

# LINKERO

A Brand Monitoring Platform



This report is submitted in partial fulfilment of the requirements  
for the  
B.Sc. (Honours) Information Systems and Information Technology (DT249)  
to the  
School of Computing  
College of Sciences & Health  
Dublin Institute of Technology

**Author:** Stefano Richiardi

**Student ID:** D12124118

**Supervisor:**

**Date:** January 2019



# Abstract

In this project I aim to build a data aggregator service for investigations and monitoring of counterfeit products online. The service will be implemented with a web interface for clients' access, an http server and specialized databases depending on data and performance requirements.

The core mission of this project is to try to simplify daily tasks of counterfeit investigations, and leverage server technologies to allow investigators to see the bigger picture painted when data from one case connect to one or multiple other cases previously unconnected.

Equally important will be the ability to deliver a product that is ready for production environment, that is: running on a virtual server accessible from the internet, protected from unauthorized intrusions and easily scalable should the user population grow significantly.

In this project I will also discuss pros and cons of data collection from online sources: from open API offered by online services to the community of developers, to unstructured data that can be scraped and schematized, and which legal implications need to be taken into account in light of the new European General Data Protection Regulation (aka GDPR).



# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Project background . . . . .	1
1.2 Background . . . . .	1
1.3 Objectives . . . . .	2
1.4 Scope . . . . .	2
1.5 Challenges and learning requirements . . . . .	3
1.6 Deliverables . . . . .	3
<b>2 Literature review</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 Open Source Intelligence . . . . .	5
2.3 GDPR . . . . .	5
2.4 Server security standards . . . . .	5
2.5 Ergonomy of software interface . . . . .	5
2.6 Code development principles . . . . .	6
2.7 Existing brand monitoring services . . . . .	6
<b>3 System design</b>	<b>7</b>
3.1 Technologies . . . . .	7
3.2 System architecture . . . . .	7
3.3 Django page requests flow . . . . .	7
3.4 Django ORM . . . . .	7
3.5 Development methodology . . . . .	7
3.6 Requirements . . . . .	7
3.7 User cases . . . . .	7
3.8 Sequence diagrams . . . . .	7
3.9 Ergonomics design principles . . . . .	7
<b>4 Project implementation</b>	<b>9</b>
4.1 Security settings . . . . .	9
4.2 Functions . . . . .	9
4.3 Exception handling . . . . .	9

<b>5</b>	<b>Testing and deployment</b>	<b>11</b>
5.1	Unit testing . . . . .	11
5.2	Integration testing . . . . .	11
5.3	System testing . . . . .	11
5.4	Usability testing . . . . .	11
5.5	Deployment . . . . .	11
<b>6</b>	<b>Evaluation</b>	<b>13</b>
6.1	Lessons learned . . . . .	13
6.2	Future work . . . . .	13
6.3	Conclusions . . . . .	13
6.4	Deployment . . . . .	13
	<b>Bibliography</b>	<b>15</b>

# Chapter 1

## Introduction

### 1.1 Project background

**WHY COUNTERFEIT IS BAD** Every year thousands of millions are spent by consumers on counterfeit products. Counterfeit is an unfair commercial practice which takes advantage of the brand identity and presence on the market built by the brand owner, without incurring in the same costs of brand development (product design, quality standards and marketing).

E-commerce has been rapidly expanding since its early days in mid '90 when online marketplaces like eBay and Amazon were launched. Although it is hard to make a global estimate of market share for e-commerce companies, if we take the US market, the Census Bureau estimated that in 2009 online sales accounted for 4% of all retail sales, whereas in 2017 the e-commerce market share went up to 9%. As legitimate retailers increase their presence online, shops selling counterfeit follow suit. In 2017 the Guardia custom identified xxx worth of counterfeit goods entering the Irish border, many of those destined for the European continental market.

This tool is designed to help users to extract data related to a specific brand from online platforms such as eBay, Mercadolibre, Allegro, in a tabular format. Online marketplaces offer API open to developers in order to enable automated interaction with their platforms. This tool will focus on extracting sales and business registration data. The data will be stored in a database so that users can keep historical records of all their queries. At the same time, users will be able to leverage the growing dataset to link new investigations to old cases across all platforms whose data has already been stored in the database. For instance, searching the email address of a shop that deals counterfeit Diesel jeans may return details of multiple businesses registered on eBay and Mercadolibre at different times, as well as information about the administrator of a facebook page about counterfeit Raiban glasses. Users can use this tool to estimate brand damage caused based on the volume of sales, as well as a forensics platform to facilitate identity attribution of potential counterfeiters.

### 1.2 Background

Identifying online counterfeit items starts with researches of a brand or product online presence. The aim is to identify sellers that offer a branded product sold at a price point be-

low average retail price. An investigator would ideally prioritize sellers with higher business turnover, and ideally located in jurisdictions where legal action is a dependable and impartial option. For instance countries like Russia and China, to name few, do not always offer adequate protection for European and North American companies, making it more difficult to pursue compensation from actors located within their borders.

There are already different online companies that offer brand monitoring services: they range from keyword web-crawlers, to consumer sentiment analytics based on social platforms, to anti-counterfeit detection. We will review those companies and their services in details in the next chapter.

Private investigation legislation S.I. No. 195/2015 - Private Security (Licensing and Standards) (Private Investigator) Regulations 2015. Private Security Services Act 2004 (No.12 of 2004).

GDPR is the latest European Union legislation about processing of personal identifiable, privacy protected data belonging to European citizens. GDPR affects the implementation of this project in two ways: first, and most obvious, because users will submit limited personal data in order to be able to access Linkero services. Something as simple as an email address required to get notification for password recovery or data collection completed confirmation, is considered a PII (personal identifiable information) and thus is protected under GDPR. But GDPR affects also the collection of public personal details available online. This is the compliance requirement that is specific to companies providing *web scraping* services. When we deal with open source intelligence and attribution, PII are the most valuable data for anti-counterfeit investigations, therefore we will talk about how GDPR affects services like Linkero and what are the requirements to guarantee compliance (<https://blog.scrapinghub.com/web-scraping-gdpr-compliance-guide/>).

### 1.3 Objectives

The goals set for this project fall in two main categories: business goals and personal goals.

In relation to the first type, the goal is to build a business ready web application for brand monitoring investigations. Being *business ready* means presenting a final application that runs on a virtual server directly connected to the public internet, that implements all industry standard security features to guarantee that only authorized users can access its functions, and is scalable, designed to accommodate an exponentially growing number of users. The other aspect of the first goal refers to the ability to provide basic case management, data extraction and keyword searches capabilities tailored for brand monitoring and anti-counterfeit investigations.

On a personal level, I aim at gaining a hands-on understanding of specific web technologies (e.g. JQuery, Django framework, NGINX, uWSGI), NoSQL databases (MongoDB, Redis), asynchronous and non-blocking programming techniques (multi-threading, AJAX, Python Celery), development methodologies and testing practices.

### 1.4 Scope

This project will consist of a ready to deploy and use system for case management and web scraping data using eBay public API.



Regular users will be able to login, change password, set their preferred email address for general notifications and report delivery, launch queries, download unlimited times completed reports from the web interface, delete old queries, and search via keyword within all data collected and stored by any user in the database. Also the system will be configured to provide security protection against external unauthorized access and use of the system. Site administrators will manage the registration, password reset and deletion of users profiles.

## **1.5 Challenges and learning requirements**

Young family committments, full time job and travel arrangements, SSL configuration, DKIM configuration, app specific issues.

## **1.6 Deliverables**

Loren Ipsum

- full documentation of the system design, development and functions;
- a working implementation of the system;
- a PowerPoint presentation of the project.



## Chapter 2

# Literature review

### 2.1 Introduction

This chapter will provide an overview of different topics that informed design decisions made during the initial stages of the project. We will start looking at the discipline of Open Source Intelligence, which refers to all those protocols and techniques used by government and private agencies to piece together intelligence reports using publicly available sources, since Linkero is a tool that facilitates the structured collection and analysis of a limited portion of open source data. The legal aspect of data collection and analysis will be explored with a reasoned summary of what is GDPR and how it impacts similar online services. Three sections will be dedicated to industry level best practices in relation to the security of servers facing the public internet, the usability of software interfaces and current guidelines to build reusable, maintainable and expandable code. Finally we will review briefly what are some of the current services already offering brand monitoring tools and how they differ from one another.

### 2.2 Open Source Intelligence

Get books from work!

### 2.3 GDPR

Loren Ipsum ...

### 2.4 Server security standards

Loren Ipsum ...

### 2.5 Ergonomy of software interface

Loren Ipsum ...

## **2.6 Code development principles**

Loren Ipsum ...

## **2.7 Existing brand monitoring services**

Loren Ipsum ...

## **Chapter 3**

# **System design**

- 3.1 Technologies**
- 3.2 System architecture**
- 3.3 Django page requests flow**
- 3.4 Django ORM**
- 3.5 Development methodology**
- 3.6 Requirements**
- 3.7 User cases**
- 3.8 Sequence diagrams**
- 3.9 Ergonomics design principles**



## **Chapter 4**

# **Project implementation**

**4.1 Security settings**

**4.2 Functions**

**4.3 Exception handling**





## **Chapter 5**

# **Testing and deployment**

**5.1 Unit testing**

**5.2 Integration testing**

**5.3 System testing**

**5.4 Usability testing**

**5.5 Deployment**



## **Chapter 6**

# **Evaluation**

**6.1 Lessons learned**

**6.2 Future work**

**6.3 Conclusions**

**6.4 Deployment**



# Bibliography

[Sanjeev Jaiswal(2015)] Ratan Kumar Sanjeev Jaiswal. *Learning Django Web Development*.  
PACKT Publishing, 2015.