

**СИСТЕМА КОРРЕЛЯЦИОННОГО АНАЛИЗА  
СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ О ЗАБОЛЕВАЕМОСТИ КОРОНАВИРУСОМ**

© 2021 И.М. Янников, М.В. Телегина, А.Н. Исенбаев, Е.Н. Исенбаева

Ижевский государственный технический университет имени М.Т. Калашникова, г. Ижевск, Россия

Статья поступила в редакцию 16.07.2021

Статья посвящена проведению исследования о влиянии различных социологических, экономических, экологических и иных факторов на состояние заболеваемости и распространению коронавируса в мире. Авторами предложена схема получения информации из Интернет-ресурсов с возможностью проведения корреляционного анализа данных о причинах, темпах и масштабах пандемии, факторах, влияющих на её распространение. Во введении показана актуальность темы, проведён подробный анализ Интернет-ресурсов. Проведена систематизация размещаемых в них данных, сделаны необходимые заключения и выводы. В качестве источника статистической информации выбран интернет-сайт [coronavirus-monitor.ru](http://coronavirus-monitor.ru). В качестве инструментария предлагается разработка программы *Coronavirus Stat*, которая необходима для просмотра статистических данных на ПК в офлайн режиме и проведения расчетов по проверке гипотез о влиянии внешних факторов на распространение и течение COVID-19. Авторами подробно рассмотрены методы получения информации из интернет-источников, их преимущества и недостатки, выбран метод автоматического поиска с разработкой собственного варианта парсера. Сформулированы требования к отбору факторов для корреляционного анализа и проверяемые гипотезы. Приведены примеры проверки гипотез с представлением зависимостей числа заболевших от различных факторов и полей корреляции в виде графиков. Проведено подробное описание разработанной программы, состоящей из frontend части программы, парсера для получения новой информации, базы данных для хранения старой информации, файлов для хранения статической информации. Сформулированы требования к парсеру, приведена блок-схема его алгоритма. Показаны требования, учтённые при разработке программы и приведены примеры её работы. Тестирование программы проведено путём проведения экспериментов по проверке выдвинутых гипотез. Результаты проведения экспериментов сведены в таблицу. В заключении сделаны выводы по дальнейшему использованию разработанной программы.

*Ключевые слова:* методы поиска информации, Интернет-ресурсы, статистическая информация, корреляционный анализ, зависимости от различных факторов, структура программы *Coronavirus Stat*, парсер, коэффициенты, графики и расчёты корреляции.

DOI: 10.37313/1990-5378-2021-23-4-133-144

**ВВЕДЕНИЕ**

В последние годы высказывание Н. Ротшильда «Кто владеет информацией – тот владеет миром» становится всё более актуальным. Человечество ежедневно потребляет громадное количество различной разного рода информации не поддающееся никаким подсчётам. С развитием техники и технологий количество информации передаваемой непосредственно от человека к человеку снижается, передаваемой же с помощью различных средств массовой информации непрерывно растёт. Среди указанных средств в свою очередь постепенно снижается роль печатных СМИ и возрастает роль электронных. Главную роль в передаче информации в настоящее

время играет Интернет – сеть в которой существует множество различных многопрофильных и специализированных сайтов, способных удовлетворить любые запросы пользователя. Особое место среди них занимают вопросы здравоохранения и, в первую очередь, связанные с пандемией коронавируса. На прошедшей 30 июня 2021 года Прямой линии Президента РФ В.В. Путина с россиянами, самые популярные вопросы от россиян к президенту были связаны с вакцинацией и борьбой с COVID-19 [1]. Россиянам, как и всему человечеству необходимо оперативно получать достоверную информацию о масштабах пандемии, скорости её распространения, мерах борьбы и пр., поэтому вопрос соответствия сведений реальной реально складывающейся обстановке имеет особо важное значение.

*Янников Игорь Михайлович, доктор технических наук, профессор. E-mail: [bvd@mail.ru](mailto:bvd@mail.ru)*

*Телегина Марианна Викторовна, кандидат технических наук, доцент. E-mail: [asoiu@istu.ru](mailto:asoiu@istu.ru)*

*Исенбаева Елена Насимьяновна, старший преподаватель. Исенбаев Алексей Николаевич, магистрант.*

**АКТУАЛЬНОСТЬ**

Из проведённого авторами сравнительного анализа данных по количеству заболевших и

умерших от COVID-19 в Российской Федерации по состоянию на 28.03.21 г. рядом источников информации [2-7] выявлено что данные источников не совпадают. Из них имеют наименьшее расхождение с официально подтверждёнными данными Минздрава России [6], данные ВОЗ [7].

К наиболее достоверным источникам информации о коронавирусе в Интернете Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН) [8], кроме упомянутых Европейского регионального бюро ВОЗ и Минздрава России относит материалы государственных и международных органов [9-13]: Правительства Российской Федерации, Минобрнауки, Роспотребнадзора, официального сайта Мэра Москвы, канала Оперативного штаба Москвы по коронавирусу. К научно-информационным ресурсам отнесены [14-17]: База данных медицинских и биологиче-

ских публикаций PubMed, Издательств Elsevier и Springer, Интернет-платформы Web of Science (Clarivate Analytics).

По мере возрастания темпов и масштабов пандемии коронавируса возрастает и количество запросов в сети интернет о данном инфекционном заболевании. В настоящее время существует множество Интернет-ресурсов, отслеживающих и обрабатывающих статистические данные по пандемии. Естественно, что каждый из них имеет собственный алгоритм сбора данных и интервал их обновления [18, 19]. Кроме того каждый из них либо хранит данные о пандемии на своих серверах либо отображает данные с других сайтов.

Результат сравнения некоторых Интернет-ресурсов, предоставляющих информацию о заболеваемости коронавирусом представлен в таблице 1.

**Таблица 1.** Результат сравнения некоторых Интернет-ресурсов

Интернет-ресурс	Карта распространения COVID-19	Статистика (табл., графики)	Возможность задать вопрос о COVID-19	Карта вакцинации	Рекомендации населению
ВОЗ. Европейский регион [9]	+	+	+	-	+
Правительство РФ [10]	-	-	-	-	-
Минздрав РФ Коронавирус .. [6]	+	-	+	-	+
Роспотребнадзор [11]	-	-	-	-	+
Сайт мэра Москвы [12]	-	-	+	+	+
Канал Оперативн. штаба Москвы по коронавирусу [13]	-	+	+	-	-
Стопкоронавирус.рф [20]	+	+	+	-	+
BBC News Русская служба [21]	+	-	+	-	-
ТАСС [22]	-	+	-	-	-
Яндекс Коронавирус: статистика [23]	+	+	-	+	-
Google Новости [24]	+	+	-	+	-

Если проанализировать состав информации размещённой на различных, в то числе и указанных Интернет-ресурсах можно отметить, что на некоторых из них имеется лишь правовая информация, на других состав размещённого функционала больше. Например, на многих ресурсах имеются карта распространения коронавируса и статистика в виде таблиц и графиков. Что очевидно объясняется более лёгким восприятием графической информации. В то же время наличие карты вакцинации и возможность задать вопрос о коронавирусе могут позволить себе лишь немногие ресурсы, хотя эти функции также весьма востребованы. Данное обстоятельство объясняется не только доступностью к оперативным данным, но и необходимостью задействовать в работе ресурса высококвалифицированных специалистов. В статистике зачастую отсутствует свежая информация, что же касается доступности сведений, то данное обстоятельство зависит от организации соответствующей работы в конкретном регионе.

Из проведённого анализа можно сделать вывод, что универсальной системы, отражающей большую часть запросов пользователя в настоящее время нет, хотя потребность в такой системе, несомненно, имеется.

#### ПРЕДЛАГАЕМОЕ РЕШЕНИЕ

Предлагается система корреляционного анализа статистических данных о заболеваемости коронавирусом Coronavirus Stat, совмещающая в себе функционал различных Интернет-ресурсов, удобная пользователю и позволяющая установить корреляционные зависимости различных социологических, экономических, экологических и факторов на развитие пандемии.

В качестве источника статистической информации предлагается использовать интернет-сайт coronavirus-monitor.ru, где собраны данные Всемирной организации здравоохранения. На основе анализа статистических данных на персональных компьютерах в офлайн режиме, выполняются расчеты с проверкой гипотез о влиянии внешних факторов на распространение и течение COVID-19.

**Для получения** информации могут быть использованы ручное получение и ввод информации в базу данных, автоматическое получение и ввод при помощи парсера, реализованного при помощи различных способов и получение данных при помощи API. Каждый из указанных методов имеет свои преимущества и недостатки. На основе анализа недостатков приходим к выводу, что недостатки ручного метода – высокие трудозатраты, низкая точность данных, невозможность реализации одним человеком вследствие большого объёма информации, значительно

превышают его достоинства. Основным недостатком API является отсутствие возможности получения информации по конкретному каналу выбранному пользователем, то есть источник должен представлять такую возможность. В случае со статистической информацией о коронавирусе таких источников найдено не было.

Автоматический метод (парсинг) может находить и добавлять в базу данных любую информацию по заданным критериям. Поскольку имеющиеся программные продукты являются платными предлагается разработать свой вариант парсера. Рассмотрев самые популярные, был выбран вариант с использованием регулярных выражений. Данный метод достаточно прост в разработке и не требует специфических знаний и больших ресурсов.

Таким образом для добавления в базу данных системы Coronavirus Stat выбран способ – автоматическое получение информации при помощи парсинга.

#### ПОСТАНОВКА ЗАДАЧИ КОРРЕЛЯЦИОННОГО АНАЛИЗА СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ О ЗАБОЛЕВАЕМОСТИ КОРОНАВИРУСОМ

Существует много определений понятия корреляционного анализа тем не менее их суть сводится к методу выявления взаимозависимостей между определёнными факторами, признаками, параметрами. В связи с чем к основным задачам корреляционного анализа относят: выявление факторов, оказывающих наибольшее влияние на результирующий признак, выявление неизвестных причин связей, построение корреляционной модели с оценкой ее параметров и их интервальной оценкой [25].

При отборе факторов корреляционного анализа необходимо учитывать, что он основывается на причинно-следственных связях. При построении многофакторной корреляционной модели стоит отобрать факторы, оказывающие наибольшее влияние с коэффициентом парной корреляции меньше не 0,85. Отображение результатов анализа необходимо представлять как в графическом - в качестве диаграммы разброса, так и в текстовом виде - в качестве коэффициента корреляции. Сильная связь показана в виде прямой, где  $r=1$ .

В данном исследовании в качестве основных гипотез выдвинуты следующие:

1. Существует зависимость смертности от экологии по разным странам.
2. Существует зависимость количества заболевших от экологии по разным странам.
3. Существует зависимость количества выздоровевших от экологии по разным странам.
4. Существует зависимость количества за-

раженных от плотности населения стран (предположительно, чем выше плотность населения – тем больше случаев заражения).

5. Существует зависимость количества зараженных от численности населения стран (предположительно, чем больше количество населения – тем больше случаев заражения).

6. Существует зависимость количества зараженных от площади стран.

7. Существует зависимость количества зараженных от затрат на здравоохранение по странам (предполагается четкая обратная зависимость, больше затрат на здравоохранение – меньше заболевших; возможно придется учитывать количество заболевших в процентном выражении от общего числа населения страны).

8. Существует зависимость количества выздоровевших от затрат на здравоохранение по странам (предполагается прямая зависимость, больше затрат – больше выздоровевших).

9. Существует зависимость количества зараженных от миграции населения (можно разные показатели использовать: коэффициент чистой миграции, просто число мигрантов или количество прибывших туристов).

10. Существует зависимость количества зараженных от индекса процветания стран (The Legatum prosperity index [26]).

11. Существует зависимость количества выздоровевших от индекса процветания стран (The Legatum prosperity index).

Для составления алгоритма для проверки гипотез необходимо проанализировать каждую из гипотез вручную.

Рассматриваемый метод корреляции применяют для наглядного изображения формы связи между изучаемыми экономическими показателями. Для этого в прямоугольной системе координат строят график, по оси ординат откладывают индивидуальные значения результативного признака  $Y$ , а по оси абсцисс – индивидуальные значения факторного признака  $X$ .

Данный метод подходит для расчета корреляции для всех выдвинутых гипотез.

Рассчитаем ковариацию:

$$cov(x,y) = \bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y} = 2829.812 - 53.927 \cdot 53.686 = -65.29, \quad (1)$$

где  $x$  – индивидуальные значения факторного признака  $X$ ;

$y$  – индивидуальные значения результативного признака  $Y$ .

Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции. Линейный коэффициент корреляции принимает значения от  $-1$  до  $+1$ . Связи между признаками определены по шкале Чеддока [25]. Сверив с ней коэффициент, можно вычислить тесноту связи.

Кроме того, коэффициент линейной парной корреляции может быть определен через коэффициент регрессии [25]:

$$r_{x,y} = b \cdot \frac{S(x)}{S(y)} = -0.124 \frac{22.923}{12.835} = -0.222, \quad (2)$$

где  $r_{xy}$  – коэффициент корреляции между переменными  $X$  и  $Y$ ;

$S_x$  – среднее квадратическое отклонение, подсчитанное для переменной  $X$ ;

$S_y$  – среднее квадратическое отклонение, подсчитанное для переменной  $Y$ .

Для того чтобы при уровне значимости  $\alpha$  проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе  $H_1 \neq 0$ , надо вычислить наблюдаемое значение критерия (величина случайной ошибки):

$$t_{набл} = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}. \quad (3)$$

По таблице критических точек распределения Стьюдента, по заданному уровню значимости  $\alpha$  и числу степеней свободы  $k = n - 2$  найти критическую точку  $t_{крит}$  двусторонней критической области. Если  $t_{набл} < t_{крит}$  оснований отвер-

**Таблица 2.** Теснота связи между признаками по шкале Чеддока

Интервал значений	Теснота связи
$0.1 < r_{xy} < 0.3$	слабая
$0.3 < r_{xy} < 0.5$	умеренная
$0.5 < r_{xy} < 0.7$	заметная
$0.7 < r_{xy} < 0.9$	высокая
$0.9 < r_{xy} < 1$	весьма высокая

гнуть нулевую гипотезу. Если  $|t_{набл}| > t_{крит}$  — нулевую гипотезу отвергают.

По таблице Стьюдента с уровнем значимости  $\alpha=0.05$  и степенями свободы  $k=12$  находим  $t_{крит}$  [25]:

$$t_{крит}(n-m-1; \alpha/2), \quad (4)$$

где  $m = 1$  – количество объясняющих переменных.

Если  $|t_{набл}| > t_{критич}$ , то полученное значение коэффициента корреляции признается значимым. Значит нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается.

Поскольку  $|t_{набл}| < t_{крит}$ , то принимаем гипотезу о равенстве 0 коэффициента корреляции. Другими словами, коэффициент корреляции статистически – не значим.

В парной линейной регрессии  $t_r^2 = t_b^2$  и тогда проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

Рассмотрим результат проверки гипотезы 1: существует зависимость смертности от экологии по разным странам.

Рейтинг самых экологических чистых стран мира ежегодно составляется Центром экологической политики и права при Йельском университете (Yale Center for Environmental Law and Policy [27]). Исследование экологической эффективности измеряет достижения страны с точки зрения состояния экологии и управления природными ресурсами на основе 22 показателей в 10 категориях, которые отражают различные аспекты состояния окружающей природной среды и жизнеспособности её экологических систем, сохранение биологического разнообразия, противодействие изменению климата, состояние здоровья населения, практику экономической деятельности и степень ее нагрузки на окружающую среду, а также эффективность государственной политики в области экологии. Зависимость количества заболевших от экологического рейтинга по разным странам представлена на рисунке 1.

Поле корреляции представлено на рисунке 2. Аналогичным образом проведен разбор других выдвинутых гипотез.

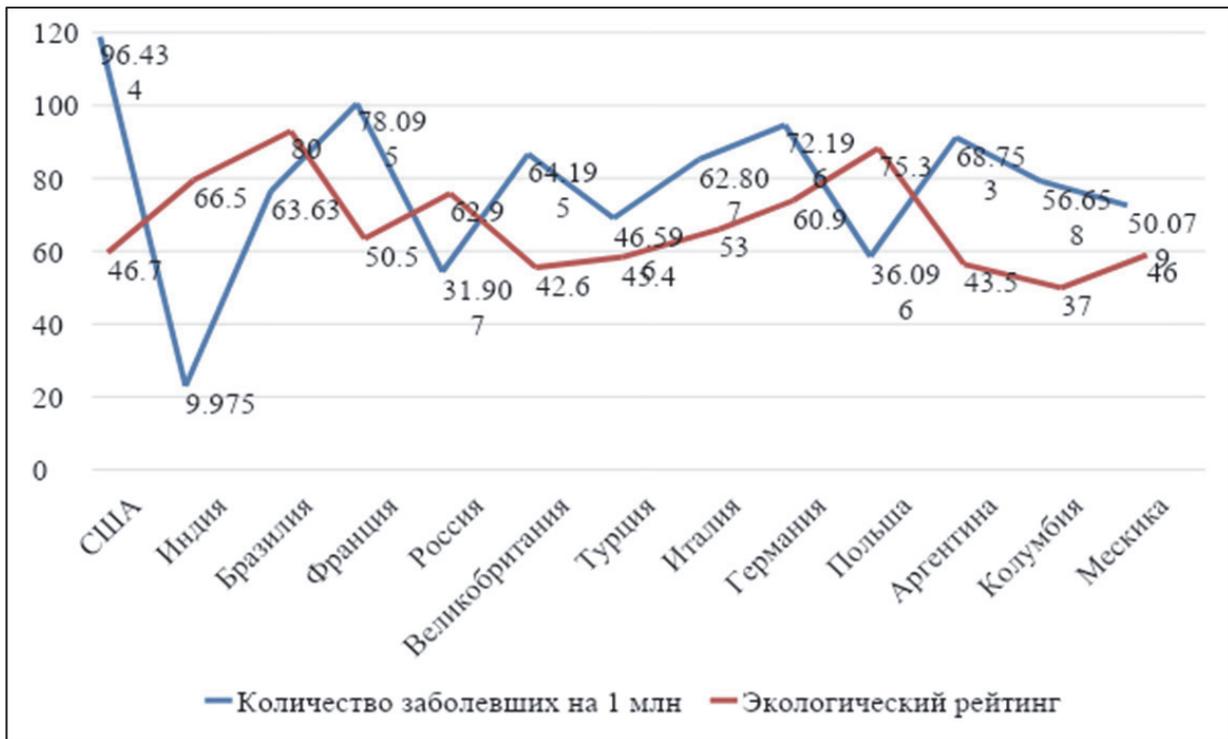


Рис. 1. Зависимость количества заболевших от экологии по разным странам

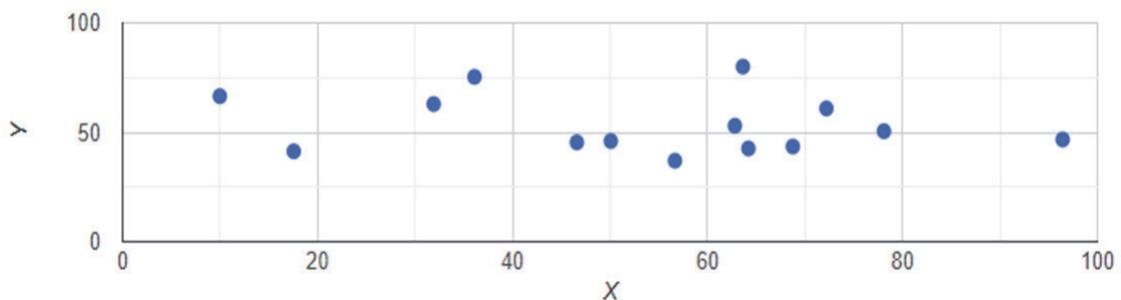


Рис. 2. Поле корреляции

### РЕАЛИЗАЦИЯ СИСТЕМЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ О ЗАБОЛЕВАЕМОСТИ КОРОНАВИРУСОМ

Для осуществления корреляционных расчетов разработана система "Coronavirus Stat". Данная разработка необходима для просмотра статистических данных на персональных компьютерах в офлайн режиме, когда нет возможности подключиться к интернету и получить прошлые статистические данные, а также для проведения необходимых расчетов, связанных с проверкой гипотез о влиянии внешних факторов на распространение и течение COVID-19.

Система состоит из следующих модулей:

- frontend часть программы;
- парсер для получения новой информации;
- база данных для хранения старой информации;

- файлы для хранения статической информации.

При разработке парсера учитывались следующие требования:

- парсер должен заходить на сайт и сканировать его на предмет новой информации, при ее появлении, добавлять новые записи в базу данных;
- парсер работает при наличии интернет-соединения;
- информация считывается построчно, ищутся регулярные выражения в html коде;
- повторный запуск парсера для обновления информации должен запускаться из программы Coronavirus Stat при нажатии кнопки "Обновить статистику".

Блок-схема алгоритма работы парсера представлена на рисунке 3.

**База данных состоит из таблицы включающей следующие столбцы:**

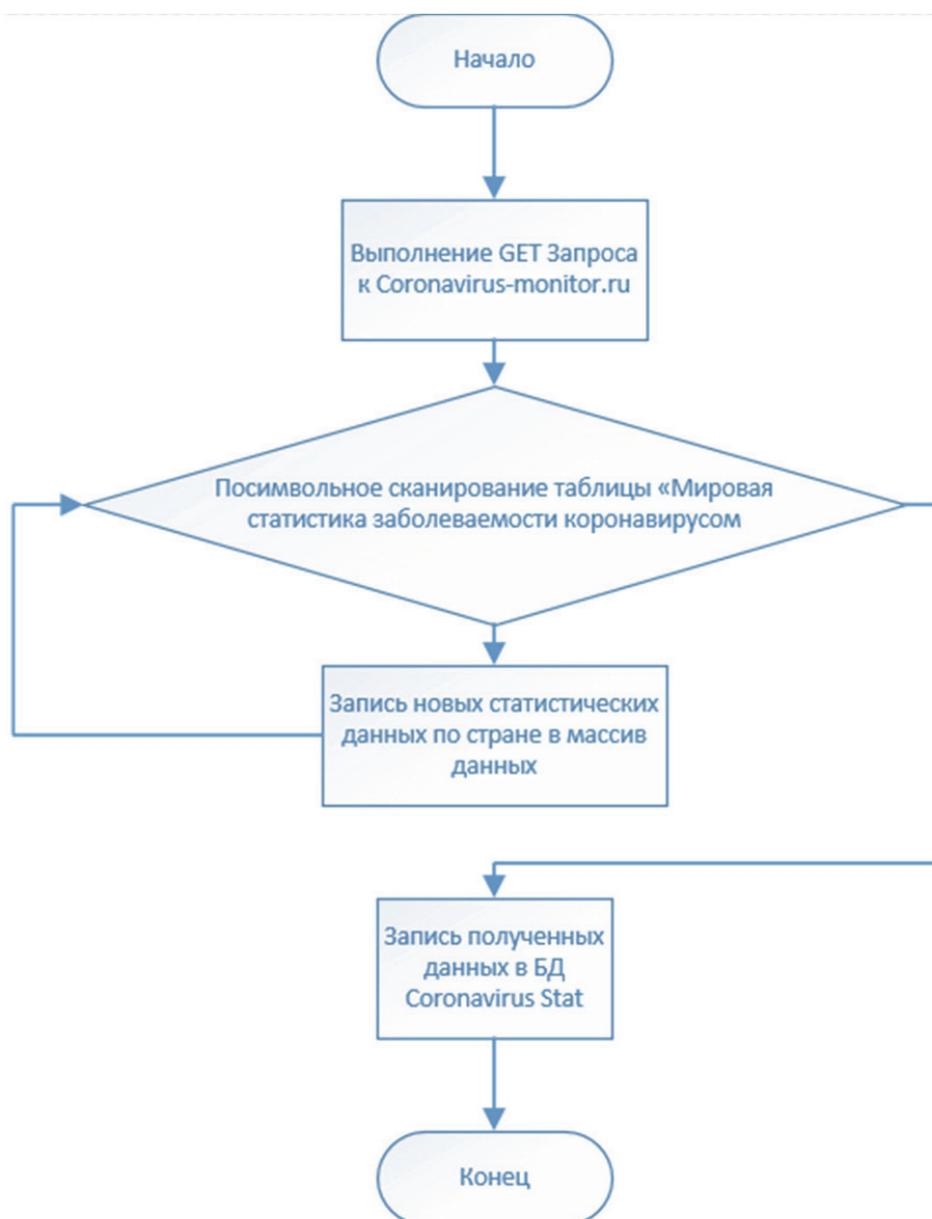


Рис. 3. Схема работы парсера

- “Rating” – место по количеству зараженных;
- “Country” – наименование страны;
- “ConfirmedStatistics” – количество зараженных;
- “CuredStatistics” – количество выздоровевших;
- “DeathStatistics” – количество смертей.

Основные функции системы Coronavirus Stat:

- вывод статистической информации по каждой из стран;
- подгрузка файлов со статической информацией (о количестве туристов, благосостоянии, площади, населенности страны);
- просмотра статистики за определенную дату;
- загрузка новой информации при помощи парсера;
- расчет и отображение списка гипотез с результатами корреляционных расчетов;
- отображение результатов с подробной информацией о расчётах до указанной в программе даты;
- отображение графика корреляции с возможностью изменения шага графика.

Для разработки был использован язык программирования C#, база данных PostgreSQL, фреймворк .NET core 3.1 и oхuplot для построения графиков корреляции.

Система Coronavirus Stat позволяет загружать и просматривать статистические данные о коронавирусе, проводить корреляционные расчеты и строить графики корреляции. На рисунке 5 показан корреляционный расчет гипотез. Гипотезы считаются, основываясь на данных о заболеваемости, начиная с определенной даты, в примере прописана дата 20.05.2020. Дату можно менять, чтобы посмотреть расчеты для другого временного отрезка. Также в данном окне показаны результаты расчетов.

На рисунке 5 показан пример расчета для гипотезы “Связь количества заболевших к экологическому рейтингу страны”, данные берутся за период с 20.05.2020 до 15.06.2021.

На рисунке 6 показан пример графика корреляции для расчета по гипотезе “Связь количества заболевших к экологическому рейтингу страны”. Шаг графика можно менять, в примере указан шаг 20.

### РЕЗУЛЬТАТЫ ПРОВЕДЕННЫХ ЭКСПЕРИМЕНТОВ ПО ПРОВЕРКЕ ВЫДВИНУТЫХ ГИПОТЕЗ

Результаты проверки гипотез о существовании зависимостей между статистическими данными о заболевших и показателями сведены в таблицу 3 ( $K_{\text{корр./лин}}$  – линейный коэффициент корреляции).

### ЗАКЛЮЧЕНИЕ

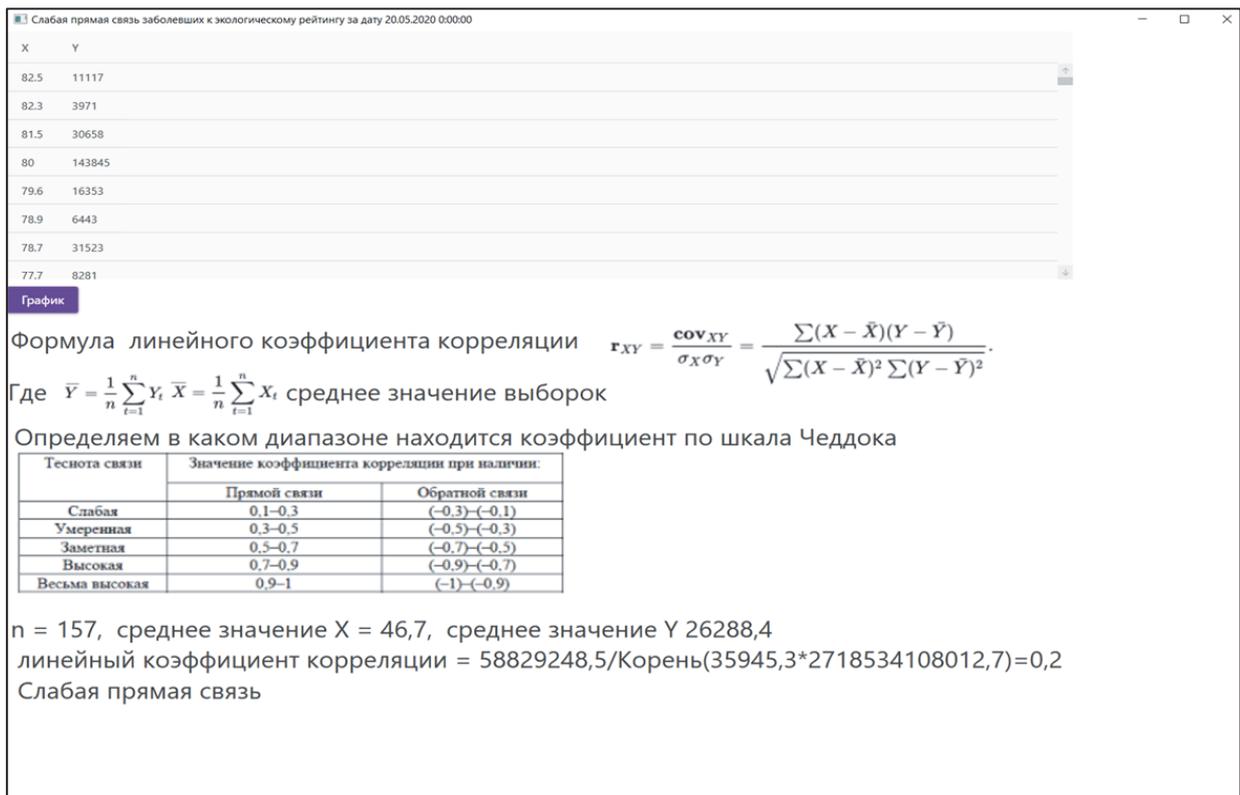
Таким образом, определен источник данных для анализа статистической информации о коронавирусе, разработан и реализован алгоритм корреляционного анализа, выдвинуты гипотезы для проведения экспериментов.

Проведены эксперименты по выявлению зависимостей между заболеваемостью, течением болезни и различными факторами соответствующих стран.

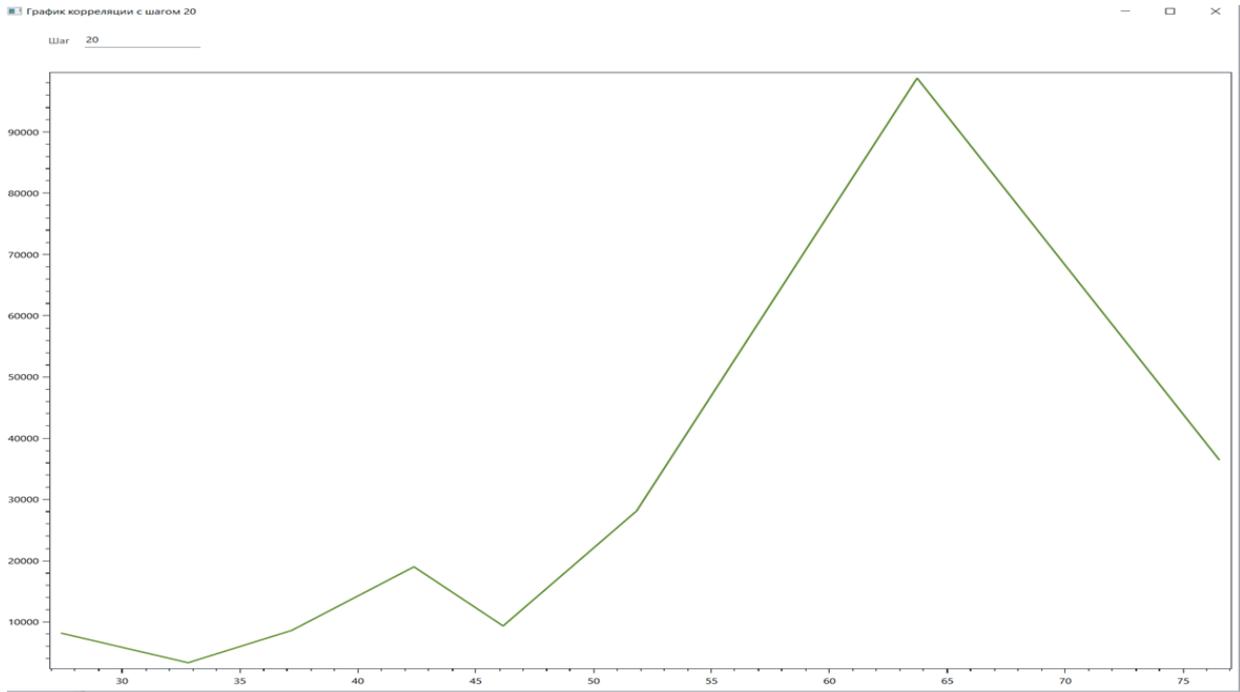
Разработанная система “Coronavirus Stat” готова для работы по отражению реальной оперативной обстановки с заболеваемостью коронавирусом и ответом на запросы пользователей, в том числе по тестированию других гипотез в целях принятия для принятия необходимых мер

343354	Слабая прямая связь заболевших к экологическому рейтингу за дату 20.05.2020	Показать	
340216	Слабая прямая связь выздоровевших к экологическому рейтингу за дату 20.05.2	Показать	
340343	Слабая прямая связь умерших к экологическому рейтингу за дату 20.05.2020 0:0	Показать	
340430	Слабая прямая связь заболевших к плотности населения за дату 20.05.2020 0:00	Показать	
340485	Слабая прямая связь заболевших к численности населения за дату 20.05.2020 0:	Показать	
341330	Заметная прямая связь заболевших к площади страны за дату 20.05.2020 0:00:00	Показать	
341509	Нулевая связь заболевших к затратам на здравоохранение за дату 20.05.2020 0:	Показать	
341712	Слабая обратная связь выздоровевших к затратам на здравоохранение за дату	Показать	
341906	Высокая прямая связь заболевших к количеству туристов за дату 20.05.2020 0:00	Показать	
339247	Нулевая связь заболевших к индексу процветания за дату 20.05.2020 0:00:00	Показать	
342032	Слабая обратная связь выздоровевших к индексу процветания за дату 20.05.202	Показать	
342243			
342415			
34263401	28251877	613465	09/06/2021
34275075	28277140	613853	10/06/2021
34287981	28305305	614413	11/06/2021
34314628	28381499	615034	12/06/2021

Рис. 4. Корреляционные расчеты гипотез



**Рис. 5.** Пример расчета по гипотезе “Связь количества заболевших к экологическому рейтингу страны” **Рис. 6.** График корреляции



**Рис. 6.** График корреляции

при борьбе с пандемией. Данную программу можно усовершенствовать добавив более точной оценки возможность проведения анализа по более узким временным промежуткам, например, за 1 месяц определенного года, в определенной стране или городе.

**СПИСОК ЛИТЕРАТУРЫ**

1. Прямая линия с Владимиром Путиным. URL: <http://kremlin.ru/events/president/news/65973> (дата обращения 10.073.2021).

Таблица 3. Результаты экспериментов по существованию корреляционных зависимостей

№ п/п	Наименование гипотезы о существовании зависимостей	Результаты эксперимента
1	смертности от экологии по разным странам	Гипотеза подтвердилась частично. К корр./лин. = 0.1. Связь найдена, но по шкале Чеддока эта связь квалифицируется как слабая прямая связь.
2	количества заболевших от экологии по разным странам	Гипотеза не подтвердилась. К корр./лин. немного ниже 0.1. По шкале Чеддока данный коэффициент не подходит ни под одну из категорий. Связь нулевая.
3	количества выздоровевших от экологии по разным странам	Гипотеза подтвердилась частично. К корр./лин. = 0.2. Связь найдена, но по шкале Чеддока эта связь квалифицируется как слабая прямая связь.
4	количества зараженных от плотности населения стран	Гипотеза подтвердилась частично. К корр./лин. = 0.1. Связь найдена, но по шкале Чеддока эта связь квалифицируется как слабая прямая связь.
5.	количества зараженных от численности населения стран	Гипотеза подтвердилась. К корр./лин. = 0.5. Связь найдена, по шкале Чеддока эта связь квалифицируется как заметная прямая связь.
6.	количества зараженных от площади стран	Гипотеза подтвердилась. К корр./лин. = 0.5. Связь найдена, по шкале Чеддока эта связь квалифицируется как умеренная прямая связь.
7.	количества зараженных от затрат на здравоохранение по странам	Гипотеза не подтвердилась. К корр./лин. немного выше - 0.1. По шкале Чеддока данный коэффициент не подходит ни под одну из категорий. Связь нулевая.
8.	количества выздоровевших от затрат на здравоохранение по странам	Гипотеза не подтвердилась. К корр./лин. немного выше - 0.1. По шкале Чеддока данный коэффициент не подходит ни под одну из категорий. Связь нулевая.
9.	количества зараженных от количества туристов	Гипотеза подтвердилась. К корр./лин. = 0.7. Связь найдена, по шкале Чеддока эта связь квалифицируется как высокая прямая связь.
10.	количества зараженных от индекса процветания стран	Гипотеза не подтвердилась. К корр./лин. немного выше - 0.1. По шкале Чеддока данный коэффициент не подходит ни под одну из категорий. Связь нулевая.
11.	количества выздоровевших от индекса процветания стран	Гипотеза не подтвердилась. К корр./лин. немного выше - 0.1. По шкале Чеддока данный коэффициент не подходит ни под одну из категорий. Связь нулевая.

2. Ситуация с COVID-19 в Европейском регионе ВОЗ. URL: <https://who.maps.arcgis.com/apps/opsdashboard/index.html#/a19d5d1f86ee4d99b013eed5f637232d> (дата обращения 28.03.2021).
3. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). URL: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (дата обращения 28.03.2021).
4. COVID-19 pandemic. URL: [https://en.wikipedia.org/wiki/COVID-19\\_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic) (дата обращения 28.03.2021).
5. Coronavirus World Map: Tracking the Global Outbreak. URL: <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html> (дата обращения 28.03.2021).
6. Сайт Министерства здравоохранения Российской Федерации. Коронавирус – симптомы, признаки, общая информация, ответы на вопросы URL: <https://covid19.rosminzdrav.ru/> (дата обращения 28.03.2021)
7. *Исинбаев А.Н., Исибаева Е.Н.* Исследование источников статистической информации о заболеваемости коронавирусом, выявление способов хранения и получения информации // Информационные технологии в науке, промышленности, образовании, Ижевск: ИЖГТУ имени М.Т. Калашникова. 2021.
8. Сайт Всероссийского института научной и технической информации Российской академии наук (ВИНИТИ РАН). Информация о коронавирусе в Интернете. <http://www.viniti.ru/covid-19/internet-info> (дата обращения 15.04.2021).
9. Сайт Всемирной организации здравоохранения. Европейское региональное бюро. URL: <https://www.euro.who.int/ru/health-topics/health-emergencies/coronavirus-covid-19>.
10. Сайт Правительства России. Санитарно-эпидемиологическая безопасность. URL: <http://government.ru/rugovclassifier/667/events/> (дата обращения 15.04.2021).
11. Сайт Роспотребнадзора. О новой коронавирусной инфекции. URL: [https://www.rosпотребнадзор.ru/region/korono\\_virus/punkt.php](https://www.rosпотребнадзор.ru/region/korono_virus/punkt.php) (дата обращения 15.04.2021).
12. Официальный сайт Мэра Москвы. Коронавирус: последние новости, симптомы коронавируса, меры профилактики. URL: <https://www.mos.ru/city/projects/covid-19/> (дата обращения 15.04.2021).
13. Канал оперативного штаба Москвы по коронавирусу. URL: [https://yandex.ru/chat/#/c/covid\\_msk](https://yandex.ru/chat/#/c/covid_msk) (дата обращения 15.04.2021)
14. База данных медицинских и биологических публикаций PubMed. Выборка научных публикаций по коронавирусу. COVID-19 Information | Public health information | (CDC) Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español. URL: <https://pubmed.ncbi.nlm.nih.gov/?term=%22COVID-19> (дата обращения 16.04.2021).
15. Издательство Elsevier. Информационный центр по новой коронавирусной инфекции COVID-19. URL: <https://www.elsevier.com/connect/coronavirus-information> (дата обращения 15.04.2021)
16. Издательство Springer Доступ к свежим научным публикациям по коронавирусу. Coronavirus (COVID-19) Research Highlights Find free access to the latest COVID-19 research papers and articles. URL: <https://www.springernature.com/gp/researchers/campaigns/coronavirusnter> (дата обращения 15.04.2021).
17. Интернет-платформа Web of Science (Clarivate Analytics). Подборка научно-информационных ресурсов Clarivate Analytics по коронавирусу URL: <https://clarivate.com/coronavirus-resources/> (дата обращения 15.04.2021).
18. *Журавлев А.Л., Китова Д.А.* Отношение жителей России к информации о пандемии коронавируса (на примере пользователей поисковых систем интернета) // Психологический журнал. 2020. Т. 41. № 4. С. 5 – 18.
19. Статистика заболеваемости коронавирусной инфекцией COVID-19. URL: [https://support.google.com/websearch/answer/9814707?p=cvd19\\_statistics&hl=ru&visit\\_id=637525181088312076-1188855498&rd=1](https://support.google.com/websearch/answer/9814707?p=cvd19_statistics&hl=ru&visit_id=637525181088312076-1188855498&rd=1) (дата обращения 25.03.2021).
20. Стопкоронавирус.рф. URL: <https://стопкоронавирус.рф/> (дата обращения 29.03.2021).
21. BBC. URL: <https://www.bbc.com/russian> (дата обращения 29.03.2021)
22. ТАСС. URL: <https://tass.ru/> (дата обращения 29.03.2021)
23. Яндекс Коронавирус: статистика. URL: <https://yandex.by/maps/covid19?ll=124.216986%2C-3.537040&z=2> (дата обращения 29.03.2021).
24. Google Новости. URL: <https://www.google.com/search?q=coronavirus> (дата обращения 29.03.2021).
25. Величина и сила коэффициента корреляции URL: <https://statpsy.ru/correlation/velicina/> (дата обращения 29.04.2021).
26. Legatum Prosperity Index 2020 URL: <https://www.prosperity.com> (дата обращения 15.04.2021).
27. Welcome | Yale Center for Environmental Law URL: <https://envirocenter.yale.edu> (дата обращения 15.04.2021).

## SYSTEM FOR CORRELATION ANALYSIS OF STATISTICAL INFORMATION ON CORONAVIRUS INCIDENCE

© 2021 I.M. Yannikov, M.V. Telegina, A.N. Isenbaev, E.N. Isenbaeva

Izhevsk State Technical University named after M.T. Kalashnikov, Izhevsk, Russia

The article is devoted to conducting a study on the influence of various sociological, economic, environmental and other factors on the state of the incidence and spread of coronavirus in the world. The authors proposed a scheme for obtaining information from Internet resources with the possibility of conducting a correlation analysis of data on the causes, rates and scale of the pandemic, and the factors affecting its spread. The introduction shows the relevance of the topic, carried out a detailed analysis of Internet resources. The systematization of the data placed in them has been carried out, the necessary conclusions and conclusions have been drawn. The website coronavirus-monitor.ru was selected as a source of statistical information. As a toolkit, it is proposed to develop the Coronavirus Stat program, which is necessary to view statistical data on a PC offline and carry out calculations to test hypotheses about the influence of external factors on the spread and course of COVID-19. The authors considered in detail the methods of obtaining information from Internet sources, their advantages and disadvantages, the method of automatic search with the development of their own version of the parser was chosen. Requirements for the selection of factors for correlation analysis and testable hypotheses are formulated. Examples of testing hypotheses with the presentation of graphs of dependences of the number of cases on various factors and correlation fields are given. A detailed description of the developed program, consisting of the frontend part of the program, a parser for obtaining new information, a database for storing old information, files for storing static information, is carried out. Requirements for the parser are formulated, a block diagram of its algorithm is presented. The requirements taken into account in the development of the program are shown and examples of its work are given. Testing of the program was carried out by conducting experiments to test the hypotheses put forward. The results of the experiments are summarized in the table. In conclusion, conclusions are drawn on the further use of the developed program.

*Keywords:* methods of information retrieval, Internet resources, statistical information, correlation analysis, dependence on various factors, structure of the Coronavirus Stat program, parser, coefficients, graphs and correlation calculations.

DOI: 10.37313/1990-5378-2021-23-4-133-144

### BIBLIOGRAPHY

1. Direct line with Vladimir Putin. URL: <http://kremlin.ru/events/president/news/65973> (date of access 10/07/2021)
2. COVID-19 situation in the WHO European Region. URL: <https://who.maps.arcgis.com/apps/opsdashboard/index.html#/a19d5d1f86ee4d99b013eed5f637232d> (date of access 03/28/2021).
3. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). URL: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (date of access 03/28/2021).
4. COVID-19 pandemic. URL: [https://en.wikipedia.org/wiki/COVID-19\\_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic) (date of treatment 03/28/2021).
5. Coronavirus World Map: Tracking the Global Outbreak. URL: <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html> (date of access 03/28/2021).
6. Site of the Ministry of Health of the Russian Federation. Coronavirus - symptoms, signs, general information, answers to questions URL: <https://covid19.rosminzdrav.ru/> (date of access 03/28/2021)
7. Isenbaev A.N. Isenbaeva E.N. Research of sources of statistical information on the incidence of coronavirus, identifying ways to store and obtain information. / Information technologies in science, industry, education, Izhevsk, 2021 / FGBOU VO "Izhevsk State Technical University named after M.T. Kalashnikov".
8. Site of the All-Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences (VINITI RAS). Information about the coronavirus on the Internet. <http://www.viniti.ru/covid-19/internet-info>.
9. Website of the World Health Organization. Regional Office for Europe. URL: <https://www.euro.who.int/ru/health-topics/health-emergencies/coronavirus-covid-19>.
10. Website of the Government of Russia. Sanitary and Epidemiological Safety. URL: <http://government.ru/rugovclassifier/667/events/>
11. Rospotrebnadzor website. About the new coronavirus infection. URL: [https://www.rospotrebnadzor.ru/region/korono\\_virus/punkt.php](https://www.rospotrebnadzor.ru/region/korono_virus/punkt.php)
12. Official site of the Mayor of Moscow. Coronavirus: latest news, coronavirus symptoms, prevention measures. URL: <https://www.mos.ru/city/projects/covid-19/>
13. Channel of the operational headquarters of Moscow on coronavirus. URL: [https://yandex.ru/chat/#/c/covid\\_msk](https://yandex.ru/chat/#/c/covid_msk).
14. Database of medical and biological publications PubMed. A selection of scientific publications on

- coronavirus. COVID-19 Information | Public health information | (CDC) Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español. URL: <https://pubmed.ncbi.nlm.nih.gov/?term=%22COVID-19>
15. Elsevier Publishing House. Information Center for the Novel Coronavirus Infection COVID-19. URL: <https://www.elsevier.com/connect/coronavirus-information>.
  16. Springer Publishing House Access to the latest scientific publications on the coronavirus. Coronavirus (COVID-19) Research Highlights Find free access to the latest COVID-19 research papers and articles. URL: <https://www.springernature.com/gp/researchers/campaigns/coronavirusnter>.
  17. Internet platform Web of Science (Clarivate Analytics). Clarivate Analytics Coronavirus Research Resource Selection URL: <https://clarivate.com/coronavirus-resources/>
  18. Zhuravlev A.L., Kitova D.A. Attitude of Russian residents to information about the coronavirus pandemic (on the example of Internet search engine users) / Zhuravlev A.L., Kitova D.A. // Psychological journal 2020, volume 41, No 4, p. 5-18.
  19. Statistics on the incidence of coronavirus infection COVID-19. URL: [https://support.google.com/websearch/answer/9814707?p=cvd19\\_statistics&hl=ru&visit\\_id=637525181088312076-1188855498&rd=1](https://support.google.com/websearch/answer/9814707?p=cvd19_statistics&hl=ru&visit_id=637525181088312076-1188855498&rd=1) (date accessed 03/25/2021).
  20. Stopkoronavirus.rf. URL: <https://stopkoronavirus.pf/> (date of access 03/29/2021).
  21. BBC. URL: <https://www.bbc.com/russian> (date of access 03/29/2021)
  22. TASS. URL: <https://tass.ru/> (date of access 03/29/2021)
  23. Yandex Coronavirus: statistics. URL: <https://yandex.by/maps/covid19?ll=124.216986%2C-3.537040&z=2> (date of circulation 03/29/2021).
  24. Google News. URL: <https://www.google.com/search?q=coronavirus> (date accessed 03/29/2021).
  25. The magnitude and strength of the URL correlation coefficient: <https://statpsy.ru/correlation/velicina/> (date of access 04/29/2021).
  26. Legatum Prosperity Index 2020 URL: <https://www.prosperity.com> (date of access 15.04.2021).
  27. Welcome | Yale Center for Environmental Law URL: <https://envirocenter.yale.edu> (date accessed 15.04.2021).

---

Igor Yannikov, Doctor of Technical Sciences, Professor.  
E-mail: [bvd@mail.ru](mailto:bvd@mail.ru)  
Marianna Telegina, Candidate of Technical Sciences,  
Associate Professor. E-mail: [asoju@istu.ru](mailto:asoju@istu.ru)  
Elena Isenbaeva, Senior Teacher.  
Alexey Isenbaev, Master's Student.