

УДК 519.688

Создание базы данных и разработка архитектуры системы для выявления участников "групп смерти"

Я. Р. Мустакимова, Е. Ю. Никитина, А. Д. Турпанова

Пермский государственный национальный исследовательский университет

Россия, 614990, г. Пермь, ул. Букирева, 15

mustakimovayr@gmail.com, neyu@psu.ru; +79028305928

Рассматривается развитие "групп смерти", описываются системы поиска игроков и системы заданий для участников таких групп. Также анализируются различные способы борьбы с "группами смерти". Описаны результаты первых этапов по разработке программного обеспечения для определения участников "групп смерти".

Ключевые слова: "группы смерти"; социальные группы; разработка программного обеспечения.

DOI: 10.17072/1993-0550-2018-3-117-123

Введение

Одной из самых тревожных и актуальных в России проблем на сегодняшний день остается деятельность "групп смерти" в социальных сетях, жертвами которых становятся дети и подростки. В 2017 г. было выявлено 1339 сетевых сообществ суицидальной направленности, общая аудитория которых составила более 12 тысяч пользователей, и насчитывающих более 200 тысяч публикаций в социальных сетях [1].

"Группы смерти" – это группы в социальных сетях, участников которых подталкивают к суициду. О явлении "групп смерти" стало широко известно благодаря материалу "Новой газеты". Согласно информации, сообщества под названиями "Тихий дом", "Море китов", f57 и другие представляли собой закрытые группы, вовлекавшие подростков в игру, финалом которой становился суицид.

Для того чтобы начать играть, подросток должен был опубликовать в сети определенную запись. После с ним связывался куратор, который давал задания, а в случае их невыполнения угрожал подростку удалением из группы или физической расправой. Куратор

мог попросить, например, вырезать кита на своей руке или проснуться в 4.20 утра. По словам "Новой газеты", "группы смерти" – это настоящие секты, заманивающие, а затем удерживающие подростков шантажом и угрозами. В расследовании говорилось о 130 подростках, погибших из-за действий подобных групп, правда, цели, которые преследовали администраторы сообществ, так и остались неясными [2].

Вступившие в "группу смерти" принимают участие в "увлекательной игре" и исполняют, как может сначала показаться, безобидные задания. На самом же деле несовершеннолетний оказывается в психологической ловушке. В сообществе пропагандируется смерть как единственный верный и красивый способ выхода из сложных жизненных ситуаций. Насилие по отношению к себе и окружающим, а также депрессивное психологическое состояние возводится в культ, тогда как семья, друзья и образование критикуются и высмеиваются.

1. Системы поиска игроков

За несколько лет системы поиска игроков претерпевали изменения.

Сначала кураторы искали новых игроков следующим образом: на странице группы

делалась закрепленная запись – текстовый пост или картинка. Те, кто хотел вступить в игру, оставляли невинный комментарий под этой записью: хочу играть, я в игре, разбудите меня в 4.20. Но Роскомнадзор быстро осознал, что это такое, и начал такие сообщества блокировать. Впоследствии схему поиска организаторов усложнили.

Сейчас игроков ищут по хештегу. Подросток уже ничего не должен писать на странице группы, он оставляет определенную запись на своей личной странице. Это серьезно осложняет работу Роскомнадзора. Ведь он не может заблокировать страницу простого пользователя – на ней нет пропаганды суицида. К тому же непонятно, к какой группе эта цепочка ведет и откуда ее куратор [3].

2. Системы заданий

С течением времени эволюционировали не только системы поиска игроков, но и сами задания. На сегодняшний день создано несколько вариантов этой "игры".

Первый вариант, можно назвать его базовым, содержит 10 уровней. Первые три уровня – безобидные. Куратор может дать задание разгадать ребус, прослушать медитационную музыку (он сам отправляет ее игроку личным сообщением), написать короткое эссе или стихотворение на тему бренности бытия. Начиная с четвертого уровня задания кураторов становятся сложнее и страшнее. Все задания для "игроков" так или иначе связаны с нанесением себе ран. Причем теперь фото просто руки с порезами не имеют ценности – надо сделать фотографию с доказательством: на фото должны быть видны не только порезы, но и лицо игрока на фоне его личной страницы в социальной сети. Периодически куратор может отправлять подопечному смс-сообщения со странными советами о том, как не проспать сеанс "игры". Например: "У тебя болит голова. Выпей семь таблеток, и я разбужу тебя в 4.20". Последний, высший уровень "игры" – якобы инсценировка собственной смерти. По условиям игры это должно быть максимально правдоподобным.

Второй вариант "игры" – более сложный, с дополнительными заданиями. Например, куратор может сказать, что пошлет к "игроку" "гида", который поможет сделать последнее задание, и, возможно, будет в роли репортера. Эти "гиды" якобы должны зафик-

сировать инсценировку самоубийства, поскольку оно ведь должно быть максимально правдоподобным.

Третий вариант "игры" – с разбивкой по возрастам. "Старшим" по возрасту "игрокам" предлагают быть "избранными". Для "игроков" "младшего" возраста – другой сценарий. Им кураторы не только пишут сообщения, но и посылают картинки с героями любимых мультфильмов, которые дают конкретные призывы к суицидам с описанием способов их совершения [3].

3. Способы борьбы

Борьба с такими группами ведется на государственном уровне. В июне 2017 г. был подписан закон об уголовной ответственности за склонение детей к суициду. Поправки в Уголовный кодекс предусматривают до шести лет заключения за создание "групп смерти" и до трех лет для тех, кто вовлекает подростков в смертельно опасные игры.

Согласно документу, в статью 110 УК РФ ("Доведение до самоубийства") добавлен пункт о совершении этого преступления в отношении несовершеннолетних и беременных через СМИ, Интернет и публично демонстрируемые произведения. За это предусматривается от пяти до восьми лет заключения. Также в кодекс добавлены новые статьи: 110.1 ("Склонение к совершению самоубийства или содействие совершению самоубийства"), 110.2 ("Организация деятельности, сопряженной с побуждением граждан к совершению самоубийства") и 151.2 ("Вовлечение несовершеннолетнего в совершение действий, представляющих опасность для жизни несовершеннолетнего"). Освобождение от уголовного преследования могут получить те, кто добровольно сдастся полиции и поможет следствию.

По статье 151.2 ответственность в виде штрафа до 80 тыс. руб. или года тюрьмы грозит тем, кто уговаривает подростка сделать что-либо потенциально опасное для жизни. А тем, кто это делает через Интернет, уже могут присудить до трех лет тюрьмы.

Несмотря на введенные меры безопасности, проблема остается. Роскомнадзор блокирует одну "группу смерти", но на ее месте тут же появляется другая. В последнее время такие группы все чаще стали делать закрытыми, а Роскомнадзор не может заблокировать

подобные сообщества. Кроме того, заблокировать можно только сайт с явно противоправным контентом. Это значит, что страница должна содержать прямые призывы к суициду. Фразы вида "жизнь потеряла смысл" или "ничего не изменится, если тебя здесь не будет" легальны. Тем более нельзя никого оштрафовать или посадить в тюрьму за картинку, будь то парящие в небе киты или фото руки с порезами. Хотя последний вариант – это уже шок-контент. Он, так или иначе, противоречит закону о защите детей от вредоносной информации. Но и тут все решается элементарно: в посте ставится метка "18+". Зритель предупрежден, что ему такое смотреть не положено. Если все же посмотрел, последствия на его совести.

Еще одна проблема в осуществлении борьбы – сроки. Уничтожить все следы о себе и своей деятельности в виртуальном мире можно за несколько секунд. Однако для того, чтобы возбудить уголовное дело, требуется не одна неделя. По обращениям на основании снимков переписки кураторов с жертвами Роскомнадзор блокируют опасные страницы в течение суток. А вот прокуратура и следственные органы обрабатывают точно такие же обращения целый месяц и только потом могут возбудить уголовное дело. А это значит, что у манипулятора есть целых 30 дней, чтобы свободно привлекать новичков в "игру" или просто раствориться среди миллионов других пользователей Интернета – для этого можно сменить фото на личной странице, переименовать группу и удалить компрометирующую переписку.

При этом тот, кто закрывает "группы смерти" в социальных сетях, всегда будет в позиции догоняющего. Однако один способ борьбы с кураторами все же найден. Данный способ заключается в следующем: необходимо создать тысячу ботов (фиктивных страниц несуществующих пользователей), боты начинают публиковать хештеги, по которым кураторы ведут поиск, раз в час или чаще.

Таким образом, заявок на "игру" от ботов становится в тысячи раз больше, чем от реальных игроков. Живые пользователи растворяются, и куратор их просто не сможет найти [3].

На сегодняшний день существует несколько способов выявления "групп смерти" и их участников.

Роскомнадзор, Роспотребнадзор, эксперты-психологи и администрация социальных сетей составили пособие, которым могут пользоваться сотрудники специализированных служб поддержки, которые проводят мониторинг социальных сетей для выявления групп, публикующих призывы совершать самоубийства. В некоторых регионах России созданы целые межведомственные группы для мониторинга социальных сетей. Одной из основных проблем в работе по выявлению "групп смерти" в сети является крайне большой объем материалов для мониторинга. Люди просто не успевают просмотреть все группы и проанализировать их на содержание противоправного контента. Другими проблемами являются закрытый характер групп и специфический язык общения между куратором и участниками.

Также Национальным центром информационного противодействия терроризму и экстремизму в образовательной среде и сети Интернет была создана памятка для школ и родителей о том, как можно выявить вовлеченность ребенка в "группы смерти". В данной памятке описаны признаки, по которым можно определить является ли ребенок участником "группы смерти", а также действия, которые надо предпринять при обнаружении опасных признаков.

Таким образом, на данный момент выявить "группу смерти" или ее участников могут только сами люди, процесс выявления происходит исключительно "вручную", никаких автоматизированных средств на данный момент не существует.

В настоящей работе мы описываем способ построения программного средства, позволяющего автоматически обнаруживать участников "групп смерти".

4. Создание базы данных

Еще одним возможным выходом из сложившейся ситуации является создание программного обеспечения для определения участников "групп смерти". Данная система может использоваться государственными органами или быть частью системы родительского контроля.

Первым этапом в разработке такой системы является создание базы данных, на основании которой можно определить является ли подросток участником "группы смерти".

Для того чтобы наполнить базу данных, было проведено пробное анкетирование. Анкета содержала в себе вопросы, затрагивающие разные ситуации, которые могут произойти с подростком и вызвать у него различные эмоции, например:

- 1) отношения с родителями;
- 2) ситуация, когда бы кто-нибудь Вас уважал/унижал, оскорблял;
- 3) ситуация, когда бы Вы разочаровались в ком-либо (в чем-либо), перестали верить кому-либо (во что-либо);
- 4) ситуация, когда бы Вас хорошо понимали/не понимали, не находили общего языка с кем-либо;
- 5) ситуация, когда бы Вы чувствовали одиночество/были "не одни" перед какой-то Вашей проблемой.

Для каждой ситуации было предложено описать ощущения, переживания, эмоции, чувства, рассуждения, мысли, решения, воспоминания, которые возникали в этот момент.

В анкетировании принимали участие подростки от 14 до 20 лет, проживающие в разных районах города Перми и Пермского края: Свердловском, Ленинском, Орджоникидзевском, Кировском, Индустриальном, Мотовилихинском, Дзержинском, Карагайском. В опросе принимали участие как девушки, так и юноши. Всего было опрошено 350 человек.

На основе полученных ответов был составлен список слов, которые подростки употребляли чаще всего. Данный список и будет являться наполнением базы данных. Двадцать первых наиболее употребляемых слов представлены в таблице.

Наиболее употребляемые слова

Слово	Вероятность, %
понимать	0,989
друг	0,951
хороший	0,878
чувствовать	0,840
родители	0,774
чувство	0,614
ситуация	0,594
отношение	0,528
хотеть	0,482
уважать	0,472
жизнь	0,437
проблема	0,434
одиночество	0,382
никогда	0,361
мочь	0,358
момент	0,354
унижать	0,333
общаться	0,312
любовь	0,292
думать	0,278

Наполнение базы данных по результатам анкетирования было выполнено в экспериментальном порядке, оно не претендует на 100 %-ную истинность с точки зрения набора выявленных ключевых слов. Целью данного этапа работы было получить хоть какое-то наполнение базы данных, чтобы можно было проводить дальнейшую разработку программной системы и пытаться анализировать, насколько корректно определяются участники "групп смерти" и определяются ли они вообще. Для реальной реализации программного обеспечения будет проводиться профессиональное анкетирование.

5. Требования к системе

Полагаем, что разрабатываемая система должна позволять собирать и анализировать открытую информацию, размещенную в социальных сетях, для автоматического определения участников "групп смерти". При этом для анализа будут использоваться открытые данные: различные записи в группах, записи на страницах пользователей, комментарии, репосты, "лайки". Поиск в текстах будет осуществляться в соответствии с полученной базой данных.

Система должна удовлетворять следующим требованиям:

1. Скорость работы системы. К примеру, база данных социальной сети Facebook на сегодняшний день содержит более 2 млрд пользовательских аккаунтов. Каждую минуту пользователи добавляют более 300 тыс. статусов и оставляют более 500 тыс. комментариев к различным объектам сети [4]. На сегодняшний день большинство существующих алгоритмов, позволяющих эффективно решать актуальные задачи, не способны обрабатывать данные подобной размерности за приемлемое время. В связи с этим возникает потребность в решениях, позволяющих осуществлять распределенную обработку и хранение данных без существенной потери качества результатов. Размерность данных обуславливает необходимость в параллельном методе сбора данных [5].

2. Расширяемость системы. Например, при необходимости анализа содержимого новой социальной сети должна быть возможность сравнительно легкого добавления в систему нового сборщика данных из этой социальной сети.

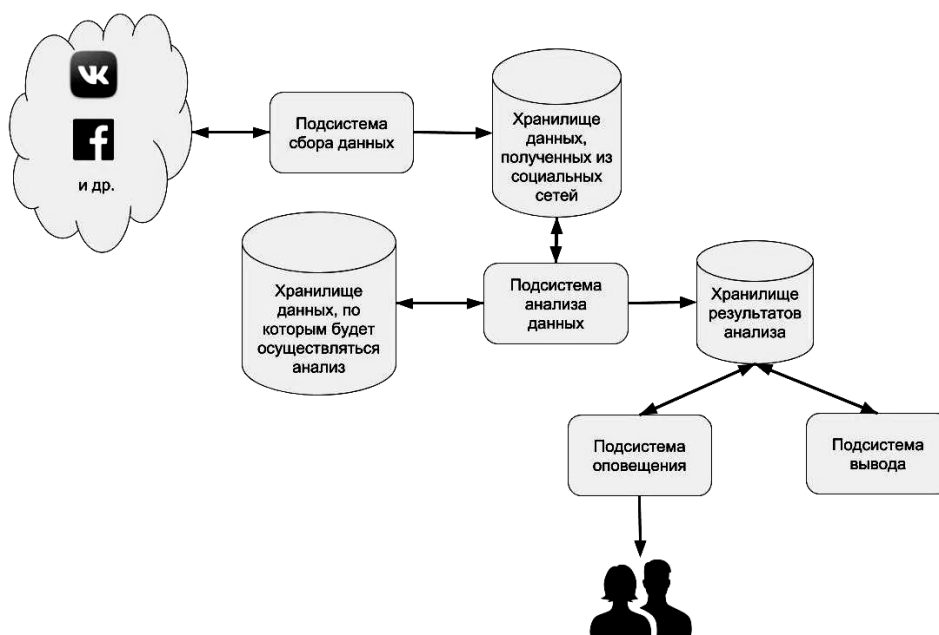
6. Общая архитектура приложения

Для того чтобы разрабатываемая система соответствовала заявленным требованиям, в качестве основы была выбрана сервис-ориентированная архитектура.

Сервис-ориентированная архитектура (англ. Service-Oriented Architecture, SOA) – это подход к разработке программного обес-

печения, основанный на использовании сервисов (служб) со стандартизированными интерфейсами. Компоненты программы могут быть распределены по разным узлам сети и предлагаются как независимые, слабо связанные, заменяемые сервисы приложения [6].

Схема архитектуры разрабатываемой системы представлена на рисунке.



Составные части системы

Рассмотрим более подробно основные подсистемы разрабатываемой системы:

1. Подсистема сбора данных

Эта подсистема осуществляет сбор информации из социальных сетей. Поскольку разрабатываемая система должна обладать возможностью расширения функционала, а именно возможностью интеграции с различными социальными сетями, имеет смысл реализовать эту подсистему как набор сервисов, каждый из которых интегрируется с конкретной социальной сетью.

Сбор информации таким сервисом может происходить с использованием API социальной сети при наличии, либо с помощью разбора HTML-страниц. Внутренняя архитектура такого сервиса будет определяться индивидуально, в соответствии со способом интеграции.

Каждый сервис интеграции с социальной сетью должен записывать полученную

информацию в хранилище с помощью вызовов API этого хранилища.

2. Подсистема хранения

Подсистема хранения решает задачу промежуточного хранения данных. Эта подсистема предоставляет программный интерфейс, который может использоваться другими подсистемами.

Так, указанный интерфейс должен предоставлять возможности для записи данных для анализа, эти возможности будут использованы сервисами подсистемы сбора. Для сервисов подсистемы анализа этот интерфейс должен предоставлять возможности чтения данных для анализа.

3. Подсистема анализа данных

Задача этой подсистемы сводится к тому, чтобы получить данные из подсистемы хранения, проанализировать их и записать результаты в подсистему хранения результатов

анализа. Анализ будет производиться в соответствии с ключевыми словами базы данных.

Эта подсистема может быть реализована в виде различных независимых программ, работающих в разных процессах и предоставляющих различные способы анализа. Такие модули могут выполняться независимо друг от друга на разных компьютерах кластера, что позволит горизонтально масштабировать систему, обеспечивать отказоустойчивость и внедрять новые методы анализа в режиме реального времени (используя механизмы непрерывной интеграции).

4. Подсистема оповещения

Задача подсистемы оповещения сводится к уведомлению администраторов системы и пользователей о событиях, генерируемых системой.

Например, о том, что выявлен потенциальный участник "группы смерти". Она может быть реализована также в виде независимых модулей, которые с помощью запросов к API хранилища результатов анализа получают результаты анализа и оповещают внешних акторов с помощью различных механизмов. Например, возможна реализация отправки письма на почту администратора системы, запрос к внешней программной системе, отправка уведомлений в специальные чаты и т.д.

5. Подсистема вывода

Эта подсистема представляет собой интерфейс (или несколько интерфейсов) для взаимодействия пользователей с системой. Возможна реализация web-приложения, которое сможет отображать различные отчеты в понятном пользователю виде.

Например, можно отображать информацию о конкретном пользователе (статистика по комментариям, "лайкам", записям), информацию о конкретной записи в группе (кто поставил "лайк", кто написал комментарий, кто переслал запись), статистику по определенным ключевым словам.

7. Выбор технологий

Способ взаимодействия между модулями разрабатываемого комплекса должен представлять собой стандартизированный протокол обмена сообщениями. Было решено использовать паттерн RESTful API, использующий JSON для представления объектов.

При этом подходе интерфейс основных сервисов (API хранилищ) будет реализован с помощью глаголов HTTP, передача контента будет осуществляться с помощью JSON. JSON выбран потому, что он обладает минимальной, в сравнении с аналогичными форматами, избыточностью. При этом он предоставляет полные возможности для сериализации различных объектов.

В качестве СУБД для организации хранилища имеет смысл рассмотреть специализированные нереляционные хранилища, предоставляющие функции полнотекстового поиска. Примером такого хранилища является ElasticSearch.

API хранилища исходных данных должен удовлетворять следующим требованиям:

- 1) возможность одновременной обработки большого количества запросов. Поскольку предстоит обрабатывать большие объемы данных, количество запросов к API со стороны подсистемы сбора и подсистемы анализа будет достаточно большим;
- 2) простота интеграции с ElasticSearch;
- 3) предоставлять интерфейс в терминах предметной области.

Исходя из перечисленных требований, было решено использовать Node.js, так как он реализует асинхронную событийно-ориентированную модель, благодаря чему становится возможным одновременная обработка большого количества запросов в одном потоке.

Таким образом, снижаются затраты на обслуживание потоков и количество одновременно обрабатываемых запросов растет в сравнении с аналогами, которые обрабатывают каждый запрос в отдельном потоке (Java, C#, PHP).

Для реализации модуля анализа имеет смысл использовать язык, обладающий развитой инфраструктурой для реализации различных стратегий искусственного интеллекта, интеллектуального анализа.

На сегодняшний день таким языком является Python – он предоставляет огромное количество библиотек для реализации семантического и синтаксического анализа текста, построения и работы с нейронными сетями и глубоким обучением, библиотеки для работы с генетическими алгоритмами и другими стратегиями искусственного интеллекта.

Заключение

Задача настоящего этапа разработки программного обеспечения – разработка его принципиального проекта – выполнена. Задача следующего этапа – собственно реализация программного обеспечения и наполнение его реальными данными.

Список литературы

1. *В России за год выявили почти 1,4 тысячи "групп смерти"*. URL: <https://ria.ru/society/20171003/1506087897.html> (дата обращения: 31.01.2018).
2. *Группы смерти*. URL: <https://www.novayagazeta.ru/articles/2016/05/16/68604-gruppy-smerti-18> (дата обращения: 31.01.2018).
3. *Охотники на "группы смерти" вооружились хакерским арсеналом*. URL: <http://www.mk.ru/social/2017/03/01/okhotniki-na-gruppy-smerti-vooruzhilis-khakerskimarsenalom.html> (дата обращения: 31.01.2018).
4. *Слишком большие данные: сколько информации хранится в интернете?* URL: <https://ain.ua/special/skolko-vesit-internet/> (дата обращения: 1.02.2018).
5. *Анализ социальных сетей: методы и приложения*. URL: http://www.ispras.ru/proceedings/docs/2014/26/1/isp_26_2014_1_439.pdf (дата обращения: 1.02.2018).
6. *Сервис-ориентированная архитектура*. URL: http://plmpedia.ru/wiki/Сервис-ориентированная_архитектура (дата обращения: 1.02.2018).

Creating a database and developing the architecture of a system to identify "death groups" participants

Ya. R. Mustakimova, E. Yu. Nikitina, A. D. Turpanova

Perm State University; 15, Bukireva st., Perm, 614990, Russia
mustakimovayr@gmail.com, neyu@psu.ru; +79028305928

This article deals with the activity of "death groups", it describes the systems for finding players and systems of tasks given to members of such groups. Various methods of combatting "death groups" activity are analyzed. The results of the first stages of software development for identifying the participants of "death groups" are presented.

Keywords: "death groups"; social groups; software development.