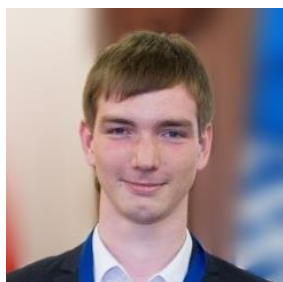


УДК 004.62+339.13

ИСПОЛЬЗОВАНИЕ ИНСТРУМЕНТОВ ЯЗЫКА R И СРЕДСТВ ПОИСКА ПО ТЕКСТУ ДЛЯ АНАЛИЗА РЫНКА ДРАГОЦЕННЫХ МЕТАЛЛОВ



В.А. Савёнок

Магистрант кафедры ПОИТ БГУИР,
инженер-программист
ООО «Незабудка Софтвер»



А.И. Парамонов

Доцент кафедры ПОИТ БГУИР,
кандидат технических наук, доцент

Белорусский государственный университет информатики и радиоэлектроники,
Республика Беларусь
E-mail: savionak@gmail.com.

В.А. Савёнок

Магистрант кафедры программного обеспечения информационных технологий БГУИР. Специальность «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей». Получил высшее образование в БГУИР по специальности «Программное обеспечение информационных технологий». Работает в ООО «Незабудка Софтвер» в должности инженера-программиста.

А.И. Парамонов

Доцент кафедры программного обеспечения информационных технологий БГУИР.

Аннотация. Рассмотрено применение инструментов языка R и средств поиска параметризованных шаблонов в тексте в качестве комплексного решения задачи анализа динамики цен на драгоценные металлы. Описано взаимодействие с открытым API для получения данных. Рассмотрены возможности пакетов языка R для разведочного статистического анализа и декомпозиции временных рядов. Представлены комплексные графики агрегированных данных. С помощью средства поиска параметризованных шаблонов в тексте проведен текстологический анализ обзоров рынка драгметаллов. В частности, изучен вопрос динамики роста цены на палладий. Приведен пример построения параметризованного шаблона для поиска сфер применения палладия.

Ключевые слова: анализ динамики цен, драгметаллы, палладий, параметризованные шаблоны, язык R, текстологический анализ.

Введение. Диверсификация рисков фондового рынка путем вложения средств в драгоценные металлы - золото, серебро, платину и палладий - надежный способ инвестирования, целью которого выступает долговременное сохранение покупательной способности сбережений. Для осуществления удачного вложения средств необходимо провести анализ рынка драгоценных металлов, выявить основные тенденции и сезонность изменений. Следует обратить внимание на выбросы и по возможности установить их причины. Это позволит строить более гибкие стратегии инвестирования в твердую валюту. Также при анализе рынка драгоценных металлов отдельной страны необходимо учитывать локальные экономические события для приведения получаемых данных к однородному виду [1, 2].

В данной работе проведен анализ рынка драгоценных металлов Республики Беларусь. Для анализа взяты данные с официального портала Национального банка Республики

Беларусь о цене одного грамма драгоценных металлов за период с 2009 по 2020 года. Для этого использованы следующие инструменты:

- доступ к интернет-ресурсам: `curl`;
- преобразование данных: `jsonlite`;
- работа с датами: `lubridate`;
- визуализация данных: `ggplot2`;
- встроенные возможности языка R.

Получение данных для анализа. Доступ к порталу Национального банка Республики Беларусь (НБРБ) осуществляется по специальному интерфейсу [3].

Для получения списка наименований драгоценных металлов и информации об их цене использовалась библиотека `curl` [4]. Данный инструмент позволяет осуществлять широко настраиваемые запросы к удаленным ресурсам и предоставляет простой интерфейс для обработки ответов. Для получения списка наименований драгоценных металлов с портала НБРБ использовался адрес: «<http://www.nbrb.by/api/metals>». Для преобразования полученных данных во внутренний формат, использована библиотека `jsonlite` [5]. Программная реализация доступа к API НБРБ по получению списка наименований драгметаллов на языке R приведена на рисунке 1.

```
con <- curl::curl("http://www.nbrb.by/api/metals")
metal.names.raw <- readLines(con)
close(con)
metal.names <- jsonlite::fromJSON(metal.names.raw)
```

Рисунок 1. – Получение наименований металлов с API НБРБ

Специфика API НБРБ заключается в том, что максимальный период, для которого возможно получить данные за один запрос, составляет один календарный год. В связи с этим потребовалось запрашивать данные в цикле, после чего было проведено объединение полученных фреймов. Данные о стоимостях драгметаллов за указанный период можно получить через запрос: «<http://www.nbrb.by/api/bankingots/prices?startdate=...&enddate=...>». Программный код, осуществляющий данные действия, приведен на рисунке 2.

```
years = c(2009:2020)
rows <- list()
i <- 1
for (y in years) {
  url <- sprintf("http://www.nbrb.by/api/bankingots/prices?
    startdate=%d-1-1&enddate=%d-12-31", y, y)
  con <- curl::curl(url)
  rows[[i]] <- readLines(con)
  close(con)
  i <- i + 1
}
metal.prices <- data.frame()
for (r in rows) {
  metal.prices <- rbind(metal.prices, jsonlite::fromJSON(r))
}
```

Рисунок 2. – Запрос данных о ценах драг. металлов с API НБРБ

Коррекция данных. Первого июля 2016 года в Республике Беларусь была проведена деноминация официальной денежной единицы – белорусского рубля. Находившиеся в обращении банкноты образца 2000 года были постепенно заменены на банкноты и монеты образца 2009 года в соотношении 10 000 к 1 [6]. В связи с этим потребовалось провести коррекцию входных данных, реализация которой приведена на рисунке 3.

```
denomination.date <- as.Date("2016-07-01")
denomination.factor = 10000
denominated_rows <- metal.prices$Date < denomination.date
metal.prices[denominated_rows, ]$Value <-
metal.prices[denominated_rows, ]$Value / denomination.factor
```

Рисунок 3. – Коррекция данных

Статистический анализ. При помощи встроенной функции *summary* произведён первичный статистический анализ загруженных данных. Результаты анализа приведены в таблице 1. Как видно из таблицы, наибольший разброс min-max отмечается для палладия, на втором месте – золото.

Таблица 1. – Статистический анализа данных

Наименование	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Золото	6.118	36.018	46.827	52.645	80.310	120.330
Серебро	0.0766	0.5859	0.8078	0.7469	1.0400	1.3400
Платина	6.359	37.341	46.808	43.660	57.810	74.600
Палладий	1.298	15.453	27.371	36.778	48.880	200.750

Учитывая специфику данных, проведена декомпозиция временного ряда на закономерную и случайную составляющую. Для этого использованы встроенные возможности языка R по анализу временных рядов: *ts()* и *decompose()*. Визуализация декомпозированного временного ряда произведена при помощи встроенной функции *plot()*.

Для удобства анализа проведена предварительная агрегация данных с усреднением по месяцам для каждого наименования. При этом коэффициент сезонности принят равный 12 месяцам, т. е. одному году. Полученные графики представлены на рисунках 4 – 7.

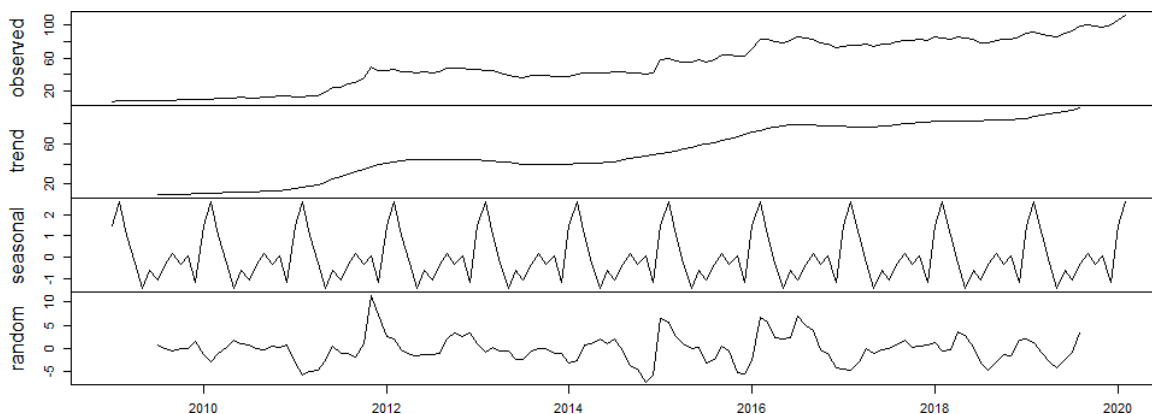


Рисунок 4. – Декомпозиция временного ряда динамики цены на золото

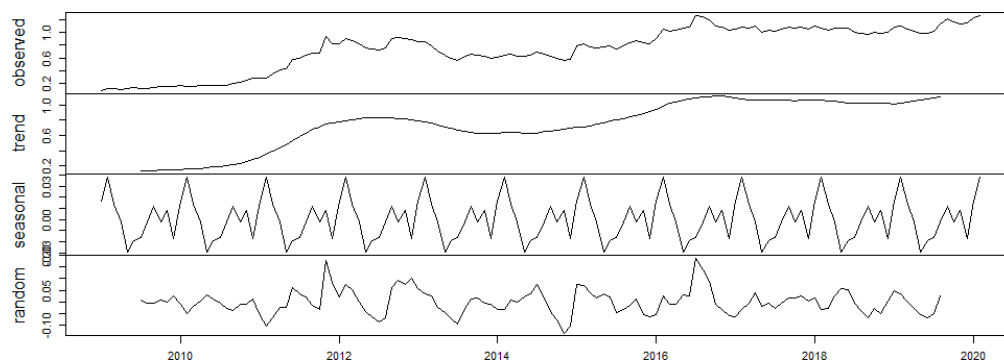


Рисунок 5. – Декомпозиция временного ряда динамики цены на серебро

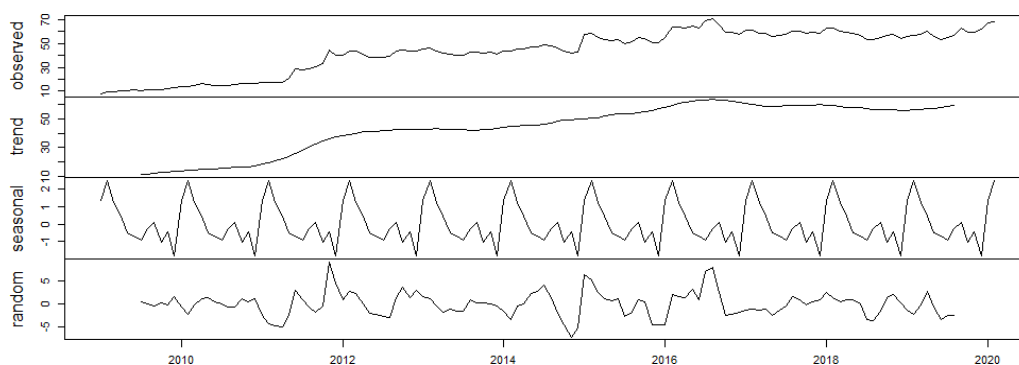


Рисунок 6. – Декомпозиция временного ряда динамики цены на платину

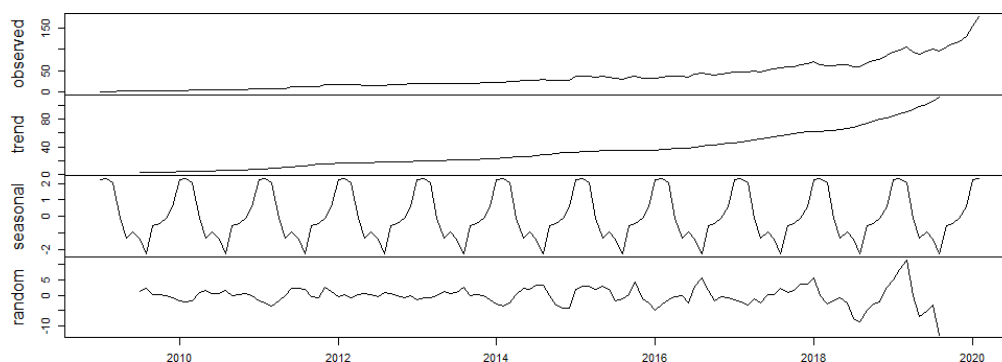


Рисунок 7. – Декомпозиция временного ряда динамики цены на палладий

Исходя из полученных графиков, наибольший сезонный пик приходится на начало календарного года. Наблюдается практически экспоненциальный рост цены на палладий.

Также выполнено сопоставление графиков динамики цены каждого наименования драгоценных металлов. Для решения данной задачи использована графическая библиотека *ggplot2* [7]. Графики динамики изменения цены драгоценных металлов, построенные на основании исходных, а также агрегированных (усреднённых) данных по месяцам, представлены на рисунках 8 и 9. При агрегации для корректной работы с датами использовалась библиотека *lubridate* [8].

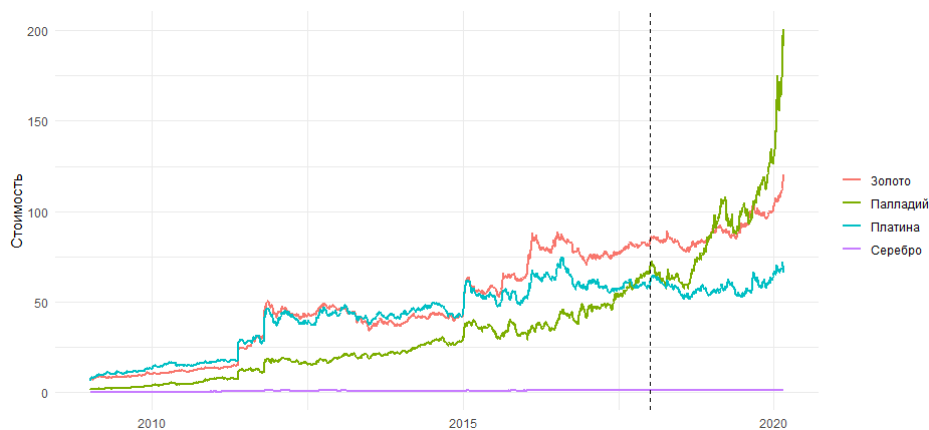


Рисунок 8. – Динамика цен на драгоценные металлы

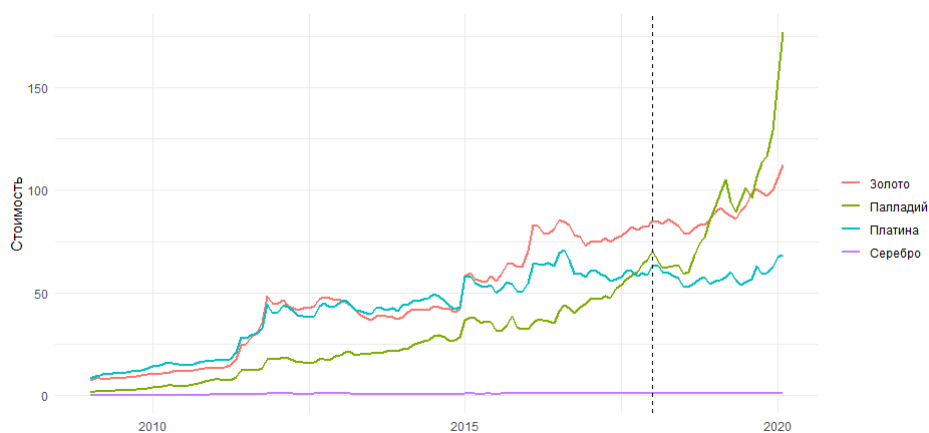


Рисунок 9. – Динамика цен на драгоценные металлы (по месяцам)

На основании полученных графиков можно сделать вывод, что наибольший рост цены наблюдается у палладия: начиная от 1.298 BYN за 1 грамм в начале 2009 года его цена поднялась до 200.750 BYN на начало 2020 года. Также стоит заметить, что в период с 2018 года (отмечен вертикальной пунктирной линией на рисунках 5 и 6) по настоящее время наблюдается ощутимый рост цены на данный драгметалл: прирост составил порядка 200%. Следует отметить, что еще в начале 2019 года цена палладия превысила цену золота. И если до 2018 года колебания цен на драгметаллы были связаны с трендами в экономике (например, экономические кризисы 2011 и 2015 годов) и эти колебания были одинаковы для всех металлов, то обозначенный взлет палладия явно выбивается из этого ряда.

Установление причин аномального роста. Для изучения причин описанного феномена проведено изучение аналитических источников при помощи средства для поиска параметризованных шаблонов в тексте [9].

После выбора ряда аналитических обзоров на тему палладия, полученная выборка была обработана при помощи инструмента поиска параметризованных шаблонов в тексте. Данный инструмент обладает мощным языком описания шаблонов, что позволяет осуществлять гибкий поиск разнообразных шаблонов [10].

Из многочисленных возможностей языка шаблонов использованы следующие: вариации (Variations), промежутки (Spans) для поиска слов на расстоянии, а также поиск лексем с заданным префиксом. Результат определен как Предложение, содержащее упоминание шаблона ПрименениеПалладия. Данный шаблон в свою очередь определен как варианты порядка следования шаблонов палладий и применение друг за другом на расстоянии

до пяти слов. Шаблоны *палладий* и *применение* представлены вариациями лексем, имеющих заданные префиксы. Итоговый шаблон поиска представлен на рисунке 10. Результаты анализа одного из обзоров приведены на рисунке 11.

```
#Результат = Предложение @having ПрименениеПалладия;  
#ПрименениеПалладия = {  
  палладий .. [0-5] .. применение,  
  применение .. [0-5] .. палладий  
};  
Предложение = Word ... {".", "~".}+Word, [2 LineBreak]};  
палладий = {"паллад"*};  
применение = {"использ"*, "примен"*};
```

Рисунок 10. – Шаблон для поиска сфер применения палладия

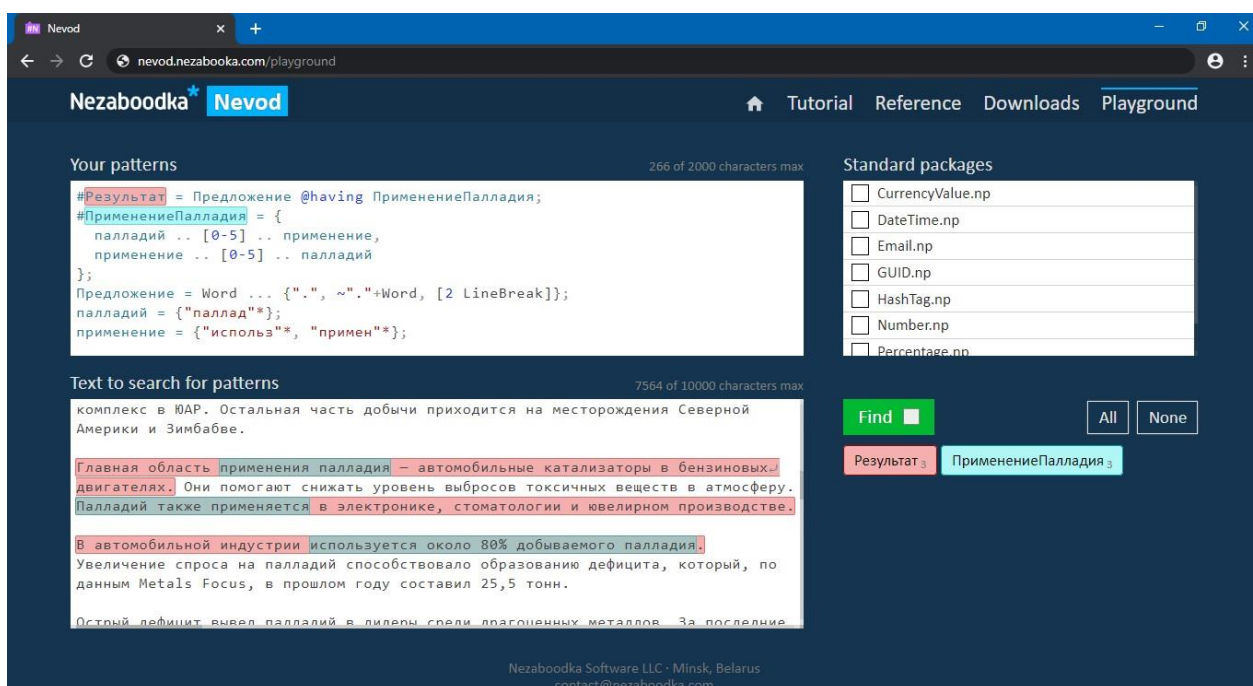


Рисунок 11. –Результат поиска заданного шаблона в тексте статьи

По данным анализа удалось выяснить, что основной сферой применения палладия является автомобилестроение. Доля используемого в нем металла составляет около 80% от добываемого объема [11]. Основной компонент, в изготовлении которого применяется палладий, – каталитические конвертеры в выхлопной системе автомобилей.

Известно, что в конце 2018 года представители Еврокомиссии, Европарламента, а также стран-участниц ЕС пришли к решению по ужесточению нормы выбросов углекислого газа автомобилями. 27 марта 2019 года Европарламент одобрил указанное решение, установив порядок снижения выбросов вплоть до 2030 года [12].

Заключение. В рамках данной работы при помощи возможностей языка R и его дополнительных библиотек был проведен анализ динамики цен на драгоценные металлы в Республике Беларусь в период с 2009 по 2020 года. Данные средства позволили выявить аномальный рост цены на палладий начиная с 2018 года. При помощи средства поиска параметризованных шаблонов в тексте был проведен текстологический разбор аналитических

обзоров на тему палладия и установлены причины указанного роста. Таким образом, использованные инструменты доказали свою состоятельность и эффективность при анализе данных и информационных источников.

Список литературы

- [1.] Паевой фонд ВТБ – Фонд Драгоценных металлов. Инвестиционная стратегия. Паевые инвестиционные фонды (ПИФы) // ВТБ Капитал Управление активами [Электронный ресурс]. – Режим доступа: https://www.vtbcapital-am.ru/products/pif/opif/vtbfdm/investment_strategy/. – Дата доступа: 05.02.2020.
- [2] InfoBank.by | Аналитика | Обзор рынка драгметаллов [Электронный ресурс]. – Режим доступа: <https://infobank.by/obzor-rynka-dragmetallov/>. – Дата доступа: 05.02.2020
- [3] API получения цен на драгоценные металлы, устанавливаемые Национальным банком | Национальный банк Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://www.nbrb.by/APIHelp/BankIngots>. – Дата доступа: 10.12.2019.
- [4] CRAN - Package curl // CRAN [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/curl/index.html>. – Дата доступа: 03.12.2019.
- [5] CRAN - Package jsonlite // CRAN [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/jsonlite/index.html>. – Дата доступа: 03.12.2019.
- [6] Деноминация | Национальный банк Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://www.nbrb.by/coinsbanknotes/denomination>. – Дата доступа: 10.12.2019.
- [7] CRAN - Package ggplot2 // CRAN [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/ggplot2/index.html>. – Дата доступа: 03.12.2019.
- [8] CRAN - Package lubridate // CRAN [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/lubridate/index.html>. – Дата доступа: 03.12.2019.
- [9] Nevod // Nezaboodka Nevod [Электронный ресурс]. – <https://nevod.nezaboodka.com/>. – Дата доступа: 15.12.2019.
- [10] Nevod Reference // Nezaboodka Nevod [Электронный ресурс]. – <https://nevod.nezaboodka.com/tutorial>. – Дата доступа: 15.12.2019.
- [11] Палладий дороже золота. Как распознать пузырь на рынке металлов :: Новости :: РБК Quote – [Электронный ресурс] – Режим доступа: <https://quote.rbc.ru/news/article/5d35b8d39a794798ae44d4ae>. – Дата доступа: 05.02.2020.
- [12] В Европе ужесточены нормы выбросов углекислого газа новыми авто // ИА "Финмаркет" [Электронный ресурс]. – Режим доступа: <http://www.finmarket.ru/database/news/4969004>. – Дата доступа: 05.02.2020.

USING R LANGUAGE TOOLS AND TEXT SEARCH TOOLS FOR ANALYZING THE PRECIOUS METALS MARKET

V.A. Savionok

*Master student of Software for
Information Technologies Department of
BSUIR, software engineer at
Nezaboodka Software LLC*

A.I. Paramonov

*PhD (Candidate of Computer Sciences),
Associate Professor at Software for
Information Technologies Department of
BSUIR, Associate Professor*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail: savionak@gmail.com.*

Abstract. The application of R language tools and means for searching for parameterized patterns in the text is considered as a complex solution to the problem of analyzing the dynamics of prices for precious metals. Interaction with an open API to retrieve data is described. The possibilities of R language packages for exploratory statistical analysis and decomposition of time series are considered. Complex graphs of aggregated data are presented. Using the search tool for parameterized patterns in the text, a textual analysis of precious metals market reviews has been carried out. In particular, the question of the dynamics of the growth in palladium prices has been studied. An example of constructing a parametrized pattern for finding applications for palladium is given.

Keywords: price dynamics analysis, precious metals, palladium, parameterized patterns, R language, textual analysis.