

УДК 004.89

ПРОГНОЗИРОВАНИЕ ПОВЕДЕНЧЕСКИХ РЕАКЦИЙ ЧЕЛОВЕКА НА ОСНОВЕ АНАЛИЗА ЕГО ЦИФРОВОГО СЛЕДА В СОЦИАЛЬНОЙ СЕТИ

К. Р. Мудрова, А. Н. Жданова

В данной работе представлено описание процесса и результатов разработки автоматизированной системы анализа цифрового следа пользователя социальной сети «ВКонтакте», позволяющей HR-менеджеру определить характер поведенческих реакций соискателей на должность на основе обработки их публичных данных. Обоснована востребованность программных решений, предназначенных для скрининга социальных сетей, в процессе автоматизации работы отделов по подбору персонала в условиях нехватки информации о личностях соискателей. Рассмотрен подход к решению задачи автоматизации анализа цифрового следа пользователя путем определения количественных показателей проявления отклоняющегося поведения и социальной открытости. Разработанная система представляет собой веб-приложение на основе клиент-серверной архитектуры, которое позволяет авторизованным пользователям HR-менеджерам добавлять и сохранять анкеты соискателей, просматривать сформированные системой отчеты об их публичной активности.

Ключевые слова: публичные данные; HR-системы; подбор персонала; оценка публикаций; обработка текстовых данных; веб-приложение.

Успех крупных компаний в условиях нарастающей конкуренции складывается из использования актуальных технологических решений и подбора высококвалифицированных специалистов, готовых и способных развивать свои собственные навыки и, как следствие, вывести достижения компании на новый уровень.

Оптимизировать работу с таким важным активом предприятия, как человеческие ресурсы (Human Resources – HR), удастся путем автоматизации работы HR-отделов, широкий спектр деятельности которых требует постоянного и планомерного внедрения современных HR-систем. Решить одну из наиболее значимых задач – подбор кандидатов, имеющих необходимые навыки, неподдельный интерес к деятельности компании и способных влиться в рабочий коллектив – менеджерам по персоналу призваны помочь программные решения, в основе которых лежат средства рекрутинговой

аналитики, выявляющие паттерны поиска, подбора и найма на основе данных компаний и резюме соискателей [1].

Очевидно, что резюме не дает исчерпывающих ответов на вопросы о личности и используемых в неформальной обстановке паттернах поведения соискателя. Согласно результатам опроса, заказанного компанией Express Employment Professionals, и опроса, проведенного группой компаний HeadHunter, в качестве инструмента для получения недостающей информации о характере поведенческих реакций, HR-специалисты широко используют социальные сети, анализируя опубликованный кандидатом на должность контент на предмет проявления им подозрительного поведения [2; 3]. С ростом популярности такой практики, возникла необходимость автоматизации процесса сбора и анализа публичных данных в социальных сетях – создания программных решений, позволяющих HR-менеджеру за короткий срок выявить

© Мудрова К. Р., Жданова А. Н., 2023.

Мудрова Ксения Родионовна (mudrova.x@gmail.com),
студент IV курса института информатики и кибернетики;
Жданова Александра Николаевна (zhdan.aleksandra@gmail.com),
доцент кафедры программных систем Самарского университета,
443086, Россия, г. Самара, Московское шоссе, 34.

«красные флаги» в цифровом следе потенциальных кандидатов на должности.

Среди наиболее популярных решений можно выделить такие продвинутые зарубежные программные системы, как Fama, Ferrettly, Neotas и Good egg, предназначенные для скрининга социальных сетей соискателей и сотрудников компаний: анализа текстовых и графических публикаций [4–7]. Однако все эти системы объединяет один общий недостаток – их ориентированность на работу с зарубежными платформами, затрудняющая использование представленных средств отечественными компаниями для оценки публикаций в российских социальных сетях.

Главная цель представленной работы – спроектировать и реализовать автоматизированную систему анализа цифрового следа в социальной сети, позволяющую определить характер поведенческих реакций на основе обработки публичных данных. Описанная в данной работе система представляет собой веб-приложение, разработанное для отображения пользователям-HR-менеджерам результатов комплексного системного анализа публичных данных соискателей в российской социальной сети «ВКонтакте».

Условия и методы исследования

Решение поставленной задачи автоматизации анализа цифрового следа соискателя включает в себя два основных этапа: определение критериев оценки и методов интерпретации опубликованной соискателем информации и реализация системы для непосредственного получения информации из профиля соискателя в социальной сети, ее завершающего анализа согласно полученному на первом этапе алгоритму.

Автоматизировать сбор опубликованного в социальных сетях контента разработчикам удастся путем использования предоставляемых самими платформами социальных сетей программных интерфейсов (API – Application Programming Interface), позволяющих получать и обрабатывать определенный набор публичных данных их пользователей.

В качестве основного инструмента в работе с публичными данными на каждом из этапов решения поставленной задачи был использован набор средств программного интерфейса социальной сети «ВКонтакте» –

«API ВКонтакте», предоставляемого сторонним разработчикам и позволяющего получать информацию напрямую из базы данных vk.com [8].

Выбранный программный интерфейс «API ВКонтакте» включает в себя методы для работы с такими внутрисистемными объектами как «Пользователь», «Сообщество», «Приложение», «Беседа», медиаконтентом (фотографиями, аудиозаписями, видеозаписями), их вспомогательными объектами и наборами значений. Для использования каждого из методов требуется с помощью протокола гипертекстовой разметки HTTP (HyperText Transfer Protocol) составить соответствующий GET или POST-запрос с явным указанием личного ключа доступа, имени метода и тех свойств объекта, с которыми необходимо работать [9].

Анализ цифрового следа соискателя в рамках разработанного приложения сводится к получению обособленных результатов анализа количественных характеристик опубликованной информации (количества фотографий, заполненных полей в описании профиля пользователя, общего числа публикаций), выраженных в оценке «открытости» кандидата на должность, и результатов анализа его текстовых публикаций, выраженных в оценке уровней проявления в письменной речи делинквентного, девиантного и позитивного поведения.

На первом-подготовительном этапе с целью определения количественных показателей проявления отклоняющегося поведения и социальной открытости соискателей было принято решение разработать вспомогательную программу на языке Python, взаимодействующую с «API ВКонтакте» и позволяющую анализировать открытые публичные данные случайной выборки пользователей социальной сети. С ее помощью были собраны публичные данные тысячи случайных пользователей «ВКонтакте»: получены списки и общее число их текстовых публикаций, определено количество опубликованных фотографий и количество заполненных полей в описании профиля пользователя в разделах «Карьера», «Образование» и «Жизненная позиция».

По завершении сбора данных тексты авторских публикаций и тексты публикаций, на

которые ссылались пользователи, удалось проанализировать путем составления «мешка слов» каждого из пользователей, представляющего собой текст в виде мультимножества его слов без учета грамматики и их порядка, но содержащего информацию об их количестве [10]. Для наиболее точного анализа текстов перед составлением «мешков слов» была проведена их предварительная подготовка, основанная на использовании методик обработки текстовых данных, включающих в себя [11]:

- очистку данных (Data Cleaning) – удаление из исходных данных особых знаков, символов, пунктуации, HTML-тэгов;

- токенизацию – процесс разбиения текстового документа на составляющие отдельные слова, которые называются токенами;

- удаление стоп-слов – процесс удаления часто используемых слов, которые не несут никакой дополнительной информации (местоимений, частиц и некоторых общеупотребительных глаголов);

- стемматизацию – процесс приведения слов к их корню или основе;

- лемматизацию – процесс приведения слов к их начальной форме.

Также на первом этапе для выявления в «мешках слов» терминов, указывающих на проявление девиантного, делинквентного или позитивного поведения на основе тональных словарей русского языка [12–14], трудов по лингвистической экспертизе текстов [15–17], а также словарей разговорной экспрессивной речи [18] и бранной лексики [19] были составлены три системных словаря. С использованием вспомогательной программы было посчитано количество совпадающих слов в «мешках слов» пользователей и системных словарях, а также были составлены для каждой категории слов соответствующие графики кумулятивной вероятности. Каждое значение на оси ординат построенных графиков характеризовало вероятность того, что в «мешке слов» пользователя, выбранного случайным образом, количество обнаруженных совпадающих слов не превышает соответствующую величину, отложенную на оси абсцисс. На основе графиков совокупный анализ

всех профилей позволил соотнести количественные оценки каждого из критериев с одним из трех уровней проявления: низким, средним и высоким. Аналогичным образом были проанализированы общее число текстовых публикаций, количество опубликованных фотографий и заполненных полей в описании профиля (рис. 1).

В отличие от оценок уровней проявления девиантного, делинквентного и позитивного поведения, оценка уровня «открытости» является комплексной и складывается из оценок количества опубликованных фотографий, заполненных полей в описании профиля кандидата, общего числа текстовых публикаций. В рамках работы уровни по первым двум критериям оценивались в 1, 2 и 3 балла соответственно, уровни по критерию количества опубликованных фотографий, как более приоритетному, в 2, 4 и 6 баллов. Уровень «открытости» определялся на основе суммы баллов по всем трем критериям: сумма в 4–5 баллов соответствовала низкому уровню, 6–9 – среднему, 10–12 – высокому.

Используемая на этапе анализа программа в дальнейшем легла в основу центрального аналитического модуля разработанной автоматизированной системы анализа цифрового следа соискателей.

Разработка системы

По завершении этапа анализа была спроектирована и реализована система, позволяющая авторизованным пользователям-HR-менеджерам добавлять и сохранять анкеты кандидатов на должности, просматривать сформированные системой отчеты о характере их поведенческих реакций, основанные на результатах анализа их цифрового следа в социальной сети «ВКонтакте», а также получать уникальные отчеты на основе личного пользовательского словаря.

Реализованная система направлена на сокращение времени, необходимого HR-специалисту для проверки онлайн-активности соискателя, путем определения и отображения в сформированных отчетах исключительно тех публикаций кандидата, которые могут указать на проявление им неприемлемого, подозрительного поведения, или напротив – активности, амбициозности и профессионализма.

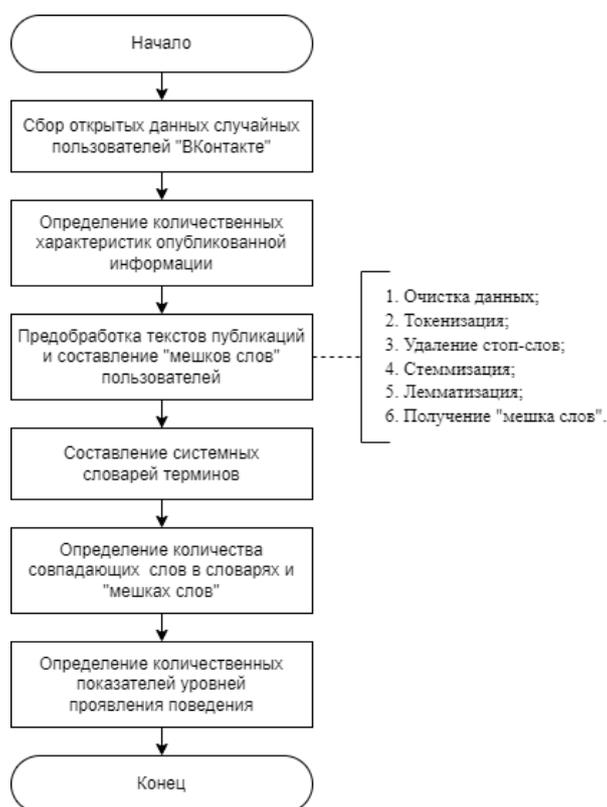


Рис. 1. Блок-схема алгоритма определения количественных показателей проявления отклоняющегося поведения и социальной открытости

Спроектированная система представляет собой веб-приложение, реализованное на основе клиент-серверной архитектуры, в рамках которой компьютеры-клиенты (удалённые системы) имеют возможность посылать запросы и получать услуги от централизованной служебной машины – сервера [20]. Серверная часть системы была разработана на языке JavaScript с помощью платформы Node.js, превращающей JavaScript из узкоспециализированного языка в язык общего назначения, фреймворка Express, предназначенного для создания веб-приложений на основе Node.js, средств объектно-реляционной системы управления базами данных PostgreSQL, а также отдельно реализованного аналитического модуля на языке Python. Для осуществления обработки текстов в аналитическом модуле были выбраны библиотеки, входящие в состав платформы для процессинга естественного языка NLTK и морфологический анализатор Rymorphy2. Клиентская часть системы была реализована с помощью библиотеки React.js для разработки пользовательских интерфейсов и дополнительных библиотек для отображения графиков и динамических элементов.

Взаимодействие между клиентской и серверной частями разработанной системы происходит посредством протокола передачи данных HTTP, согласно принципам организации которого клиентское приложение формирует запрос и отправляет его на сервер, после чего сервер обрабатывает полученный запрос, формирует ответ и передает его обратно клиенту [21].

В серверной части системы сосредоточена логика по сбору и обработке публичных данных соискателей в социальной сети «ВКонтакте». При поступлении соответствующего запроса от клиента в аналитическом модуле, проводится сбор необходимых открытых данных заданного кандидата, обработка и анализ всех текстовых публикаций, определяется частота публикаций на основе данных за последние полгода и количественные показатели проявления отклоняющегося поведения и социальной открытости, после чего формируется его итоговая анкета – уникальная запись в серверной базе данных.

Для каждой анализируемой текстовой публикации программа проводит

предварительную обработку текста и составляет «мешок слов», состав которого сравнивает с составами сохраненных на сервере системных словарей и личного пользовательского словаря менеджера, привязанного к его учетной записи. В случае, если системой в тексте публикации выделяются слова из словарей девятнадцатой, делинквентной или позитивной лексики, система создает запись о публикации в «системном» отчете, в случае если выделенные слова совпадают со словами из пользовательского словаря – в «пользовательском» отчете. Каждая запись в отчете содержит исходный текст, дату и время публикации, список и количество выделенных системой слов, а также признак авторства, указывающий является ли кандидат автором текста или ссылается на публикацию другого человека.

Наряду с подсистемой формирования анкет кандидатов на должности, на сервере реализованы файловая подсистема, подсистемы аутентификации пользователей-менеджеров и взаимодействия с базой данных. Наибольшая часть данных, необходимых для корректной работы системы: учетные записи пользователей, сохраненные ими в системе анкеты кандидатов, системные и пользовательские отчеты, – хранятся в развернутой на сервере базе данных. Составленные системные словари также хранятся на сервере в виде текстовых файлов.

Принцип взаимодействия пользователя и системы

Запуск и взаимодействие с разработанной системой на стороне клиента осуществляется пользователем в веб-браузере после перехода по ссылке при помощи пользовательского интерфейса. Клиентская часть системы главным образом служит для отображения полученных с сервера данных, к которым есть доступ у конкретного пользователя: каждый менеджер имеет в системе свою личную учетную запись. К учетной записи менеджера привязываются данные для персональной настройки интерфейса, его личный словарь терминов, все добавленные им в систему анкеты соискателей, а также сохраненные в

отчетах публикации. Таким образом, в случае если несколько пользователей будут анализировать цифровой след одного и того же кандидата, для каждого менеджера будет создана уникальная копия анкеты соискателя и его «системного» отчета, и при редактировании или удалении одной из уникальных анкет другие не подвергнутся изменению.

До начала работы с системой пользователю необходимо пройти последовательные этапы регистрации и авторизации. Авторизованному пользователю становятся доступны все функции разработанного приложения: работа с анкетами соискателей, редактирование личного пользовательского словаря, просмотр справочной информации о системе, ее возможностях и используемых алгоритмах анализа публичных данных в социальной сети «ВКонтакте».

Перед началом работы с анкетами на панели управления пользователю предлагается заполнить персональный словарь терминов, на основе которого формируются уникальные пользовательские отчеты, содержащие информацию исключительно о тех публикациях, в текстах которых встречаются указанные пользователем слова (рис. 2).

Работа с анкетами соискателей включает в себя добавление новых анкет кандидатов на должности, сохранение и просмотр результатов системного анализа их цифрового следа, управление списками ранее сохраненных анкет.

Для добавления новой анкеты кандидата менеджеру необходимо заполнить форму, предполагающую указание основной информации о личности кандидата, его опыте работы, должности, на которую он претендует, а также его идентификатор в социальной сети для проведения последующего анализа. По завершении заполнения соответствующей формы система анализирует публичные данные соискателя в социальной сети «ВКонтакте», не скрытые настройками приватности, и отображает пользователю-HR-специалисту персональную страницу кандидата, содержащую результаты анализа, сформированные системой отчеты по публикациям, набор специальных элементов управления отображением отчетов и график активности кандидата – частоты его публикаций в каждый из дней недели (рис. 3).

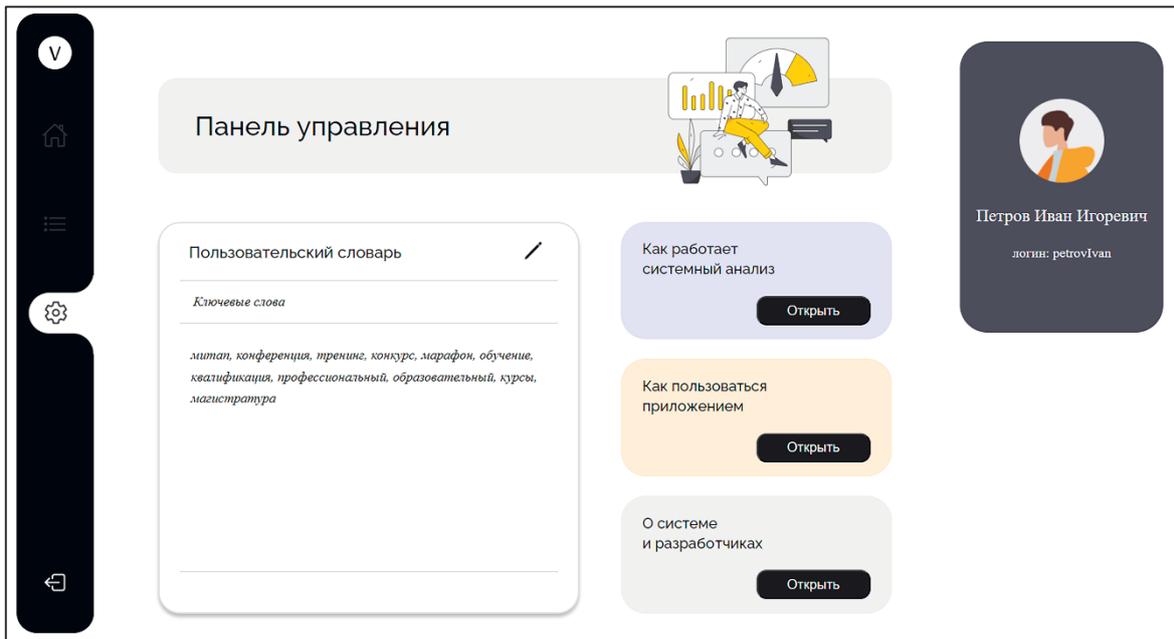


Рис. 2. Экранная форма панели управления

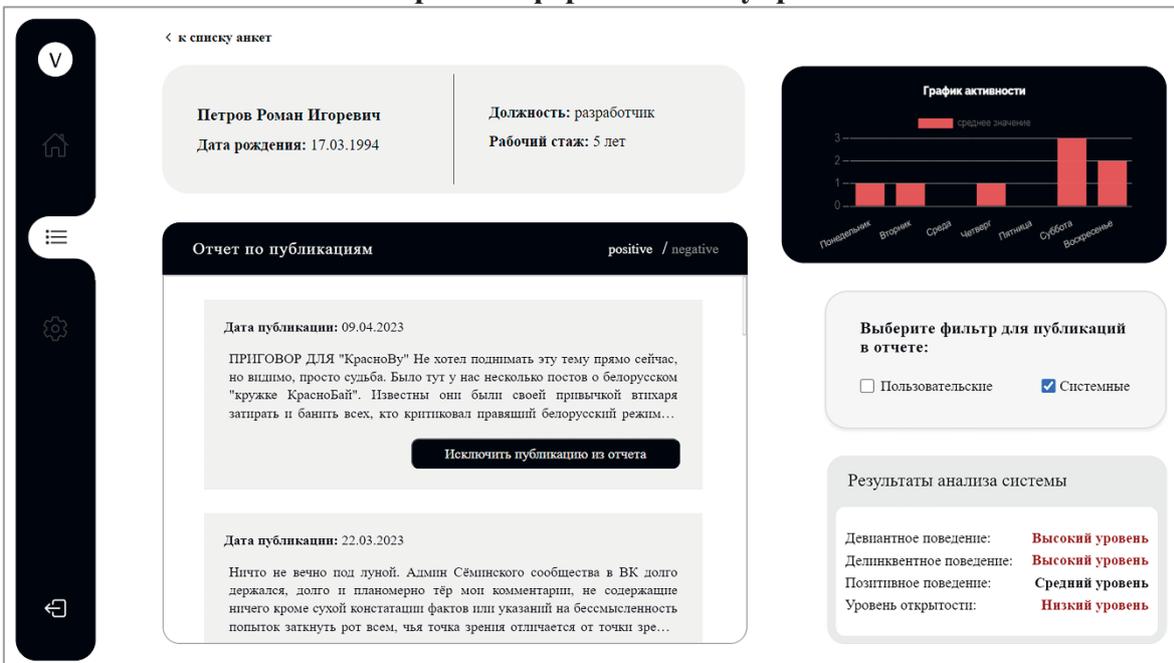


Рис. 3. Экранная форма анкеты соискателя

Содержание страницы анкеты также зависит от того, был ли заполнен пользовательский словарь терминов: если на момент проведения анализа словарь пользователя заполнен, наряду с «системным» отчетом система также сформирует «пользовательский» отчет по публикациям, на отображение которого пользователь может переключиться при просмотре анкеты.

При просмотре сформированного «системного» отчета в качестве результатов анализа пользователю отображаются соответствующие уровни проявления поведения и уровень «открытости» соискателя, а также

становится доступной настройка отображения текстовых публикаций, указывающих на проявление либо позитивного, либо негативного (девиантного и делинквентного) поведения соискателя. При просмотре «пользовательского» отчета в качестве результатов анализа система отображает общее число выделенных в текстовых публикациях слов из пользовательского словаря.

Приложение позволяет пользователю просматривать описание и полный текст сохраненных в отчетах публикаций, список выделенных системой ключевых слов. В случае если после просмотра текста публикации

пользователь не согласен с тем, что она указывает на проявление того или иного поведения, обладая иной семантикой, он может удалить выбранную публикацию из отчета. Менеджерам обеспечена возможность редактирования как «системного», так и «пользовательского» отчетов. При редактировании отчетов система сразу перепроверяет количество выделенных слов во всех текстовых публикациях и обновляет соответствующие результаты анализа.

Заключение

В работе были рассмотрены роль программных систем, предназначенных для скрининга социальных сетей соискателей должности, и актуальность их внедрения в процессе автоматизации работы отделов по подбору персонала. Рассмотрен один из возможных подходов к решению задачи автоматизации анализа цифрового следа пользователя социальной сети путем создания специального веб-приложения. Приведен алгоритм и результаты исследования данных случайных пользователей социальной сети, позволившего определить количественные показатели проявления отклоняющегося поведения и социальной открытости, используемые при реализации соответствующего приложения. Описаны структура и функциональность разработанной системы, призванной помочь HR-менеджеру в подборе кандидатов на должности, посредством анализа их публичной активности в социальной сети «ВКонтакте».

В качестве вектора развития приложения рассматривается расширение возможностей аналитического модуля: внедрение средств анализа графического контента, а также дополнительных средств обработки естественного языка, позволяющих анализировать не только отдельно взятые из текстовых публикаций слова, но и сложные грамматические конструкции, с учетом их семантики, что позволит повысить точность уже реализованного анализа.

Литература

1. Рекрутинговая аналитика: три уровня оптимизации рекрутинга [Электронный ресурс]. URL: <https://www.aihr.com/blog/implement-hris-human-resources-information-system> (дата обращения: 6.11.2022).

2. 71% of Hiring Decision-Makers Agree Social Media is Effective for Screening Applicants [Electronic resource]. URL: <https://www.expresspros.com/Newsroom/America-Employed/71-of-Hiring-Decision-Makers-Agree-Social-Media-is-Effective-for-Screening-Aplicants.aspx> (accessed: 6.11.2022).

3. Не пригласили на работу? Возможно, проблема в вашем профиле в соцсетях [Электронный ресурс]. URL: <https://samara.hh.ru/article/24311> (дата обращения: 6.11.2022).

4. Software for Screening Social Media Fama.io [Electronic resource]. URL: <https://fama.io/product> (accessed: 15.11.2022).

5. The Ferretly Platform [Electronical resource]. URL: <https://www.ferretly.com/product> (accessed: 15.11.2022).

6. Neotas [Electronic resource]. URL: <https://www.neotas.com> (accessed: 16.11.2022).

7. Good Egg Pricing, Features, Reviews and Alternatives [Electronic resource]. URL: <https://www.getapp.com/hr-employee-management-software/a/good-egg> (accessed: 16.11.2022).

8. Использование API. Быстрый старт. VK для разработчиков [Электронный ресурс]. URL: <https://dev.vk.com/api/getting-started> (дата обращения: 14.11.2022).

9. Справочник API. Объекты. Список объектов. VK для разработчиков [Электронный ресурс]. URL: <https://dev.vk.com/reference/objects> (дата обращения: 14.11.2022).

10. Мешок слов [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Мешок_слов (дата обращения: 10.12.2022).

11. Плавное введение в Natural Language Processing (NLP) [Электронный ресурс]. URL: <https://datastart.ru/blog/read/plavnoe-vvedenie-v-natural-language-processing-nlp> (дата обращения: 10.12.2022).

12. Тональный словарь русского языка КартаСловСент [Электронный ресурс]. URL: <https://github.com/dkulagin/kartaslov/tree/master/dataset/kartaslovsent> (дата обращения: 16.12.2022).

13. NRC Word-Emotion Association Lexicon [Electronic resource]. URL: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm> (accessed: 16.12.2022).

14. Словарь оценочных слов и выражений русского языка РуСентиЛекс [Электронный ресурс]. URL: <https://www.labinform.ru/pub/rusentilex/index.htm> (дата обращения: 16.12.2022).
15. Выявление признаков унижения чести, достоинства, умаления деловой репутации и оскорбления в лингвистической экспертизе текста / И. А. Стернин и [др.]. Ярославль: Канцлер, 2013. 35 с.
16. Еленская К. В. Лингвистический паспорт лексемы «насилие» // Преподаватель XXI век. 2020. № 4–2. С. 386–392.
17. Сердюкова В. М. Практикум по специальному переводу (внешнеполитические отношения) // Сетевое электронное учебное издание. УО «Минский государственный лингвистический университет», 2016 г.
- [Электронный ресурс]. URL: <http://elearning.mslu.by/assignments/88/about/> (дата обращения: 20.12.2022).
18. Химик В. В. Большой словарь русской разговорной экспрессивной речи. СПб.: Норинт, 2004. 768 с.
19. Мокиенко В. М., Никитина Т. Г. Словарь русской брани (матизмы, обценнизмы, эвфемизмы). СПб.: Норинт, 2003. 448 с.
20. Архитектура «Клиент-Сервер» [Электронный ресурс]. URL: <https://itelon.ru/blog/arkhitektura-klient-server/> (дата обращения: 22.02.2023).
21. Простым языком об HTTP [Электронный ресурс]. URL: <https://habr.com/ru/articles/215117/> (дата обращения: 24.02.2023).

PREDICTING INDIVIDUAL BEHAVIOR REACTIONS WITH SOCIAL MEDIA DIGITAL FOOTPRINTS

K. R. Mudrova, A. N. Zhdanova

This paper presents the development and description of automated system for analyzing the digital footprint of a VKontakte social network user. The system enables HR-managers to process the applicant's publicly available data in order to determine his/her behavioral reactions. The work substantiates the demand to use social media screening software in order to automate recruitment processes under the lack of applicant's personal information. It presents an approach to automation of user's digital footprint analysis through determining the quantitative indicators for deviant behavior and social openness. The developed system is a web application based on a client-server architecture, which allows HR managers as authorized users to add and save applicants profiles, view their social media activity reports generated by the system.

Key words: publicly available data; HR systems; recruitment; posts analysis; text processing; web application.

Статья поступила в редакцию 08.06.2023 г.

© Mudrova K. R., Zhdanova A. N., 2023.

Mudrova Ksenia Rodionovna (mudrova.x@gmail.com),

IVth year student of the Institute of Computer science and cybernetics;

Zhdanova Aleksandra Nikolaevna (zhdan.aleksandra@gmail.com),

associate professor of the Department of software systems of Samara University,
443086, Russia, Samara, Moskovskoye shosse, 34.