Московский государственный университет имени М.В.Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра Системного Программирования

Лазарев Владимир Александрович

Исследование методов OSINT для поиска информации о человеке

Курсовая работа

Научный руководитель:
к.ф.-м.н.
Турдаков Денис Юрьевич
Научный консультант:
Яцков Александр Константинович

Москва, 2021

Аннотация

Исследование методов OSINT для поиска информации о человеке

Лазарев Владимир Александрович

Данная работа посвящена исследованию и разработке методов OSINT для поиска информации о человеке. Данная курсовая содержит описание реализованных методологий и повествует о созданных приемах извлечения информации.

В ходе работы были изучены и представлены существующие различные методы как по способу взаимодействия с сервисами: извлечение данных с web-страницы и посредством скрытого или открытого api; так и по типу сервиса: поисковый агрегатор и социальные сети.

Содержание

1 Введение **4**

2 Постановка задачи **5**

3 Обзор существующих решений **7**

3.1 Поиск данных в поисковых сервисах 7

3.1.1 Google Dorks (Google Hacking) 7

3.1.2 Carrot2 8

3.1.3 Yippy 9

3.2 Поиск данных в социальных сетях 10

3.2.1 Maltego 10

3.2.2 ITools 10

3.2.3 FindThatLead 10

3.2.4 Palantir 10

3.3 Универсальные приложения 11

3.3.1 Виток OSINT 11

4 Исследование и построение решения задачи **12**

5 Описание практической части **13**

6 Заключение **14**

Список литературы **15**

1 Введение

В разделе 1 сформулирована постановка задачи. В разделе 2 приведен анализ существующих решений методов поиска, сбора и анализа информации из открытых источников. В разделе 3 описано исследование и построение решения задачи. В разделе 4 приведено описание практической части курсовой работы. В конце документа сформулировано заключение.

2 Постановка задачи

Целью данной курсовой работы является исследование и разработка методов OSINT для поиска информации о человеке. Для решения задачи, ее можно разбить на несколько подзадач: сбор информации при помощи поисковых сервисов, сбор информации с помощью социальных сетей. В свою очередь каждую из подзадач также можно поделить на следующие части: определение структуры web-страницы и извлечение данных непосредственно из страницы, поиск более быстрого доступа к информации посредством открытого или закрытого api.

В итоге для достижения постановленной цели необходимо решить следующие задачи:

- Поиск данных в поисковых сервисах:
  - Провести анализ литературы и существующих решений для извлечения данных из поисковых систем;
  - Разработать методы поиска и сбора информации из поисковых систем:
    - Проанализировать структуру web-страниц поискового сервиса;
    - Реализовать метод поиска и извлечения информации при помощи атрибутов web-страницы;
    - Провести исследование о возможности получения данных из ресурса посредством открытого или закрытого api;
    - Если api реализовано на стороне сервиса, то реализовать метод поиска и сбора посредством api;
  - Получить тестовые данные от реализованных методов и провести анализ, исследование полученной информации;
- Поиск данных в социальных сетях:
  - Провести анализ литературы и существующих решений для извлечения данных из социальных сетей;
  - Разработать методы поиска и сбора информации из социальных сетей:
    - Проанализировать структуру web-страниц социальных сетей;

    * Реализовать метод поиска и извлечения информации при помощи атрибутов web-страницы;
    * Провести исследование о возможности получения данных из ресурса посредством открытого или закрытого api;
    * Если api реализовано на стороне соц. сети, то реализовать метод поиска и сбора посредством api;
- Получить тестовые данные от реализованных методов и провести анализ, исследование полученной информации;

3 Обзор существующих решений

3.1 Поиск данных в поисковых сервисах

3.1.1 Google Dorks (Google Hacking)

Google Dorks[1] - это по сути та же самая поисковая система от Google. Отличие заключается только в том, что обычный пользователь вбивает типовые запросы а-ля "Какая погода в Москве? то Google Dorks позволяет использовать специальные запросы для получения конкрентной информации. Google Dorks имеет множество операторов, которые можно использовать для составления очень гибких и точных запросов [1]. По факту, это запросы, с помощью которых можно проверить безопасность того или иного сайта, найти IP-адреса сервисов, камер. Весьма эффективна для поиска документации по ключевым словам, а также поиску людей с помощью тех же самых Google Dorks Queries.

Плюсы данный системы:

- быстрый и объемный поиск по ключевым словам.

Из недостатков системы можно определить следующее:

- составленный запрос выдаст перечень ссылок в интерфейсе поисковой системы, а не сами данные;
- перед использованием необходимо изучить синтаксис запросов;
- нет накопления собранной информации, нельзя отслеживать изменения (дельты);
- нет построения графа зависимостей объекта.

[1] https://www.google.com/

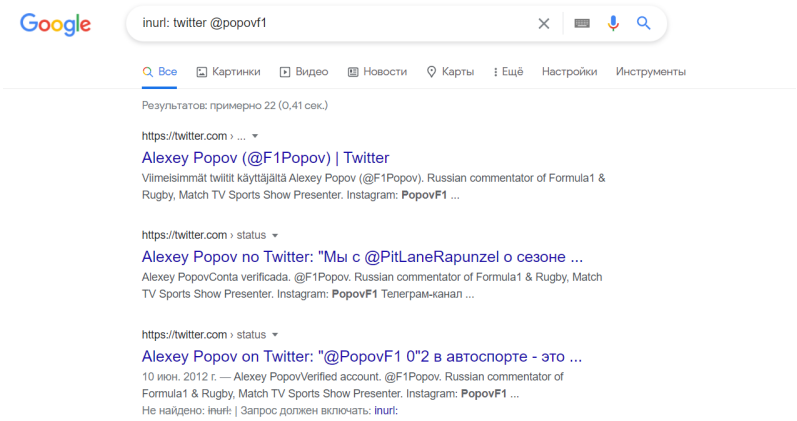

Рис. 1: Пример использования GDQ для поиска человека.

3.1.2 Carrot2

Carrot2 - движок кластеризации результатов поисковых запросов с открытым исходным кодом. Carrot2 может самостоятельно группировать по категориям найденные документы или данные. Работает в свою очередь как обычный поисковик, то есть по указанному ключевому слову возвращает некоторое множество ссылок, затем которые группируются по категориям [2].

Преимущества:

- быстрый и обширный поиск по ключевым словам;
- автоматическая группировка данных в соответствии с категориями;
- наличие удобного интерфейса с возможностью просмотра древовидной карты и круговидной диаграммы.

Недостатки:

- как и в случае с Google Dorks, Carrot2 возвращает нам перечень ссылок на источники данных, а не сами данные непосредственно;
- невозможно произвести точечный поиск файлов и данных, как это реализовано в Google Dorks. Как следствие - большое количество лишней информации.

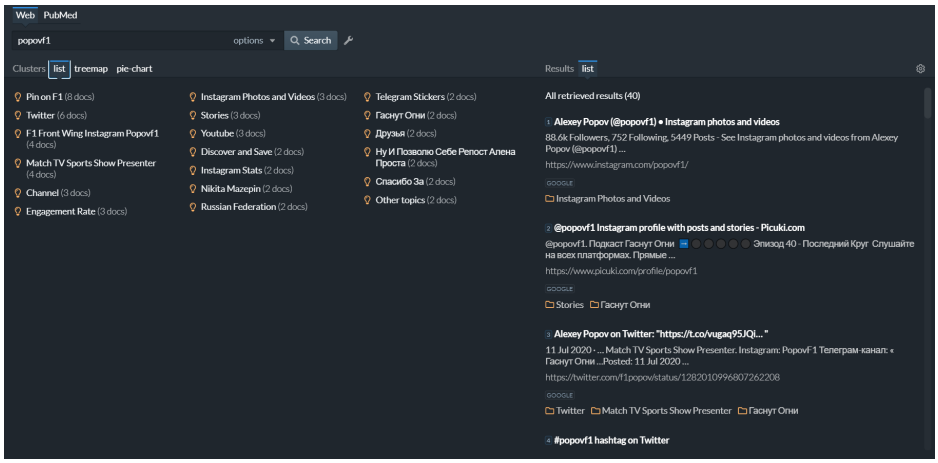

Рис. 2: Пример использования Carrot2 с разбиением результатов на группы.

3.1.3 Yippy

Yippy[2] - это метапоисковый движок, который группирует результаты поиска на категориям в группы. Наделен обширным функционалом: позволяет искать по ключевым словам новости, вакансии, правительственную информацию и блоги. Также позволяет вручную настраивать источники данных для собственного уникального метапоиска. [3]

Преимущества:

- группирует данные по тематическим категориям;
- есть возможность поиска не только ссылок в web-пространстве, но и непосредственно новостей, изображений и видео;

Недостатки:

- сервис недоступен на территории РФ;
- нет поддержки GDQ.

[2]http://yippy.com/

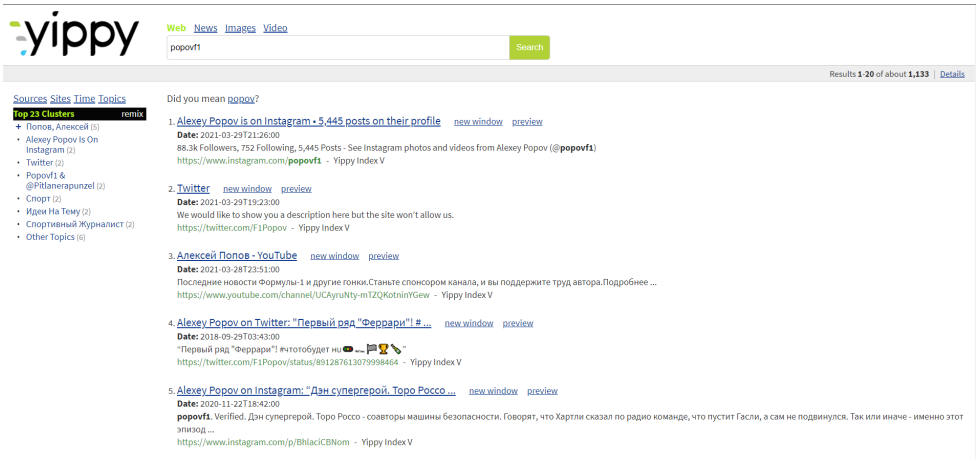

Рис. 3: Пример использования Yippy.

3.2 Поиск данных в социальных сетях

3.2.1 Maltego

Maltego[3] - это

3.2.2 ITools

ITools[4] - это

3.2.3 FindThatLead

FindThatLead[5] - это

3.2.4 Palantir

Palantir[6] - это

[3] https://www.maltego.com/
[4] http://itools.com/search/people-search
[5] https://findthatlead.com/en
[6] https://www.palantir.com/solutions/intelligence/

3.3 Универсальные приложения

3.3.1 Виток OSINT

Виток OSINT[7] - это

[7] https://norsi-trans.ru/catalog/vitok-osint/

4 Исследование и построение решения задачи

Здесь надо декомпозировать большую задачу из постановки на подзадачи и продолжать этот процесс, пока подзадачи не станут достаточно простыми, чтобы их можно было бы решить напрямую (например, поставив какой-то эксперимент или доказав теорему) или найти готовое решение.

5 Описание практической части

Если в рамках работы писался какой-то код, здесь должно быть его описание: выбранный язык и библиотеки и мотивы выбора, архитектура, схема функционирования, теоретическая сложность алгоритма, характеристики функционирования (скорость/память).

6 Заключение

Здесь надо перечислить все результаты, полученные в ходе работы. Из текста должно быть понятно, в какой мере решена поставленная задача.

Список литературы

[1] *ru.wikipedia.org*. Google hacking. — 2020. — Ноябрь. https://ru.wikipedia.org/wiki/Google_hacking.

[2] *en.wikipedia.org*. Carrot2. — 2021. — Март. https://en.wikipedia.org/wiki/Carrot2.

[3] *en.wikipedia.org*. Yippy. — 2021. — Февраль. https://en.wikipedia.org/wiki/Yippy.

[4] *Ольга, Дзюба*. OSINT: что это, кому он нужен, какие методы сбора и типы информации использует? — 2020. — Август. https://yushchuk.livejournal.com/1451268.html.

[5] *Карев, Антон*. SHODAN: САМЫЙ СТРАШНЫЙ ПОИСКОВИК ИНТЕРНЕТА. — 2018. http://samag.ru/archive/article/3714.

[6] *Шагаев, Иван*. Поисковая система Shodan не то, чем кажется. — 2018. — Май. https://www.anti-malware.ru/analytics/Threats_Analysis/Shodan.

[7] *kali.tools*. theHarvester. https://kali.tools/?p=2286#:~:text=theHarvester.

[8] *Опанюк, Игорь*. Maltego. Нароет все. — 2009. — October. https://habr.com/ru/post/73306/.

[9] *https://www.spiderfoot.net/*. SpiderFoot: OSINT Automation. — 2019. — Сентябрь. https://ai-news.ru/2019/09/spiderfoot_osint_automation.html#:~:text=SpiderFoot.

[10] *geocreepy*. Creepy. https://www.geocreepy.com.

[11] *https://jivoi.github.io/*. Awesome OSINT. — 2021. https://github.com/jivoi/awesome-osint.

[12] *Kozhuh*. Что такое Google Dorks? https://spy-soft.net/gugl-dorki/.

[13] *Goossens, Michel*. The LaTeX Companion / Michel Goossens, Frank Mittelbach, Alexander Samarin. — Reading, Massachusetts: Addison-Wesley, 1993.