

# ОТ ТЕОРИИ К ЭКОНОМИЧЕСКОЙ ПОЛИТИКЕ

**В.С. Щербаков**

*к.э.н, Банк России (Екатеринбург), Омский государственный университет им. Ф.М. Достоевского*

## ОЦЕНКА ИНФЛЯЦИОННЫХ ОЖИДАНИЙ НАСЕЛЕНИЯ РОССИИ НА ОСНОВЕ ПОИСКОВЫХ ЗАПРОСОВ В СЕТИ ИНТЕРНЕТ (ПОДХОД «СВЕРХУ ВНИЗ»)

**Аннотация.** По своей природе инфляционные ожидания выступают ненаблюдаемой переменной. В рамках экономической теории и практики прокси-показатели инфляционных ожиданий (преимущественно на основе опросов) используются в качестве важнейших переменных для анализа и прогнозирования инфляционных процессов. Наряду с этим, при реализации режима таргетирования инфляции регуляторы преимущественно посредством коммуникаций в области денежно-кредитной политики нацелены на управление инфляционными ожиданиями. В этом аспекте проявляется их особый, дуальный характер. Сегодня продолжает расти популярность использования альтернативных оценок инфляционных ожиданий, включая статистику поисковых запросов. Концептуальным вопросом остается отбор ключевых слов для квантификации ожиданий населения. Целью исследования выступает разработка методологически обоснованного подхода к отбору ключевых слов для поисковых запросов, статистика по которым может использоваться в качестве прокси-переменных инфляционных ожиданий. В рамках статьи поставленная цель достигается на основе текстового анализа коммуникаций Банка России с применением моделей машинного обучения (в особенности NLP). На основе проведенного частотного анализа (Baseline-подход), а также использования дообученных NLP-моделей (семейство моделей T5 («Text-to-Text Transfer Transformer»)) были выделены четыре группы ключевых слов («инфляция», «Центральный банк», «курс», «ключевая ставка»), используя которые регулятор может формировать инфляционные ожидания населения России (подход «сверху вниз»). Ввиду последних изменений в политике доступности исторических данных, а также популярности поисковой сети среди жителей России, особый акцент сделан на данных поисковой сети Яндекс. Предполагается, что отслеживание динамики запросов по группам «инфляция» и «Центральный банк» дают оперативную информацию повсеместно, а по группам «курс», «ключевая ставка» — в кризисных и/или изменяющихся экономических условиях. Проведена апробация полученных результатов по поисковой статистике выделенных ключевых слов в качестве прокси-показателей в рамках прогнозирования инфляции в Российской Федерации на основе набора моделей семейства ARIMAX. Результаты свидетельствуют о целесообразности использования статистики по ключевым словам в качестве объясняющих переменных для минимизации ошибок прогнозов в рамках моделей прогнозирования инфляции.

**Ключевые слова:** *денежно-кредитная политика, таргетирование инфляции, инфляционные ожидания, поисковые запросы, текстовый анализ, методы машинного обучения, Яндекс.*

JEL: C82; C88; E31; E52

УДК: 336.7, 338.57

DOI: 10.52342/2587-7666VTE\_2025\_4\_68\_90

© В.С. Щербаков, 2025

© ФГБУН Институт экономики РАН «Вопросы теоретической экономики», 2025

**ДЛЯ ЦИТИРОВАНИЯ:** *Щербаков В.С. Оценка инфляционных ожиданий населения России на основе поисковых запросов в сети Интернет (подход «сверху вниз») // Вопросы теоретической экономики. 2025. №4. С. 68–90. DOI: 10.52342/2587-7666VTE\_2025\_4\_68\_90.*

**FOR CITATION:** *Shcherbakov V.S. Measuring Russian Public Inflation Expectations Using Internet Search Data: A Top-Down Approach // Voprosy teoreticheskoy ekonomiki. 2025. No. 4. Pp. 68–90. DOI: 10.52342/2587-7666VTE\_2025\_4\_68\_90.*

Настоящая статья отражает личную позицию автора. Содержание и результаты данного исследования не следует рассматривать, в том числе цитировать в каких-либо изданиях, как официальную позицию Банка России или указание на официальную политику или решения регулятора. Любые ошибки в данном материале являются исключительно авторскими.

## Введение

В самом общем виде инфляционные ожидания могут быть определены в качестве агрегированных предположений экономических субъектов относительно будущего уровня инфляции. Несмотря на весь достаточно богатый накопленный мировой опыт в области исследований природы инфляционных ожиданий (см., например: [Carlson, Parkin, 1975; Mankiw, Reis, Wolfers, 2003; Carroll, 2003; Batchelor, 2009; Жемков, Кузнецова, 2017, Винокуров, Медведь, 2023]) и использования данного показателя для анализа и прогнозирования инфляции (см., например: [Bernanke, 2007; Fuhrer, 2012; Banbura, Leiva-Leon, Menz, 2021]), по-прежнему остаётся актуальным вопрос их измерения.

Это обусловлено тем, что, как правило, инфляционные ожидания экономических агентов выступают ненаблюдаемой величиной. Одним из самых распространённых подходов для сбора информации об ожидаемой населением инфляции выступают разного рода опросы<sup>1</sup>. Впервые такие опросы стали проводиться более 50 лет назад и на сегодняшний день являются мейнстримом в области сбора данных об инфляционных ожиданиях. Для получения количественных оценок собранных данных применяются различные статистические приемы — методы квантификации инфляционных ожиданий [Хазанов, 2015; Перевышин, Рыкалин, 2018]. В России регулярные исследования в данной области проводятся с 2009 г. по заказу Банка России. Полученные данные<sup>2</sup> используются на регулярной основе в рамках реализации текущей денежно-кредитной политики.

В последние годы усилилась работа в области получения альтернативных оценок инфляционных ожиданий, в особенности с использованием широкого спектра методов машинного обучения [Голощанова, Андреев, 2017; Larsen, Thorsrud, Zhulanova, 2021]. Особый интерес к данному вопросу обусловлен возможностью квантификации инфляционных ожиданий на основе общедоступной информации в режиме реального времени с минимальными затратами в отличие от опросных подходов [Петрова, 2022]. В первую очередь речь идёт об использовании различных неструктурированных данных из сети Интернет, включая публикуемые новости, данные социальных сетей и так далее [Aromí, Llada, 2020, Angelico et al., 2022; Шуляк, 2022; Евстигнеева, Карпов, 2023; Shcherbakov, Karpov, 2024].

В этом отношении отдельным блоком выделяется корпус исследований, посвящённый использованию поисковых запросов в качестве прокси-показателей для различных экономических индикаторов [Щербаков Харламова, Гартвич, 2022; Федюнина, Юревич, Городный, 2024]. В области инфляционных ожиданий сделан достаточно большой задел с точки зрения применения подобной информации для анализа и прогнозирования инфляции (Приложение 1).

В целом поисковое поведение в Интернете можно интерпретировать в качестве показателя выявленных ожиданий ввиду того, что люди ищут информацию по темам, о которых хотят узнать больше, или о вещах, которые вызывают у них беспокойство. В рамках одной из пионерных работ в исследуемой области установлено, что данная логика в полной мере относится и к оценке инфляционных ожиданий [Guzman, 2011]. При этом, учитывая значительно возросшее количество пользователей сети Интернет за последнее

<sup>1</sup> Подробнее о проводимых по заказу Банка России опросах инфляционных ожиданий и применяемых методах их квантификации можно ознакомиться в специализированном разделе официального сайта «Инфляционные ожидания и потребительские настроения». Банк России. URL: [https://cbr.ru/analytics/dkp/inflationary\\_expectations/infl\\_exp\\_23-12/#highlight=инфляционные%7Сожидания%7Спотребительски-е%7Снастроения%7Синфляционных%7Сожиданиях%7Спотребительских%7Снастроениях](https://cbr.ru/analytics/dkp/inflationary_expectations/infl_exp_23-12/#highlight=инфляционные%7Сожидания%7Спотребительски-е%7Снастроения%7Синфляционных%7Сожиданиях%7Спотребительских%7Снастроениях) (дата обращения: 06.09.2024)

<sup>2</sup> Накопленные статистические данные по инфляционным ожиданиям на основе опросов «инФОМ». Банк России. URL: [https://cbr.ru/statistics/ddkp/inflationary\\_expectations/](https://cbr.ru/statistics/ddkp/inflationary_expectations/) (дата обращения: 20.07.2025)

десятилетие<sup>3</sup>, можно говорить о том, что получаемая таким образом информация является репрезентативной с точки зрения характеристики поведения как широких слоёв населения, так и общества в целом.

При этом в рамках обсуждаемого направления одним из важнейших вопросов остаётся определение ключевых слов, статистика по поисковым запросам которых может свидетельствовать об изменениях инфляционных ожиданий. Этот момент является архиважным ввиду агрегированной природы показателя инфляции как такового. В частности, индекс потребительских цен (далее ИПЦ), лежащий в основе расчёта инфляции, измеряется на основе потребительской корзины, состоящей из более 550 товаров и услуг. Учёт всё многообразие возможных словоформ и словосочетаний на основе входящих в его расчёт товаров и услуг является проблематичным.

С одной стороны, логично предположить, что если мы хотим получить прокси-показатель для инфляционных ожиданий, то в первую очередь необходимо использовать в качестве ключевого слова термин «инфляция» и его производные. Это на практике и наблюдается. С другой стороны, нет никаких свидетельств о том, что все пользователи сети Интернет правильно трактуют данное понятие и до конца понимают его природу, а поэтому используют его корректно.

Остаётся открытым вопрос обоснования выбора других значимых ключевых слов, которые потенциально могут отражать изменение инфляционных ожиданий через статистику их поисковых запросов как в разрезе регионов, макрорегионов, так и страны в целом. Как правило, исследователи в данной области подходят к определению ключевых слов экспертным методом или своеобразным перебором, например, через включение статистики по широкому спектру экономических терминов и конструированию наборов словосочетаний на их основе с дальнейшим исключением незначимых рядов данных.

В связи с этим в рамках данного исследования целью *выступает разработка методологического подхода к отбору ключевых слов для поисковых запросов, статистика по которым может использоваться в качестве прокси-переменных инфляционных ожиданий.*

Как известно, ключевой целью денежно-кредитной политики (далее ДКП) Центрального банка Российской Федерации является поддержание ценовой стабильности, другими словами — стабильно низкой инфляции (вблизи 4% постоянно). Для достижения поставленной цели Банк России, начиная с 2015 года, полноценно перешёл к реализации режима инфляционного таргетирования (далее ИТ).

В рамках данной политики Банк России, с одной стороны, принимает решения по ДКП на основе макроэкономического прогноза, который строится с учётом большого массива разносторонней информации, в том числе замеров и инкорпорирования в свои модели инфляционных ожиданий профессиональной и непрофессиональной (широкой) аудитории. В этом случае инфляционные ожидания выступают как один из факторов принятия решения об изменении ключевой ставки как основного инструмента ДКП.

С другой стороны, осуществление ИТ напрямую связано с проведением коммуникаций, направленных на снижение инфляционных ожиданий. При таком подходе уже инфляционные ожидания могут рассматриваться в качестве зависимой переменной, которая подвержена влиянию со стороны Банка России.

Чтобы меры регулятора были более эффективными, необходимо привязать инфляционные ожидания населения и бизнеса к цели по инфляции («заякорить»). В рамках данного процесса важнейшей составляющей выступает доверие со стороны экономических агентов Банку России. Оно формируется при успешном достижении обозначенной цели,

<sup>3</sup> Согласно Всемирному банку (World Bank), удельная доля населения мира, пользующаяся сетью Интернет, возросла с 29% в 2010 г. до 67% в 2023 г. Для России этот показатель составил 42 и 92% соответственно. Всемирный Банк. URL: <https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=RU> (дата обращения: 20.01.2025).

а также понимании обществом сути проводимой ДКП. В данном отношении Банк России стремится быть максимально открытым, в том числе посредством проводимых коммуникаций. Это выступает одним из ключевых принципов ДКП при таргетировании инфляции.

Необходимо отдельно отметить, что в России инфляционные ожидания профессиональной аудитории были заякорены на цели по инфляции начиная с 2017 г. При этом инфляционные ожидания широкой аудитории остаются не заякоренными. Представляется, что это не должно выступать препятствием для таргетирования. Мировой опыт показывает, что даже в странах с более длительным опытом поддержания ценовой стабильности инфляционные ожидания реального сектора экономики являются в значительной мере адаптивными и следуют за фактической динамикой инфляции [Банк России, 2024].

Таким образом, своеобразный дуализм в использовании инфляционных ожиданий со стороны Банка России, в особенности его вторая составляющая, приводит к тому, что сами коммуникации регулятора могут содержать ключевые смыслы, направленные на управление инфляционными ожиданиями населения. Представляется, что со временем данная логика подлежит усилению, в том числе с учётом повышения эффективности коммуникаций Банка России, накопления необходимого опыта [Евстигнеева, 2023]. Данные суждения легли в основу разрабатываемых подходов.

## Методология исследования

Основная гипотеза исследования состоит в том, что комплекс коммуникаций Банка России в области ДКП содержит специальную информацию, направленную на управление инфляционными ожиданиями агентов. В связи с этим необходимо сделать ряд дополнительных комментариев. Является ли поисковая активность в сети Интернет отражением инфляционных ожиданий населения в целом либо только Интернет-пользователей? Принято считать, что основным потребителем информации Банка России является профессиональная аудитория.

Во-первых, с учётом сегодняшнего уровня проникновения и доступности Интернета для граждан России, поведение Интернет-пользователей видится вполне репрезентативным с точки зрения отражения поведения населения в целом.

Во-вторых, как уже говорилось, Банк России нацелен на коммуникацию как с профессиональной, так и непрофессиональной аудиторией. Установлено, что уровень коммуникаций Банка России остаётся относительно сложным для понимания именно широкой аудиторией. Но выявлены улучшения удобочитаемости материалов начиная с 2018 г. [Евстигнеева, Sidorovskiy, 2021]. В контексте данной работы важно учесть, что даже если население напрямую не часто или вовсе не сталкивается с коммуникациями в области ДКП посредством официального сайта Банка России, который в качестве одного из основных источников информации находится в фокусе исследования, то сами коммуникации являются инфоповодом и ретранслируются далее различными СМИ, в социальных сетях, в том числе в адаптированном, упрощённом виде. Имеет место своеобразный мультипликативный эффект коммуникаций, осуществляемых Банком России.

В-третьих, хотя и невозможно выделить из общего количества поисковых запросов именно те, которые были сделаны профессиональной аудиторией, но, учитывая их значимость при малочисленности самого экспертного сообщества, мы придерживаемся мнения, что поиск по ключевым словам именно населением является определяющим в наблюдаемых тенденциях.

Основная идея применяемого метода заключается в том, чтобы упростить исходный текст до версии, которая содержала бы ключевую суть или общий смысл изложенного. В контексте данной работы нас будет интересовать возможность «сворачивания» официальных текстов коммуникаций Банка России в области ДКП до нескольких ключевых слов,

поисковая статистика по которым дальше может быть квантифицирована с помощью данных поисковых запросов.

Другими словами, в рамках исследования проводятся эксперименты с так называемой саммаризацией текстов (summarization) в области ДКП. Сама задача саммаризации текста существует уже более 50 лет [Дауит, Кемалов, Джаксылыкова, 2020; Yadav D., Desai, Yadav A., 2022] и является одним из направлений NLP<sup>4</sup> (natural language processing, обработка естественного языка). При этом она получила «второе дыхание» только в последнее десятилетие в связи с внедрением продвинутых методов машинного обучения для её решения<sup>5</sup>.

Особый импульс или даже популяризацию получила данная отрасль в связи с внедрением больших языковых моделей (LLM), таких как зарубежный ChatGPT<sup>6</sup>, отечественные GigaChat<sup>7</sup>, YandexGPT<sup>8</sup> и других. Безусловно, внедрение данных моделей для решения огромного спектра задач обладает огромным потенциалом, в том числе с точки зрения саммаризации текста. При этом в силу значительного масштаба они могут обладать некоторыми особенностями или даже недостатками использования с точки зрения их локализации под определённые точечные вопросы, которые среди прочего рассматриваются в данной работе.

В этом отношении значимой вехой в области использования методов машинного обучения, в том числе в NLP, является разработка и внедрение механизма «передачи знаний» (transfer learning) [Thrun, Pratt, 1998]. Это позволяет использовать сложные, предварительно обученные на большом массиве информации модели (pre-trained models) для решения смежных задач через точную настройку/дообучение модели (fine-tuning) под конкретную цель [Han et al., 2021]. С практической точки зрения это помогает экономить значительные временные, вычислительные и интеллектуальные ресурсы без существенной потери эффективности.

В свою очередь, с точки зрения решения целого комплекса задач в области NLP существенным шагом последних лет стала, например, презентация семейства моделей T5 («Text-to-Text Transfer Transformer»). Подобные модели имеют мультизадачную природу и предназначены для решения широкого спектра задач, включая саммаризацию текста. В целом T5 модели являются представителями класса предобученных encoder-decoder моделей, в основе которой лежит трансформер архитектура (Transformer) [Vaswani et al., 2017].

Результаты использования модификаций данных моделей на стандартизированных тестах превзошли существующие на тот момент аналоги [Raffel et al, 2020]. Сопоставимость получаемых результатов с более новыми моделями (Pegasus, различные модификации BERT-based, CNN-based и другие) также подтвердилась в ряде последующих исследований для различных стран [Ay et al., 2022, Wang et al, 2023, Guan, Zhu, Yuan, 2024].

Для решения поставленной в рамках данной работы задачи и, учитывая экспериментальный характер подхода, после дополнительного анализа были отобраны две модели:

<sup>4</sup> Natural Language Processing — это область искусственного интеллекта, главная задача которой — дать машинам возможность анализировать, понимать и интерпретировать человеческий язык. NLP выступает основой для больших языковых моделей (large language model, LLM), которые призваны также генерировать очеловеченный текст.

<sup>5</sup> Подробнее с прогрессом работ в области саммаризации текста можно ознакомиться на сайте NLP-progress. URL: <http://nlpprogress.com/english/summarization.html> (дата обращения: 15.01.2025)

<sup>6</sup> Чат-бот с генеративным искусственным интеллектом, созданный компанией OpenAI. URL: <https://openai.com/chatgpt> (дата обращения: 20.02.2025)

<sup>7</sup> Русскоязычная нейросеть от Сбера. URL: <https://developers.sber.ru/gigachat/login> (дата обращения: 20.02.2025)

<sup>8</sup> Генеративная модель Яндекса. URL: <https://ya.ru/ai/gpt-3> (дата обращения: 20.02.2025)

- ▶ ruT5-base<sup>9</sup> [Zmitrovich et al., 2023]
- ▶ keyt5-base<sup>10</sup>

При их выборе мы руководствовались несколькими критериями: 1) модели базируются на архитектуре и логике T5; 2) модели изначально предназначены/предобучены для работы с текстами на русском языке; 3) модели не являются «тяжёлыми» и их возможно использовать с точки зрения необходимых вычислительных мощностей даже на персональном компьютере.

Дополнительно в рамках данной работы в качестве определённого бенчмарка (*Baseline-подход*) для сравнения с полученными результатами указанных моделей NLP будет использоваться частотный подход к анализу текстовых коммуникаций Банка России. Суть данного подхода заключается в следующем: необходимо определить ключевые слова, которые наиболее часто (относительное количество употреблений) использовались регулятором.

С одной стороны, логика такого подхода заключается в том, что Банк России может более часто употреблять те ключевые слова, которые имеют наибольшую информационную силу в рамках проводимых коммуникаций. С другой стороны, этот метод является относительно простым.

Для этого на первом этапе вся собранная база коммуникаций регулятора будет объединена в единый текст с удалением лишних пунктуационных знаков. Далее он подвергнется токенизации (разделению на отдельные слова) и очистки от общеупотребительных, так называемых, стоп-слов (например: он, это, ещё и т.д.). На следующем шаге будет произведена лемматизация оставшихся слов. Это поможет устранить разные вариации одних и тех же слов, связанные с различиями в спряжении, падежах, употреблении множественного и/или единственного числа и т.д. В итоге будет получен список наиболее часто встречающихся слов в коммуникациях Банка России, который будет сопоставлен с результатами саммаризации текстов с использованием отобранных языковых моделей.

## Сбор и обработка данных

Для решения поставленной в рамках данного исследования задачи на первом этапе возникла необходимость формирования двух баз данных: 1 — «База для извлечения смыслов» и 2 — «База для обучения моделей машинного обучения». Остановимся на логике их формирования подробнее.

«База для извлечения смыслов» представляет собой набор текстовых коммуникаций Банка России в области ДКП, осуществлённых через официальный сайт регулятора ([www.cbr.ru](http://www.cbr.ru)) в период с 1 января 2014 г. по 1 января 2025 г. Всего в базу попало 286 коммуникаций в области ДКП, которые можно подразделить на следующие разновидности: пресс-релиз по итогам заседания Совета директоров Банка России, материалы пресс-конференций по итогам заседания Совета директоров Банка России, материалы сессий «вопрос-ответ» после заседаний Совета директоров Банка России, интервью по вопросам ДКП для широкого круга изданий, выступления (на форумах, в Государственной Думе РФ и на других площадках).

С одной стороны, такая логика разделения коммуникаций регулятора по типам уже задана на официальном сайте в тематическом разделе, посвящённом новостям ДКП<sup>11</sup>.

<sup>9</sup> Описание модели ruT5-base. Hugging Face. URL: <https://huggingface.co/ai-forever/ruT5-base> (дата обращения: 24.01.2025).

<sup>10</sup> Описание модели 0x7194633/keyt5-base. Hugging Face. URL: <https://huggingface.co/0x7194633/keyt5-base> (дата обращения: 24.01.2025).

<sup>11</sup> Специализированный раздел, посвящённый новостям в области ДКП на официальном сайте Банка России. URL: <https://www.cbr.ru/dkp/news/> (дата обращения: 20.02.2025).

С другой стороны, представляется целесообразным выделить обсуждения в рамках сессий «вопрос-ответ» после заседаний Совета директоров Банка России в отдельную категорию ввиду особой значимости прямой коммуникации со СМИ, возможности уточнения озвученных позиций в пресс-релизах и выступлениях Председателя напрямую в режиме диалога.

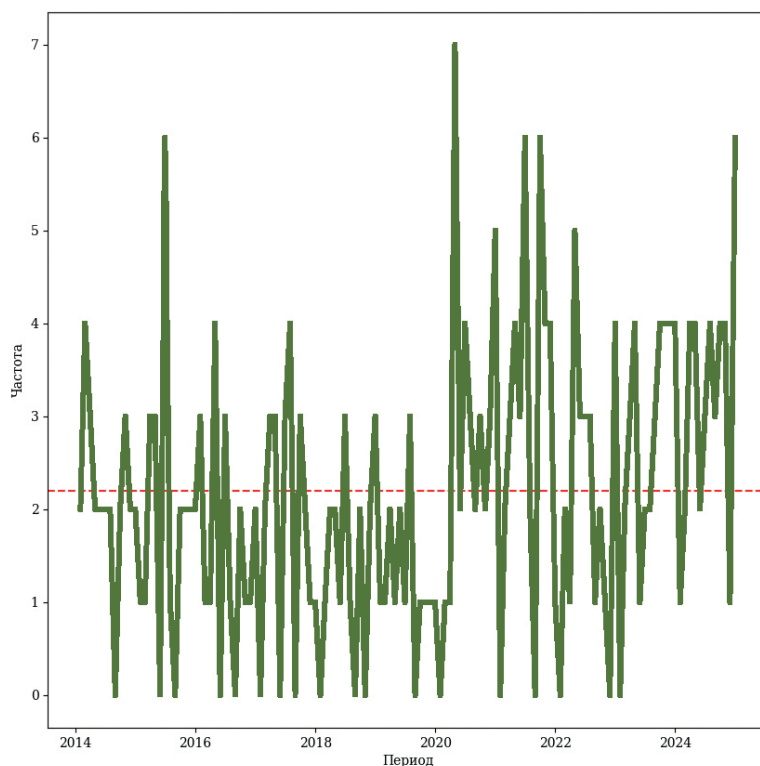


Рис. 1. Коммуникации Банка России в области ДКП на официальном сайте Банка России  
Источник: составлено автором.

Исходя из представленных данных (рис. 1) видно, что в среднем на официальном сайте Банка России размещалось два материала в области ДКП в месяц, отнесённых к вышеобозначенным категориям. Начиная с 2020 г. среднее значение подобных коммуникаций превысило 2,5.

Представляется, что применяемый подход требует ряда пояснений, которые раскрывают суть содержания собранной базы.

Во-первых, необходимо понимать, что информационная политика Банка России в области ДКП не является статичной. На протяжении всего наблюдаемого периода регулятор ведёт работу в рамках проводимых коммуникаций, экспериментирует, вводит новые форматы взаимодействия. Например, в плановом режиме заседания Совета директоров Банка России проходят 8 раз в год. При этом на начальном этапе реализации режима таргетирования инфляции только по итогам четырёх заседаний Совета директоров публиковались развёрнутые материалы, включающие оценки экономической ситуации в стране, прогноз макроэкономического развития, которые влияют на сохранение или изменение ключевой ставки. В свою очередь Э.С. Набиуллина, Председатель Банка России, осуществляла коммуникации в режиме пресс-конференций по результатам принятых решений. В остальных случаях осуществлялась только публикация пресс-релизов на сайте. Далее перешли к проведению пресс-конференций по итогам всех заседаний. Кроме того, новацией последних лет является участие в сессии «вопрос-ответ» после заседаний Совета директоров Банка России не только Председателя, но и заместителя Председателя Банка России (А.Б. Заботкина), курирующего блок ДКП в целом (рис. 2).

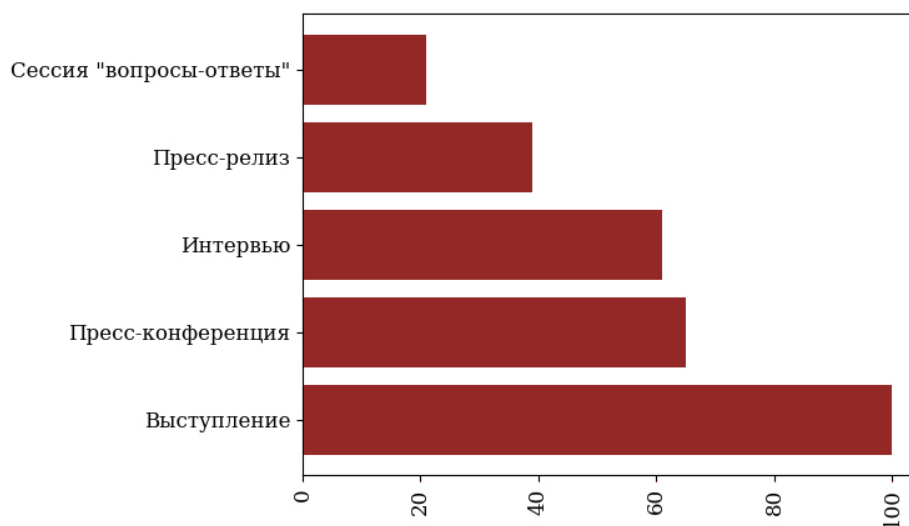


Рис. 2. Структура коммуникаций Банка России в области ДКП на официальном сайте Банка России по основным видам

Источник: составлено автором.

Во-вторых, в силу принятого подхода коммуникации в области ДКП, попавшие в сформированную базу, осуществлялись ключевыми руководителями Банка России: от Председателя до директора департамента ДКП. В том случае, когда материал (преимущественно пресс-релизы по итогам ранних заседаний Совета директоров) публиковался пресс-службой Банка России, авторство материала значилось за данным структурным подразделением (не персонифицировано). Отметим, что в 59% случаев в рамках собранной базы в качестве спикера выступала индивидуально Председатель Банка России Э.С. Набиуллина.

В-третьих, вне зависимости от источников и авторов публикаций по теме ДКП в базу включались только те материалы, которые были размещены/перепечатаны на официальном сайте Банка России. Например, выступления на пленарном заседании Совета Федерации, на совместном заседании профильных комитетов Государственной Думы, на встрече Ассоциации банков России и так далее. Представляется, что размещение данных коммуникаций на официальном ресурсе регулятора может свидетельствовать об их значимости и актуальности с точки зрения проводимой ДКП.

Безусловно, в базу попали далеко не все материалы Банка России в области ДКП. Однако собранной информации вполне достаточно для формирования общего представления о проводимых коммуникациях, в особенности с учётом реализуемой политики «единого голоса»<sup>12</sup>.

Данная база будет использоваться для выделения ключевых слов, содержащихся в официальных коммуникациях Банка России в области ДКП. Квантификация инфляционных ожиданий на основе статистики поисковых запросов выделенных слов даёт возможность создать соответствующие прокси-показатели.

Как было отмечено ранее, механизм «передачи знаний» (transfer learning) позволяет использовать сложные, предварительно обученные на большом массиве информации модели для решения целого ряда фундаментальных и прикладных задач. Это открывает широкие возможности для пользователей, не обладающих большими вычислительными ресурсами, с точки зрения адаптации этих моделей. В рамках использования данного механизма важным моментом является адаптация обученных моделей под конкретные цели исследователя.

<sup>12</sup> Банк России. URL: [https://cbr.ru/dkp/information\\_policy/](https://cbr.ru/dkp/information_policy/) (дата обращения: 01.12.2024).



В связи с этим возникла необходимость создания второй базы: «База для обучения моделей машинного обучения». Информация, содержащаяся в данной базе, должна позволить выбранным NLP-моделям более качественно решать задачу суммаризации с учётом того, что все тексты имеют финансово-экономическую природу, опубликованы на русском языке. Другими словами, такая база позволяет «затачивать» избранные, заранее обученные на широком массиве информации модели именно под экономическую тематику.

Для формирования данной базы использовались ведущие российские новостные порталы: РБК и РИА Новости. Финальный отбор данных масс-медиа произошёл по следующим критериям:

1) они являются общепризнанными федеральными изданиями, что подтверждается, например, постоянным нахождением в топах рейтингов Медиалогии<sup>13</sup>, системы автоматического мониторинга анализа СМИ и соцмедиа;

2) издания не были созданы «вчера» и имеют накопленный пул публикаций по финансово-экономической тематике, по крайней мере начиная с 2014 г. — основного подготовительного этапа в преддверии начала таргетирования инфляции Банком России;

3) каждая тематическая публикация снабжена набором специальных тегов, которые как раз могут выступать в роли ключевых слов, характеризующих содержание опубликованных статей.

На последнем критерии требуется остановиться подробнее. Все сайты в сети Интернет нацелены на максимизацию количества посетителей своих ресурсов. Одним из значимых источников перехода на порталы являются поисковые системы, которые, в свою очередь, заинтересованы в выдаче релевантных ссылок в ответ на запросы по тем или иным ключевым словам. В связи с этим сайты проводят поисковую оптимизацию для поднятия своих позиций в результатах выдачи поисковых систем по определённым запросам пользователей с целью увеличения трафика. Одним из элементов оптимизации выступает использование тегов к материалам, размещаемым на сайтах.

Таким образом, представляется, что информационные агентства, в том числе РБК и РИА Новости, крайне заинтересованы в подборе надлежащих тегов, наиболее полно характеризующих содержание публикуемых новостей. Использование подобной информации из двух источников позволит избежать смещения (bias), которое могло возникнуть при фокусировке только на одном портале или использовании несбалансированной выборки. В данном контексте идёт речь о смещении, вызванном тем, что каждое из информационных агентств имеет свою специфику, а также логику разметки тематических статей с помощью тегов. Дальнейшее расширение базы посредством включения других Интернет-ресурсов является нецелесообразным ввиду несоответствия одному или нескольким критериям, выделенным выше.

Как результат, в указанный период в «Базу для обучения моделей машинного обучения» вошло 26 932 новости финансово-экономической тематики.

Отметим, что частота публикационной активности на порталах РБК и РИА Новости разная (рис. 3). Тем не менее новости в итоговую базу отбирались путём случайной выборки в соотношении 50 на 50. Другими словами, полученная база является сбалансированной относительно источников информации за весь период в целом. При этом разбивка по годам может отличаться с точки зрения включаемых новостей из различных источников в определённые годы, как это и показано на рис. 3.

Отдельного внимания заслуживают комментарии относительно использования данных поисковых систем в качестве потенциального источника для квантификации инфляционных ожиданий населения. Устоявшимся стандартом в области использования данных поисковых запросов для целей наукастинга и краткосрочного прогнозирования различных экономических показателей является использование данных Google. Начало использования

<sup>13</sup> Медиалогия. URL: <https://www.mlg.ru/ratings/media/federal/12702/#internet> (дата обращения: 02.12.2024).

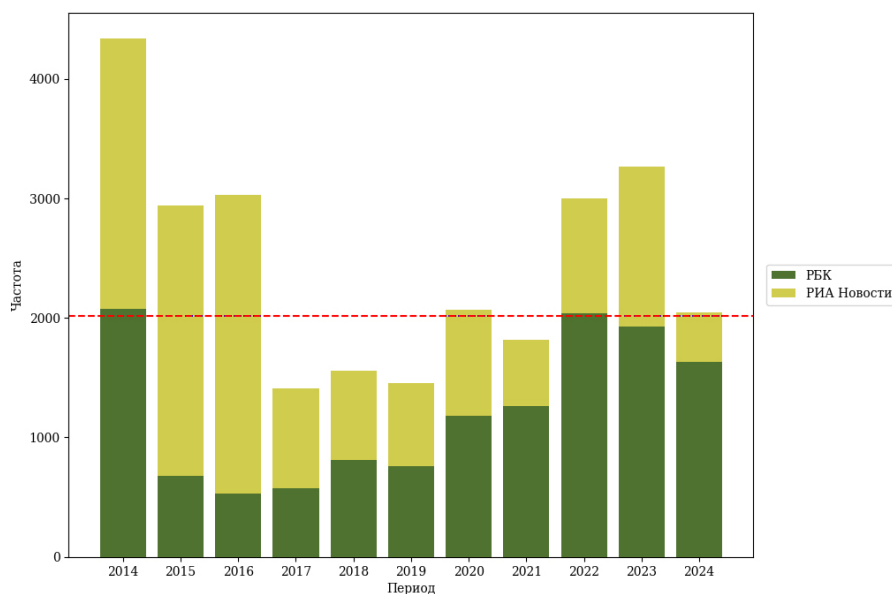


Рис. 3. База для обучения моделей машинного обучения на основе финансово-экономических новостей РБК и РИА Новости (янв.14 — дек.24)

Источник: составлено автором.

данного подхода было заложено более 15 лет назад, после запуска сервиса Google Trends<sup>14</sup> [Choi, Varian, 2009]. Исследование опубликованных работ в области использования поисковых данных для целей анализа инфляционных процессов в различных работах подтверждают данный тезис — превалирует Google (Приложение).

Сложившаяся ситуация является обоснованной ввиду того, что на поисковую систему Google в масштабах мира приходится более 90% всех пользовательских запросов<sup>15</sup>. При этом в российской практике сложилась уникальная ситуация, когда до середины 2022 г. рынок поисковых запросов был практически полностью поделён между двумя системами — Google и Яндекс — в примерном соотношении 50 на 47% [Щербаков, Харламова, Яковина, 2022. С. 4483]. Начиная с указанного периода, Яндекс занял лидирующие позиции. На текущий момент на него приходится 74% запросов<sup>16</sup>, против примерно 25% у Google. Представляется, что данное соотношение будет сохраняться в ближайшее время.

С одной стороны, важным преимуществом специализированного сервиса от Яндекс<sup>17</sup> по сравнению с конкурирующим и общепринятым в данной сфере Google Trends выступает возможность получения доступа к информации не в индексной форме, а в абсолютном значении, то есть количестве запросов по тем или иным ключевым словам. Хотя надо отметить, что до недавнего времени существенным ограничением в работе с открытыми данными Wordstat была короткая история запросов по ключевым словам — за 24 последних месяца на день обращения для данных частотой в месяц и за 12 последних месяцев — для недельных [Щербаков, Харламова, Яковина, 2022, С. 4484]. С марта 2024 г. стала доступна новая версия сервиса, которая расширила горизонт доступной статистики о динамике ключевых запросов до 7 лет (с 2018 г.)<sup>18</sup>. Представляется, что этот шаг значительно увеличивает возможности использования данных Яндекс для решения исследовательских задач.

<sup>14</sup> Google Trends. URL: <https://trends.google.ru/trends/> (дата обращения: 15.05.2025).

<sup>15</sup> По данным аналитического агентства Statcounter по состоянию на 10.03.2025. Statcounter Global Stats. URL: <https://gs.statcounter.com/search-engine-market-share> (дата обращения: 12.03.2025).

<sup>16</sup> По данным аналитического агентства Statcounter по состоянию на 10.03.2025. Statcounter. URL: <https://gs.statcounter.com/search-engine-market-share/all/russian-federation> (дата обращения: 12.03.2025)

<sup>17</sup> «Яндекс. Подбор слов» (Wordstat)». URL: <https://wordstat.yandex.ru/> (дата обращения: 02.06.2025)

<sup>18</sup> Подробности на: Яндекс Вебмастер. URL: <https://webmaster.yandex.ru/blog/s-20-marta-na-wordstat-yandex-ru-zapustitsya-obnovlennyy-vordstat> (дата обращения: 14.08.2024)

## Результаты отбора ключевых слов для квантификации инфляционных ожиданий населения

На первом шаге был проведён анализ ключевых слов на основе *Baseline-подхода*, описание которого приведено ранее. В течение всего наблюдаемого периода (январь 2014 г. — декабрь 2024 г.) в топ-3 наиболее часто употребляемых слов в рамках коммуникаций Банка России попали: *инфляция, банк и ставка*<sup>19</sup>. При этом необходимо отметить, что слово «инфляция» было на первом месте по частотности использования с 2014 г. по 2018 г. включительно, далее — в 2021 г. В периодах 2019–2020 гг., 2022–2024 гг. лидировало слово «ставка»<sup>20</sup>.

Как отмечалось ранее, данные результаты являются отчасти предсказуемыми и интуитивно понятными в силу того, что если исследуются инфляционные процессы, то нужно в первую очередь использовать корневое слово «инфляция». Такой логики придерживаются большинство авторов работ по данному направлению (Приложение 1).

При использовании в качестве дополнительного фильтра сервиса Яндекса, позволяющего оценить абсолютное число запросов, становится понятным, что далеко не все производные и очень близкие словоформы к «инфляции» выступают релевантными в рамках подобных исследований (в том числе с точки зрения относительных масштабов самих пользовательских запросов). Так, если среднемесячное количество запросов через поисковую систему Яндекс по ключевому слову «инфляция» в 2024 г. в рамках всей страны составило 747 544, то по словоформе «инфляционный» — только 27 652, т.е. в 27 раз меньше. Это отсылка к тому, что множественный перебор различных словосочетаний с корневым словом, о котором говорилось ранее, не всегда может выступать лучшей стратегией отбора ключевых запросов.

На втором шаге был произведён анализ ключевых слов на той же базе текстовых коммуникаций Банка России («База для извлечения смыслов») с использованием избранных NLP моделей (на основе T5 архитектуры)<sup>21</sup>.

Необходимо отметить, что без предварительного дообучения на собранной новостной базе («База для обучения моделей машинного обучения») и использования моделей «из коробки»<sup>22</sup> в топ-3 ключевых слов массива коммуникаций Банка России вошли:

- ▶ финансы;
- ▶ аналитика;
- ▶ аналитика рынка.

Указанные термины имеют весьма широкий характер и очевидно не могут быть использованы для дальнейшей квантификации инфляционных ожиданий. При таком подходе слово «инфляция» оказалось на четвёртом месте. Это ещё раз подчёркивает необходимость и важность дообучения моделей под тематическую сферу применения.

<sup>19</sup> В составе наиболее употребляемых словосочетаний: Банк России/Центральный банк и ключевая ставка.

<sup>20</sup> В составе наиболее употребляемого словосочетания: ключевая ставка.

<sup>21</sup> NLP-анализ был выполнен с помощью языка программирования Python.

Отдельные параметры дообучения избранных моделей:

- `optimizer = optim.AdamW(model.parameters(), lr=1e-5)`
- `lr_scheduler = optim.lr_scheduler.MultiStepLR(optimizer=optimizer, milestones=[3, 6], gamma=0.1, verbose=True)`
- `batch_size = 10`
- `report_steps = 500`
- `NUM_EPOCH = 10`

Дообучение моделей с помощью GPU: Tesla T4

<sup>22</sup> В качестве базовых, необученных моделей («из коробки») использовались образцы со специализированного ресурса Hugging Face. URL: <https://huggingface.co/> (дата обращения: 02.02.2025)

После дообучения используемые модели (ruT5-base и keyT5-base) продемонстрировали относительно схожие с точки зрения ключевых слов — лидеров коммуникаций регулятора результаты. При этом возникла дополнительная необходимость объединения синонимичных понятий, например: ЦБ, Центробанк, Центральный Банк, Банк России, которые в виде различающихся ключевых слов были выделены моделями.

Отдельно стоит обратить внимание на то, что в рамках полученных результатов используемых моделей в перечень наиболее частотных ключевых слов вошли также такие термины, как: ФРС, ОПЕК, Минфин, пенсионная система и ряд других. Важно понимать, что Банк России в своих коммуникациях данные понятия напрямую не упоминает, хотя, безусловно, анализирует и учитывает их при принятии решений. По всей видимости, размеченные новости на финансово-экономическую тематику, даже относящиеся к денежно-кредитной политике и деятельности Банка России в целом, имеют широкий спектр тегов, установленных СМИ (для больших охватов пользователей поисковых систем), поэтому дообучение моделей на собранном массиве привело к неожиданным эффектам.

После очистки финальных результатов от выявленных «выбросов» и объединения синонимичных терминов к топ-3 ключевых слов отнесены следующие группы:

- ▶ Центральный банк (включая поисковые запросы: ЦБ, Банк России, Центробанк, Центральный банк);
- ▶ инфляция;
- ▶ курс (курс доллара, курс рубля, курс валюты, курс евро)<sup>23</sup>.

Как мы видим, Baseline-подход и избранные NLP-модели имеют по два пересечения в рамках выделенных ключевых слов. Таким образом, было принято решение совместить полученные согласно этим подходам результаты и далее анализировать статистику поисковых запросов по следующим агрегированным блокам: инфляция, курс, ключевая ставка, Центральный банк. Представляется, что в своих коммуникациях Банк России на системной основе делал акцент на данных группах слов, которые в конечном счёте могут повлиять на состояние инфляционных ожиданий населения России.

Для сравнения: большая языковая модель (YandexGPT) суммировала несколько последних (сентябрь, октябрь, декабрь) заседаний Совета директоров Банка России в 2024 г. до следующих ключевых слов: инфляция, ключевая ставка, денежно-кредитные условия. В целом результаты соответствуют полученным нами в рамках исследования выводам на более длительной выборке (2014–2024 гг.), что в том числе может свидетельствовать об относительной устойчивости и стабильности информационного сигнала регулятора.

В следующей части работы произведён анализ поисковых запросов по выделенным группам слов на базе данных российской поисковой системы Яндекс, которая, как было отмечено выше, на текущий момент обладает рядом преимуществ для исследовательских целей, в том числе за счёт расширения доступного периода наблюдения.

### **Анализ поисковых запросов на основе выявленных ключевых слов, характеризующих инфляционные ожидания населения России**

К самым популярным запросам среди российских пользователей Яндекс на протяжении 2018–2024 гг. из выделенных блоков были запросы по теме «курс», следом идут запросы по теме «Центральный банк». За 5 лет интерес вырос по всем тематикам, пиковый интерес наблюдался в шоковом 2022 г. — 370,3 млн шт. В целом за 6 лет число запросов по данным тематикам выросло на 75,7% (табл. 1).

<sup>23</sup> Словосочетания «валютный рынок» и «валютный курс» были исключены из перечня ключевых слов, хотя и выделены моделями, после дополнительного фильтра в «Яндекс. Подбор слов» было обнаружено относительно незначительное количество запросов ТАКИХ СЛОВСОЧЕТАНИЙ по сравнению с включёнными, синонимичными понятиями (аналогично словоформе «инфляционный»).

Таблица 1

Статистика по поисковым запросам Яндекс (абсолютные значения)

Млн запросов в год	2018	2019	2020	2021	2022	2023	2024	Темпы прироста 2024/2018, %
Курс	102,2	85,4	138,2	109,6	301,4	189,8	147,7	44,4
г/г, %	—	-16,5	61,8	-20,7	175,0	-37,0	-22,2	
Центральный банк	23,3	20,7	27,4	28,3	53,9	58,5	65,0	178,7
г/г, %	—	-11,3	32,7	3,1	90,6	8,5	11,1	
Инфляция	3,6	4,0	6,4	7,7	9,8	8,1	9,0	150,3
г/г, %	—	10,8	61,6	19,3	28,2	-17,5	10,9	
Ключевая ставка	1,2	1,2	1,8	2,7	5,2	4,6	7,3	529,9
г/г, %	—	0,3	54,9	49,3	95,4	-12,3	58,6	
Всего	130,3	111,2	173,8	148,2	370,3	261,0	228,9	75,7
г/г, %	—	-14,6	56,3	-14,7	149,9	-29,5	-12,3	

Источник: Яндекс. Вордстат (Wordstat). URL: <https://wordstat.yandex.ru/>, расчёты автора.

Аналогичная динамика просматривается и с точки зрения нормированных значений поисковых запросов в системе Яндекс. Анализ таких данных позволяет исключить выводы о росте абсолютного количества запросов только за счёт роста пользователей поисковой системы (табл. 2).

Таблица 2

Статистика по поисковым запросам Яндекс (относительные значения<sup>24</sup>)

Средняя ежемесячная доля запросов	2018	2019	2020	2021	2022	2023	2024	Темпы прироста 2024/2018, %
Курс	0,1178	0,0926	0,1231	0,0900	0,2246	0,1499	0,1218	3,4
г/г, %	—	-21,4	32,9	-26,9	149,5	-33,2	-18,7	
Центральный банк	0,0266	0,0221	0,0240	0,0230	0,0396	0,0458	0,0539	103
г/г, %	—	-16,7	8,6	-4,3	72,2	15,8	17,7	
Инфляция	0,0041	0,0043	0,0055	0,0063	0,0073	0,0063	0,0073	79,5
г/г, %	—	5,1	29,6	12,9	16,0	-13,5	16,3	
Ключевая ставка	0,0013	0,0013	0,0016	0,0022	0,0040	0,0037	0,0061	359
г/г, %	—	-3,9	27,8	36,8	78,3	-7,4	65,5	
Всего	0,1498	0,1202	0,1543	0,1215	0,2754	0,2057	0,1891	26,3
г/г, %	—	-19,7	28,3	-21,3	126,7	-25,3	-8,1	

Источник: «Яндекс. Подбор слов» (Wordstat), расчёты автора.

<sup>24</sup> Абсолютные значения нормируются на количество показов результатов поиска Яндекса за соответствующий месяц.

Далее приведена визуализация динамики поисковых запросов отдельно по каждому блоку ключевых слов.

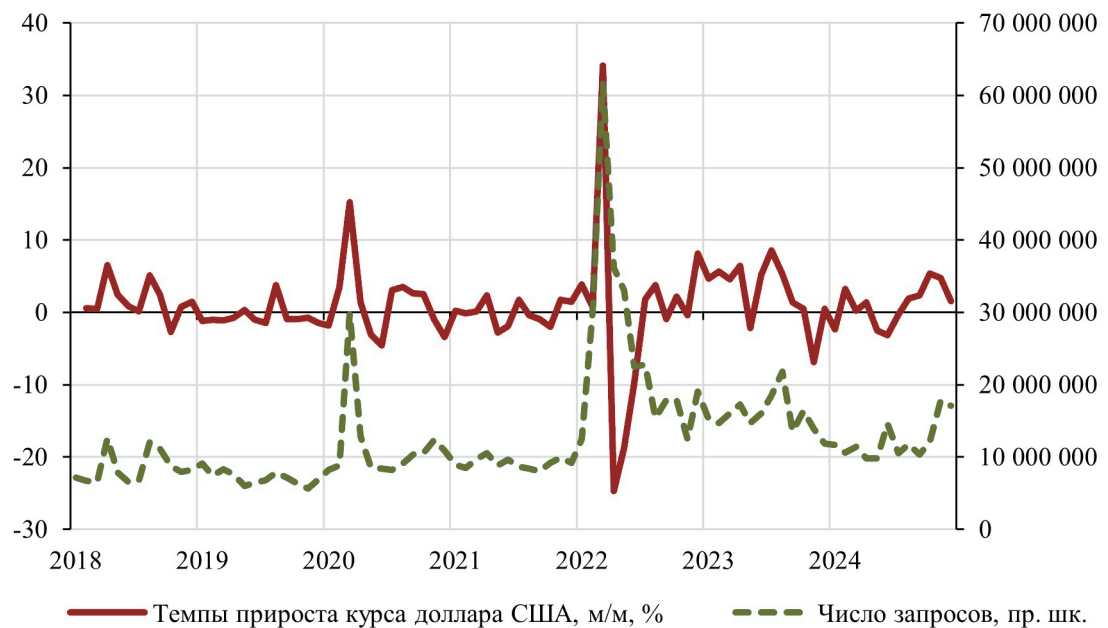


Рис. 4. Запросы по теме «Курс»  
 Источник: Банк России, «Яндекс. Подбор слов (Wordstat)», расчёты автора.

Запросы по теме «Курс» являются самыми популярными среди российских пользователей поисковой системы Яндекс. Всплески интереса наблюдаются в месяцы высоких темпов прироста (убыли) курса доллара США (март 2020, март — июнь 2022, рис. 4).



Рис. 5. Запросы по теме «Центральный банк» (абсолютные значения)  
 Источник: «Яндекс. Подбор слов» (Wordstat)», расчёты автора.

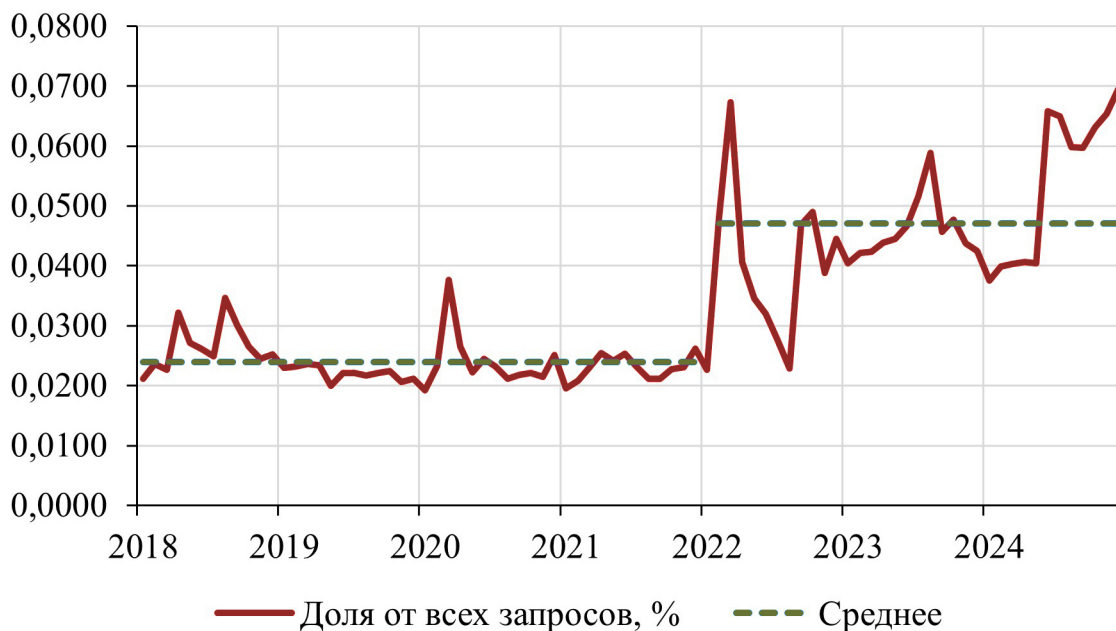


Рис. 6. Запросы по теме «Центральный банк» (относительные значения)  
 Источник: «Яндекс. Подбор слов (Wordstat)», расчёты автора.

В 2022–2024 гг. среднемесячное число запросов на тему «Центральный банк» выросло в 2,4 раза (при этом в 2 раза выросла доля запросов по данной теме) по сравнению с 2018–2021 гг. (рис. 5 и 6). Это может указывать на некоторый сдвиг («level shift») в уровне интереса российских пользователей по данной теме. С одной стороны, это может быть следствием информационно-коммуникационной политики Банка России по ДКП. С другой стороны, отражением органического роста и сохранения интереса к данной тематике после событий 2022 г.

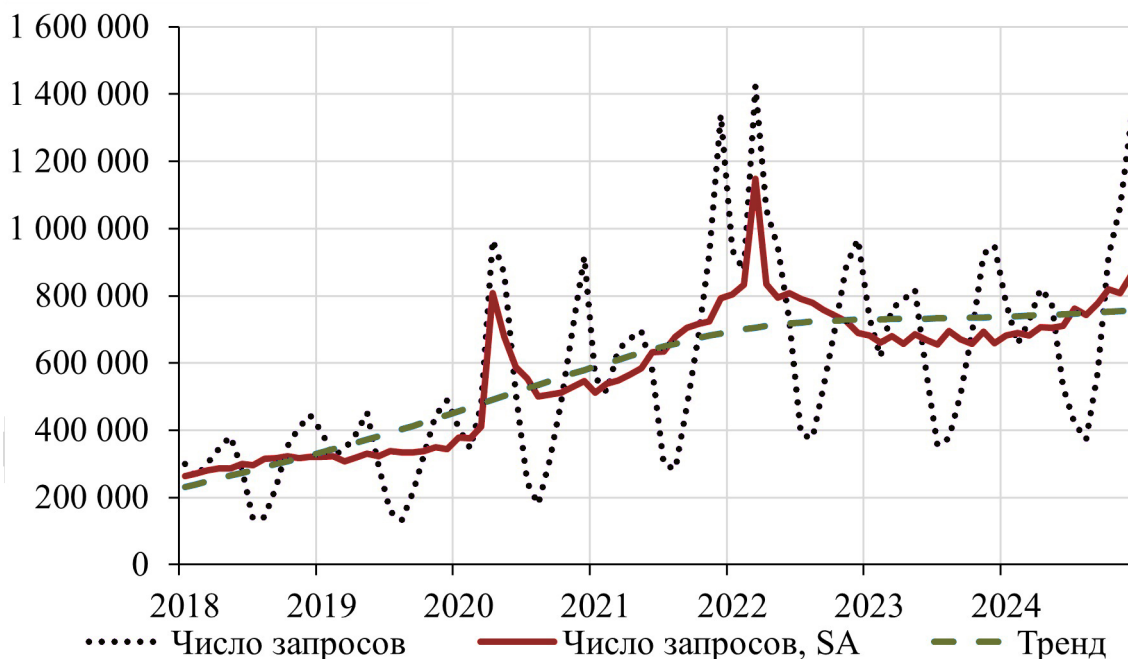


Рис. 7. Запросы по теме «Инфляция» (абсолютные значения)  
 Источник: «Яндекс. Подбор слов (Wordstat)», расчёты автора.

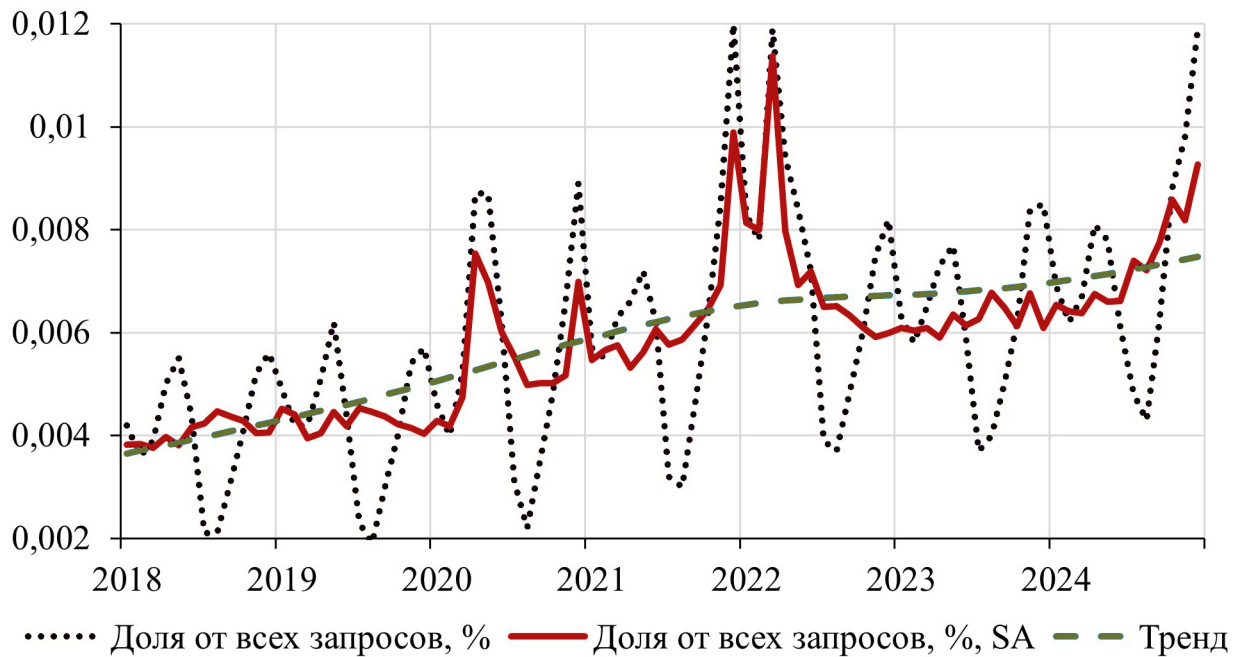


Рис. 8. Запросы по теме «Инфляция» (относительные значения)

Источник: «Яндекс. Подбор слов (Wordstat)», расчёты автора.

В целом интерес (тренд по запросам НР фильтр<sup>25</sup>) к теме инфляции стабильно рос на протяжении 2018–2021 гг., в 2022–2024 гг. он достиг своих пиков и остаётся примерно на этом уровне (рис. 7–8). Возрастающая статистика по поисковым запросам по теме «инфляция» может свидетельствовать об увеличении озабоченности населения вопросами общего роста цен. Это в конечном счёте может выступать проекцией высоких инфляционных ожиданий.

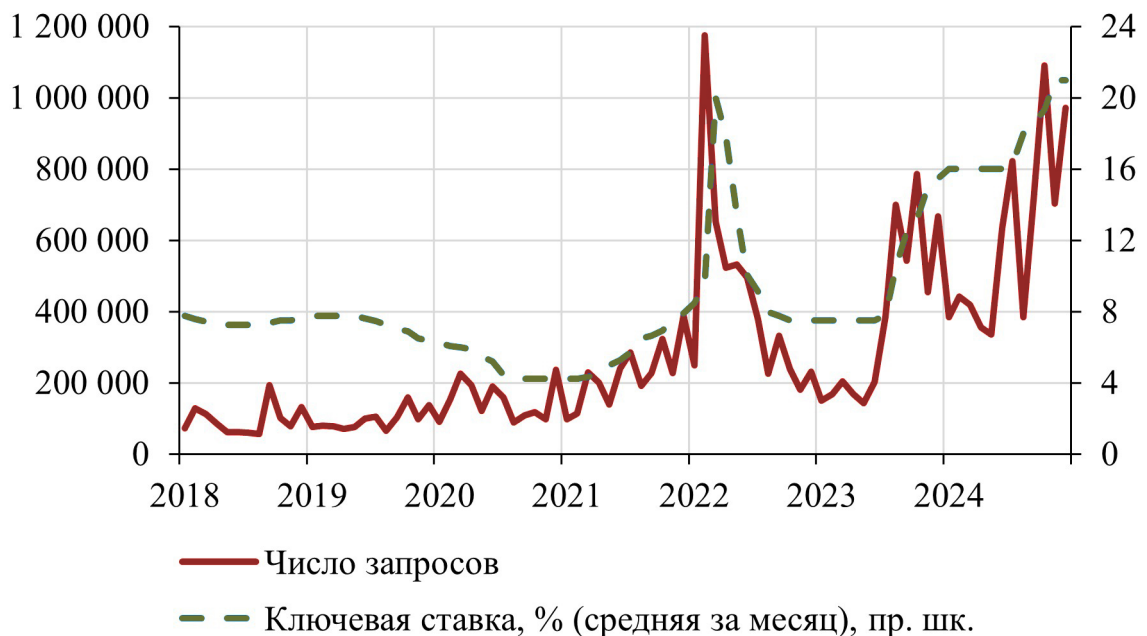


Рис. 9. Запросы по теме «Ключевая ставка»

Источник: «Яндекс. Подбор слов (Wordstat)», расчёты автора

<sup>25</sup> Нами использован фильтр Ходрика-Прескотта (НР), предназначенный для сглаживания временного ряда (устранения циклической компоненты и выделения именно трендовой составляющей)



По большей части всплески запросов по ключевой ставке происходят в периоды её резкого повышения (2022–2024 гг.). Вероятно, пользователи пытаются узнать новости о её изменении и причинах этого, вводя соответствующие поисковые запросы. В остальные периоды они в целом имеют относительно слабый интерес к теме. К тому же часть запросов по ключевой ставке может отражаться в динамике запросов по теме «Центральный банк». Отделить данные эффекты на этом этапе является проблематичным.

### **Апробация прокси-показателей инфляционных ожиданий в рамках моделей прогнозирования инфляции в России**

Для апробации полученных результатов по поисковой статистике выделенных ключевых слов в качестве прокси-показателей в рамках прогнозирования инфляции использовались модели ARIMAX<sup>26</sup>, где выбор структуры осуществлялся автоматически по информационному критерию Акаике (AIC). В каждой из моделей в качестве объясняемой переменной использовались значения сезонно сглаженных месячных темпов инфляции (SA)<sup>27</sup>, а объясняющей переменной — индикаторы запросов по ключевым словам, а также показатели инфляционных ожиданий на основе опросов «инФОМ»<sup>28</sup>, среднемесячный курс доллара США<sup>29</sup>. В качестве альтернативной модели (бенчмарка) была взята ARIMA-модель с автоматическим выбором ARIMA-структуры по информационному критерию Акаике (AIC) без дополнительных объясняющих переменных.

Модели рассматривались как без временного сдвига, так и со сдвигами на один месяц. Другими словами, если в модели использовались данные без сдвига, то для прогнозирования инфляции в текущем периоде применялись значения объясняющей переменной за тот же период. В свою очередь, если использовался сдвиг на один месяц, то для прогнозирования инфляции применялись данные по объясняющим переменным за предыдущий период.

Для оценки прогнозной точности моделей использовались месячные данные с января 2018 г. по декабрь 2024 г.: обучающая выборка (train) с января 2018 г. по декабрь 2023 г., тестовая (осуществлялись вневыборочные прогнозы на один месяц вперёд) с января по декабрь 2024 г. (12 прогнозов). При необходимости временные ряды были приведены к стационарному виду путём соответствующих преобразований. Прогнозировались темпы инфляции по России (SA) на месяц вперёд и сравнивались с фактическими значениями.

В качестве метрики для сопоставления построенных моделей использовался показатель RMSE (root mean square error, среднеквадратическая ошибка). Данный показатель демонстрирует то, насколько хорошо предсказанные значения в рамках рассматриваемых моделей совпадают с фактическими наблюдаемыми значениями конкретного набора данных. Соответственно, чем ниже значение RMSE, тем лучше (точнее) та или иная модель способна предсказывать объясняемую переменную, т.е., в контексте данной работы — инфляцию. Исходя из данной логики, в табл. 3 отражены полученные результаты по ТОП-10 моделям.

Полученные данные свидетельствуют о том, что статистики по поисковым запросам выделенных в рамках данного исследования ключевых слов могут использоваться в качестве объясняющих переменных, направленных на минимизацию ошибок прогнозов в рамках моделей прогнозирования инфляции на уровне Российской Федерации. Как показано

<sup>26</sup> Всего было построено 26 моделей ARIMAX с различными вариантами объясняющей переменной

<sup>27</sup> База данных по сезонно сглаженным месячным темпам инфляции. Банк России. URL: <https://www.cbr.ru/statistics/ddkp/aipd/> (дата обращения: 24.02.2025)

<sup>28</sup> Банк России. URL: [https://www.cbr.ru/analytics/dkp/inflationary\\_expectations/#highlight=инфляционные%7Сожидания%7Синфляционных%7Сожиданий](https://www.cbr.ru/analytics/dkp/inflationary_expectations/#highlight=инфляционные%7Сожидания%7Синфляционных%7Сожиданий) (дата обращения: 24.02.2025)

<sup>29</sup> Банк России. URL: [https://www.cbr.ru/currency\\_base/dynamics/](https://www.cbr.ru/currency_base/dynamics/) (дата обращения: 24.02.2025)

Таблица 3

**Выбор лучшей модели при использовании данных поисковых запросов по выделенным ключевым словам**

№	Тип модели	Объясняющая переменная	Преобразование	Временной сдвиг	RMSE
1	ARIMAX	Статистика по запросам по всем выделенным группам ключевых слов (абсолютные значения)	dlog	1 месяц	0,249
2	ARIMAX	Статистика по запросам «Инфляция» (относительные значения)	—	1 месяц	0,254
3	ARIMAX	Статистика по запросам по всем выделенным группам ключевых слов (относительные значения)	dlog	1 месяц	0,257
4	ARIMAX	Статистика по запросам «Инфляция» (абсолютные значения)	—	1 месяц	0,259
5	ARIMAX	Статистика по запросам «Центральный банк» (абсолютные значения)	—	—	0,271
6	ARIMAX	Статистика по запросам «Центральный банк» (абсолютные значения)	—	1 месяц	0,271
7	ARIMAX	Статистика по запросам «Курс» (абсолютные значения)	dlog	1 месяц	0,272
8	ARIMAX	Статистика по запросам «Курс» (относительные значения)	dlog	1 месяц	0,277
9	ARIMAX	Значения наблюдаемой инфляции по данным «инФОМ»	—	—	0,290
10	ARIMAX	Значения наблюдаемой инфляции по данным «инФОМ»	—	1 месяц	0,298

Источник: расчёты автора.

выше, наиболее значимые с точки зрения прогнозирования инфляции результаты (относительно минимальные значение RMSE в рамках представленных выше моделей) дала агрегированная статистика по всем выделенным ключевым словам: «Центральный банк», «Курс», «Ключевая ставка», «Инфляция» (абсолютные и относительные значения), а также «Инфляция» (относительное и абсолютное значения).

## Выводы

Инфляционные ожидания являются важным параметром, учитываемым в рамках реализации такого режима денежно-кредитной политики, как таргетирование инфляции. На сегодняшний момент во всём мире основным направлением, способствующим сбору и формированию агрегированных данных об инфляционных ожиданиях населения, выступает проведение специализированных опросов. В России подобные исследования проводятся с 2009 г. Данный подход обладает как рядом преимуществ, так и недостатков, одним из которых является дороговизна получения подобных данных (особенно детализированных), а также отсутствие возможности их обновления «в моменте». Ввиду этого, а также с учётом развития современных информационных технологий, в последнее десятилетие

происходит активное тестирование альтернативных способов замеров инфляционных ожиданий. Одним из них выступает определение инфляционных ожиданий населения на основе поисковых запросов пользователей в сети Интернет. При этом остаётся открытым вопрос отбора ключевых слов для квантификации ожиданий населения.

Таким образом, в рамках данной работы был применён экспериментальный подход на основе NLP-анализа коммуникаций Банка России в области ДКП для выявления ключевых слов, потенциально формирующих инфляционные ожидания населения России. В работе заложены концептуальные основы обоснования отбора ключевых слов, поисковая статистика которых может использоваться в рамках квантификации инфляционных ожиданий, их дальнейшего отслеживания и использования в рамках анализа инфляционных процессов, прогнозирования и наукастинга. В отличие от большинства других исследований в данной области отбор осуществлялся через призму проводимых регулятором коммуникаций, которые являются одной из важнейших составляющих режима таргетирования инфляции.

Предлагаемый подход отражает логику «сверху-вниз», когда в основу положены ключевые слова, через которые регулятор доносит необходимую информацию в рамках проводимой коммуникационной политики до населения страны [Щербаков, Харламова, Яковина, 2022]. Безусловно, нельзя утверждать, что данный список выделенных групп слов является исчерпывающим ввиду того, что часть населения может использовать ряд других слов-маркеров, отражающих инфляционные настроения в обществе (подход «снизу-вверх»). Такой анализ возможен, например, на основе больших данных социальных сетей [Shcherbakov, Karov, 2024]. Поэтому отдельным блоком можно выделить исследование пересекающихся множеств ключевых слов в рамках данных подходов, дальнейшая работа по их сближению.

Статистика по поисковым запросам выделенных групп слов («инфляция», «курс», «Центральный банк», «ключевая ставка») позволяет получить дополнительную информацию, характеризующую инфляционные процессы, в том числе с точки зрения поведенческих характеристик пользователей поисковых систем. Отслеживание динамики запросов по группам «инфляция» и «Центральный банк» дают оперативную информацию повсеместно, а по группам «курс», «ключевая ставка» — в кризисных и/или изменяющихся экономических условиях. Ввиду последних изменений в политике доступности исторических данных, а также популярности поисковой сети среди жителей России, особый акцент сделан на данных системы Яндекс.

Как было показано в разделе, посвящённом апробации использования данных показателей для прогнозирования месячной инфляции с помощью базовых моделей семейства ARIMAX, такие индикаторы могут повысить точность прогноза. При этом мы не призываем к отказу от использования опросных данных по инфляционным ожиданиям. Посыл работы состоит в том, чтобы использовать их совместно с альтернативными, более быстрыми с точки зрения появления данных, показателями, что должно привести синергетический эффект в рамках изучения инфляционных процессов.

Вместе с тем разработка данной тематики показала, что есть потенциал дальнейшего исследовательского развития, в том числе по таким направлениям:

- ▶ Дальнейшие эксперименты с NLP-моделями, а также LLM-моделями.
- ▶ Поиск дополнительных и/или создание собственных баз с размеченными данными для дообучения моделей.
- ▶ Дальнейшее формирование агрегированных показателей ключевых слов на основе выделенных групп. Во-первых, для сопоставления полученных показателей с замерами инфляционных ожиданий на основе проводимых «инФОМ» опросов по заказу Банка России. Во-вторых, для тестирования в более сложных эконометрических моделях, направленных на краткосрочное прогнозирование и наукастинг инфляции, в том числе в субфедеральном и региональном разрезах.

## ПРИЛОЖЕНИЕ

№	Исследование	Источник статистики запросов	Период	Поисковый образ	Локация
1	Винокуров, Медведь, 2023	Google	2016–2021	«Инфляция», «экономика»	Россия
2	Петрова, 2022	Google	2015–2021	«Инфляция», «рост цен», «цены выросли», «цены растут», «повышение цен», «цена на газ», «цена на бензин», «цены на продукты», «индекс потребительских цен», «индекс цен», «ипц».	Россия
3	Юревич, 2021	Google	2011–2021	«Инфляция»	Россия
4	Sahu, Chattopadhyay, 2020	Google	2006–2018	Словосочетания с «инфляция» и «цена»	Индия
5	Niesert et al., 2020	Google	2004–2016	20 экономических категорий (тем)	США, Великобритания, Канада, Германия и Япония
6	Ержан, 2019	Google	2016–2019	Словосочетания с «инфляция» и «цена»	Казахстан
7	Петрова, 2019	Google	2004–2019	75 слов и словосочетаний по экономической тематике	Россия
8	Bichhal, Raja, 2019	Google	2006–2018	«Инфляция», «рост цен» и «цены на топливо»	Индия
9	Hassani, Silva, 2018	Google	2006–2018	«Инфляция»	Великобритания
10	Wei et al., 2017	Google	2012–2017	«Инфляция», ИПЦ	13 стран с крупной экономикой, включая Россию
11	Seabold, Coppola, 2015	Google	2004–2014	22 словосочетания, включая «инфляция», «цена», «заработная плата»	Гондурас, Коста-Рика, Сальвадор
12	Li et al., 2015	Baidu, Google	2004–2012	Набор словосочетаний со словом цена	Китай
13	Zhang et al., 2012	Google	2004–2009	44 словосочетания, объединенные в категории «макроэкономика» и «спрос–предложение»	Китай
14	Guzman, 2011	Google	2004–2008	«Инфляция»	США

Источник: адаптировано автором на основе [Юревич, 2021].

## ЛИТЕРАТУРА / REFERENCES

- Банк России (2024). Основные направления единой государственной денежно-кредитной политики на 2025 год и период 2026 и 2027 годов. [Bank of Russia (2024). Monetary policy guidelines for 2025-2027]. URL: [https://cbr.ru/Content/Document/File/164702/on\\_2025\(2026-2027\).pdf](https://cbr.ru/Content/Document/File/164702/on_2025(2026-2027).pdf) (дата обращения: 30.06.2024).
- Винокуров С.С., Медведев А.А. (2023). Инфляция, информационный фон и ожидания домохозяйств [Vinokurov S.S., Medved A.A. (2023). Inflation, informational environment and expectations of households] // *Статистика и экономика*. Т. 20, № 1. С. 37–52. 10.21686/2500-3925-2023-1-37-52.
- Голощапова И.О., Андреев М.Л. (2017). Оценка инфляционных ожиданий российского населения методами машинного обучения [Goloshchayeva I.O., Andreev M.L. (2017). Measuring inflation expectations of the Russian population with the help of machine learning] // *Вопросы экономики*. № 6. С. 71–93.
- Дауит Д.М., Кемалов М.М., Джаксылыкова А.Б. (2020). Обзор различных методов обобщения текста [Dauit D., Ketalov M., Jaxylykova A. (2020). Overview of the different text summarization methods] // *Вестник Казхстанско-Британского технического университета*. Т. 17. № 2. С. 163–168.
- Евстигнеева А. (2023). Коммуникация как инструмент денежно-кредитной политики [Evstigneeva A. (2023). Communication as a monetary policy tool] // *Аналитическая записка*. Банк России. С. 1–33. URL: [https://cbr.ru/StaticHtml/File/146496/research\\_policy\\_notes\\_b\\_4\\_1.pdf](https://cbr.ru/StaticHtml/File/146496/research_policy_notes_b_4_1.pdf) (дата обращения: 27.01.2025)
- Евстигнеева А., Карпов Д. (2023). Влияние негативных новостей на восприятие инфляции населением [Evstigneeva A., Karpov D. (2023). The impact of negative news on the perception of inflation by the population] // *Серия докладов об экономических исследованиях*. Банк России. №111. С 1–33. URL: [https://cbr.ru/Content/Document/File/144918/wp\\_111.pdf](https://cbr.ru/Content/Document/File/144918/wp_111.pdf) (дата обращения: 20.02.2025)
- Ержан И.С. (2019). Апробация альтернативных методов оценки инфляционных ожиданий в Казахстане [Yerzhan I.S. (2019). Approbation of alternative methods for assessing inflation expectations in Kazakhstan.] // *Экономическое обозрение*. Национальный Банк Республики Казахстан. № 2–3. С. 4–15.
- Жемков М.И., Кузнецова О.С. (2017). Измерение инфляционных ожиданий участников финансового рынка в России [Zhetkov M., Kuznetsova O. (2017). Measuring inflation expectations in Russia using stock market data] // *Вопросы экономики*. № 10. С. 111–122.
- Перевышин Ю.Н., Рыкалин А.С. (2018). Моделирование инфляционных ожиданий в российской экономике [Perevyshin Y., Rykalin A. (2018). Modeling Inflation Expectations in the Russian Economy] // URL: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3149565](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3149565) (дата обращения: 20.09.2024)
- Петрова Д.А. (2019). Прогнозирование инфляции на основе интернет-запросов [Petrova D.A. (2019). Inflation forecasting based on Internet search queries] // *Экономическое развитие России*. Т. 26. № 11. С. 55–62.
- Петрова Д.А. (2022). Оценка инфляционных ожиданий на основе интернет-данных [Petrova D. (2022). Assessment of inflation expectations based on Internet data] // *Прикладная эконометрика*. Т. 66. № 2. С. 25–38.
- Федюнина А.А., Юревич М.А., Городный Н.А. (2024). Пандемия, санкции и беспокойство в регионах России: наукастинг ожиданий деловой активности [Fedyunina A.A., Yurevich M.A., Gorodny N.A. (2024). Pandemic, sanctions and anxiety in Russia's regions: business expectations nowcasting] // *Вопросы экономики*. № 3. С. 96–119.
- Хазанов А.А. (2015). О квантификации инфляционных ожиданий Банком России [Khazanov A. (2015). On quantification of inflation expectations by the Bank of Russia] // *Деньги и кредит*. № 3. С. 59–63.
- Шуляк Е. (2022). Макроэкономическое прогнозирование с использованием данных социальных сетей [Shulyak E. (2022). Macroeconomic forecasting using data from social media] // *Деньги и кредит*. Т. 81. № 4. С. 86–112.
- Щербаков В.С., Харламова М.С., Гартвич Р.Е. (2022). Методы и модели наукастинга экономических показателей с помощью поисковых запросов [Shcherbakov V.S., Kharlamova M.S., Gartvich R.E. (2022). Methods and models for nowcasting economic indicators with help of search queries] // *Мат-лы Межрегиональной научно-практ. онлайн-конф. «Развитие экономики регионов: пространственная трансформация, глобальные вызовы и перспективы экономического роста»*. — Красноярск. С. 117–127.
- Щербаков В.С., Харламова М.С., Яковина М.Ю. (2022). Статистика поисковых запросов как прокси-показатель региональной ценовой динамики [Shcherbakov V.S., Kharlamovs M.S., Yakovina M.Yu. (2022). Search query statistics as a proxy indicator of regional price dynamics] // *Креативная экономика*. Т. 16. №11. С. 4475–4490.
- Юревич М.А. (2021). Инфляционные ожидания и инфляция: наукастинг и прогнозирование [Yurevich M.A. (2021). Inflation expectations and inflation: nowcasting and forecasting] // *Journal of Economic Regulation*. Т. 12. № 2. С. 22–35.
- Angelico C., Marcucci J., Miccoli M., Quarta F. (2022). Can we measure inflation expectations using Twitter? // *Journal of Econometrics*. Vol. 228. No. 2. Pp. 259–277.
- Aromí D., Llada M. (2020). Forecasting inflation with twitter // *Asociación Argentina de Economía Política*. Working Papers. No. 4308.
- Ay B., Ertan F., Fidan G., Aydin G. (2023). Turkish abstractive text document summarization using text to text transfer transformer // *Alexandria Engineering Journal*. No. 68. Pp. 1-13. <https://doi.org/10.1016/j.aej.2023.01.008>.
- Banbura M., Leiva-Leon D., Menz J-O. (2021). Do Inflation Expectations Improve Model-based Inflation Forecasts? // *Banco de Espana Working Paper*. No. 2138.

- Batchelor R. (2009). How Robust are Quantified Survey Data? Evidence from the United States // *Inflation Expectations* / Ed. by P. Sinclair. — Routledge. Pp. 8-33.
- Bernanke B. (2007). Inflation Expectations and Inflation Forecasting // *Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute* / Cambridge, Massachusetts.
- Bicchal M., Raja Sethu Durai S. (2019). Rationality of inflation expectations: an interpretation of Google Trends data // *Macroeconomics and Finance in Emerging Market Economies*. Vol. 12. No. 3. Pp. 229–239.
- Carlson J.A., Parkin M. (1975). Inflation Expectations // *Economica*. Vol. 42. No. 166. Pp. 123–138.
- Carroll C.D. (2003). Macroeconomic expectations of households and professional forecasters // *The Quarterly Journal of Economics*. Vol. 118. No. 1. Pp. 269–298.
- Choi H., Varian H. (2009). Predicting initial claims for unemployment benefits. URL: <http://static.googleusercontent.com/media/research.google.com/en/us/archive/papers/initialclaimsUS.pdf> (access date: 15.11.2024).
- Evtigneeva A., Sidorovskiy M. (2021). Assessment of Clarity of Bank of Russia Monetary Policy Communication by Neural Network Approach // *Russian Journal of Money and Finance*. Vol. 80. No. 3. Pp. 3–33. DOI: 10.31477/rjmf.202103.03.
- Fuhrer J. (2012). The Role of Expectations in Inflation Dynamics // *International Journal of Central Banking*. No. 8. Pp. 137–165.
- Guan B., Zhu X., Yuan Sh. (2024). A T5-based interpretable reading comprehension model with more accurate evidence training // *Information Processing & Management*. Vol. 61. No. 2. DOI: 10.1016/J.IPM.2023.103584.
- Guzman G. (2011). Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations // *Journal of Economic and Social Measurement*. Vol. 36. No. 3. DOI:10.3233/JEM-2011-0342.
- Han X., Zhang Zh., Ding N., Gu Y. (2021). Pre-trained models: Past, present, and future // *AI Open*. No. 2. Pp. 225–250. DOI: 10.1016/J.AIOPEN.2021.08.002.
- Hassani H., Silva E.S. (2018). Forecasting UK consumer price inflation using inflation forecasts // *Research in Economics*. Vol. 72. No. 3. Pp. 367–378.
- Larsen V. H., Thorsrud L. A., Zhulanova J. (2021). News-driven inflation expectations and information rigidities // *Journal of Monetary Economics*. No. 117. Pp. 507–520. DOI: 10.1016/j.jmoneco.2020.03.004.
- Li X., Shang W., Wang S., Ma J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating Google search data // *Electronic Commerce Research and Applications*. Vol. 14. No. 2. Pp. 112–125.
- Mankiw G., Reis R., Wolfers J. (2003). Disagreement about inflation expectations // *NBER Macroeconomic Annual*. No. 18. Pp. 209–248.
- Niesert R.E., Oorschot J.A., Veldhuisen C.P., Brons K., Lange R-J. (2020). Can Google search data help predict macroeconomic series? // *International Journal of Forecasting*. Vol. 36. No. 3. Pp. 1163–1172.
- Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer // *Journal of Machine Learning Research*. No. 21. Pp. 1–67.
- Sahu S., Chattopadhyay S. (2020). Epidemiology of inflation expectations and internet search: an analysis for India // *Journal of Economic Interaction and Coordination*. No. 15. Pp. 649–671.
- Seabold S., Coppola A. (2015). Nowcasting Prices Using Google Trends: An Application to Central America // *World Bank Policy Research Working Paper*. No. 7398. Pp. 1–40.
- Shcherbakov V.S., Karpov I.A. (2024). Regional Inflation Analysis Using Social Network Data // *Economy of regions*. Vol. 20. No. 3. Pp. 930–946. DOI: 10.17059/EKON.REG.2024-3-21
- Thrun S., Pratt L. (1998). Learning to learn: Introduction and overview // *Springer Science & Business Media*.
- Vaswani A., Shazeer N.M., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. (2017). Attention is All you Need // *ArXiv*, abs/1706.03762.
- Wang M., Xie P., Du Y., Hu X. (2023). T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions // *Applied Sciences*. Vol. 13. No. 12. Pp. 1–16. DOI: 10.3390/AP13127111-
- Wei Y., Zhang X., Wang S. (2017). Can search data help forecast inflation? Evidence from a 13-country panel // 2017 IEEE International Conference on Big Data (Big Data). DOI: 10.1109/BigData.2017.8258442
- Yadav D., Desai J., Yadav A.K. (2022). *Automatic Text Summarization Methods: A comprehensive Review* // *arXiv:2204.01849*.
- Zhang C., Lv B., Peng G., Liu Y., Yuan Q. (2012). A study on correlation between web search data and CPI // *In Recent Progress in Data Engineering and Internet Technology*. Pp. 269–274.
- Zmitrovich D., Abramov A., Kalmykov A., Tikhonova M., Taktasheva E., Astafurov D., Baushenko M., Snegirev A., Shavrina T., Markov S., Mikhailov V., Fenogenova A. (2023). *A Family of Pretrained Transformer Language Models for Russian*. DOI: 10.48550/arXiv.2309.10931.

**Щербаков Василий Сергеевич**

*shcherbakovvs@mail.ru*

**Vasilii Shcherbakov**

*shcherbakovvs@mail.ru*

*PhD (Economy), head of economic department, the Ural Main Branch of the Central Bank of the Russian Federation; associate professor, Faculty of Economics, Psychology, Management, Dostoevsky Omsk State University.*

#### **ASSESSMENT OF INFLATION EXPECTATIONS OF THE RUSSIAN POPULATION BASED ON INTERNET SEARCH QUERIES (TOP-DOWN APPROACH)**

**Abstract.** By their nature, inflation expectations are an unobservable variable. In the framework of economic theory and practice, proxy indicators of inflation expectations (mainly based on surveys) are used as the most important variables for analyzing and forecasting inflationary processes. At the same time, when implementing the inflation targeting regime, regulators primarily focus on managing inflation expectations through monetary policy communications. In this respect, their special, dual character is manifested. Today, the use of alternative estimates of inflation expectations, including search query statistics, continues to grow in popularity. The selection of keywords for quantifying the expectations of the population remains a conceptual issue. The purpose of the study is to develop a methodologically sound approach to selecting keywords for search queries, statistics on which can be used as proxy variables of inflationary expectations. Within the framework of the article, this goal is achieved on the basis of text analysis of communications of the Bank of Russia using machine learning models (especially NLP). Based on the frequency analysis (Baseline approach), as well as the use of advanced NLP models (the T5 family of models (“Text-to-Text Transfer Transformer”), four groups of keywords (“inflation”, “Central Bank”, “exchange rate”, “key rate”) were identified using which the regulator can shape the inflation expectations of the Russian population (top-down approach). Due to recent changes in the policy of accessibility of historical data, as well as the popularity of the search network among residents of Russia, special emphasis is placed on the data of the Yandex search network. It is assumed that tracking the dynamics of requests for the “inflation” and “Central Bank” groups provides operational information everywhere, and for the “exchange rate” and “key rate” groups — in crisis and/or changing economic conditions. The results obtained on the search statistics of the selected keywords were tested as proxy indicators in the framework of forecasting inflation at the level of the Russian Federation based on a set of ARIMAX family models. The results indicate that it is advisable to use keyword statistics as explanatory variables to minimize forecast errors within the framework of inflation forecasting models.

**Keywords:** *monetary policy, inflation targeting, inflation expectations, search queries, text analysis, machine learning methods, Yandex.*

**JEL:** C82, C88, E31, E52.