции точек в плоскость первой и второй, а также второй и третьей главных компонент. В результате было обнаружено, что гены фотосистем образуют два основных кластера, соответствующих прямому и обратному стренду. Аналогичные результаты были получены для генов фотосистем цианобактерий и высших растений. При этом не обнаружено группирования точек внутри основных кластеров ни по видам организмов, ни по типам генов, как это характерно для хлоропластов наземных растений и цианобактерий. Распределение по значениям GC-состава неоднородно. Часть надтипов имеют градиентное распределение, как цианобактерии, часть не обнаруживают выраженного порядка распределения, как высшие растения. Таким образом, структура множества точек, соответствующих генам фотосистем I и II водорослей, позволяет подтвердить промежуточное положение водорослей в эволюционной цепи между цианобактериями и высшими растениями.

FEATURES OF THE SPATIAL STRUCTURE OF THE PHOTOSYNTHETIC SYSTEM GENES IN ALGAE CHLOROPLASTS

M.Yu. Senashova

*Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences,
Krasnoyarsk*

Algae are an intermediate link in the evolution of photosynthetic organisms between cyanobacteria and higher plants. Many works are devoted to the analysis of the similarities and differences between the photosystems of cyanobacteria, algae, and higher plants from the biology and genetics point of view. This work attempts to compare the photosystem genes of all these species in terms of bioinformatics. Does the information contained in the symbolic sequences corresponding to the photosystem genes reflect the evolution process of the organisms under consideration? To answer this question, the spatial structure of the algal chloroplasts photosystems I and II genes was considered. The spatial structure is understood as the distribution of points corresponding to the genes frequency dictionaries in the space of triplet frequencies. The points resulting set is projected from the triplet frequencies space into the first three principal components space for visualization. The freely distributed program ViDaExpert is used here to analyze and visualize the results.

Projections of points into the plane of the first and second and second and third principal components are considered. Photosystem genes form two main clusters corresponding to the forward and reverse strands. Similar results were obtained for the photosystem genes of cyanobacteria and higher plants. At the same time, no grouping of points within the main clusters was found either by species of organisms or by types of genes, as is typical for chloroplasts of higher plants and cyanobacteria. The distribution of GC-composition values is non-uniform. Some supertypes have a gradient distribution, like cyanobacteria, some do not show a pronounced distribution order, like higher plants. Thus, the structure of the set of points corresponding to the genes of photosystems I and II of algae makes it possible to confirm the intermediate position of algae in the evolutionary chain between cyanobacteria and higher plants.

УСИЛЕНИЕ АРТЕФАКТОВ НА КАРТАХ ГЛУБИНЫ ИЗОБРАЖЕНИЙ ЛИЦ ПРИ ОБНАРУЖЕНИИ ПРЕЗЕНТАЦИОННЫХ АТАК

М.Н. Фаворская, А.И. Пахирка

*Сибирский государственный университет науки и технологий им. М.Ф. Решетнева,
г. Красноярск*

Распознавание лиц в системах биометрической идентификации занимает ведущее положение среди других способов идентификации, таких как идентификации личности по отпечаткам пальцев и радужной оболочке глаза, которые требуют наличия дорогостоящего оборудования. Однако изображения лиц могут быть легко украдены, например, из социальных сетей, что вызывает появление презентационных атак на системы распознавания лиц. Поэтому разработка методов и средств защиты от поддельных изображений лиц (face anti-spoofing) является важным направлением исследований. В данном исследовании рассматриваются атаки печати и воспроизведения.

В статье приведен обзор антиспуфинговых методов. Показано, что методы глубокого обучения способны автоматически извлекать богатую семантическую информацию по сравнению с традиционными методами обработки изображений. Приведены основные артефакты, связанные с искажением цвета, а именно, цветовая гамма, распределение цвета, цветовое искажение, цвет текстуры и артефакты захвата изображения. Построена модель искажений яркости и цветности при наличии презентационных атак.

Зачастую задача обнаружения поддельных изображений лиц рассматривается как задача бинарной классификации. Однако при этом не учитывается природа спуфинговых изображений (потеря контурной информации, искажение цвета, появление муара и других артефактов). Один из способов преодоления таких проблем основан на разработке методов обнаружения презентационных атак с использованием глубины сцены. В данном исследовании при построении карты глубины за основу взята глубокая сеть, которая состоит из блоков свертки центральной разности (central difference convolution). Сеть упрощена за счет удаления многомасштабного модуля внимания (multiscale attention fusion module). Отличие предлагаемого подхода состоит в том, что на вход сети подаются не RGB-изображения, а изображения, преобразованные в другие цветовые пространства. Приведены визуальные примеры построения карт глубины подлинных и поддельных RGB- и HSV-изображений из собственного набора, набора KITTI и набора Cityscapes. Визуальная оценка свидетельствует, что в большинстве случаев (в зависимости от устройств захвата) карты глубины поддельных изображений, переведенных в цветовое пространство HSV, фиксируют более заметные отличия от карт глубины подлинных изображений, также переведенных в цветовое пространство HSV. При этом точность обнаружения поддельных изображений лиц возросла на 3-7% в зависимости от устройств захвата, условий освещения и настроек самого алгоритма.

IMPROVING ARTIFACTS ON DEPTH MAPS OF FACE IMAGES UNDER PRESENTATION ATTACKS

M.N. Favorskaya, A.I. Pahirka

Reshetnev Siberian State University of Science and Technology, Krasnoyarsk

Face recognition in biometric identification systems occupies a leading position among other identification methods, such as fingerprint and iris recognition, which require expensive equipment. However, facial images can be easily stolen from, for example, social networks that causes presentational attacks on facial recognition systems. Therefore, the development of methods and means of protection against fake faces called as face anti-spoofing is an important area of research. This study focuses on print and replay attacks.

The article provides a short review of anti-spoofing methods. It is shown that deep learning methods are able to automatically extract rich semantic information compared to traditional image processing methods. The main artifacts associated with color distortion are given, namely, color gamut, color distribution, color distortion, texture color and image capture artifacts. A model of intensity and color distortions under presentation attacks is proposed.

The problem of detecting fake face images is usually considered as a binary classification problem. However, the nature of fake images (loss of contour information, color distortion, appearance of moiré and other artifacts) does not take into account. One of ways to overcome such problems is based on the development of methods for detecting presentational attacks using scene depth. In this study, a deep network with central difference convolution layers was used to build a depth map. The network is simplified by removing the multiscale attention fusion module. The originality is that not RGB images are fed to the network input, but images converted to other color spaces. Visual examples of depth maps for genuine and fake RGB and HSV images from our own dataset, the KITTI dataset, and the Cityscapes dataset are given. Visual evaluation shows that in most cases, depending on the capture devices, the depth maps of the fake images in HSV color space capture more noticeable differences compared to the depth maps of the real images also converted to the HSV color space. At the same time, the accuracy of detecting fake facial images increased by 3-7%, depending on the capture devices, lighting conditions, and the settings of the algorithm.

СЕГМЕНТАЦИЯ ОБЪЕМНЫХ ОБРАЗОВАНИЙ НА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ И МЕТОДА ПОРОГОВОЙ ОБРАБОТКИ

Ю.А. Хамад¹, А.В. Медиевский², А.Г. Зотин³, К.В. Симонов⁴, А.С. Кругляков⁴

¹*Сибирский федеральный университет*

²*Красноярский государственный медицинский университет
имени профессора В.Ф. Войно-Ясенецкого,*

³*Сибирский государственный университет науки и технологий им. ак. М.Ф. Решетнева*

⁴*Институт вычислительного моделирования СО РАН,
Красноярск*

В работе предложена модифицированная вычислительная методика для сегментации на примере обработки и анализа объемных образований (опухоли молочной железы) и непораженных областей на снимках с использованием комбинации средств нечеткой кластеризации и сегментации на основе порогового значения (метод FCMT). В рамках апробации использованы две базы данных изображений маммограмм молочной железы. Перед применением сегментации использовались методы предварительной обработки, нацеленные на снижение шумов и увеличение локального контраста. Для оценки эффективности рассчитаны среднеквадратическая ошибка, коэффициент DCS, индекс структурного сходства, точность и чувствительность. Используя предложенный метод сегментации FCMT, было достигнуто среднее пересечение по объединению (IoU) 93,85. Согласно результатам экспериментов, предлагаемый метод является устойчивым и точным в сегментации объемных образований (опухоли) на медицинских изображениях.

Цель исследования: разработать алгоритмическое обеспечение для решения задачи обнаружения границ и сегментации опухолей молочной железы. Изучено 50 медицинских снимков с поражением ткани из коллекции DICOM, представляющую набор цифровых изображений. Представлены результаты исследований на нескольких примерах сегментации и обнаружения опухоли молочной железы. Дифференциация нормальной и новообразованной ткани также помогает вычислить площадь пораженных опухолью участков, при этом отображается рассчитанная площадь в пикселях.

Выполнено сравнение модифицированной методики с известными методами. Предлагаемый метод способен обнаруживать и успешно определять опухолевые и интактные области молочной железы на исходном входном изображении с цветовым кодированием результатов. Представлены результаты анализа эффективности обработки изображений с обнаружением краев, а также расчета площади зон опухоли и нормальных участков молочной железы. Следующие тесты проведены, чтобы показать, что предлагаемая техника имеет хорошую степень производительности и может выдерживать средние уровни шума. Все параметры оценки исследования рассчитаны с использованием эталонных изображений маммограммы молочной железы. Представленные результаты экспериментальных исследований данные дают надежные оценки с хорошим качеством изображения для медицинского профессионального анализа. В ходе исследований медицинские работники (эксперты) оценили карты краев и обнаружили, что в ситуациях опухолевой патологии молочной железы полученная точность разработанного метода сегментации и измерения на 8-10% выше, чем аналогичные экспертные оценки. Экспериментальное исследование подтвердило устойчивость детектирования краев, созданного в рамках разработанного подхода FCMT, и при воздействии шума.

SEGMENTATION OF VOLUMETRIC LESIONS ON MEDICAL IMAGES USING A COMBINATION OF FUZZY CLUSTERING AND THRESHOLDING

Yu.A. Hamad¹, A.V. Medievsky², A.G. Zotin³, K.V. Simonov⁴, A.S. Kruglyakov⁴

¹ *Siberian Federal University*

² *Krasnoyarsk State Medical University named after Prof. V.F. Voino-Yasenetsky,*

³ *Reshetnev Siberian State University of Science and Technology*

⁴ *Institute of Computational Modelling of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk*

The paper proposes a modified computational technique for segmentation using the example of processing and analyzing volumetric lesions (breast tumors) and unaffected areas on images. It uses a combination of fuzzy clustering and threshold-based segmentation (FCMT method). As part of the experimental study the two databases of images of mammograms of the breast were used. Prior to applying segmentation pre-processing techniques were used to reduce a noise and increase local contrast. In order to evaluate the proposed technique effectiveness the mean square error, DCS coefficient, structural similarity index, accuracy and sensitivity values were calculated. Using the proposed FCMT segmentation method an average Intersection over Union (IoU) of 93.85 was achieved. According to the experimental results the proposed technique is stable and accurate in the segmentation of volumetric lesions (tumors) on medical images.

The purpose of the study: to develop algorithmic support for solving the problem of boundary detection and segmentation of breast tumors. We studied 50 medical images with tissue lesions from the DICOM collection. The results of studies on several examples of segmentation and detection of breast tumors are presented. Differentiation of tumor and non-tumor tissues also helps to calculate the area affected by the tumor, and the calculated area in pixels is displayed.

The modified technique is compared with the known methods. The proposed technique is able to detect and successfully identify tumor and intact areas of the breast on the original input image with color-coded results. The results of the analysis of the efficiency of image processing with boundary detection as well as the calculation of the area of tumor zones and normal areas of the mammary gland are presented. The tests have been carried out to show that the proposed equipment has a good degree of performance and can withstand average noise levels. All study evaluation parameters are calculated using reference breast mammogram images. The presented results of experimental research data provide reliable estimates with good image quality for medical professional analysis. In the course of the research medical workers (experts) evaluated the boundary maps and found that in situations of breast tumor pathology, the accuracy of the developed segmentation and measurement technique is 8-10% higher than similar expert estimates. An experimental study confirmed the stability of the edge detection created within the framework of the developed FCMT approach under the influence of noise.

**III Сибирский научный семинар
Data Analysis Technologies with Applications
SibDATA-2022**

Сборник трудов

17 июня 2022 г.

Редакторы:
Татьяна Геннадьевна Пенькова
Анна Владимировна Коробко

Институт вычислительного моделирования СО РАН

660036, г. Красноярск, Академгородок, 50 стр. 44