DOI 10.36622/VSTU.2022.25.2.005
УДК 004.02 : 004.05

ПОДХОДЫ К ОБНАРУЖЕНИЮ СОЦИАЛЬНЫХ ИНТЕРНЕТ-БОТОВ

А.О. Логинова

В данной статье представлен обзор подходов к обнаружению социальных интернет-ботов. Рассматриваются методики, основанные на выявлении автоматических или автоматизированных социальных акторов по следующим группам признаков: метаданные контролируемого ботом аккаунта, активность аккаунта, - также рассматриваются подходы к обнаружению ботов по совокупности признаков, в основе которых заложено машинное обучение (искусственный интеллект). Представленный обзор позволит выявить преимущества и недостатки существующих на сегодняшний день подходов к обнаружению социальных ботов, оценить возможность использования конкретного подхода в проектировании системы обнаружения ботов для определённого интернет-средства массовой коммуникации, оценить перспективы развития производства готовых отечественных решений по обнаружению социальных ботов. Статья подготовлена при поддержке Министерства науки и высшего образования Российской Федерации в рамках выполнения государственного задания в сфере науки № FSFU-2020-0020.

Ключевые слова: социальные интернет-боты, методика обнаружения ботов, информационно-психологическая безопасность, интернет-средства массовой коммуникации.

Введение

Особенности распространения информации в сети Интернет – возможность анонимизации источника информации – возрастающий интерес к Интернету, как к площадке для политической коммуникации, широкое применение практики информационного воздействия делают актуальной задачу обнаружения Интернет-ботов. Разработка методик обнаружения автоматических или автоматизированных акторов – одна из задач обеспечения цифрового суверенитета государства.

Особую опасность не только для индивидуума, но и для целых наций, государств и их отношений представляют собой социальные боты. Деятельность таких ботов оказывает большое влияние на участников интернет-коммуникации. Социальные боты могут выступать в качестве инструментов политической дестабилизации общества. Внедрение социальных ботов в коммуникативные пространства сети Интернет, позволяет как дезинформировать общество, так и изменять существующие в обществе коммуникативные установки [7].

До недавнего времени вопрос обнаружения Интернет-ботов не находил отклика среди исследователей, но вышеперечисленные факторы способствовали разработке методик обнаружения ботов, имитирующих поведение человека.

Начиная с 2016 года, представителями различных исследовательских коллективов публиковались работы, систематизирующие накопленные знания в области обнаружения интернет-ботов [1, 3, 5, 6, 7]. Описанные в этих работах подходы к обнаружению автоматических или автоматизированных социальных акторов основаны на выявлении групп признаков, проявляющихся в:

а) оформлении страницы аккаунта (далее – страница, персональная страница) социальной сети, который управляется ботом;

б) активности аккаунта: количестве сообщений, отправляемых за единицу времени, времени отправки сообщений, количестве пользователей, получивших/прочитавших сообщение.

Также существуют методики выявления социальных ботов, основанные на пересечении вышеуказанных групп признаков.

Работы, посвящённые обнаружению Интернет-ботов по содержательной части электронного сообщения, в открытых

источниках опубликованы не были. При этом разработка методики обнаружения ботов по лингвистическим характеристикам текстов сообщений представляется возможной, на что указывает ряд исследований, направленных на решение задачи идентификации интернет-пользователей по коротким электронным сообщениям [8, 9, 10, 11, 12].

Рассмотрим подробнее существующие подходы к обнаружению социальных ботов.

Обнаружение интернет-ботов по метаданным контролируемого аккаунта

Для создания аккаунта в социальной сети или другом интернет-средстве массовой коммуникации (далее – интернет-СМК) новому пользователю необходимо внести информацию о себе в анкету для регистрации, при этом достаточно заполнить только обязательные поля, игнорируя остальные. Обычно обязательными для заполнения, кроме логина и пароля, являются следующие поля: имя пользователя, ник, который будет использоваться для регистрируемого аккаунта, дата рождения, контактные данные необходимые для процедуры регистрации или восстановления доступа к аккаунту в случае утраты логина или пароля.

Стоит отметить, что не вся указанная при регистрации информация, метаданные, демонстрируется другим пользователям. Объём информации об аккаунте, с которым могут ознакомиться другие пользователи, зависит от самого интернет-СМК, наборе персональных настроек аккаунта.

Среди методик обнаружения социальных ботов можно выделить группу, имеющих в основе анализ метаданных аккаунтов пользователей в социальных сетях. Признаками, по которым распознаётся бот, в методиках такого типа являются (рис. 1):

– слово «bot», составляющее часть имени аккаунта (данный признак бота наблюдается чаще у полезных ботов);

– имя пользователя, состоящее из набора символов, не имеющего значения;

– наличие некоторой ссылки в ознакомительной части персональной страницы, позволяющей боту увеличивать количество посетителей сайтов, других аккаунтов социальной сети, количество скачиваний приложений или другого продукта;

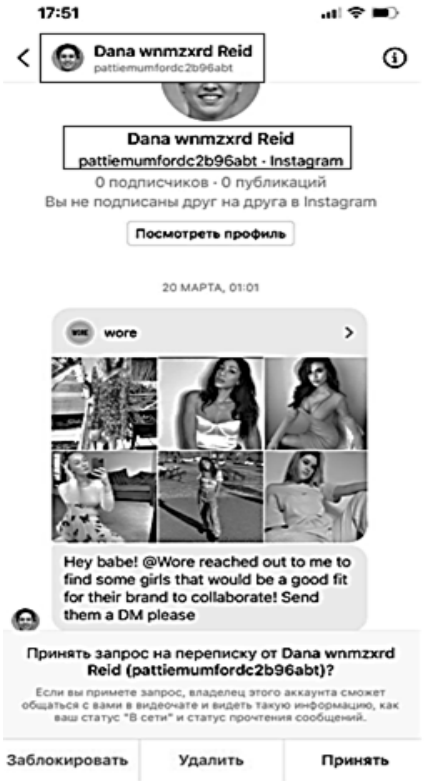

Рис. 1. Специфика оформления страницы контролируемого ботом аккаунта в Instagram[3]

– заполнение минимального количества полей для регистрации аккаунта, в связи с этим: отсутствие персонифицированной информации в ознакомительной части страницы, отсутствие фото владельца профиля, а также использование изображения или чужого фото (определить является ли изображение единственным в своём роде можно, выполнив поиск «по изображению» в Интернете посредством инструментов браузера);

– отсутствие соответствия между именем пользователя аккаунта и указанным полом;

[1] Социальная сеть Instagram запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

– отсутствие соответствия между именем пользователя и именем, используемым в ознакомительной части персональной страницы (используются разные имена).

Группа исследователей из Южного методического университета[4] разработала инструмент CATS для обнаружения ботов среди пользователей социальной сети Twitter[5], основанную на использования двух групп признаков ботов: особенности оформления персональной страницы и поведенческие признаки [1].

Авторы выделяют 18 идентификационных признаков, отличающих спам-бота от человека. Использование этих идентификационных признаков даёт точность обнаружения спам-ботов, равную 96%. Инструмент CATS позволяет обнаружить до 90% активных спам-ботов в социальной сети Twitter[3] по 5 опубликованным постам и до 50% - по одному посту.

Обнаружение интернет-ботов по активности, проявляемой контролируемым аккаунтом

Бот – автоматическая или автоматизированная программа, поэтому генерирование текстов сообщений и постов, создание рассылок, а также выполнение действии по выполнению подписки на тот или иной аккаунт не занимает много времени. Этот факт обуславливает высокую активность контролируемых ботами аккаунтов. Такие «пользователи» социальных сетей обычно подписаны на большое количество других аккаунтов, при этом сами они не имеют собственных подписчиков или лишь малое их количество. Также ситуация может быть обратной: бот популярен среди пользователей сети, тогда он имеет несоразмерное с количеством собственных подписок количество фолловеров[6] (рис. 2).

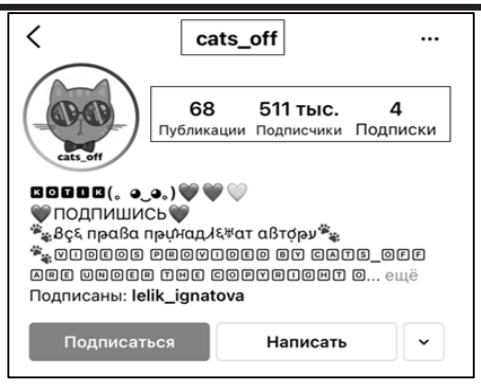

Рис. 2. Скриншот страницы аккаунта в Instagram[5], контролируемой ботом

Необходимо отметить, что речь не идёт о персональных страницах в социальных сетях, принадлежащих известным людям, на которых подписано большое количество почитателей их работ. Для сравнения на рис. 3 представлен скриншот страницы в социальной сети Instagram[7] Татьяны Владимировны Черниговской, доктора биологических наук, доктора филологических наук, член-корреспондента РАО.

Отличительной чертой бот-аккаунтов также является частота публикаций. Результаты исследования, проведённого представителями Оксфордского института Интернета, показали, что среднее ежедневное количество твиттов, превышающее 50 единиц – это свидетельство подозрительной активности [2]: в расчёте на 12 часов (длительность дня)

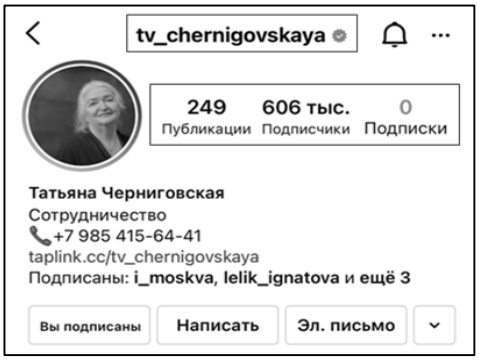

Рис. 3. Скриншот страницы аккаунта в Instagram [5] Т. В. Черниговской

[2] Southern Methodist University, Техас, США

[3] Социальная сеть Twitter запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

[4] Фолловер (англ. follower) – читатель, подписчик, друг, наименование зависит от используемой социальной сети. Пользователь социальной сети, подписанный на обновления страницы конкретного аккаунта

[5] Социальная сеть Instagram запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

выходит, что новая запись на странице в социальной сети публикуется каждые 15 минут. Стоит отметить, что такая запись в большинстве случаев содержит информацию о других аккаунтах, что способствует увеличению числа их подписчиков.

В качестве примеров подходов с использованием поведенческих идентификаторов интернет ботов можно отметить следующие:

– разработанная исследователями Южного методического университета методика обнаружения социальных ботов в Twitter[8], основанная на теории машинного обучения [3]. Алгоритм обнаружения бота включает анализ аккаунта в социальной сети и его активности для дальнейшего построения матрицы данных и выявления на её основе автоматизированного социального актора среди реальных пользователей социальной сети. В матрице данных, представленной в двоичной системе, фиксируется наличие у проверяемого аккаунта следующих признаков: отсутствие фото профиля, наличие метки геолокации, достижение отметки в более чем 1000 друзей/подписчиков, наличие постов и др. Погрешность в обнаружении бота при использовании такой методики составляет всего 2,25%;

– адаптивный алгоритм BotWalk, разработанный группой исследователей из Университета Нью-Мексико и Университета штата Мичиган, направленный на обнаружение социальных ботов в Twitter[6], поведение которых изменяется, в целях сокрытия своей активности. Работа этого алгоритма не нуждается в контроле человеком [4]. Данный алгоритм учитывает вектор развития поведенческих признаков ботов. Алгоритм позволяет регистрировать аномалии поведения аккаунтов в социальной сети и использовать сформированные шаблоны поведения или их комбинации для обнаружения ботов. Применение BotWalk делает возможным определение бота в социальной сети с точностью до 90%.

Комбинированные подходы к обнаружению интернет-ботов

Рассмотрим подробнее подходы к обнаружению ботов по совокупности признаков.

Группа исследователей из Университета Дрекселя[9] выделяет три категории подобных методик:

- основанные на теории сетей,
-имеющие в основе краудсорс анализ,
-использующие модели машинного обучения [5].

Такое разделение можно считать условным, поскольку методики очень близки по своей сути, отличие заключается в материале, используемом для исследования.

Методики поиска ботов, основанные на теории сетей – алгоритмы анализа социальных графов – учитывают связи пользователей социальных сетей. Считается, что люди чаще взаимодействуют с людьми, согласно этой логике, боты чаще взаимодействуют с ботами, распространяя информацию посредством репостов или рассылок. Так обнаружение одного социального бота может стать шагом на пути к раскрытию ботнета.

Примером реализации методики с применением анализа социальных графов является работа учёных Сеульского национального университета[10]. Для тестирования и отработки методики исследователи использовали базу данных CRESCI-2018, которая содержит 25 987 аккаунтов пользователей Twitter[6], подвергнутых бинарной классификации (бот, человек). Исследователи расширили данную базу за счёт аккаунтов, которые были подписаны или на которых были подписаны уже имеющиеся в базе пользователи. Так исследовательская база стала составлять 4,6 миллионов аккаунтов пользователей Twitter.

Учёными были отобраны атрибуты, на основе которых была создана модель социального графа исследуемых аккаунтов пользователей. В число атрибутов вошли: ID

[6] Социальная сеть Twitter запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

[7]Drexel University, Филадельфия, США

[8]Seoul National University, Сеул, Корея

пользователя, имя пользователя, ник пользователя, число подписчиков, число аккаунтов, на которых подписан пользователь, описание профиля, информация о локации и другие атрибуты. Сравнительный анализ значений атрибутов позволил выявить взаимосвязи исследуемых аккаунтов. С помощью визуализации полученной модели социального графа исследователям удалось с высокой точностью выделить аккаунты, контролируемые ботами во всём множестве аккаунтов исследуемой базы. [6]

Подходы к обнаружению ботов, имеющими в основе краудсорс анализ, во многом напоминают те, в которых используется анализ социальных графов. Отличие заключается в том, что применяется расширенный перечень атрибутов. Значения атрибутов характеризуют не только взаимосвязи пользователей социальных медиа, но и особенности самих профилей пользователей и их «подписчиков».

В качестве примера методики, в основу которой заложен краудсорс анализ, рассмотрим концепцию исследовательского коллектива, представленную в статье «Online human-bot interactions: Detection, estimation, and characterization» [7].

Исследователи использовали более тысячи атрибутов для оценки аккаунтов пользователей. Сами авторы выделяют среди них шесть групп, характеризующих: особенности аккаунта пользователя, «друзей» пользователя, сеть (авторы подразумевают характер взаимодействия пользователей: использование «лайков», упоминание аккаунта пользователя на своей странице, репост, оставление комментариев под постом), темпоральные особенности активности аккаунта, особенности контента, публикуемого пользователем в целом и уровень эмоциональности постов.

В ходе исследования авторы делают вывод о том, что с каждым годом социальные боты обретают всё более сложную конфигурацию; отсюда использование статической базы данных для отслеживания активности новых ботов не представляется возможной.

Специфика работы с большим данными предполагает использование методов машинного обучения (искусственный интеллект). В случае, когда речь идёт о выявлении высокоорганизованного бота в сети среди множества аккаунтов реальных людей, анализ большого объёма данных является неотъемлемой частью процедуры обнаружения.

В целях повышения доверия пользователей к социальной сети Facebook[11] исследователи Университета Дрекселя создали аннотированную базу политических социальных ботов, собранную из профилей пользователей, оставлявших свои комментарии в интернет-изданиях, освещавших президентскую гонку в США в 2015-2016 годах. Дифференциация социальных ботов и реальных людей, проявляющих интерес к статье, опубликованной интернет-изданием, сначала проводилась исследователями вручную, дабы отработать механизм классификации. Авторы работы выделяли три группы пользователей социальных медиа: люди, спаммеры, киборги (высокоорганизованные боты). Сложность возникала при отделении сообщений людей от сообщений киборгов.

В результате проделанной работы была создана модель машинного обучения, учитывающая особенности публикации комментария: время размещения, скорость ответов, средняя длина публикуемого текста, количество комментариев в день, частота использования ссылок на другие интернет-ресурсы и другое.

Обнаружение интернет-бота по содержательной части электронного сообщения

В основу интернет-СМК заложены различные концепции взаимодействия людей, это отражается на интерфейсе, архитектуре сети, возможностях презентации

[9] Социальная сеть Facebook запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

текста сообщения (имеется в виду возможность использования поликодового текста), оформлении профиля пользователя и других особенностях. Опираясь на это обстоятельство, исследователи отмечают, что на сегодняшний день не представляется возможной разработка единого алгоритма обнаружения активности интернет-бота для всех интернет-СМК [5]. Это обуславливает необходимость разработки методики обнаружения социальных ботов с учётом специфики интернет-СМК, в частности характере публикуемых сообщений.

Так текст сообщения/поста, сгенерированного ботом, можно рассматривать на предмет оригинальности: является ли оно репостом[12] или отдельно созданным оригинальным текстом.

В первом случае отличительным признаком аккаунта, контролируемого ботом, является полное или практически полное отсутствие оригинальных текстов сообщений. Такие «пользователи» часто используют опцию репоста чужих публикаций. Заимствованные таким образом публикации могут быть на нескольких языках (рис. 4).

Во втором случае речь идёт об оригинальных сообщениях, сгенерированных социальными ботами. Можно выделить особенности таких сообщений:

– односложность и примитивность предложений;

– наличие словосочетаний, используемых в мемах;

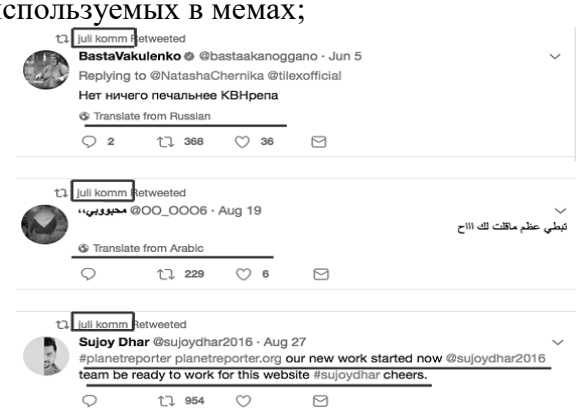

Рис. 4. Скриншот страницы социального бота в Twitter[13] [2]

– использование коротких фраз, вызывающих интерес, часто незаконченных, позволяющих владельцу бота повысить кликабельность[14] показатель эффективности рекламируемого сообщения, отражающий отношение числа переходов по указанной ссылке к количеству показов этого рекламного сообщения;

– подмена букв русского алфавита идентичными из латинского алфавита;

– отсутствием связи между контекстом публикации и комментарием бота к нему;

– наличие грубых грамматических, лексических и т.д. ошибок.

Идентификация пользователей по содержательной части электронных сообщений остаётся актуальной проблемой. Её разработкой занимались: Романов А.С., Мещеряков Р.В. [8], А.А. Воробьева [9, 10, 11], М.Е. Сухопаров [12]. Работы исследователей подтверждают необходимость применения междисциплинарного подхода к решению вопроса атрибуции текстов. Недостаточная интегрированность методов, применяемых в автороведческих исследованиях, может стать причиной недостаточной проработки вопроса и, как следствие, формирования ложных заключений.

Возвращаясь к вопросу идентификации ботов по содержательной части электронного сообщения, отметим следующее:

– обнаружение ботов посредством атрибуции текстов публикуемых ими коротких электронных сообщений возможна;

– данная задача усложняется ввиду особенностей сообщений ботов, которыми также могут отличаться сообщения человека с низкой культурой письменной речи.

Дублирование и распространение одних и тех же записей большим количеством аккаунтов может говорить о работе ботнета только в том случае, если аккаунты выборки имеют признаки Интернет-ботов, относящиеся как минимум к двум различным

[10]Репост (англ. repost) – повторное опубликование чего-либо чужой записи на своей странице в социальной сети.

[11] Социальная сеть Twitter запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

[12] CTR – click through rate.

группам. В противном случае это проявление естественной активности пользователей социальной сети.

Заключение

В статье рассмотрены группы подходов к обнаружению социальных ботов по следующим признакам: метаданные контролируемого ботом аккаунта, аномальная активность, проявляемая аккаунтом, комбинированные подходы к обнаружению социальных ботов, основанные на использовании технологии машинного обучения. Стоит отметить, что анализ и классификация существующих подходов к обнаружению ботов проводились только на материале исследований, которые проводились за рубежом, поскольку в открытых источниках нет информации об отечественных разработках.

Комбинированные подходы к обнаружению ботов демонстрируют наибольшую эффективность. Результаты обнаружения ботов по нескольким атрибутам с использованием технологии машинного обучения отличаются незначительным количеством ошибок 1-го и 2-го рода.

В статье также приводится оценка перспективы разработка методики обнаружения социальных ботов по содержательной части генерируемого им электронного сообщения.

Результаты анализа состояния проблемы обнаружения интернет-ботов показали её ежегодную эскалацию [13]. Несмотря на большое количество разнообразных методов идентификации социальных ботов, проблема их обнаружения не утратила своей актуальности из-за постоянно развивающихся технологий. Это обуславливает необходимость разработки новых методик обнаружения автоматических или автоматизированных акторов для повышения киберустойчивости и поддержания цифрового суверенитета, как одного из направлений реализации информационной функции государства [14].

Своевременное обнаружение активности социального бота позволит оградить пользователей социальных сетей, форумов, блогов и др. интернет-СМК от информационно-психологического воздействия, имеющего целью дестабилизацию общества, введения его в состояние, при котором оно поддаётся манипуляциям. Блокировка социального бота, чья деятельность направлена на искажение восприятия людьми событий, происходящих в особенности в сфере финансов, образования и государственного управления, позволит снизить внутренние угрозы киберустойчивости государства.

Список литературы

1. Amleshwaram A.A., Reddy N., Yadav S., Gu G., CATS: Characterizing automation of Twitter[15] spammers // Fifth International Conference on Communication Systems and Networks (COMSNETS). 2013. – С. 10.1109/COMSNETS.2013.6465541.

2. Двенадцать способов распознать бота//Беллингкэт URL:https://ru.bellingcat.com/materialy/putevoditeli/2017/09/22/botspot/ (дата обращения: 20.05.2020).

3. Efthimion P. G., Payne S., Proferes N. Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter[13] Bots // SMU Data Science Review: Vol.1: No.2, Article 5.

4. Minnich A., Chavoshi N., Koutra D., Mueen A. BotWalk: Efficient Adaptive Exploration of Twitter[13] Bot Networks // IEEE/ACM International Conference on Advances in Social Networks and Mining – 2017. Doi.org/10.1145/3110025.3110163/

5. Santia G. C., Mujib M. I., Williams J. R. (2019). Detecting Social Bots on Facebook in an Information Veracity Context. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 463-472.

6. Bebensee B, Nazarov N, Zhang B-T Leveraging node neighborhoods and egograph topology for better bot detection in social graphs // Social Network Analysis and Mining 2021, – 11:10, – p.1–14.

[13] Социальная сеть Twitter запрещена в России. Meta признана экстремистской организацией, её деятельность в России запрещена.

7. Varol O., Ferrara E., Davis C. A., Menczer F., Flammini A. Online human-bot interactions: Detection, estimation, and characterization // arXiv preprint arXiv:1703.03107 (2017)

8. Романов А.С., Мещеряков Р.В. Определение пола автора короткого электронного сообщения // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог» (Бекасово, 25–29 мая 2011 г.). – М: Изд-во РГГУ. – 2011. – Вып. 10 (17). – С. 620–626

9. Воробьева А.А. Отбор информативных признаков для идентификации интернет-пользователей по коротким электронным сообщениям // Научно-технический вестник информационных технологий, механики и оптики. – 2017. – Том 17, № 1. – С. 117-128.

10. Воробьева А.А. Методика идентификации интернет-пользователя на основе стилистических и лингвистических характеристик коротких электронных сообщений // Информация и космос. – 2017. – № 1. – С. 127-130.

11. Vorobeva A.A. Forensic linguistics: automatic web author identification // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. – 2016. – Т. 16. № 2. – С. 295-302.

12. Сухопаров М.Е. Методика идентификации пользователей порталов сети Интернет на основе методов математической лингвистики: дис. ... канд. тех. наук: 05.13.19. - СПб, 2015.

13. Логинова А.О. Анализ существующих подходов к классификации и типологии ботов // Инновационные технологии: теория, инструменты, практика. – 2020. – Том 1 – С. 462-467.

14. Волков С.Д., Царегородцев А.В. Один из подходов к обеспечению защиты от компьютерных атак при реализации информационной функции государства на внутреннем уровне // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2020. – № 36. – С. 159-174.

Московский государственный лингвистический университет
Moscow State Linguistic University

Поступила в редакцию 26.04.2022

Информация об авторе

Логинова Алина Олеговна – аспирант кафедры международной информационной безопасности Института информационных наук, эксперт отдела научного менеджмента и наукометрии, Московский государственный лингвистический университет, e-mail: loginova@linguanet.ru

APPROACHES TO DETECTING SOCIAL INTERNET BOTS

A.O. Loginova

This article provides an overview of existing approaches to detecting social Internet bots. The methods based on identification of automatic or automated social actors by the following groups of signs: metadata of an account controlled by a bot, account activity; approaches to detecting bots by a set of signs based on machine learning (artificial intelligence) are also considered. This review reveals advantages and disadvantages of existing approaches to social bots detection, promotes a possibility to assess using a specific approach to design a bot detection system for a certain social net, a possibility to assess prospects for the development of production of solutions for social bots detection.

Keywords: social bots, methods of detecting bots, information-psychological security, Internet mass media.

Submitted 26.04.2022

Information about the author

Alina O. Loginova – postgraduate student of the International Information Security Department of the Information Sciences Institute of Moscow State Linguistic University, expert of the Department of Scientific Management and Scientometrics of Moscow State Linguistic University, e-mail: loginova@linguanet.ru