

Вестник Череповецкого государственного университета. 2023. № 5 (116). С. 75–86.
Cherepovets State University Bulletin, 2023, no. 5 (116), pp. 75–86.

Научная статья

УДК 004.896

<https://doi.org/10.23859/1994-0637-2023-5-116-6>

EDN: GTWOLD

Разработка интеллектуальной системы поиска, обработки, генерации контекстной информации

Алексей Александрович Суконщиков^{1✉}, Александр Николаевич Егоров²,
Анатолий Николаевич Швецов³

^{1, 2, 3}Вологодский государственный университет,
Вологда, Россия

¹avt@vogu35.ru, <https://orcid.org/0000-0001-8673-3314>

²avt@vogu35.ru

³smithv@mail.ru, <https://orcid.org/0000-0002-7021-5184>

Аннотация. В современном мире сайт – это уже давно не просто страницы в интернете, сейчас это важная составляющая любого бизнеса. Но чтобы на сайт заходили, он должен находиться на первых позициях в поисковой выдаче. Этого можно достичь при помощи контекстной рекламы. Чтобы она начала работать, ее нужно настроить, написать тексты, указать конечные страницы сайта, загрузить на площадку, ее нужно анализировать, вносить изменения в объявления и т. д. Сейчас это ручной процесс, но его можно автоматизировать с помощью нейронных сетей.

Ключевые слова: контекстная реклама, Яндекс, ключевые слова, нейронные сети, метод опорных векторов

Для цитирования: Суконщиков А. А., Егоров А. Н., Швецов А. Н. Разработка интеллектуальной системы поиска, обработки, генерации контекстной информации // Вестник Череповецкого государственного университета. 2023. № 5 (116). С. 75–86.
<https://doi.org/10.23859/1994-0637-2023-5-116-6>

Development of an intelligent system for searching, processing, generating context information

Aleksey A. Sukonshchikov^{1✉}, Aleksandr N. Egorov², Anatoly N. Shvetsov³
^{1, 2, 3}Vologda State University,

Vologda, Russia

¹avt@vogu35.ru, <https://orcid.org/0000-0001-8673-3314>

²avt@vogu35.ru

³smithv@mail.ru, <https://orcid.org/0000-0002-7021-5184>

Abstract. In the modern world, a site is no longer just pages on the Internet, now a site is an important component of any business. But for the site to be visited, it must be in the first three positions in the search results. This is where contextual advertising comes to the rescue. For it to start

working one needs to set up, write texts, specify the final pages of the site, upload it to the site, analyze it, make changes to ads, etc. Now it is a manual process, but it can be automated using neural networks.

Keywords: contextual advertising, Yandex, keywords, neural networks, Support Vector Machines

For citation: Sukonshchikov A. A., Egorov A. N., Shvetsov A. N. Development of an intelligent system for searching, processing, generating context information. *Cherepovets State University Bulletin*, 2023, no. 5 (116), pp. 75–86 (In Russ.). <https://doi.org/10.23859/1994-0637-2023-5-116-6>

Введение

В настоящее время у бизнеса есть множество инструментов для продажи своих товаров и услуг. Одним из таких инструментов является веб-сайт. С ростом количества интернет-ресурсов увеличилась конкуренция в среде Интернет, потому что основной поток пользователей приходит из поисковой выдачи. Наибольшее количество кликов и показов приходится на первые 3 позиции, а дальше, чем ниже позиция показа, тем ниже процент кликов, и уже на десятой позиции он близок к 0 %. То есть, чтобы привести на сайт наибольшее количество людей, требуется, чтобы сайт выдавался в поисковой выдаче на первом месте.

Существует два основных способа, чтобы «перевести» пользователя из поисковой выдачи на свой ресурс:

- 1) естественная поисковая выдача,
- 2) контекстная реклама.

Основная часть

Разработка эффективного бизнес-процесса по созданию и управлению контекстной рекламой – сложная задача, так как для предсказания результата требуется учесть множество факторов, таких как: проработка семантики, правильное определение целевой аудитории, сегментация, составление текстов, скорость работы сайта, мобильное отображение, репутация организации, конкуренция и мн. др. В том числе влияние на процесс оказывает уровень специалиста по ведению контекстной рекламы, способность оперативно отреагировать и проанализировать критичную ситуацию.

Человек многие процессы автоматизирует и уходит от ручного труда. Делается это для того, чтобы ускорить производительность создания какого-либо продукта. В нашем случае – создания кампании для контекстной рекламы.

Целью работы является разработка интеллектуальной системы, основная задача которой заключается в повышении эффективности и автоматизации процесса создания кампании и ведения контекстной рекламы.

Задача – разработать программный комплекс, который автоматизирует процесс создания рекламных кампаний в Яндекс.Директ, так как на данный момент большая часть процесса выполняется вручную и занимает от 10 рабочих дней.

Основной задачей системы является снижение стоимости привлечения посетителей из поисковой выдачи в зависимости от отслеживаемых результатов. Для этого есть специальный показатель CPC (Cost Per Click) — цена за каждый клик пользователя по рекламному объявлению с последующим переходом на сайт или одну из его

страниц. Рассчитывается как отношение стоимости рекламной кампании к числу кликов. Чем ниже показатель, тем эффективнее работает реклама.

$$CPC = \frac{\sum_{i=0}^n AD_i}{\sum_{i=0}^n C_i},$$

где AD_i – затраты на рекламу за период i ; C_i – количество кликов за период i .

Были составлены следующие требования к системе:

- самостоятельно принимает решение о том, по каким запросам требуется осуществить поиск, какие запросы отнести к ключевым фразам, а какие к минус-словам, какие запросы на какие страницы распределить;
- на основе математических грамматик самостоятельно генерирует тексты в соответствии с деятельностью организации и ключевым запросом;
- все тексты соответствуют требованиям Яндекс.Директ.

Процесс создания кампании для тематической площадки делится на три этапа: сбор поисковых запросов и их обработка, создание сниппетов для этих запросов и формирование файла и его загрузка на площадку. Так как на каждом этапе есть конечная цель и требуется автономность, реализация системы будет производиться с помощью агентов¹. В данной статье представлена общая схема интеллектуальной системы, которая состоит из следующих агентов (см. рис. 1).

¹ Егоров А. Н., Суконщиков А. А. Разработка интеллектуальной системы поиска, обработки, генерации контекстной информации // Автоматизация и энергосбережение машиностроительного и металлургического производств, технология и надежность машин, приборов и оборудования: материалы XIII Международной научно-технической конференции. 2018. С. 102–106; Суконщиков А. А., Швецов А. Н. Мультиагентные интеллектуальные системы и сети. Вологда: Вологодский государственный университет, 2019. 171 с.; Суконщиков А. А. Проектирование интеллектуальных агентно-ориентированных систем // Интеллектуально-информационные технологии и интеллектуальный бизнес (ИНФОС – 2019): материалы десятой международной научно-технической конференции (Вологда, 28–29 июня 2019 года). Вологда: Вологодский государственный университет, 2019. С. 244–247.

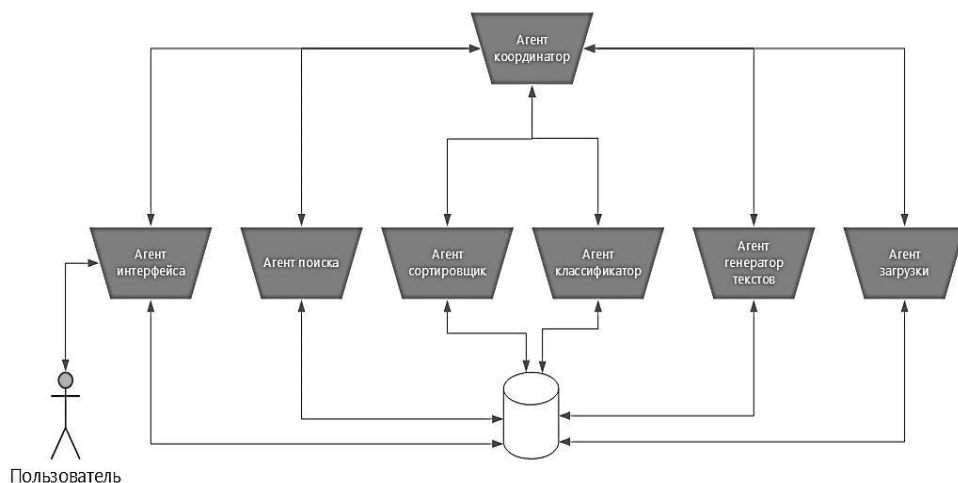


Рис. 1. Схема многоагентной интеллектуальной системы

Взаимодействие пользователя с системой будет происходить с помощью графического интерфейса, поэтому первым агентом будет агент интерфейса, который отвечает за двустороннюю связь и принимает начальные данные.

$$\left\{ \begin{array}{l} tgc_1, tgc_2, tgc_3 \dots tgc_i \\ kw_1, kw_2, kw_3 \dots kw_i \\ nkw_1, nkw_2, nkw_3 \dots nkw_i \\ lnk_1, lnk_2, lnk_3 \dots lnk_i \\ syn_1, syn_2, syn_3 \dots syn_i \\ rgn_1, rgn_2, rgn_3 \dots rgn_i \\ add_1, add_2, add_3 \dots add_i \end{array} \right\},$$

где tgc (target customer) – список параметров целевого клиента,

kw (keywords) – список ключевых слов,

nkw (negative keywords) – список минус-слов,

lnk (links) – список ссылок на сайт, куда будут распределяться готовые объявления,

syn (synonyms) – список слов синонимов для ключевых слов,

rgn (region) – список регионов, где будет производиться показ,

add (additional information) – дополнительная информация, которую введет пользователь.

Изначально человеку будет предложен выбор тематики, по которой он хочет создать кампанию, после чего ему будут представлены на выбор некоторые значения, необходимые для создания кампании (ключевые фразы, минус-слова / фразы, синонимы, регион / регионы, на который будет осуществляться показ объявлений и т. д.). Эти значения будут браться из базы данных. Если пользователь не нашел нужных ему значений, он вводит их вручную, значения записываются в базу данных (для

дальнейшего анализа), и в случае, когда системе непонятно, что имел пользователь в виду, она задает ему уточняющие вопросы. Также для повышения точности и скорости работы будет создана база для синонимов и генератора запросов из поисковой системы.

После того, как вся вводная информация получена, агент интерфейса сообщает о завершении своей работы агенту-координатору, который запускает следующего агента – агента поиска. Задачей этого агента является получить все возможные запросы из поисковой системы и записать их в свою базу данных, чтобы в дальнейшем отслеживать все изменения и пополнять ее. О завершении своей работы он сообщает агенту-координатору.

Следующим этапом будет разделение поисковых запросов на ключевые фразы и на минус-фразы, а также их запись в БД. Этой задачей занимается агент-сортировщик. На данный момент предполагается, что для выполнения данной задачи будет применяться многослойная нейронная сеть с обратным распространением на основе персептрона. Пороговое значение будет определяться:

$$y_i = \begin{cases} 1, & \sum_i W_{ij} \cdot x_i > \theta_j, \\ 0, & \sum_i W_{ij} \cdot x_i \leq \theta_j. \end{cases}$$

Функцией активации будет являться логистическая функция:

$$f(x) = \frac{1}{1 + e^x}.$$

Для расчета ошибки используется формула:

$$e = y - d,$$

где e – ошибка нейронной сети на текущем шаге, y – желаемое значение на текущем шаге, d – результат выполнения на текущем шаге.

Вычисление локального градиента (в точке) будет происходить по формуле:

$$\delta = \frac{dE}{dw} = e \cdot f'(y_{\text{вых}}).$$

Формула для корректировки связей методом градиентного спуска:

$$\Delta W = -\alpha \cdot \delta,$$

где α – скорость обучения, ΔW – величина, на которую необходимо изменить величину связи, δ – градиент в этой точке.

О завершении своей работы сообщает агенту-координатору.

Агент-кластеризатор распределяет ключевые запросы по смысловым группам по определенным параметрам и заносит эти группы в свою базу данных для дальнейшего анализа и изменения. На вход принимает:

$$\begin{cases} kw_1, kw_2, kw_3 \dots kw_i \\ nk w_1, nk w_2, nk w_3 \dots nk w_i, \\ rgn_1, rgn_2, rgn_3 \dots rgn_i \end{cases}$$

где kw (keywords) – список ключевых слов,

$nk w$ (negative keywords) – список минус-слов,

rgn (region) – список регионов, где будет производиться показ.

Значим и уровень вложенности, до которого необходимо производить группировку.

Агент-генератор текста на основе кластеризации и данных от агента-интерфейса о предполагаемом пользователе создаст тексты заголовка и описания для сниппета. После генерации текст будет записан в базу данных.

$$\begin{cases} tgc_1, tgc_2, tgc_3 \dots tgc_i \\ nk w_1, nk w_2, nk w_3 \dots nk w_i \\ clu_1, clu_2, clu_3 \dots clu_i \\ rgn_1, rgn_2, rgn_3 \dots rgn_i \\ add_1, add_2, add_3 \dots add_i \end{cases},$$

где tgc (target customer) – список параметров целевого клиента,

$nk w$ (negative keywords) – список минус-слов,

clu (clustering) – результаты кластеризации,

rgn (region) – список регионов, где будет производиться показ,

add (additional information) – дополнительная информация, которую ввел пользователь.

Для оценки качества работы агента будет использоваться показатель CTR – частота кликабельности объявления. Рассчитывается по формуле

$$CTR = \frac{\sum_{i=0}^n C_i}{\sum_{i=0}^n I_i} \cdot 100 \%,$$

где C_i – общее количество кликов за период, I_i – общее количество показов за период.

Если показатель будет возрастать по мере работы кампании, значит, агент правильно работает.

Для загрузки на площадку требуется оформить все в соответствии с требованиями поисковой системы: распределить объявления по посадочным страницам, прописать ключевые и минус-слова, распределить по регионам и т. д. и в итоге успешно произвести загрузку. С этой задачей будет справляться агент загрузки.

$$\begin{cases} \text{adv}_1, \text{adv}_2, \text{adv}_3 \dots \text{adv}_i \\ \text{lnk}_1, \text{lnk}_2, \text{lnk}_3 \dots \text{lnk}_i \\ \text{rgn}_1, \text{rgn}_2, \text{rgn}_3 \dots \text{rgn}_i \end{cases},$$

где adv (advertisement) – список готовых объявлений,

lnk (links) – список ссылок на сайт, куда будут распределяться готовые объявления,

rgn (region) – список регионов, где будет производиться показ.

После того, как компания поработала какое-то время, агент сбора статистики выгружает ее, выбирает статистические данные, записывает в базу данных.

Агент анализа обрабатывает полученные данные. Структурно-функциональная схема агента анализа приведена на рис. 2.¹

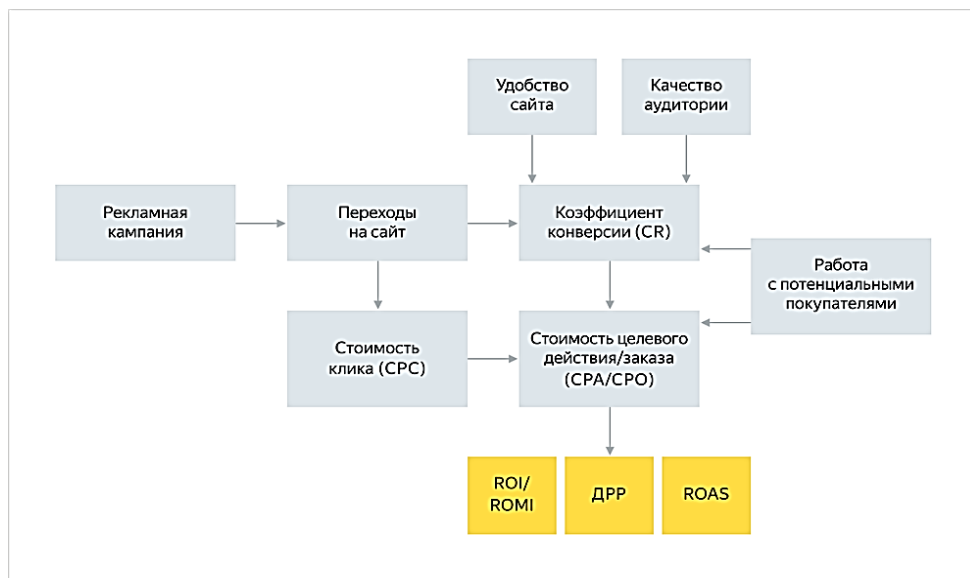


Рис. 2. Структурно-функциональная схема

¹ Егоров А. Н. Интеллектуальная система для создания контекстной информации на основе многоагентной системы // Инновационное развитие. 2019. № 1 (28). С. 13–14.

Для оценки эффективности работы кампании будут использоваться коэффициенты DPP, ROAS¹.

ROAS – показывает окупаемость рекламы. Рассчитывается по формуле:

$$ROAS = \frac{\sum_{i=0}^n MR_i}{\sum_{i=0}^n MC_i} \cdot 100 \%,$$

где MR_i – прибыль от рекламы за определенный период; MC_i – затраты на рекламу за определенный период.

Когда $ROAS < 100 \%$ – от рекламы понесены убытки; $ROAS = 100 \%$ – реклама окупилась; $ROAS > 100 \%$ – от рекламы получена прибыль.

DPP – это доля рекламных расходов.

$$DPP = \frac{\sum_{i=0}^n MC_i}{\sum_{i=0}^n MR_i} = \frac{1}{ROAS},$$

где MR_i – прибыль от рекламы за определенный период; MC_i – затраты на рекламу за определенный период.

Чем ниже DPP, тем эффективнее рекламная кампания. ROAS зависит от коэффициентов CPA и CPO.

Цель рекламного объявления – привести пользователя на сайт для совершения целевого действия, которое заключается либо в оформленном заказе, либо в другом целевом действии (например: сообщение с формы обратной связи, звонок по телефону с сайта, регистрация на сайте и др.). Коэффициенты CPA и CPO – это и есть стоимость цели.

Стоимость обращения (cost per action, CPA) показывает, сколько стоит целевой визит посетителя. CPA – отношение рекламного бюджета к количеству обращений.²

$$CPA = \frac{\sum_{i=0}^n AD_i}{\sum_{i=0}^n CA_i},$$

где AD_i – затраты на рекламу за определенный период, CA_i – количество целевых визитов.

¹ Юзвович Л. И. Инвестиции и инвестиционная деятельность. Екатеринбург: Уральский государственный экономический университет, 2021. 498 с.

² Коэффициент возврата инвестиций (ROI). URL: <https://yandex.ru/adv/edu/direct/pokazateli-effektivnosti/koeffitsiyent-vozvrata-investitsiy-roi> (дата обращения: 20.05.2023).

CPA – широкое понятие. Здесь оцениваются значимые целевые действия, например: скачивание прайс-листа, заполнение анкеты или заявки на обратный звонок. Параметр подходит для сайтов, где нет продаж либо цепочка от обращения до продажи достаточно длинная. Поэтому важно оценивать не только продажи, но и промежуточные метрики. Пример сайтов: сайт визитка, сайт каталог.

CPO оценивает только стоимость привлечения заказа. Этот параметр подходит для интернет-магазинов.

CPO (Cost Per Order) показывает, в какую сумму обходится каждый заказ. В отличие от CPC, в CPO учитываются абсолютно все заказы, даже неоплаченные. Рассчитывается как отношение затрат на рекламу к количеству полученных заказов. Если маржинальность меньше CPO, реклама «работает в минус», если нет регулярных повторных продаж.

$$CPO = \frac{\sum_{i=0}^n AD_i}{\sum_{i=0}^n CO_i},$$

где AD_i – затраты на рекламу за определенный период, CO_i – количество оформленных заказов.

На коэффициенты CPA и CPO влияют показатели стоимости клика (CPC) по объявлению в поисковой выдаче и количество этих переходов (CR).

Cost per click (CPC) – это цена за клик, которую рекламодатель отдает системе контекстной рекламы за клик по объявлению.

$$CPC = \frac{\sum_{i=0}^n AD_i}{\sum_{i=0}^n C_i},$$

где AD_i – затраты на рекламу, C_i – количество кликов по рекламе.

Чем ниже CPC, тем дешевле стоимость привлечения клиента. В рамках рекламной кампании удобнее делать это через аргумент C_i .

Конверсия сайта (CR) – это отношение количества посетителей сайта (SV), совершивших целевое действие, к общему количеству привлеченных посетителей (AV) за определенный период.

$$CR = \frac{\sum_{i=0}^n SV_i}{\sum_{i=0}^n AV_i},$$

где SV_i – количество посетителей сайта за промежуток n , AV_i – количество привлечённых посетителей за промежуток n .

Из показателя CR можно сделать вывод, насколько целевой трафик приходит на сайт. Чем выше данный показатель, тем дешевле стоимость заказа. Также от значения данного показателя напрямую зависит значение показателя CPA. То есть, чтобы снизить CPA, необходимо увеличить CR.

Исходя из значений коэффициентов, агент может делать определенный аудит и давать рекомендации. Например, для повышения показателя CR:

- проработать уникальное торговое предложение,
- проанализировать объявление и контент на посадочной странице, дать рекомендации по его изменению,
- внести изменение в элементы объявления (быстрые ссылки, виртуальная визитка и др.),
- корректировка списков минус-фраз и ключевых запросов,
- проанализировать работоспособность форм, ссылок и других элементов сайта, которые необходимы для показателя конверсии.

Для повышения показателя CPC, например, можно выполнить следующее: сравнить ключевые запросы и заголовки в объявлениях, сделать заголовки максимально приближенными к ключевому понятию. Если качество трафика зависит от сезона, можно сделать корректировку на данную особенность.

По завершении работы агент передает информацию агенту-координатору, который уже решает, что делать с ней, какие задачи ставить и кто из агентов будет их решать. Отметим, что агент-координатор управляет деятельностью всей системы.

Выводы

Разработана архитектура многоагентной интеллектуальной системы. Приведено описание каждого агента с его задачами, параметрами и результатом. Составлена структурно-функциональная схема для оценки эффективности работы кампании с описанием ключевых параметров и их влияния на результат оценки работы.

В случае успешной реализации проекта будет экономиться время сотрудников по работе с контекстной рекламой, сократится бюджет организации в связи с экономией затрат на некоторые направления работы.

Список источников

Егоров А. Н. Интеллектуальная система для создания контекстной информации на основе многоагентной системы // Инновационное развитие. 2019. № 1 (28). С. 13–14.

Егоров А. Н., Суконщиков А. А. Разработка интеллектуальной системы поиска, обработки, генерации контекстной информации // Автоматизация и энергосбережение машиностроительного и металлургического производств, технология и надежность машин, приборов и оборудования: материалы XIII Международной научно-технической конференции. 2018. С. 102–106.

Коэффициент возврата инвестиций (ROI). URL: <https://yandex.ru/adv/edu/direct/pokazateli-effektivnosti/koeffitsiyent-vozvrata-investitsiy-roi> (дата обращения: 20.05.2023).

Суконщиков А. А., Швецов А. Н. Мультиагентные интеллектуальные системы и сети. Вологда: Вологодский государственный университет, 2019. 171 с.

Суконщиков А. А. Проектирование интеллектуальных агентно-ориентированных систем // Интеллектуально-информационные технологии и интеллектуальный бизнес (ИНФОС-2019): материалы десятой международной научно-технической конференции (Вологда, 28–29 июня 2019 года). Вологда: Вологодский государственный университет, 2019. С. 244–247.

Юзович Л. И. Инвестиции и инвестиционная деятельность. Екатеринбург: Уральский государственный экономический университет, 2021. 498 с.

References

Egorov A. N. *Intellektual'naia sistema dlia sozdaniia kontekstnoi informatsii na osnove mnogooagentnoi sistemy* [Intelligent system to generate the context information based on multi-agent system]. *Innovatsionnoe razvitiie* [Innovative development], 2019, no. 1 (28), pp. 13–14.

Egorov A. N., Sukonshchikov A. A. *Razrabotka intellektual'noi sistemy poiska, obrabotki, generatsii kontekstnoi informatsii* [Development of an intelligent system for search, processing, generation of contextual information]. *Avtomatizatsiia i energosberezhenie mashinostroitel'nogo i metallurgicheskogo proizvodstv, tekhnologiya i nadezhnost' mashin, priborov i oborudovaniia: materialy XIII Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii* [Automation and energy saving of machine-building and metallurgical production, technology and reliability of machines, devices and equipment: Proceedings of the XIII International Scientific and Technical Conference], 2018, pp. 102–106.

Koeffitsient vozvrata investitsii (ROI) [Return on investment (ROI) coefficient]. Available at: <https://yandex.ru/adv/edu/direct/pokazateli-effektivnosti/koeffitsiyent-vozvrata-investitsiy-roi> (accessed: 20.05.2023).

Sukonshchikov A. A., Shvetsov A. N. *Mul'tiagentnye intellektual'nye sistemy i seti* [Multi-agent intelligent systems and networks]. Вологда: Вологодский государственный университет, 2019. 171 p.

Sukonshchikov A. A. *Proektirovanie intellektual'nykh agentno-orientirovannykh sistem* [Developing intelligent agent-based systems]. *Intellektual'no-informatsionnye tekhnologii i intellektual'nyi biznes (INFOS-2019): materialy desiatoi mezhdunarodnoi nauchno-tekhnicheskoi konferentsii (Vologda, 28–29 iyunia 2019 goda)* [Intellectual Information Technologies and Intelligent Business (INFOS-2019): Proceedings of the Tenth International Scientific and Technical Conference (Vologda, 28–29 June 2019)]. Вологда: Вологодский государственный университет, 2019, pp. 244–247.

Iuzvovich L. I. *Investitsii i investitsionnaia deiatel'nost'* [Investments and investment activity]. Екатеринбург: Уральский государственный экономический университет, 2021. 498 p.

Сведения об авторах

Алексей Александрович Суконщиков – кандидат технических наук, доцент, Вологодский государственный университет (д. 1, ул. Галкинская, 160000 Вологда, Россия); **Aleksey A. Sukonshchikov** – Candidate of Technical Sciences, Associate Professor, Vologda State University (1, ul. Galkinskaia, 160000 Vologda, Russia).

Александр Николаевич Егоров – аспирант, Вологодский государственный университет (д. 1, ул. Галкинская, 160000 Вологда, Россия); **Aleksandr N. Egorov** – Postgraduate student, Vologda State University (1, ul. Galkinskaia, 160000 Vologda, Russia).

Анатолий Николаевич Швецов – доктор технических наук, профессор, Вологодский государственный университет (д. 1, ул. Галкинская, 160000 Вологда, Россия); **Anatoly N. Shvetsov** – Doctor of Technical Sciences, Professor, Vologda State University (1, ul. Galkinskaia, 160000 Vologda, Russia).

Заявленный вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Статья поступила в редакцию 29.05.2023; одобрена после рецензирования 26.06.2023; принята к публикации 30.07.2023.

The article was submitted 29.05.2023; Approved after reviewing 26.06.2023; Accepted for publication 30.07.2023.