УДК 004.8, 004.056.53
DOI: 10.15827/0236-235X.137.045-053
Дата подачи статьи: 21.09.21
2022. Т. 35. № 1. С. 045–053

Прототип программного комплекса для анализа аккаунтов пользователей социальных сетей: веб-фреймворк Django

В.Д. Олисеенко [1], *младший научный сотрудник, vdo@dscs.pro*
М.В. Абрамов [1], *к.т.н., руководитель лаборатории, mva@dscs.pro*
А.Л. Тулупьев [1,2], *д.ф.-м.н., профессор, главный научный сотрудник, alt@dscs.pro*
К.А. Иванов [2], *бакалавр, mail@dscs.pro*

[1] *Санкт-Петербургский федеральный исследовательский центр РАН, лаборатория теоретических и междисциплинарных проблем информатики, г. Санкт-Петербург, 199178, Россия*
[2] *Санкт-Петербургский государственный университет, кафедра информатики, г. Санкт-Петербург, 199034, Россия*

В статье рассматриваются вопросы реализации прототипа исследовательско-практического комплекса для автоматизации анализа аккаунтов пользователей в социальных сетях. Данный прототип используется в качестве инструмента для косвенной оценки выраженности психологических особенностей пользователей, их уязвимостей к социоинженерным атакам и выработки рекомендаций по защите от них. Прототип разработан на языке программирования Python 3.8 с применением веб-фреймворка Django 3.1, а также PostgreSQL 13.2 и Bootstrap 4.6.

Цель работы заключается в повышении оперативности процесса извлечения информации из размещаемых в социальных сетях данных, позволяющей косвенно оценить психологические, поведенческие и иные особенности пользователей, и достигается через автоматизацию извлечения указанных данных и разработку инструментария для их анализа.

Предметом исследования являются методы автоматизированного извлечения, предобработки, унификации и представления данных из аккаунтов пользователей социальных сетей в контексте их защиты от социоинженерных атак.

Предложенный прототип приложения на основе веб-фреймворка Django решает задачу автоматизированного извлечения, предобработки, унификации и представления данных со страниц пользователей социальных сетей, что является одним из важных этапов в построении системы анализа защищенности пользователей от социоинженерных атак, опирающейся, в свою очередь, на синтез профиля пользователей.

Теоретическая значимость работы заключается в комбинировании и апробации через автоматизацию разработанных ранее методов и подходов для восстановления пропущенных значений атрибутов аккаунта и сопоставления аккаунтов пользователей социальных сетей на предмет их принадлежности одному пользователю.

Практическая значимость состоит в разработке прикладного инструмента, размещенного на поддомене sea.dscs.pro и позволяющего производить первичный анализ аккаунтов пользователей социальных сетей.

Ключевые слова: *социальные сети, социоинженерные атаки, информационная безопасность, веб-фреймворк Django, анализ аккаунта пользователя в социальной сети*.

В публикациях по информационной безопасности отмечается тенденция к увеличению числа социоинженерных атак, а также ущерба от таких киберпреступлений [1]. Субъектом социоинженерных атак является пользователь, а одна из причин роста числа успешных инцидентов социальной инженерии заключается в недостаточной осведомленности пользователей информационных систем в области информационной безопасности. Согласно исследованию [2], около 39 % рисков безопасности связаны с человеческим фактором, а по данным компании Purplesec, 98 % кибератак основаны на применении методов социальной инженерии [3]. Существующие превентивные решения защиты информационных систем (DLP-системы, системы разграничения доступа, антивирусные программы и т.д.) могут не учитывать ключевые личностные особенности пользователей, влияющие на выраженность их уяз-

вимостей к социоинженерным атакующим воздействиям. Работа над повышением осведомленности пользователей о возможных атаках и способах противодействия им может проводиться недостаточно эффективно (тренинг по информационной безопасности раз в 5 лет, семинар одновременно для специалистов в разных областях и т.п.). Несмотря на смещающийся в направлении защиты пользователей от социоинженерных атак фокус, в СМИ по-прежнему появляются сведения об инцидентах. Таким образом, проблема защиты пользователей от социоинженерных атак продолжает оставаться нерешенной и актуальной.

Для обеспечения защиты пользователей информационных систем от социоинженерных атак прежде всего необходимо оценить степень их защищенности. Один из подходов к оценке основан на построении профиля уязвимостей пользователей. По данной тематике существует ряд теоретических и прикладных наработок. Так, в [4] был предложен подход к построению профиля уязвимостей пользователя, учитывающий набор пар «уязвимость–выраженность уязвимости». В свою очередь, оценки выраженности уязвимостей пользователя ассоциированы с оценками выраженности их личностных особенностей [5]. Информация о выраженности личностных особенностей пользователя может быть получена из различных источников: психологических тестов, анкетирования, HR-служб, аккаунтов в социальных сетях и т.п. На основе полученных оценок защищенности пользователей от социоинженерных атак могут быть выработаны рекомендации, способствующие ее повышению. Такие рекомендации могут сводиться к конфигурированию политик разграничения доступа сотрудников компании к важным документам/информации, прохождению рядом сотрудников соответствующих тренингов, в том числе включающих обучающие или симуляционные компьютерные игры [6], таргетированным советам по работе со специалистами и т.д.

Сейчас основной фокус в исследованиях сосредоточен на получении информации из данных, извлекаемых из аккаунтов пользователей в социальных сетях, для оценки выраженности их личностных особенностей. Детализация ряда начальных задач в этом контексте может быть сформулирована следующим образом: автоматизировать извлечение данных из аккаунтов пользователей в социальных сетях, при этом выборка данных должна осуществляться по атрибутам аккаунта, отнесенным экспертом к существенным с точки зрения оценки психологических особенностей пользователя и одновременно его поведенческих особенностей, степени выраженности уязвимостей к социоинженерным атакам; восстановить недостающие пропущенные данные (город, возраст) [7] с целью сопоставления и поиска аккаунтов пользователей в других социальных сетях (для повышения объема собираемой о них информации) [8]; создать удобный пользовательский интерфейс и способы визуализации результатов применения разработанных ранее методов. Стоит отметить, что уже существуют соответствующие практические наработки [4, 5, 7, 8], однако они разрозненные, реализованы с использованием разных технологий, их сложно применять на практике, что в совокупности затрудняет полноценную автоматизацию процесса анализа защищенности пользователей. Таким образом, актуальна задача создания единого комплекса, агрегирующего результаты из перечисленных исследований как базы для создания рекомендательной системы по защите от социоинженерных атак. Данный комплекс внесет вклад, в частности, в оптимизацию процесса управления персоналом в компаниях и послужит инструментом для снижения ущерба от социоинженерных атак.

Цель данного исследования – повысить оперативность процесса извлечения информации из размещаемых пользователями в социальных сетях данных, которая позволит косвенно оценить их психологические, поведенческие и иные особенности. Она достигается автоматизацией извлечения сведений и разработкой инструментария для их анализа. Повышение оперативности процесса извлечения возможно за счет автоматизации агрегирования результатов уже имеющихся практических наработок, что позволит существенно уменьшить затрачиваемое время в сравнении с ручными отбором данных и агрегированием результатов [4, 5, 7, 8]. Теоретическая значимость работы заключается в комбинировании и апробации через автоматизацию разработанных ранее методов и подходов для восстановления пропущенных значений атрибутов аккаунта, а также сопоставления аккаунтов пользователей социальных сетей на предмет их принадлежности одному пользователю. Практическая значимость заключается в разработке прикладного инструмента, размещенного на поддомене sea.dscs.pro и позволяющего производить первичный анализ ак-

каунтов пользователей социальных сетей. Представленный комплекс впоследствии планируется использовать в качестве основы для оценки выраженности психологических особенностей, уязвимостей к социоинженерным атакам пользователей и рекомендаций по защите от них.

Релевантные работы

На данный момент можно выделить три основных вектора исследований в области защиты пользователей от социоинженерных атак. Первый вектор связан с изучением вопросов работы с персоналом через тренинги и инструктажи [9–12], однако в данном случае за пределами рассмотрения часто оказываются проблемы так называемых инсайдеров (злоумышленников, работающих в компании или иначе проникающих в офис). Кроме того, тренинги бывают дорогостоящими и отнимают существенное время у сотрудников, из-за чего проводятся нерегулярно и редко. Второй вектор характеризуется применением программно-технических наработок к задачам защиты пользователей от социоинженерных атак (DLP-системы) [13]. Но DLP-система не сможет, например, застраховать от фотографирования злоумышленником экрана с содержащейся на нем критичной информацией. Третий вектор исследований направлен на использование социоинженерных пентестов для выявления основных уязвимостей пользователей и системы в целом [14]. Особенность пентестов заключается в том, что они, как правило, невоспроизводимы. Один и тот же сотрудник в разные моменты времени может по-разному себя повести при одном и том же социоинженерном атакующем воздействии.

В работах [9, 10] подчеркивается необходимость регулярного проведения профилактических мер (тренинги, имитации атак, внутренний аудит) в компании для снижения рисков доступа социальных инженеров к конфиденциальной информации. В исследованиях [11, 12] обоснована необходимость интеграции программ повышения осведомленности персонала об опасности социоинженерных атак в процесс управления рисками компании, приведены примеры информационных систем, нацеленных на обучение сотрудников в области информационной безопасности. Однако данные рекомендации имеют общий характер, не адаптированы для конкретного сотрудника, что может значительно снизить эффективность принятых мер.

В работе [15] дан анализ социальных сетей как источника полезной для злоумышленника информации о сотрудниках компании. Показано, что злоумышленник может планировать свои действия на основе этой информации для создания персонализированных фишинговых email-рассылок или сайтов. В работе также предложены принципы и политики обеспечения безопасности организации в контексте социальных сетей, которые заключаются в объединении существующих стандартов ISO, IEC и др. в сфере информационной безопасности и предложенной авторами системы SESM. Таким образом, можно заключить, что социальные сети являются доступным богатым источником сведений о пользователях, что позволяет злоумышленникам использовать их в своих целях. Следует отметить, что существуют разработки, позволяющие строить психологический портрет человека на основе анализа его речевого поведения [5, 16].

В исследовании [17] рассмотрено применение аппарата нейронных сетей в задаче определения подверженности пользователя социальной сети негативным воздействиям. В [18] предложена модель оценки выраженности уязвимостей пользователей к социоинженерным атакам, основанная на четырех компонентах: социопсихологическом, социоэмоциональном, факторе привычки и восприятии. Помимо этого, также описан сценарий выявления наиболее слабых мест пользователя в контексте социоинженерных атак и даны рекомендации по принятию превентивных мер (разграничение прав доступа, тренинги, работа с психологами и т.д.). В исследованиях [19–22] описаны подходы, позволяющие определять эмоциональное состояние и давать оценку степени выраженности психологических особенностей пользователей социальных сетей на основе опубликованных ими данных.

Постановка задачи

Таким образом, множество проведеннных исследований доказывают необходимость создания единого инструмента для анализа аккаунтов пользователей социальных сетей с целью построения оценок их защищенности от социоинженерных атак. Однако на сегодня такого инструмента нет.

В данной статье рассматривается вопрос разработки вычислительного инструмента для извлечения информации из размещаемых пользователями в социальных сетях данных, позво-

ляющей косвенно оценить их психологические, поведенческие и иные особенности. Такая разработка будет включать автоматизацию извлечения, предобработки, унификации и представления данных со страниц пользователей социальных сетей «ВКонтакте» и «Одноклассники». С учетом представленной задачи текущие результаты должны стать основой для создания системы по выявлению уязвимостей к социоинженерным атакам по аккаунтам пользователей в социальных сетях и рекомендаций по защите от них. Выбор обозначенных социальных сетей обусловлен их высокой популярностью в России [23].

Для достижения поставленной цели необходимо разработать прототип веб-приложения, имеющий следующие функциональные и нефункциональные качества:

– получать данные о пользователях социальных сетей («ВКонтакте» и «Одноклассники»);

– восстанавливать пропущенные значения атрибутов аккаунта пользователя (город, возраст);

– сопоставлять аккаунты пользователей социальных сетей «ВКонтакте» и «Одноклассники», выявлять оценку вероятности их принадлежности одному пользователю с последующей целью поиска аккаунтов, которые могут принадлежать одному пользователю;

– визуализировать социальный граф друзей пользователя;

– предоставлять удобный пользовательский интерфейс.

Используемые методы и подходы

Для достижения необходимых качеств разрабатываемого прототипа веб-приложения нужно выбрать способы экстракции данных из социальных сетей и их представления (в том числе с использованием социального графа), методы для восстановления пропущенных значений атрибутов аккаунтов пользователей, методы сопоставления аккаунтов социальных сетей, а также удобное представление пользовательского инерфейса.

Извлечение данных из социальных сетей. Для получения данных о пользователях в рассматриваемых социальных сетях («ВКонтакте» и «Одноклассники») можно использовать два подхода: загружать HTML-страницы пользователей с последующим выделением нужной информации либо воспользоваться методами API, предоставляемыми социальной сетью. Преимуществом первого подхода является его относительная устойчивость к изменениям API социальных сетей, однако он требует написания отдельного парсера для каждой социальной сети и может нарушать политику их конфиденциальности. Второй подход (использование API социальных сетей) позволяет сократить срок разработки (за счет готовых функций API) и не нарушать политику конфиденциальности социальных сетей. Таким образом, для экстракции данных будет использоваться подход на основе API-инструментария социальных сетей.

Восстановление недостающих атрибутов аккаунтов. Часто некоторые атрибуты аккаунтов пользователей бывают незаполненными или искаженными, что усложняет дальнейший анализ, связанный с сопоставлением и поиском аккаунтов пользователей в различных социальных сетях, или идентификацию аккаунтов сотрудников компаний. Именно поэтому необходимо восстановление недостающих значений. В данной работе для восстановления недостающих значений атрибутов «город» и «возраст» использован подход, описанный в исследовании [7].

Сопоставление аккаунтов пользователей социальных сетей. Задача сопоставления двух аккаунтов в разных социальных сетях заключается в определении оценки вероятности их принадлежности одному пользователю. Решение данной задачи необходимо для создания единого профиля пользователя [7], содержащего информацию из разных социальных сетей, что позволит предоставить большее количество данных для дальнейшего анализа. В данной работе используются практические наработки из [8], которые сводят задачу сопоставления аккаунтов в социальных сетях к задаче бинарной классификации. Для определения вероятности беруются анкетные данные пользователя (Ф.И.О., город, возраст) в социальной сети и данные его социального окружения (друзей пользователя). Модель построена на основе логистической регрессии [8], которая принимает на вход результат сравнения, лежащий на отрезке [0, 1]: имени, фамилии, города проживания (используя словарь синонимов или метрику строковой схожести Джаро–Винклера [24]), возраста (точное сопоставление при наличии возраста, иначе два возраста считаются равными, если абсолютная разница меньше или равна 3), списка друзей пользователя (используя коэффициент Браун–Бланке). На выход данная модель предоставляет оценку

вероятности принадлежности двух аккаунтов одному пользователю.

Построение социального графа. Задачу построения социального графа друзей пользователя можно решить, представляя вершинами друзей данного пользователя, а ребрами дружеские связи. В данном случае под друзьями понимаются пользователи, которые обозначены соответствующим образом в анализируемых социальных сетях. Таким образом, социальный граф друзей пользователя – это тройка вида $G = (U_0, F, L)$, где U_0 – целевой пользователь, для которого строится социальный граф; F – друзья пользователя U_0 в данной социальной сети – вершины графа; $L = \{\{f_1, f_2\}: f_1, f_2 \in F\} \cup \{\{U_0, f\}: f \in F\}$, где f_1 и f_2 – друзья в социальной сети, а $f \in F$ – ребра графа. Построение графа необходимо как с функциональной точки зрения (проводить поиск сотрудников компаний, находить наиболее близких друзей и т.д.), так и с нефункциональной (предоставлять понятный для человека результат работы модулей).

Реализация прототипа веб-приложения

Для реализации прототипа веб-приложения использовались Python 3.8.0, Django 3.2, PostgreSQL 13.2, Bootstrap 4.6. Система имеет клиент-серверную архитектуру. Диаграмма пакетов веб-приложения приведена на рисунке 1.

Рассмотрим компоненты диаграммы. Controller представляет собой диспетчер адресов веб-страниц, используемых в приложении. Данный пакет отвечает за выбор нужного обработчика адреса и передачу запроса этому обработчику. Views представляет собой набор функций, обрабатывающих запрос от контроллера. Он также отвечает за отображение веб-страниц, формируя их на основе запроса классов модели и HTML-шаблонов. Model содержит классы, описывающие информацию в БД, и логику заполнения атрибутов аккаунта по короткой ссылке или идентификатору «ВКонтакте» или «Одноклассники». Utils содержит логику взаимодействия модели с API социальных сетей, а также код, отвечающий за восстановление недостающих атрибутов аккаунта, визуализацию социального графа и сопоставление аккаунтов в социальных сетях «ВКонтакте» и «Одноклассники». Templates содержит HTML-шаблоны страниц сервиса, написанные с применением шаблонизатора Django. Database – БД, хранящая информацию о пользователях социальных сетей с восстановленными атрибутами.

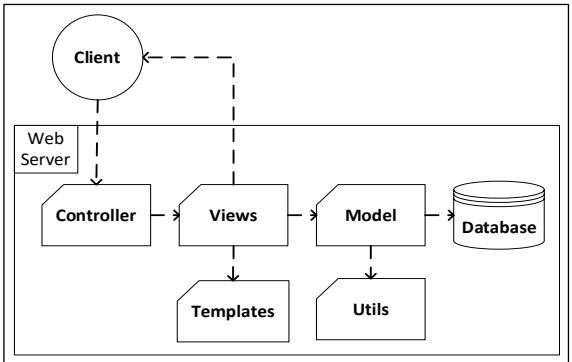

Рис. 1. Диаграмма пакетов

Fig. 1. Package diagram

Функциональность и графическое представление комплекса реализуются следующими модулями: поиск пользователей, сопоставление аккаунтов, построение социального графа.

Страница для отображения личной информации из аккаунтов пользователей (имя, фамилия, город, дата рождения, список друзей) социальных сетей содержит HTML-форму для ввода уникального идентификатора пользователя. После отправки формы происходит перенаправление на сформированную страницу пользователя (рис. 2), содержащую интерактивный социальный граф друзей пользователя (см. http://www.swsys.ru/uploaded/image/2022-1/2022-1-dop/9.jpg), анкетные данные в социальной сети, а также восстановленные недостающие значения атрибутов «город» и «возраст» [7]. Помимо этого, рассчитывается и выводится результат метрики «ближайшие друзья пользователя», которая определяется как соотношение пользователей-друзей, имеющих наибольшее количество связей в социальном графе, с другими друзьями пользователя.

Страница сопоставления аккаунтов пользователей «ВКонтакте» и «Одноклассники» содержит поля для ввода идентификатора пользователя в данных социальных сетях (см. http://www.swsys.ru/uploaded/image/2022-1/2022-1-dop/13.jpg). После отправки формы пользователь получает анкетную информацию из аккаунтов данных социальных сетей и оценку вероятности их принадлежности одному пользователю, выраженную в вероятности [8].

Заключение

Таким образом, в результате исследования был предложен прототип приложения на основе веб-фреймворка Django, решающий задачу автомати-

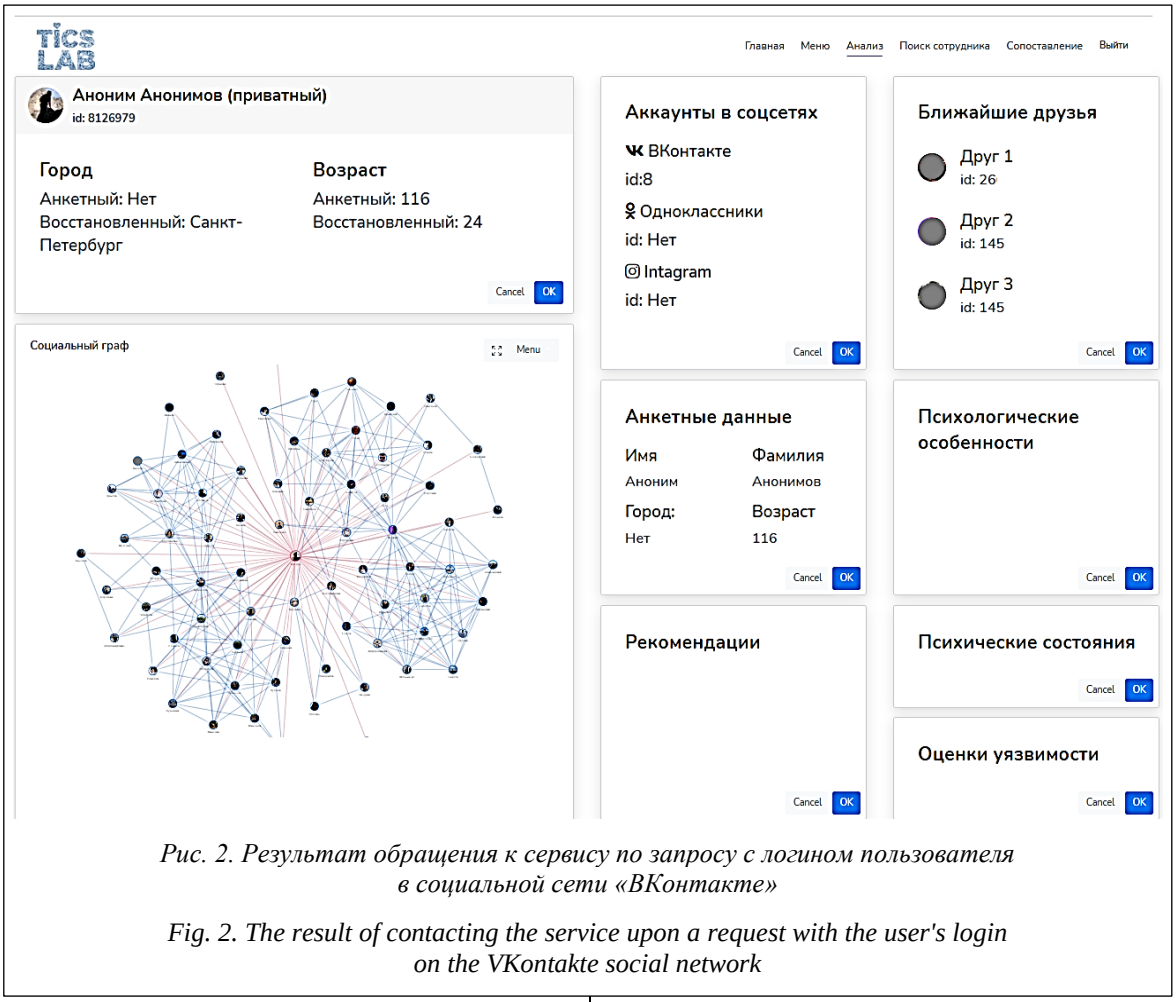

Рис. 2. Результат обращения к сервису по запросу с логином пользователя в социальной сети «ВКонтакте»

Fig. 2. The result of contacting the service upon a request with the user's login on the VKontakte social network

зированного извлечения, предобработки, унификации и представления размещаемых пользователями в социальных сетях данных, которые позволяют косвенно оценить их психологические, поведенческие и иные особенности.

Дальнейшими направлениями исследования являются разработка и внедрение функциональности модели оценки выраженности психологических особенностей пользователей по их постам, а также синтез методов и подходов для создания рекомендаций по защите от социоинженерных атак, в том числе для реализации системы рекомендаций по управленческим решениям для защиты предприятий.

Работа выполнена в рамках проекта по государственному заданию СПБ ФИЦ РАН СПИИРАН № FFZF-2022-0003; при финансовой поддержке РФФИ, проект № 20-07-00839.

Литература

1. Statista. Amount of Monetary Damage Caused by Reported Cybercrime to the IC3 from 2001 to 2020. URL: https://www.statista.com/statistics/267132/total-damage-caused-byby-cyber-crime-in-the-us/ (дата обращения: 15.08.2021).

2. Alsharif M., Mishra S., AlShehri M. Impact of human vulnerabilities on cybersecurity. Computer Systems Science and Engineering, 2021, no. 40 (3), pp. 1153–1166. DOI: 10.32604/CSSE.2022.019938.

3. PURPLESEC. 2021 Cyber Security Statistics. The Ultimate List of Stats, Data and Trends. URL: https://purplesec.us/resources/cyber-security-statistics/ (дата обращения: 21.08.2021).

4. Абрамов М.В., Тулупьев А.Л., Тулупьева Т.В. Психологические особенности, психические состояния пользователя и профиль его уязвимостей в контексте социоинженерных атак // Психология психических состояний. 2019. № 13. С. 312–317.

5. Тулупьева Т.В., Суворова А.В., Азаров А.А., Тулупьев А.Л., Бордовская Н.В. Возможности и опыт применения компьютерных инструментов в анализе цифровых следов студентов – пользователей социальной сети // Компьютерные инструменты в образовании. 2015. № 5. С. 3–13.

6. Krylov B., Abramov M., Khlobystova A. Automated player activity analysis for a serious game about social engineering. Recent Research in Control Engineering and Decision Making, 2020, pp. 587–599. DOI: 10.1007/978-3-030-65283-8_48.

7. Абрамов М.В., Тулупьев А.Л., Тулупьева Т.В. Агрегирование данных из социальных сетей для восстановления фрагмента метапрофиля пользователя // XVI конф. по искусственному интеллекту с междунар. уч. КИИ-2018. 2018. С. 189–197.

8. Корепанова А.А., Олисеенко В.Д., Абрамов М.В., Тулупьев А.Л. Применение методов машинного обучения в задаче идентификации аккаунтов пользователя в двух социальных сетях // Компьютерные инструменты в образовании. 2019. № 3. С. 29–43. DOI: 10.32603/2071-2340-2019-3-29-43.

9. Ревенков П.В., Бердюгин А.А. Социальная инженерия как источник рисков в условиях дистанционного банковского обслуживания // Национальные интересы: приоритеты и безопасность. 2017. Т. 13. № 9. С. 1747–1760. DOI: 10.24891/ni.13.9.1747.

10. Ghafir I., Prenosil V., Alhejailan A., Hammoudeh M. Social engineering attack strategies and defence approaches. Proc. IV Intern. Conf. FiCloud, 2016, pp. 145–149. DOI: 10.1109/FiCloud.2016.28.

11. Salahdine F., Kaabouch N. Social engineering attacks: A survey. Future Internet, 2016, vol. 11, no. 4, p. 89. DOI: 10.3390/FI11040089.

12. Aldawood H., Skinner G. Contemporary cyber security social engineering solutions, measures, policies, tools and applications: A critical appraisal. IJS, 2019, vol. 01, no. 10, p. 1. URL: https://www.cscjournals.org/manuscript/Journals/IJS/Volume10/Issue1/IJS-151.pdf (дата обращения: 22.08.2021).

13. Faiz M.F., Arshad J., Alazab M., Shalaginov A. Predicting likelihood of legitimate data loss in email DLP. Future Generation Computer Systems, 2020, vol. 110, pp. 744–757. DOI: 10.1016/j.future.2019.11.004.

14. Bertoglio D., Zorzo A. Overview and open issues on penetration test. J. of the Brazilian Computer Society, 2017, vol. 23, art. 2. DOI: 10.1186/s13173-017-0051-1.

15. Wilcox H., Bhattacharya M. A framework to mitigate social engineering through social media within the enterprise. Proc. ICIEA, 2016, pp. 1039–1044. DOI: 10.1109/ICIEA.2016.7603735.

16. Татарникова Т.М., Богданов П.Ю. Построение психологического портрета человека с применением технологий обработки естественного языка // Науч.-технич. вестн. информационных технологий, механики и оптики. 2021. Т. 21. № 1. С. 85–91. DOI: 10.17586/2226-1494-2021-21-1-85-91.

17. Дойникова Е.В., Браницкий А.А., Котенко И.В. Использование нейросетей для прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям // Информационно-управляющие системы. 2020. № 1. С. 24–33. DOI: 10.31799/1684-8853-2020-1-24-33.

18. Albladi S.M., Weir G.R.S. User characteristics that influence judgment of social engineering attacks in social networks. Hum. Cent. Comput. Inf. Sci., 2018, vol. 8, art. 5. DOI: 10.1186/s13673-018-0128-7.

19. Xue D., Hong Z., Guo S., Gao L., Wu L., Zheng J., Zhao N. Personality recognition on social media with label distribution learning. IEEE Access, 2017, vol. 5, pp. 13478–13488. DOI: 10.1109/ACCESS.2017.2719018.

20. Wei X., Xu G., Wang H., He Y., Han Z., Wang W. Sensing users' emotional intelligence in social networks. IEEE Transactions on Computational Social Systems, 2020, vol. 7, no. 1, pp. 103–112. DOI: 10.1109/TCSS.2019.2944687.

21. Wang W., Li Y., Huang Y., Liu H., Zhang T. A Method for identifying the mood states of social network users based on cyber psychometrics. Future Internet, 2017, vol. 9, no. 2, art. 22. DOI: 10.3390/fi9020022.

22. Andreassen C.S., Pallesen S., Griffiths M.D. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. Addictive Behaviors, 2017, vol. 64, pp. 287–293. DOI: 10.1016/j.addbeh.2016.03.006.

23. Statista. Leading Social Media Platforms in Russia as of 3rd Quarter of 2020, by Penetration Rate. URL: https://www.statista.com/statistics/867549/top-active-social-media-platforms-in-russia/ (дата обращения: 22.08.2021).

24. Winkler W.E. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. Washington, 1990, 8 p.

Software & Systems
DOI: 10.15827/0236-235X.137.045-053

Received 21.09.21
2022, vol. 35, no. 1, pp. 045–053

A software package prototype for analyzing user accounts in social networks: Django web framework

V.D. Oliseenko [1], *Junior Researcher, vdo@dscs.pro*
M.V. Abramov [1], *Ph.D. (Engineering), Head of Laboratory, mva@dscs.pro*
A.L. Tulupyev [1,2], *Dr.Sc. (Physics and Mathematics), Professor, Chief Researcher, alt@dscs.pro*
K.A. Ivanov [2], *Bachelor of Science, mail@dscs.pro*

[1] *St. Petersburg Federal Research Center of the Russian Academy of Sciences,*
Laboratory of Theoretical and Interdisciplinary Problems of Informatics,
St. Petersburg, 199178, Russian Federation
[2] *Petersburg State University, Computer Science Department, St. Petersburg, 199034, Russian Federation*

Abstract. The paper considers the issues implementing a prototype of a research and practical complex to automate the analysis of user accounts in social networks. Such prototype is used as a tool to indirectly assess users' psychological features manifestation, their vulnerabilities to social engineering attacks as well as to develop recommendations for protection against these attacks. The prototype is developed in the Python 3.8 programming language using the Django 3.1 web framework and PostgreSQL 13.2, Bootstrap 4.6.

This paper aims to increase the efficiency of extracting information from data posted by users in social networks, which allows indirect assessment of psychological, behavioral and other characteristics of users. The goal is achieved by automating data extraction and developing tools for their analysis. The subject of the study is the methods of automated extraction, pre-processing, unification, and presentation of data from users' accounts in social networks to protect them against social engineering attacks.

A prototype application based on the Django web framework solves the problem of automated extraction, preprocessing, unification, and presentation of data from user accounts in social networks. The solution of this problem is one of the essential steps to build a system for analyzing the security of users from social engineering attacks. The theoretical significance of the work is in the combination and validation through the automation of previously developed methods and approaches to recover missing values of the attributes of the account, the comparison of online social networks users' accounts for their belonging to the same user.

The practical significance comes from the development of an application tool located on the subdomain sea.dscs.pro, which allows performing primary analysis of users' accounts in social networks.

Keywords: social networks, social engineering attacks, information security, web framework Django, Analysis of the user's social network account.

Acknowledgements. *The work was carried out within the framework of the project under the state assignment of SPC RAS SPIIRAS no. FFZF-2022-0003; with the financial support of the RFBR, project no. 20-07-00839.*

References

1. *Statista. Amount of Monetary Damage Caused by Reported Cybercrime to the IC3 from 2001 to 2020*. Available at: https://www.statista.com/statistics/267132/total-damage-caused-byby-cyber-crime-in-the-us/ (accessed August 15, 2021).

2. Alsharif M., Mishra S., AlShehri M. Impact of human vulnerabilities on cybersecurity. *Computer Systems Science and Engineering*, 2021, no. 40 (3), pp. 1153–1166. DOI: 10.32604/CSSE.2022.019938.

3. *PURPLESEC. 2021 Cyber Security Statistics. The Ultimate List of Stats, Data and Trends*. Available at: https://purplesec.us/resources/cyber-security-statistics/ (accessed August 21, 2021).

4. Abramov M.V., Tulupyev A.L., Tulupyeva T.V. User's psychological traits, mental states, and vulnurabilities profile in the context of social engeneering attacks. *Psychology of Psychological States*, 2019, no. 13, pp. 312–317 (in Russ.).

5. Tulupyeva T.V., Syvorova A.V., Azarov A.A., Tulupyev A.L., Bordovskaya N.V. Computer tools in the analysis of students" digital footprints in social network: possibilities and primary results. *Computers Tools in Education*, 2015, no. 5, pp. 3–13 (in Russ.).

6. Krylov B., Abramov M., Khlobystova A. Automated player activity analysis for a serious game about social engineering. *Recent Research in Control Engineering and Decision Making*, 2020, pp. 587–599. DOI: 10.1007/978-3-030-65283-8_48.

7. Abramov M.V., Tulupyev A.L., Tulupyeva T.V. Aggregating data from social networks to restore a fragment of a user's meta-profile. *Proc. XVI Conf. on Artificial Intelligence with Int. Participation RCAI-2018*, 2018, pp. 189–197 (in Russ.).

8. Korepanova A.A., Oliseenko V.D., Abramov M.V., Tulupyev AL. Application of machine learning methods in the task of identifying user accounts in two social networks. *Computer Tools in Education*, 2019, no. 3, pp. 29–43. DOI: 10.32603/2071-2340-2019-3-29-43 (in Russ.).

9. Revenkov P.V., Berdyugin A.A. Social engineering as a source of risks in online banking services. *National Interests: Priorities and Security*, 2017, vol. 13, no. 9, pp. 1747–1760. DOI: 10.24891/ni.13.9.1747 (in Russ.).

10. Ghafir I., Prenosil V., Alhejailan A., Hammoudeh M. Social engineering attack strategies and defence approaches. *Proc. IV Intern. Conf. FiCloud*, 2016, pp. 145–149. DOI: 10.1109/FiCloud.2016.28.

11. Salahdine F., Kaabouch N. Social engineering attacks: A survey. *Future Internet*, 2016, vol. 11, no. 4, p. 89. DOI: 10.3390/FI11040089.

12. Aldawood H., Skinner G., Contemporary cyber security social engineering solutions, measures, policies, tools and applications: A critical appraisal. *International Journal of Security (IJS)*, 2019, vol. 10, no. 1, p. 1 Available at: https://www.cscjournals.org/manuscript/Journals/IJS/Volume10/Issue1/IJS-151.pdf (accessed August 22, 2021).

13. Faiz M.F., Arshad J., Alazab M., Shalaginov A. Predicting likelihood of legitimate data loss in email DLP. *Future Generation Computer Systems*, 2020, vol. 110, pp. 744–757. DOI: 10.1016/j.future.2019.11.004.

14. Bertoglio D., Zorzo A. Overview and open issues on penetration test. *J. of the Brazilian Computer Society*, 2017, vol. 23, art. 2. DOI: 10.1186/s13173-017-0051-1.

15. Wilcox H., Bhattacharya M. A framework to mitigate social engineering through social media within the enterprise. *Proc. ICIEA*, 2016, pp. 1039–1044. DOI: 10.1109/ICIEA.2016.7603735.

16. Tatarnikova T.M., Bogdanov P.Yu. Human psyche creation by application of natural language processing technologies. *Sci. Tech. J. Inf. Technol. Mech. Opt.*, 2021, vol. 21, no. 1, pp. 85–91 (in Russ.). DOI: 10.17586/2226-1494-2021-21-1-85-91.

17. Doynikova E.V., Branitskiy A.A., Kotenko I.V. Use of neural networks for forecasting of the exposure of social network users to destructive impacts. *Information and Control Systems*, 2020, no. 1, pp. 24–33. DOI: 10.31799/1684-8853-2020-1-24-33 (in Russ.).

18. Albladi S.M., Weir G.R.S. User characteristics that influence judgment of social engineering attacks in social networks. *Hum. Cent. Comput. Inf. Sci*., 2018, vol. 8, art. 5. DOI: 10.1186/s13673-018-0128-7.

19. Xue D., Hong Z., Guo S., Gao L., Wu L., Zheng J., Zhao N. Personality recognition on social media with label distribution learning. *IEEE Access*, 2017, vol. 5, pp. 13478–13488. DOI: 10.1109/ACCESS.2017.2719018.

20. Wei X., Xu G., Wang H., He Y., Han Z., Wang W. Sensing users' emotional intelligence in social networks. *IEEE Transactions on Computational Social Systems*, 2020, vol. 7, no. 1, pp. 103–112. DOI: 10.1109/TCSS.2019.2944687.

21. Wang W., Li Y., Huang Y., Liu H., Zhang T. A Method for identifying the mood states of social network users based on cyber psychometrics. *Future Internet*, 2017, vol. 9, no. 2, art. 22. DOI: 10.3390/fi9020022.

22. Andreassen C.S., Pallesen S., Griffiths M.D. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addictive Behaviors*, 2017, vol. 64, pp. 287–293. DOI: 10.1016/j.addbeh.2016.03.006.

23. *Statista. Leading Social Media Platforms in Russia as of 3rd Quarter of 2020, by Penetration Rate*. Available at: https://www.statista.com/statistics/867549/top-active-social-media-platforms-in-russia/ (accessed August 22, 2021).

24. Winkler W.E. *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage*. Washington,1990, 8 p.

Для цитирования

Олисеенко В.Д., Абрамов М.В., Тулупьев А.Л., Иванов К.А. Прототип программного комплекса для анализа аккаунтов пользователей социальных сетей: веб-фреймворк Django // Программные продукты и системы. 2022. Т. 35. № 1. С. 045–053. DOI: 10.15827/0236-235X.137.045-053.

For citation

Oliseenko V.D., Abramov M.V., Tulupyev A.L., Ivanov K.A. A software package prototype for analyzing user accounts in social networks: Django web framework. *Software & Systems*, 2022, vol. 35, no. 1, pp. 045–053 (in Russ.). DOI: 10.15827/0236-235X.137.045-053.