

УДК 33

Инструментарий эффективной статистической обработки экономических данных

Костиков Юрий Александрович

Кандидат физико-математических наук,
заведующий кафедрой 812,

Московский авиационный институт (национальный исследовательский университет),
125993, Российская Федерация, Москва, Волоколамское шоссе, 4;
e-mail: jkostikov@mail.ru

Романенков Александр Михайлович

Кандидат технических наук,
доцент кафедры 812,

Московский авиационный институт (национальный исследовательский университет),
125993, Российская Федерация, Москва, Волоколамское шоссе, 4;
e-mail: romanaleks@gmail.com

Аннотация

Статья посвящена рассмотрению возможности применения современных информационных технологий для автоматизации и повышения эффективности работы с данными интернет-порталов на примере портала Федеральной службы государственной статистики (Росстат). Для обработки статистических данных используется универсальный программный инструментарий, обеспечивающий гибкую настройку под индивидуальные потребности пользователя. Предлагаемый подход позволяет оперативно обрабатывать данные, проводить статистические расчеты и анализ данных. Функционал автоматического отслеживания обновлений на веб-портале позволяет поддерживать данные в актуальном состоянии. Возможности статистической обработки и анализа данных предоставляют удобное средство для подготовки отчетной документации, для проведения сравнительного анализа данных. Имеется богатый функционал визуализации данных.

Для цитирования в научных исследованиях

Костиков Ю.А., Романенков А.М. Инструментарий эффективной статистической обработки экономических данных // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 8А. С. 416-425.

Ключевые слова

Обработка данных РОССТАТ, автоматизация обработки статистических данных, Федеральная служба государственной статистики (Росстат), информационные технологии.

Введение

В повседневной деятельности многих фирм регулярно возникает задача оперативного получения и графической визуализации данных с использованием общероссийских баз статистических данных. Применение современных информационных технологий позволяет автоматизировать многие рабочие процессы фирмы. Представленный в данной статье программно-аналитический комплекс дает возможность проводить статистическую обработку данных. Он имеет интуитивно понятный интерфейс и позволяет представлять данные в удобном для пользователя виде.

Источником надежных и корректных данных по экономической детальности в Российской Федерации является Федеральная служба государственной статистики (Росстат).

Организация портала представляет собой иерархическую систему знаний и данных. Единичный поиск документов обеспечивается поисковой системой, которая основана на сервисе yandex.ru, и не составляет особых трудностей. Результатом поиска являются ссылки на web-ресурсы, по которым можно загрузить файлы формата .doc (MSWord) или .xls (MSExcel). Данные из загруженных файлов могут быть использованы для дальнейшей обработки.

Однако данные, которые располагаются на портале Росстата, являются разноформатными, и для работы со многими файлами приходится преодолевать ряд трудностей, связанных с приведением данных к единому формату. Помимо того, что присутствуют данные двух различных офисных приложений, сами данные в рамках одного типа имеют разную структуру: разные заголовки таблиц, разное количество столбцов, разные форматы в ячейках и так далее. Работа по приведению данных к одному формату не предъявляет высоких требований к квалификации, но на нее затрачивается рабочее время, которое может быть использовано с большей эффективностью. Отметим тот факт, что нет универсального формата табличных данных, и даже в рамках одного отдела фирмы его сотрудники могут иметь различное представление об универсальном формате данных.

Разработанное приложение предлагает решения по повышению производительности и эффективности труда сотрудников при работе с порталом Федеральной службы государственной статистики. Стоит отметить, что приложение не дублирует функционал портала, а является дополнительным инструментарием для работы с ним. При проектировании и разработке приложения использовались подходы, предложенные в [Коваленко, Костиков, 2018]. Следуя идеям, предложенным в [Костиков, Павлов, Романенков, Терновсков, 2017], реализован функционал представления данных и общая структура комплекса.

Основная часть

Рассматриваемое приложение является многопользовательским оконным приложением, разработанным с использованием современных методов [Гамма, Хелм, Джонсон, Влиссидес, 2016] и технологий программирования на языках python, PyQt [Прохоренок, Дронов, 2016].

Функционал, обеспечивающий многопользовательский режим работы, позволяет разграничить данные и текущую работу пользователей. В разработанном приложении присутствует механизм верификации документов, который основан на современных надежных алгоритмах цифровой подписи. Тем самым данные без ведома пользователей не перепутаются, не потеряются и можно будет подтвердить, что обработкой того или иного документа занимался конкретный сотрудник.

Рабочая область приложения представляет собой окно, разделенное на две части (рис. 1).

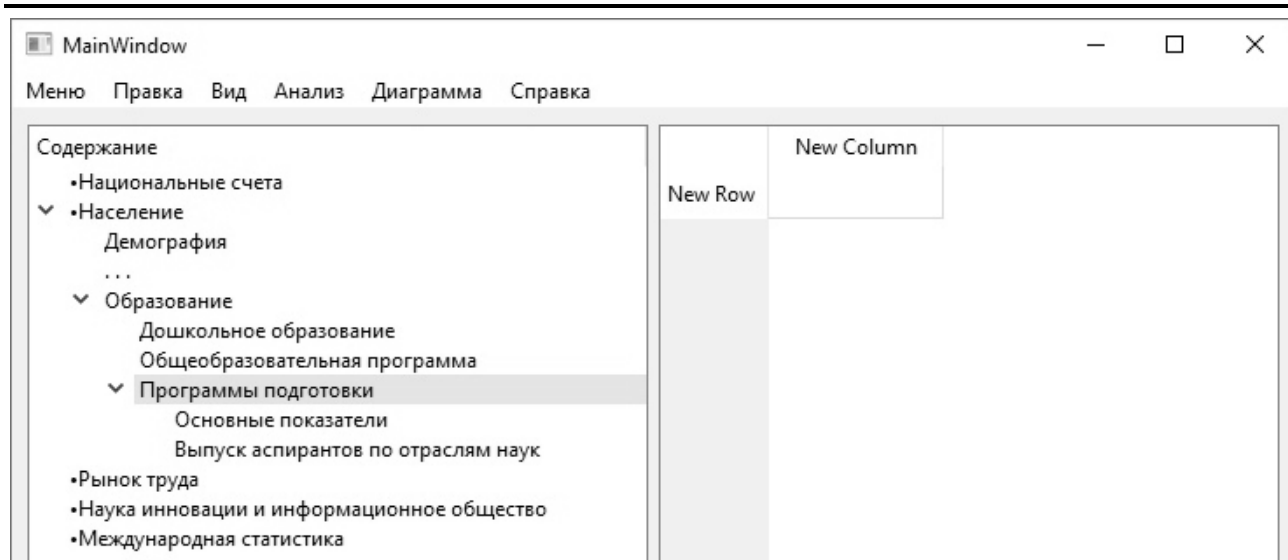


Рисунок 1 - Рабочая область программного комплекса

В левой части отображается иерархия документов с портала Федеральной службы государственной статистики в том виде, какой она имеет на веб-странице. При поиске документа и при переходе между разделами документов формируются запросы и отправляются на исполнение в веб-портал. Результаты запросов – полученные ссылки, скачанные документы, просмотренные разделы каталога документов – сохраняются локально и индивидуально для каждого пользователя.

В правой части отображаются данные в табличной форме, которые загружены из файла, который выбрал пользователь.

В приложении реализована возможность отслеживания обновлений документов на веб-портале. Такая опция позволяет пользователю всегда иметь документ с актуальными данными. В случае если текущая версия документа устарела, пользователю будет предложена возможность загрузить новую версию документа. При этом реализованы возможности замены документа, загрузки нового документа с сохранением старого и поиском отличий между этими документами.

Приложение позволяет пользователям как открывать загруженные документы (рис. 2) и отображать содержимое в своем окне, так и открывать их стандартными офисными приложениями. Особенностью работы с файлами от MSOffice является то, что интеллектуальная система программного комплекса в любом таком файле попытается определить таблицу с данными, в каком бы формате она ни была записана в этом файле. В автоматизированном режиме будут найдены и распознаны заголовки столбцов, определены строки с данными. Найденные данные будут отображены в табличном формате. Приложение позволяет настроить стиль и вид таблицы для вывода данных (рис. 3).

Средствами приложения можно организовать статистическую обработку данных. Реализован функционал для работы с ячейками и манипулированием данными в них. Предоставляется функционал для математических и статистических расчетов. Внутри приложения можно готовить отчетные материалы и проводить статистический анализ данных. Стоит отметить, что пользователи могут делиться результатами своей работы с другими пользователями. Для этих целей реализован механизм экспорта данных. При работе с приложением несоответствия в представлении таблиц, связанные с различными форматами

табличных данных, будут отсутствовать. В приложении реализован механизм, позволяющий привести таблицы к единому виду, что способствует тому, чтобы табличные данные выглядели одинаково у всех пользователей.

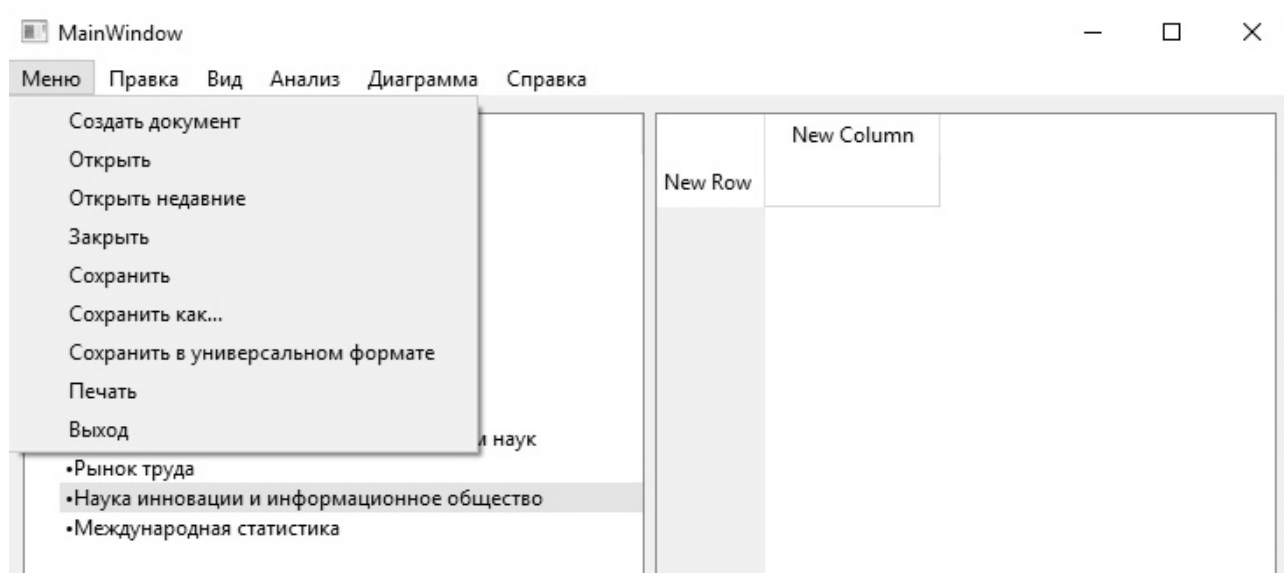


Рисунок 2 - Меню приложения

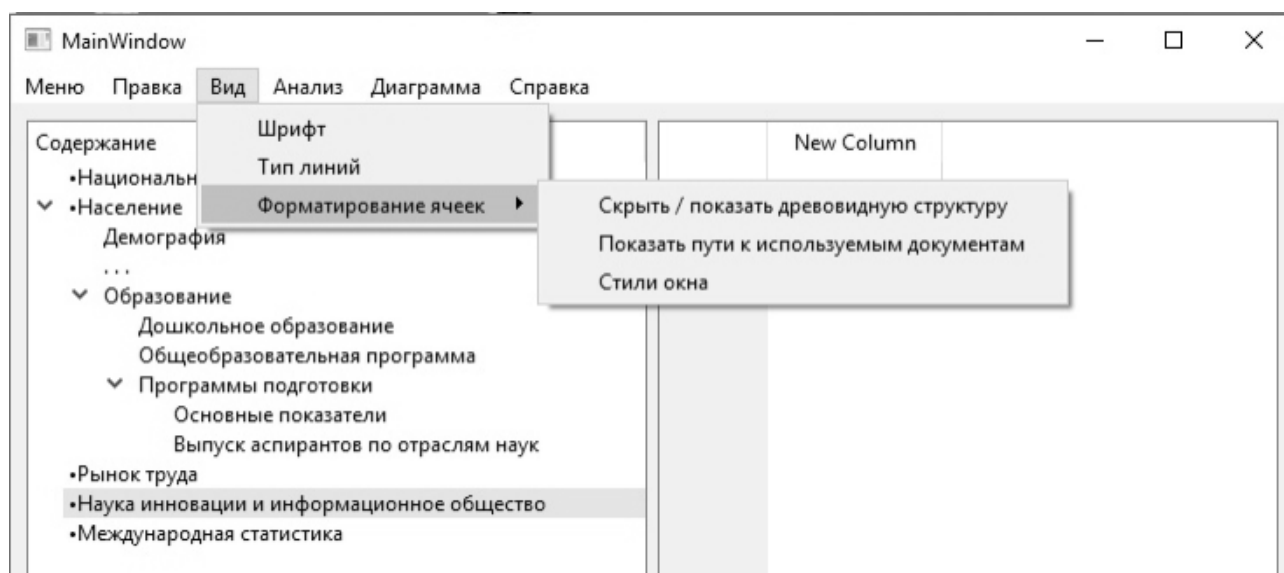


Рисунок 3 - Настройки вида отображения таблиц

Помимо таблиц и данных, приложение может определить элементы документа, которые отвечают за визуализацию данных (диаграммы, графики), и данные, на которых эти элементы основаны (рис. 4). Предоставляется возможность конструировать диаграммы и настраивать их внутри приложения по своему усмотрению.

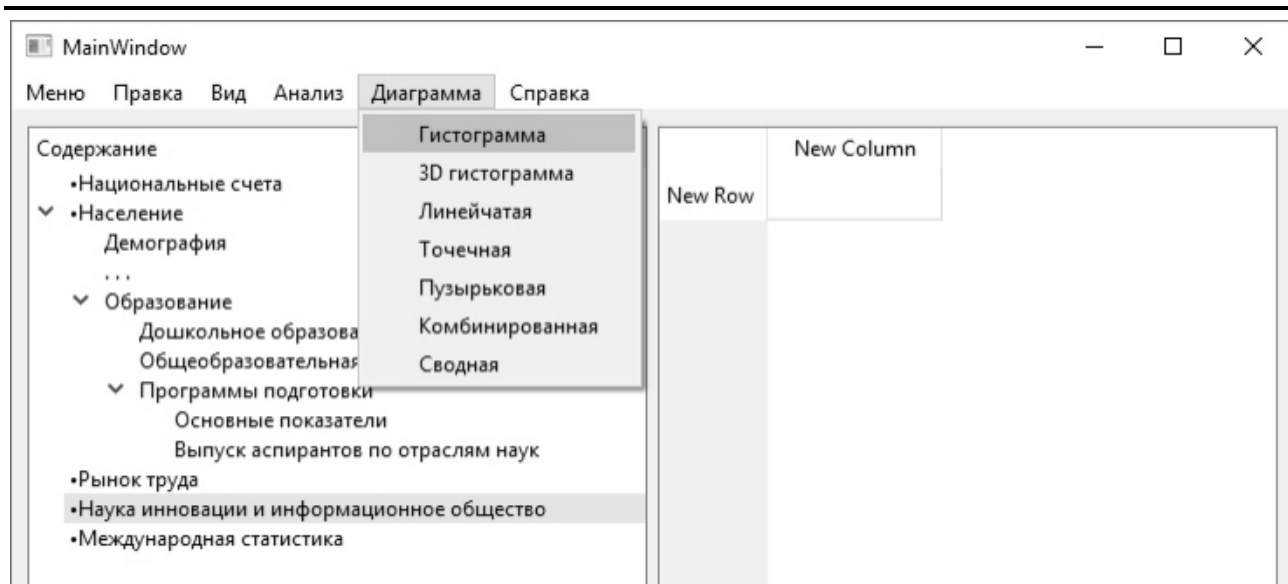


Рисунок 4 - Настройка вида диаграммы

На рисунке 5 показана динамика изменения населения России, изменения городского населения и сельского с 1917 по 2017 год.

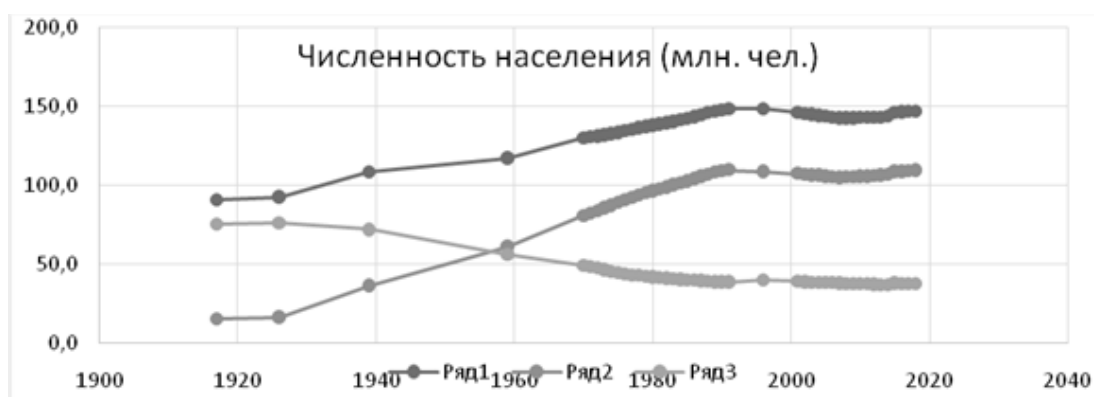


Рисунок 5 - Население России с 1917 по 2017 год

У пользователей есть широкий спектр возможностей сохранения результатов своей работы. Построенные диаграммы и графики могут быть сохранены в формате изображения (bmp, jpeg), или в формате документа Word, или в формате документа Excel с возможностью дальнейшей архивации и отправки по электронной почте. Данные, с которыми проводилась работа, могут быть сохранены в формате Excel. При этом для всех пользователей используется единый формат сохранения документа.

Сохраненный документ добавляется в рабочую папку пользователя, и другим пользователям может быть предоставлен доступ для работы с этим документом. При этом вне зависимости от наличия данного приложения на компьютере, Excel-документ будет выглядеть одинаково в любом Excel-совместимом редакторе.

Далее рассмотрим отображения данных о посевах сельскохозяйственных культур. После выбора в древовидном элементе управления, необходимого для отображения файла данных, его содержимое отображается в приложении (рис. 6).

MainWindow - [Preview] - Qt Designer

Меню Правка Вид Анализ Диаграмма Справка

Содержание

- Национальные счета
- > •Население
- ▼ •Рынок труда
 - Индекс производительности труда по России и основным отраслям экономики РФ (по ра...
 - Наука инновации и информационное общество
- > •Международная статистика

| коды | 2003 | 2004 | 2005 | 2006 |
|-------------------------------------------------------------------------------------------------------------------------|--------------|--------------|--------------|--------------|
| В целом по экономике | 107,0 | 106,5 | 105,5 | 107,5 |
| Раздел А СЕЛЬСКОЕ ХОЗЯЙСТВО, ОХОТА И ЛЕСНОЕ ХОЗЯЙСТВО | 105,0 | 102,9 | 101,8 | 104,3 |
| Раздел В РЫБОВОДСТВО, РЫБОВЫДСТВО | 102,1 | 104,3 | 99,0 | 101,0 |
| Раздел С ДОБЫЧА ПОЛЕЗНЫХ ИСКОПАЕМЫХ | 109,2 | 107,3 | 100,3 | 103,3 |
| Раздел D ОБРАБАТЫВАЮЩИЕ ПРОИЗВОДСТВА | 100,0 | 100,0 | 106,0 | 103,5 |
| Раздел E ПРОИЗВОДСТВО И РАСПРЕДЕЛЕНИЕ ЭЛЕКТРОЭНЕРГИИ, ГАЗА И ВОДЫ | 103,7 | 100,7 | 103,7 | 101,9 |
| Раздел F СТРОИТЕЛЬСТВО | 106,3 | 108,3 | 105,9 | 115,8 |
| Раздел G ОПТОВАЯ И РОЗНИЧНАЯ ТОРГОВЛЯ, РЕМОНТ АВТОТРАНСПОРТНЫХ СРЕДСТВ, МОТОЦИКЛОВ, БЫТОВЫХ ИЗДЕЛИЙ И ПРЕДМЕТОВ ЛичНОГО | 100,0 | 110,5 | 105,1 | 110,0 |
| Раздел H ГОСТИНИЦЫ И РЕСТОРАНЫ | 100,3 | 103,1 | 108,0 | 108,2 |
| Раздел I ТРАНСПОРТ И СВЯЗЬ | 107,5 | 108,7 | 102,1 | 110,7 |
| Раздел K ОПЕРАЦИИ С НЕДВИЖИМЫМ ИМУЩЕСТВОМ, АРЕНДА И ПРЕДОСТАВЛЕНИЕ УСЛУГ | 102,5 | 101,3 | 112,4 | 108,2 |

Рисунок 6 - Статистические данные по сельскому хозяйству

Загруженные данные можно визуализировать с помощью диаграммы. Стоит отметить, что разработанное программное решение не преследует цели превзойти возможности, которые предоставляет MSExcel. В приложении реализована возможность сохранения построенной диаграммы в книгу Excel с теми данными, на основании которых данная диаграмма была построена. Разработанное приложение позволяет фактически одним нажатием кнопки сгенерировать диаграмму и избавляет рядового пользователя от рутинных действий в Excel, таких как выбор страницы, выбор области данных, выбор типа диаграммы, настройка легенды, оформление подписей, что сильно упрощает и ускоряет рабочий процесс.

Обратимся к следующему примеру. Из всей статистической информации отобразим только данные по зернобобовым в виде круговой диаграммы (рис. 7).

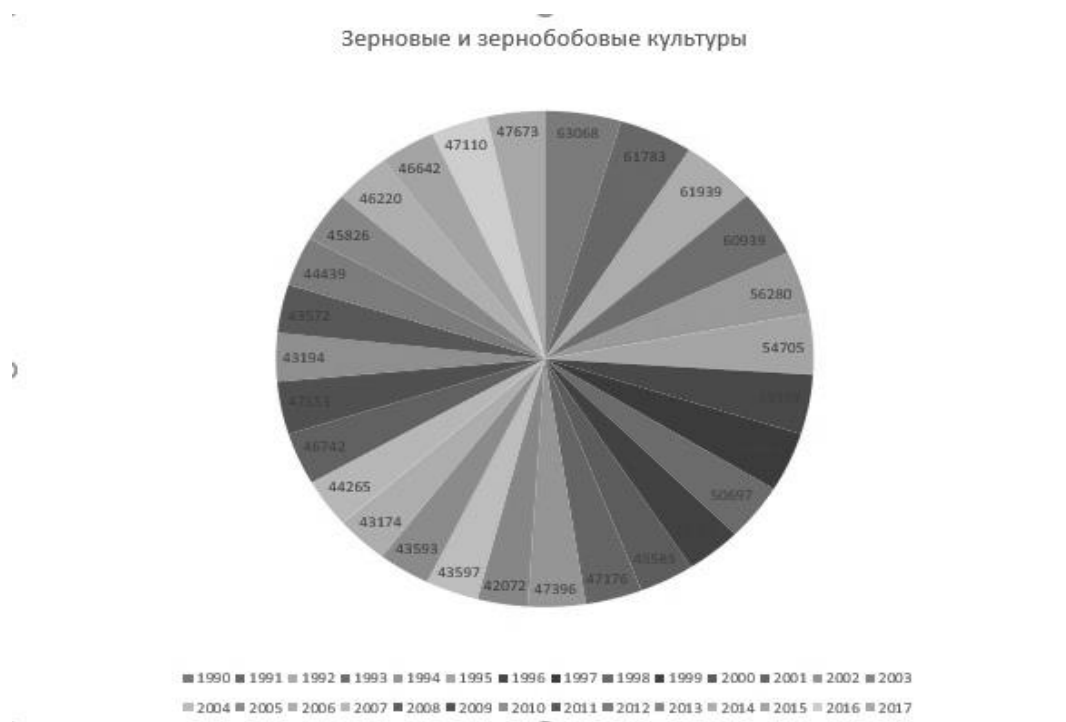


Рисунок 7 - Диаграмма посевных площадей

Отметим также, что в приложении поддерживаются настройки представления графических данных, которые согласованы с MSExcel. Иными словами, полученные графики и диаграммы могут быть интегрированы в документы Excel. На рисунке 8 в виде столбчатой диаграммы отображен индекс производительности по годам и по отраслям экономики РФ с 2003 по 2014 год. Для удобства представления выполнена группировка столбцов по годам.



Рисунок 8 - Индекс производительности труда по отраслям экономики РФ

Стоит отметить, что в приложении реализована возможность установления некоторых зависимостей между выбранными параметрами. Подходы и методы базируются на решениях, предложенных в [Вентцель, 2007; Боровков, 2007].

В качестве примера рассмотрим корреляционный анализ затрат на технологические инновации организаций по субъектам РФ (при анализе были использованы официальные данные). Был проведен корреляционный анализ годовых затрат в периоде с 2010 по 2017 год включительно по Центральному, Северо-Западному, Северо-Кавказскому, Приволжскому и Дальневосточному федеральным округам. В таблице 1 приведены соответствующие значения коэффициента корреляции между значениями затрат за 2010 и 2011 годы.

Таблица 1 - Коэффициенты корреляция затрат на инновации

| Годы | Центральный федеральный округ | Северо-Западный федеральный округ | Приволжский федеральный округ | Северо-Кавказский федеральный округ | Дальневосточный федеральный округ |
|-----------|-------------------------------|-----------------------------------|-------------------------------|-------------------------------------|-----------------------------------|
| 2010-2011 | 0,73 | 0,91 | 0,86 | 0,98 | 0,99 |
| 2011-2012 | 0,97 | 0,90 | 0,67 | 0,99 | 0,96 |
| 2012-2013 | 0,96 | 0,72 | 0,94 | 0,99 | 0,98 |
| 2013-2014 | 0,99 | 0,62 | 0,92 | 0,99 | 0,94 |
| 2015-2016 | 0,99 | 0,99 | 0,91 | 0,99 | 0,99 |
| 2016-2017 | 0,97 | 0,99 | 0,95 | 0,99 | 0,98 |

Отметим, что на основании полученных данных можно утверждать, что затраты на технологические инновации не могут произвольным образом изменяться в соседних временных периодах. Так, если уже были потрачены средства на модернизацию техники и технологии, то

резко менять инвестиционную политику нежелательно, так как необходимы затраты на внедрение, дальнейшее обслуживание и т.д.

Помимо корреляции затрат по годам, были выполнены расчеты корреляции затрат за 2010-2017 годы между разными субъектами, которые принадлежат к одному и тому же федеральному округу. Соответствующая часть полученных результатов приведена в таблице 2.

Таблица 2 - Корреляция по субъектам РФ

| Название субъектов | Коэффициент корреляции |
|------------------------------------------------------------------|------------------------|
| Республика Дагестан, Ставропольский край | 0,45 |
| Курская область, Москва | 0,05 |
| Самарская область, Саратовская область | 0,52 |
| Центральный федеральный округ, Северо-Западный федеральный округ | 0,59 |

Заключение

Таким образом, в данной статье нами были рассмотрены методология и возможности применения программного комплекса для эффективной работы с данными интернет-порталов. Использование разработанного приложения для обработки данных с портала Федеральной службы государственной статистики (Росстат) позволяет единым образом работать с различными форматами и разнотипными документами, проводить статистический анализ и делиться результатами своей работы с другими пользователями. Универсальный подход к обработке документов MSOffice позволяет разумно экономить рабочее время сотрудников, что в конечном итоге приводит к повышению эффективности их труда.

Библиография

1. Боровков А.А. Математическая статистика. М.: Физматлит, 2007. 703 с.
2. Вентцель Е.С. Теория вероятностей и ее инженерные приложения. М.: Высшая школа, 2007. 491 с.
3. Гамма Э., Хелм Р., Джонсон Р., Влиссидес Дж. Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2016.
4. Дэйт К. Дж. Введение в системы баз данных. М.: Вилльямс, 2017.
5. Коваленко И.С., Костиков Ю.А. Компьютерные технологии и эффективное построение рейтингов контрагентов малого бизнеса // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 3А. С. 372-378
6. Костиков Ю.А., Павлов В.Ю., Романенков А.М., Терновсков В.Б. Адаптивная архитектура программно-аппаратного комплекса хранения и обработки данных // Экономика: вчера, сегодня, завтра. 2017. Том 7. № 9А. С. 192-207.
7. Прохоренок Н., Дронов В. Python 3 и PyQt 5. Разработка приложений. СПб.: БХВ-Петербург, 2016.
8. Тюрин Ю.Н. Теория вероятностей и статистика. М.: МЦНМО, 2009. 256 с.

Tools for effective statistical processing of economic data

Yurii A. Kostikov

PhD in Physical and Mathematical Sciences,
 Head of the Department 812,
 Moscow Aviation Institute (National Research University),
 125993, 4 Volokolamskoe highway, Moscow, Russian Federation;
 e-mail: jkostikov@mail.ru

Aleksandr M. Romanenkov

PhD in Technical Sciences, Associate Professor,
Department 812,
Moscow Aviation Institute (National Research University),
125993, 4 Volokolamskoe highway, Moscow, Russian Federation;
e-mail: romanaleks@gmail.com

Abstract

The article is devoted to the consideration of the possibility of using modern information technologies for automation and improving the efficiency of work with the data of Internet portals on the example of the portal of the Russian Federal State Statistics Service. For the processing of statistical data, a universal software tool is used, which provides flexible adjustment to the individual needs of the user. The proposed approach allows to quickly process data, perform statistical calculations and data analysis. The functionality of automatic update tracking on the web portal allows to keep the data up to date. The possibilities of statistical processing and analysis of data provide a convenient tool for the preparation of reporting documentation, for comparative analysis of data. There is a rich functionality of data visualization. Using the developed application for data processing from the portal of the Russian Federal State Statistics Service (Rosstat) allows to work in a single way with different formats and different types of documents, conduct statistical analysis and share the results of your work with other users. A universal approach to processing MSOffice documents allows to reasonably save employees' working time, which ultimately leads to an increase in the efficiency of their work.

For citation

Kostikov Yu.A., Romanenkov A.M. (2018) Instrumentarii effektivnoi statisticheskoi obrabotki ekonomicheskikh dannykh [Tools for effective statistical processing of economic data]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (8A), pp. 416-425.

Keywords

ROSSTAT data processing, automation of statistical data processing, Russian Federal State Statistics Service (Rosstat), information technologies.

References

1. Borovkov A.A. (2007) *Matematicheskaya statistika* [Mathematical statistics]. Moscow: Fizmatlit Publ.
2. Deit K.Dzh. (2017) *Vvedenie v sistemy baz dannykh* [Introduction to database systems]. Moscow: Vill'yams Publ.
3. Gamma E., Khelm R., Dzhonson R., Vlissides Dzh. (2016) *Priemy ob"ektno-orientirovannogo proektirovaniya. Patterny proektirovaniya* [Techniques of object-oriented design. Design patterns]. Saint Petersburg: Piter Publ.
4. Kostikov Yu.A., Pavlov V.Yu., Romanenkov A.M., Ternovskov V.B. (2017) *Adap-tivnaya arhitektura programmno-apparatnogo kompleksa hraneniya i obrabotki dannykh* [Adaptive architecture of software and hardware complex for data storage and processing]. *Ekonomika: vchera, segodnya, zavtra* [Economics: yesterday, today, tomorrow], 7 (9A), pp. 192-207.
5. Kovalenko I.S., Kostikov Yu.A. (2018) *Komp'yuternye tekhnologii i effektivnoe postroenie reitingov kontragentov malogo biznesa* [Computer technologies and efficient construction of ratings of small business counterparties]. *Ekonomika: vchera, segodnya, zavtra* [Economics: yesterday, today, tomorrow], 8 (3A), pp. 372-378.

-
6. Prokhorenok N., Dronov V. (2016) *Python 3 i PyQt 5. Razrabotka prilozhenij* [Python 3 and PyQt 5. Application development]. Saint Petersburg: BHV-Peterburg Publ.
 7. Tyurin Yu.N. (2009) *Teoriya veroyatnostei i statistika* [The probability theory and statistics]. Moscow: Moscow Center for Continuous Mathematical Education.
 8. Venttsel' E.S. (2007) *Teoriya veroyatnostei i ee inzhenernye prilozheniya* [The probability theory and its engineering applications]. Moscow: Vysshaya shkola Publ.