

Оглавление

1.1	Введение	2
1.2	Основной функционал приложения	5
1.2.1	Цели и задачи дипломного проекта	5
1.3	Анализ существующих решений	6
1.3.1	smm-продукты	7
1.3.2	социальные агрегаторы	7
1.3.3	web-приложения для поиска людей	10
1.3.4	Сервисы анализа сообществ и трендов в социальных сетях	12
1.3.5	Приложения для платформы vk.com	18
1.3.6	Выводы	20
1.4	Обоснование выбора технических инструментов для реализации приложения	21
	Литература	24
1.5	Список терминов	24

1.1 Введение

В настоящий момент довольно остро стоит вопрос о сохранении тайны связи при использовании электронной почты, чата, социальных сетей и иных электронных средств коммуникаций. Закон о сохранении тайны связи не охватывает публичные сервисы.¹ Помимо этого, опубликованные Эдвардом Сноуденом данные наглядно демонстрируют, что межправительственные системы слежения (созданные для борьбы с терроризмом) используются для достижения экономических и политических целей, нарушая права граждан на тайну частной жизни и тайну переписки.

Люди часто недооценивают значение метаданных и комплексного анализа. Под комплексным анализом далее будет подразумеваться сочетание методов, подходов, инструментов по интеллектуальной добыче данных (Data Mining), обработки больших объемов неструктурированных и слабосвязанных данных (Big Data).

Зачастую остается неосознанным тот факт, что вступая в электронные сообщество человек переносит элементы реального общения в виртуальное. На первый взгляд может показаться, что изменяется лишь форма взаимодействия, но на самом деле меняется его суть. К примеру, при реальной встрече двух людей, если это встреча не была кем либо еще подслушана, то она остается известной только участникам и тем, кого они информировали об этом. В социальной же сети сервер выступает безоговорочным трентом (доверенным центром), который хранит информацию обо всех происходящих событиях.² Таким образом, личная жизнь человека становится синхронизированной с электронным сервером и человек, обладающий личностью, превращается в очередного пользователя системы. Так как одновременно происходят миллионы событий на первый план

¹ Тайна связи, электронная почта и российские суды (<http://www.securitylab.ru/blog/personal/emeliyannikov/37733.php>)

²далее по тексту я предполагаю под событием любое взаимодействие с сервером социальной сети, например факт авторизации, просмотр страницы, написание сообщения, создание группы, комментариев и т.д. - в общем все то, что можно делать с помощью социальной сети

выходят метаданные: кто вступал во взаимодействия, когда, как долго, характер взаимодействия, метайнформация касательно события (количество символов в сообщении, количество участников встречи и прочее). Значительную часть метайнформации об отдельном человеке возможно проанализировать даже не являясь трендом, потому что сама природа социальных сетей заставляет пользователей делиться информацией (конечно если пользователь специально не установил некоторый уровень приватности).

Один из аспектов дипломной работы будет демонстрировать описанный выше недостаток следующим образом: эмпирически автором было установлено, что каждая социальная группа имеет свои характерные особенности анкеты, которые можно проанализировать автоматизированным способом, а именно, пользователи в основном добавляют в друзья людей своего возраста, живущих с ними в одном городе. Так же, если проранжировать по частоте встречаемости учебные заведения в которых учатся друзья некоторой персоны, то наиболее часто встречается то учебное заведение, в котором учится исследуемая анкета. Что бы подтвердить данное наблюдение было произведено исследование тысячи случайно взятых людей из социальной сети Вконтакте. Подробнее это исследование будет рассмотрено в следующих главах. Другим интересным фактом, который был замечен, исследован и добавлен в программную часть анализатора анкеты является, то что пользователь чаще всего упоминает в своих сообщениях город в котором живет, то же касается и возраста - как правило пользователи пишут людям своего возраста, а так же теми с кем они вместе работают или учатся. Последние достижения в области обработки естественных языков позволяют автоматически извлекать факты из произвольного текста написанного на естественном языке. Тут стоит упомянуть научную работу сотрудников IBM Research во главе с Джалал Махмудом (Jalal Mahmud) в которой демонстрируют возможность опередить местонахождение человека по его постам в Twitter с точностью до 70% (определяется обычно

город или округ).³ Основная идея алгоритма придуманного сотрудниками IBM заключается в том, что само содержание твитов несет в себе информацию о местонахождении и современные инструменты позволяют ее извлечь. Так например в посте может быть ссылка на фото или пост в другой социальной сети в которой отмечена гео-информация, кроме того анализируются семантика текста для извлечения фактов, например из текста "Сергей не забудь самовар, встречаемся в Туле" Можно извлечь следующие факты: Место - Тула, Объект - Сергей. Всю необходимую информацию исследователи извлекают напрямую из Twitter с помощью Streaming API в основном используя GET statuses/firehose⁴)

Таким образом вторым косвенным признаком который будет анализироваться дипломным продуктом будут являться комментарии и записи пользователей.

Целью данного дипломного проекта является создание веб-приложения, демонстрирующего различные возможности по сбору сведений об отдельном человеке с использованием только открытых источников информации.^{5 6} Особенно интересным представляется создать веб-сервис для автоматизированного анализа страницы в социальной сети с выделением дополнительных сведений о человеке, на основе сведений о его друзьях. Веб-сервис создается как инструмент, позволяющий оценить значимость косвенных данных в процессе верификации анкеты пользователя социальной сети.

Данный сервис задуман с целью проверки оценки уровня защищенности персональной информации, которую пользователь оставляет конфиденциальной становясь участником виртуального сообще-

³Who will retweet this?: Automatically Identifying and Engaging Strangers on Twitter to Spread Information (digital.cs.usu.edu/~kyumin/pubs/lee14iui.pdf\beginingroup\let\relax\relax\endgroup[Pleaseinsert\PrerenderUnicode{\BFF}intopreamble])

⁴подробнее см. Twitter Rest API (<https://dev.twitter.com/docs/api/1.1>)

⁵Правительство США предало интернет. Нам надо вернуть его в свои руки (<http://habrahabr.ru/post/192852/>)

⁶Эдвард Сноуден (http://ru.wikipedia.org/wiki/\T2A\CYRS\T2A\cyrn\T2A\cyro\T2A\cyru\T2A\cyrd\T2A\cyre\T2A\cyrn,_T2A\CYREREV\T2A\cyrd\T2A\cyrv\T2A\cyra\T2A\cyrr\T2A\cyrd)

щества, но которая может быть получена в результате анализа косвенных источников.

Данный сервис не является социально опасным по следующим причинам:

- пользователь сервиса имеет возможность анализа только той страницы, для которой известны данные авторизации;
- сервис безопасен для пользователя т. к. авторизация происходит по средствам API социальной сети и данные авторизации не передаются на сервер приложения;
- мировой опыт показывает, что уже созданы куда более мощные средства для анализа данных. Однако, все они являются достоянием специальных служб. Данный сервис является попыткой защитить конечного пользователя, демонстрируя ему часть той информации, которую о нем могут собрать соответствующие службы.

1.2 Основной функционал приложения

Обязательный функционал позволит определить пол, возраст, ВУЗ некоторого человека в социальной сети Вконтакте, на основе данных получаемых в автоматическом режиме. Состав дополнительного функционала, сообщаящий значимую дополнительную информацию о человеке, будет определен в процессе разработки, т.к. на начальном этапе не представляется возможным определить его из-за большого размера проекта социальной сети Вконтакте.

Оценка уровня конфиденциальности закрытых персональных данных пользователя на основе активности в социальной сети

1.2.1 Цели и задачи дипломного проекта

Задачи:

- анализ легитимности функционала приложения;
- анализ существующих web-сервисов, которые предоставляют дополнительную информацию о пользователе с помощью анализа косвенных признаков;
- анализ существующих научных подходов для реализации данной задачи;
- составление описания для каждого решения;
- анализ законности существования приложений данного типа;
- анализ существующих научных подходов для реализации данной задачи;
- реализация обязательного функционала. Уточнение и реализация дополнительного функционала;
- тестирование и доработка приложения.

1.3 Анализ существующих решений

Вследствие огромной популярности социальных сетей и того что они играют большую роль в жизни современных людей, уже давно стали появляться проекты, дополняющие их функционал. Продукты которые взаимодействуют с социальными сетями можно разделить на следующие категории:

- продукты для SMM - автоматизируют работы по привлечению внимания к брендам через социальные сети.
- социальные агрегаторы - упрощают управлению несколькими аккаунтами в социальных сетях. Как правило позволяют настроить крос-постинг, иногда присутствует функция собирание ленты, сообщений, комментариев с различных аккаунтов

- сервисы анализа сообществ и тредов в социальных сетях - позволяют вести SMM на более высоком уровне, проверять эффективность собственных pr-компаний и отслеживать конкурентов.
- нативные приложения для социальной сети - дополняют функционал социальной сети, значительную долю приложений этого класса занимают игры.

В анализ существующих решений вошли продукты, которые показали интересные технические решения, новаторство в отрасли или были интересны по другим причинам. Основной целью анализа является сбор сведений о существующих решениях в отрасли и об общих тенденциях в развитии. Анализ не ставил своей целью составить список лучших или всех приложений определенного рода. Стоит понимать что на данный момент количество сервисов и приложений у которые используют интеграция с социальными сетями уже сотни тысяч и описать их всех не имеет смысла⁷

1.3.1 smm-продукты

Такие проекты автоматизируют задачи с использованием инструментария, предоставляемого непосредственно социальными сетями, например, публикация постов в определенное время, статистика популярности сообщений. Так же продукты этого класса могут автоматизировать любые другие действие упрощающие социальный медиа маркетинг (smm)⁸

1.3.2 социальные агрегаторы

Так же существуют проекты, программные продукты или сервисы, которые собирают информацию из разных социальных сетей, бло-

⁷имеется ввиду все приложения, которые имеют функцию интеграции с какой-либо социальной сетью

⁸Social media marketing (http://ru.wikipedia.org/wiki/Social_media_marketing)

гов и других ресурсов в один источник.⁹ Стоит отметить что не все сервисы четко вписываются в тот или иной класс приложений, потому как многие из них достаточно самобытны и быстро изменяются и даже зачастую перестают существовать. Так за время подготовки дипломной работы перестали функционировать ряд сервисов:

- **twinfluence**

был простым инструментом для измерения совокупного влияния твитов и их фолловеров, а также в качестве бонуса предоставляет статистику некоторых социальных сетей. В данный момент недоступен, по доменну на котором находился проект стоит переадресацию на компанию в которой работают бывшие владельцы Twinfluence;

- **TweetEffect**

– отражал изменение количества фоловеров после каждого сообщения. Сервис перестал работать после изменения в twitter API;

- **Tweetoclock.com**

- помогал отследить время использования пользователями своего твиттер-аккаунта. В данный момент недоступен.

К самым интересным социальным агрегаторам можно отнести:

- **Hootsuite**

- Один из самых надежных и доступных инструментов, HootSuite постоянно совершенствует свой интерфейс и возможности. Онлайн-

⁹20 Ways To Aggregate Your Social Networking Profiles (<http://mashable.com/2007/07/17/social-network-aggregators/>)

доступ позволяет войти в свой аккаунт с любого места, чтобы контролировать свои аккаунты. В настоящее время, есть поддержка Twitter, Facebook Pages, Facebook, LinkedIn, Ping.fm, Wordpress.com, MySpace и Foursquare. HootSuite обладает функционалом, которые позволяют настроить, отправку поста во множество источников в несколько кликов.¹⁰ Ключевыми характеристиками являются:

- Планирование. Выбор между обновлением постов он-лайн или по заранее заготовленному расписанию.
- Гибкая работа с url. Добавление ссылок-счетчиков для отслеживания кликов и получение детальной информации об аудитории.
- RSS канал. Возможность добавить отправку постов в блоги и социальные медиа по RSS каналу.
- Закладки и апплет для браузера. Возможно использовать фирменный апплет для браузера, что бы быстро поделиться информацией

Для данного исследование представляется наиболее важным выделить существующие методы получения информации и поиска в социальных сетях, в то время как остальные особенности сервисов отходят на второй план. Был проведен анализ существующих решений, выделен ряд приложений которые с помощью косвенных данных и методов автоматического анализа позволяют «вычислить» дополнительную информацию о человеке, которую он не указывал в явном виде, найти на web-ресурсах информацию не доступную обычным поисковым системам, получить релевантную информацию которая обычно слишком низко ранжируется.

¹⁰7 Social Media Aggregation Tools To Simplify Your Streams <http://socialmediatoday.com/SMC/192312>

1.3.3 web-приложения для поиска людей

В сети Интернет представлен ряд приложений для поиска аккаунтов людей сразу во множестве социальных сетей. Стоит отметить, что в данный момент количество социальных сетей уже исчисляется десятками и это только те, которые имеют значительное (более нескольких миллионов) и живое сообщество.¹¹ Существует большое количество CMS, конструкторов сайтов позволяющие достаточно быстро создать свою собственную социальную сеть или отдельный блог с интеграцией с другими блогами построенными на той же технологии.¹² Все сети имеют свои особенности, поэтому агрегация этого многообразия - задача не простая, и ее можно решить несколькими способами. К основным проблемам, которые необходимо решить таким приложениям являются:

- написание адаптеров для каждого источника информации¹³
- решение вопросов разряженности данных (социальные сети обладают различным функционалом и данными о своих пользователях)
- скорость работы - агрегатор собирает информацию с других сервисов и значит впадает в зависимость от скорости работы 3-их лиц, что не всегда может быть надежно

основными представителями являются:

- **<http://people.yandex.ru>**

people.yandex.ru – это специализированная поисковая вертикаль, с помощью которой возможно быстро находить разме-

¹¹Top 15 Most Popular Social Networking Sites (<http://www.ebizmba.com/articles/social-networking-websites>)

¹²8 Great Social Networking CMS (<http://www.cmscritic.com/8-great-social-networking-cms>)

¹³конечно существует Open API, но многие социальные сети имеют свои особенности, поэтому все таки необходим индивидуальный подход

щенные в открытом доступе профили людей в социальных сетях. Для поиска не требуется регистрация в социальных сетях. Характерной чертой является то, что сервис очень бережно относится к персональным данным пользователей:

- Не собирает и не хранит у себя никаких дополнительных данных о пользователе, лишь ищет и индексирует уже существующую информацию.
- Индексирует только те профили, индексация которых не запрещена самим пользователем.
- Индексирует только публично доступные данные, которые видны любому незалогиненному в социальной сети пользователю.
- Склеивает только те профили, которые явно и публично ссылаются друг на друга (или в двух профилях представлены взаимные ссылки друг на друга, или в одном из них есть провалидированная, т.е. требующая авторизации, ссылка на другой).

<http://qwant.com>

qwant.com — поисковая система с особым методами ранжирования и поиском по англоязычным социальным сетям (в этом она напоминает people.yandex.ru);

● **<http://spokeo.com>**

spokeo.com — сайт для поиска людей, агрегирующий информацию из множества других он-лайн и офф-лайн источников, таких как: телефонные справочники, социальные сети, фотоальбомы, маркетинговые исследования, списки рассылки, государственные переписи, безбизнес-сайты, всего — более чем из 60 источников. Основные базы для поиска на английском языке

и, как следствие, позволяет довольно точно отследить людей, пользующихся иностранными сайтами в повседневной жизни. Сервис является прекрасным примером того, насколько эффективным может быть автоматизированное использование различных источников данных.

1.3.4 Сервисы анализа сообществ и трендов в социальных сетях

В интернете содержится огромное количество книг, инструкций и примеров психологических анализов страницы из социальной сети, но сервисы для автоматизации этого процесса практически отсутствуют. Это можно объяснить тем, что на такого рода сервисы сложно манетизировать. Естественно, что у самих владельцев есть подобные и даже куда мощные средства. Так например система матрикснет от Яндекс умеет классифицировать следующим образом пользователей.

Данный класс приложений похож на мое приложение тем, что с помощью автоматических алгоритмов он анализирует состояние и изменения в сообществах и социумах, в то время как я анализирую отдельного человека. Некоторые из этих приложений уникальны и весьма интересны, и на основании этого включены в анализ. Интересно что много сервисов для анализа twitter'а являются некоммерческими и вследствие этого быстро теряли поддержку, так например в 2011 году эти сервисы еще существовали или были популярны и хорошо работали:

- <http://topsy.com>

topsy.com - realtime поисковая система, специализирующаяся на поиске и аналитике по социальным медиа, таким как блоги, twitter, google+ и другие социальные сети. Компания является сертифицированным партнером twitter и поддерживает индекс

всех сообщений начиная с момента создания twitter в 2006 году. Запуску предшествовали три года разработки. С 2012 года партнер Яндекс (используется в формировании новостной ленты), в 2013 куплена Apple за \$200 мл. Ключевые характеристики:

- Анализ миллиардов разговоров в реальном времени.
- Мгновенное получение новостей и информации об изменении в цитируемости
- Поиск наиболее влиятельных пользователей Twitter по любой тематике
- Просмотр продвижения любого хештега в Twitter. Возможность отследить искусственное раскручивание
- Интерактивный анализ по ключевым словам и авторам, каталогизация по темам, влиянию, эмоциональной окраске, языку или географии. Пользователь может узнать, наиболее релевантные твиты, ссылки, фотографии и видео для любой темы из индекса Topsy в сотни миллиардов твитов. Пользователи могут групповые термины в сохраненных темах и настройки индивидуальных оповещений и ежедневных дайджестов деятельности.

Подводя итог, можно сказать что topsy - является одним из лидеров на рынке извлечения данных из социальных сетей, но в силу того что рынок чрезвычайно разнообразен и имеет множество особенностей в разных странах мира, то topsy не является единственным представителем этого класса сервисов

• **<http://www.kribrum.ru/>**

- система мониторинга и анализа социальных медиа для управления репутацией в Интернете, позволяет отслеживать и анализировать упоминания бренда, продуктов, услуг и ключевых персон компании. Система в автоматическом режиме находит

отзывы, обрабатывает их, определяет эмоциональную окраску высказываний и выгружает информацию в виде наглядных графиков и интерактивных отчетов. Интересно, что это одна из немногих отечественных разработок на этом рынке. Продукт принадлежит компании "Ашманов и партнеры"¹⁴

- Широкий охват поиска. Порядка 700 000 отслеживаемых площадок, постоянно добавляются новые источники, в т.ч. по запросу пользователя
- Фильтрация спама, точность выборки. Система учитывает только те отзывы, которые относятся к объекту мониторинга, отсеивает спам и сообщения, в которых бренд упомянут вскользь.
- Автоматическое определение тональности и тематики сообщений
- собственная лингвистическая технология, которая позволяет системе «понимать» правила построения предложений, анализировать связи между словами и автоматически определять тональность высказывания (хорошо, плохо, нейтрально) относительно объекта мониторинга с точностью более 80
- Оперативность обновления данных. Данные попадают в систему в период от 15 минут до 2-4 часов после публикации.
- Система позволяет определить общий охват, а также «вес» каждого упоминания и его автора, что особенно важно для формирования эффективной информационной политики, выбора подходящих площадок взаимодействия с аудиторией и выявления лидеров мнений. Возможность реагирования
- Разнообразие отчетов, экспорт данных

¹⁴Крибрум | Ашманов и партнеры <http://www.ashmanov.com/services/kribrum>

- Автоматическая генерация отчетов по шаблону и рассылка по электронной почте по заданной схеме.
- Возможность заказать аналитический отчет у экспертов в области мониторинга социальных медиа.
- Ролевой доступ, система назначения заданий, журналирование действий операторов в системе¹⁵

• TweetStats

TweetStats - создает инфографику на основе постов человека в twitter по следующим направлениям:

- количество твиттов в час
- количество твиттов в месяц
- количество твиттов в зависимости от времени (день, ночь, день недели) Есть функция сохранения результатов анализа.¹⁶ Проект особенно не развивается, масштаб проекта не большой, сервис просто хорошо справляется с заявленной функциональностью. Tweets per month Tweet timeline Reply statistics

показывает количество сообщений по месяцам, частоту сообщений в зависимости от времени дня и дня недели. Проект некоммерческий, не развивается, некоторые функции работают не стабильно;

• Twitteranalyzer

Twitteranalyzer - статистика по направлениям: Пользователи, Друзья, Упоминания, Группы и более мелким подуровням, что позволяет получить довольно много информации для анализа; Так же перестал работать.

¹⁵Что такое Крибрум <http://www.kribrum.ru/about/>

¹⁶TweetStats - Graph your Twitter Stats <http://www.tweetstats.com/>

- **sleepingtime.org**

- Простой сервис с одно единственной функцией - определение времени сна по твиттам. Принцип работы достаточно прост: сервис анализирует последние 1000 твитов и по ним строит приближенное расписание сна человека. Сервис обладает красивым интерфейсом и набором людей и областей из которых можно проанализировать людей, например шоу-бизнес, it-специалисты, политики, спортсмены.

- **klout.com**

- веб-сайт и мобильное приложение, которое использует аналитику по социальным медиа для выставления ранга от 0 до 100 под названием "Klout Score" по направлениям:

- True Reach - на какое количество пользователей вы оказываете влияние;
- Amplification - охват зоны влияния. Когда вы публикуете что-то, как много людей отвечает на ваш пост или перепечатывает его. Чем больше люди реагируют на ваши посты и сообщения, тем выше зона влияния;
- Network Score - как ваша аудитория реагирует на ваше влияние. Как часто пользователи (друзья, подписчики или их друзья) делятся вашим контентом со своими читателями и как далеко он расходуется по сети? Чем больше вас упоминают, тем выше этот показатель.¹⁷

¹⁷Работаем с сервисом Klout - а как вы влияете на вашу аудиторию в социальных сетях? <https://www.facebook.com/notes/mike-ponomarenko/>

По заявлению разработчиков ранг является корреляцией между вкладом человека в контент социальных сетей и тем насколько контент, создаваемый пользователем востребован другими пользователями социальных сетей. Аналитика производится на основе данных сайтов Twitter, Facebook, Google+, LinkedIn, Foursquare, YouTube, Instagram, Tumblr, Blogger, WordPress, Last.fm и Flickr.^{18 19} Klout оценивает степень влияния, используя показатели такие показатели как:

- сколько авторов отслеживает пользователь;
- сколько авторов отслеживают пользователя;
- количество ретвитов
- упоминания в списках авторов;
- за сколькими спам/мертвыми авторами следит пользователь;
- какова степень влияния тех, кого ретвитит пользователь;
- количество приватных сообщений

Полученная информация объединяется с информацией из Facebook, комментариями, отметками о понравившейся публикации, количеством друзей. Все эти данные отображаются в «Klout Score», который показывает степень влияния пользователя в социальных сетях. У сервиса подвергается постоянной критике^{20 21 22} из-за того, во что он фактически получает власть человеческие

cyrh/233439080026653

¹⁸How can you measure influence? (<http://www.simplyzesty.com/Blog/Article/July-2010/How-can-you-measure-influence>)

¹⁹<http://klout.com/corp/about> <http://klout.com/corp/about>

²⁰Why Klout scores are possibly evil (http://money.cnn.com/2011/11/15/technology/klout_scores/index.htm)

²¹Don't Fall for this Sneaky Klout Trick Designed to Suck You In (<http://www.forbes.com/sites/anthonykosner/2012/05/08/klout-uses-this-trick-to-make-you-feel-bad-about-yourself-dont-let-it-ruin-your-life/>)

²²Klout overhauls its business model, but does it answer its critics? <http://www.businessesgrow.com/2012/08/14/klout-overhauls-its-business-model-but-does-it-answer-its-critics/>

судьбы, так в 2012 году в США одного за место специалиста с низким рейтингом klout взяли неопытного парня с высоким рейтингом.²³ Джон Скалзи (John Scalzi) из CNN описал принцип, лежащий в основе Klout как «социально зло» в результате использования klout он вызывает тревожное состояние у своих пользователей.²⁴ Британский писатель Чарльз Стресс охарактеризовал klout как "герпес для интернета". Анализ условий использования и лицензионного соглашения klout показывает, что бизнес-модель компании является незаконной в Великобритании, где она противоречит закону Data Protection Act 1998 года; Стресс советует читателям удалить их аккаунты Klout и отказаться от услуг этой компании.²⁵

1.3.5 Приложения для платформы vk.com

Отдельно стоит упоминать приложения, написанные на платформе ВКонтакте — все они реализуются по средствам flash, javascript или как iframe со стороннего сайта. Особенностью приложений под эту платформу является их относительная простота, как правило выполняются на стороне клиента или имеют не сложную серверную по сравнению с приложениями для анализа сообществ и трендов. В ходе анализа существующих решений были выявлены следующие приложения:

- Анализатор (<https://vk.com/ianaliz>) — способен проанализировать количество друзей, сколько из них женского пола, сколько мужского, сколько не сообщили такую информацию, примерная дата регистрации в vk.com
- Радар (<https://vk.com/vkradar>) — сообщают статистику по сообщениям на стене, в группе и т. д. по количеству сообщений,

²³см. подробнее wired: What Your Klout Score Really Means (http://www.wired.com/2012/04/ff_klout/)

²⁴Klout Now Measures Your Influence on Facebook <http://mashable.com/2010/10/14/facebook-klout>

²⁵Charlie Stross - Evil social networks (<http://www.antipope.org/charlie/blog-static/2011/11/evil-social-networks.html>)

по полярности сообщений. Относительно не сложное приложение, с моделью монетизации за дополнительный функционал.

- Модерация пабликов и страниц, анализ популярности (http://vk.com/public_tools) - инструмент для автоматизации повседневных задач модератора. Отличается низкой надёжностью (стало временно не доступно во время написания работы)
- Анализ Аватара (<https://vk.com/avascan>) — выдает результаты близкие к случайным но на основе аватара пользователя по таким характеристикам как: сексуальность, красота и прочее. Не так давно было заблокировано администрацией платформы.
- Лайк-машина (<http://vk.com/like.machine>) приложение, позволяющие за дополнительную плату повысить свою популярность и отследить посетителей. Больше приложений для платформы vk решающих подобные задачи выявлено не было

По итогам анализа были сделаны следующие выводы:

- представляется интересным показывать результаты работы приложения посредством публикации отчета на стене пользователя
- возможна монетизация приложения путем предоставления пользователю бесплатно, некоторого количество виртуальных денег, которых хватит, на то чтобы попробовать лишь часть функционала, в то время, как весь функционал будет стоить дополнительных денег.
- В приложение можно реализовать дополнительный(не основной) функционал, но он должен пониматься пользователем как дополнительный и предоставляться бесплатно

1.3.6 Выводы

По итогам анализа этих проектов были сделаны следующие выводы:

- большие объемы данных позволяют построить более детальную аналитику, чем локальный анализ
- так как в разных социальных сетях сидят одни и те же люди, то при отслеживании каких либо общественных изменений, как правило, достаточно глубоко анализа одной из платформ, поэтому большинство сервисов заточены на Twitter, как наиболее

Несогласованное удобную и открытую социальную сеть из всех.

предло-
жение?

- достаточно интересным оказался функционал сайта sleepingtime.org, который анализирует время публикации постов. В дальнейшем возможно развить отсюда следующие направления:
 - вычислить время, в которое пользователь активно пишет посты в социальной сети
 - вычислить время сна
 - примерно вычислить сколько часов в день пользователь проводит в социальной сети

я точно не реализую эту функциональность, стоит ли тогда писать об этом? может как то изменить предложение?)

- показательным является пример klout, который зарабатывает деньги меняя наше общество к худшему

Среди всех приложений я пытался выявить приложения которые анализируют метаданные в целях получения дополнительных сведений о пользователе. Таких приложений оказалось немного: [sleepingtime](http://sleepingtime.org), Радар, Анализатор для [vkontakte](https://vk.com/vkontakte). Приложения Радар и Анализатор не реализуют функционала который заложен в дипломный проект. Малое количество приложений можно объяснить тем, что вся мощь заключенная в метаданных раскрывается при больших объемах информации, создание приложений анализирующих большие объемы

стоит больших денег поэтому перед созданием приложения должна существовать ясная модель монетизация приложения. В гражданской сфере в основном востребованным мониторинг брендов, особняком стоит относительно новый сервис klout. Совсем другая ситуация складывается в военном секторе, особенно в области разведки, где созданы огромные системы, такие как Xkeyscore и другие, в которых основной источником анализа являются метаданные.

Можно сказать, что не было выявлено существующих приложений реализующих функционал заявленный в дипломной работе.

1.4 Обоснование выбора технических инструментов для реализации приложения

Инструменты которые будут использоваться для построения приложения:

- Python выбран в связи с большой функциональной выразительностью²⁶ и гибкостью языка²⁷. Проект не рассчитан на очень высокие нагрузки, поэтому с одной стороны производительности интерпретатора Python вполне достаточно, а с другой упрощается процесс написания и сопровождения приложения.
- MySQL выбрана как одна из самых быстрых СУБД при средних и маленьких объемах БД. Так же у проекта хорошая документация.
- Django выбран за высокую скорость написания приложения и архитектурные преимущества, по сравнению с такими фрейм-

²⁶С. Макконнелл — Совершенный код, с.60, Microsoft Press, М.:2012

²⁷<http://ru.wikipedia.org/wiki/Python>

ворками как Symfony(PHP) и Dancer (Perl), а также в связи с тем, что написание приложения обработки данных и приложения веб-клиента на одном языке упрощает сопровождение (Django написан на Python).

- Ubuntu server 12.04 выбрана среди прочих аналогов, таких как Fedora, Debian, OpenSuse по следующим причинам:

- дружелюбное сообщество
- безопасность²⁸
- легкую расширяемость с помощью ppa²⁹ и центра приложений
- активная поддержка (Ubuntu server 12.04 будет поддерживаться до апреля 2017 года)³⁰

В процессе написания приложения необходимо будет решить следующие проблемы:

- распознавание информации с сайта
- обход страниц друзей пользователя и распознавании информации на их странице
- минимизация нагрузки с одного клиента
- авторизация пользователя средствами социальной сети (по протоколу OAuth 2.0)

Приложение реализовано по клиент-серверной архитектуре. Сервер обработки данных на Python, база данных на MySQL, клиентская часть (веб-форма) на фреймворке Django. Так как планируется использовать отдельный сервис для обработки данных, то размещение

²⁸см. подробнее доклад National Technical Authority for Information Assurance (CESG) - End User Devices Security Guidance: Introduction <https://www.gov.uk/government/collections/end-user-devices-security-guidance--2#overview>

²⁹<http://help.ubuntu.ru/wiki/ppa>

³⁰<http://ru.wikipedia.org/wiki/Ubuntu>

проекта на виртуальном (shared) хостинге является недостаточным. Проект будет размещен на виртуальном выделенном сервере (VPS), ОС для сервера — Ubuntu 12.04
Все инструменты, которые я применяю, является продуктами с открытым исходным кодом. Они активно развиваются, поддерживаются и имеют живое сообщество пользователей в т.ч. русскоязычное.

безопасности¹ легкую расширяемость с помощью `ppa2` и центра приложений активная поддержка (Ubuntu server 12.04 будет поддерживаться до апреля 2017 года)³

Литература

- [1] W.T. Cathey and E.R. Dowski, “New paradigm for imaging systems”, Appl. Opt. 41, pp. 6080-6092, 2002.

1.5 Список терминов

Rest API - (сокр. англ. Representational State Transfer, «передача состояния представления»), стиль построения архитектуры распределенного приложения.

Adobe AIR - Adobe Integrated Runtime - это кроссплатформенная среда от компании Adobe для запуска приложений, позволяющая использовать HTML/CSS, AJAX, Adobe Flash и Apache Flex для переноса веб-приложений (в том числе Rich Internet Application) на ПК.