

# Chapter 3

## Alternative Data for Trading

– Categories and Use Cases

**报告人：周颜锐、梅姝贤、刘汶楚**

October 13th, 2022

# Content

1.The Alternative Data Revolution

2.Sources of Alternative Data

3.Criteria for evaluating Alternative Datasets

4.The Market for Alternative Data

5.Working with Alternative Data

# I.The Alternative Data Revolution

The definition of alternative data

The five Vs:

- Volume: The amount of data is magnitude larger.
- Velocity: Data becomes available at real-time speed.
- Variety: Data is organized in more kinds of formats.
- Veracity: it's difficult to validate the reliability of the data.
- Value: Determining the value of datasets can be time-consuming.

# I.The Alternative Data Revolution

Use cases for new data sources include the following:

- Online price data on a representative set of goods and services can be used to measure inflation.
- The number of store visits or purchases permits real-time estimates of company or economic activity.
- Satellite images can reveal agricultural yields, or activity at mines or on oil rigs earlier.

# I.The Alternative Data Revolution

The capability to process and integrate diverse datasets and apply **ML** allows for complex **insights**.

These insights create new opportunities to capture classic investment themes such as value, momentum, quality, and sentiment.

## 2.Sources of Alternative Data

Alternative datasets are generated by many sources but can be classified at a high level as predominantly produced by:

- **Individuals** who post on social media, review products, or use search engines.
- **Businesses** that record commercial transactions.
- **Sensors** that capture economic activity through images from satellites or security cameras.

## 2.Sources of Alternative Data

### **Individuals**

Data generated by individuals is frequently unstructured in text, image, or video formats, disseminated through multiple platforms, and includes:

- Social media posts, such as opinions or reactions on general-purpose sites
- E-commerce activity that reflects an interest in or the perception of products on sites

## 2.Sources of Alternative Data

### **Business processes**

Data that results from business processes often has more structure. It is very effective as a leading indicator for activity.

- Payment card transaction data from processors and financial institutions.
- Company exhaust data produced by ordinary digitized activity or record-keeping, such as banking records, cashier scanner data, or supply chain orders



## 2.Sources of Alternative Data

### **Sensors**

This category of alternative data is typically very unstructured and often significantly larger in volume than data generated by individuals or business processes, and it poses much tougher processing challenges.

## 2.Sources of Alternative Data

- **Satellites :**

Use cases include monitoring economic activity that can be captured using aerial coverage, such as agricultural and mineral production and shipments.

- **Geolocation data:**

A familiar source is smartphone, with which individuals voluntarily share their geographic location through an application, or from wireless signals such as GPS, CDMA, or Wi-Fi.

### 3.Criteria for evaluating Alternative Datasets

- The **ultimate objective** of alternative data is to provide an informational advantage in the competitive search for trading signals that produce alpha.
- In practice, the signals extracted from alternative datasets can be used on a standalone basis or **combined with other signals** as part of a quantitative strategy.

### 3.Criteria for evaluating Alternative Datasets

#### **Quality of the signal content**

- Asset classes

Data on fixed income and around interest-rate projections is a more recent phenomenon but continues to increase.

- Investment style

specific sectors and stocks;macro themes;market risk.

- Risk premiums

have a low correlation (lower than 5 percent) with traditional risk premiums

- Alpha content and quality

### 3.Criteria for evaluating Alternative Datasets

#### **Quality of the data**

The quality of a dataset is another important criterion because it impacts the **effort** required to analyze and monetize it, and the **reliability** of the predictive signal it contains.

- **Legal and reputational risks**
- **Exclusivity**: the more exclusive and harder to process the data, the better the chances that a dataset with alpha content can drive a strategy

### 3.Criteria for evaluating Alternative Datasets

#### **Quality of the data**

- **Time horizon**
- **Frequency**
- **Reliability**
- **Technical aspects**
  - Latency
  - Format

## 4.The Market for Alternative Data

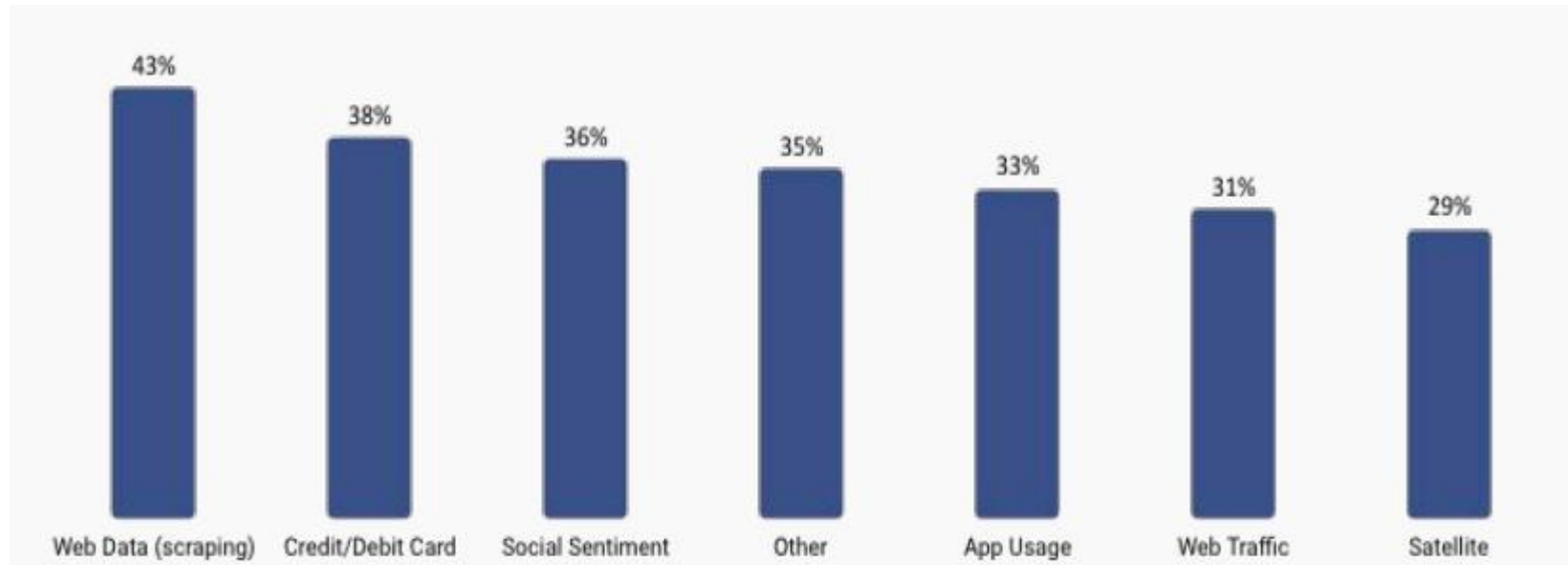
The investment industry spent an estimated \$2-3 billion on data services in 2018, and this number is expected to grow at a double-digit rate per year in line with other industries.

This expenditure includes:

- the acquisition of alternative data
- investments in related technology
- and the hiring of qualified talent.

## 4.The Market for Alternative Data

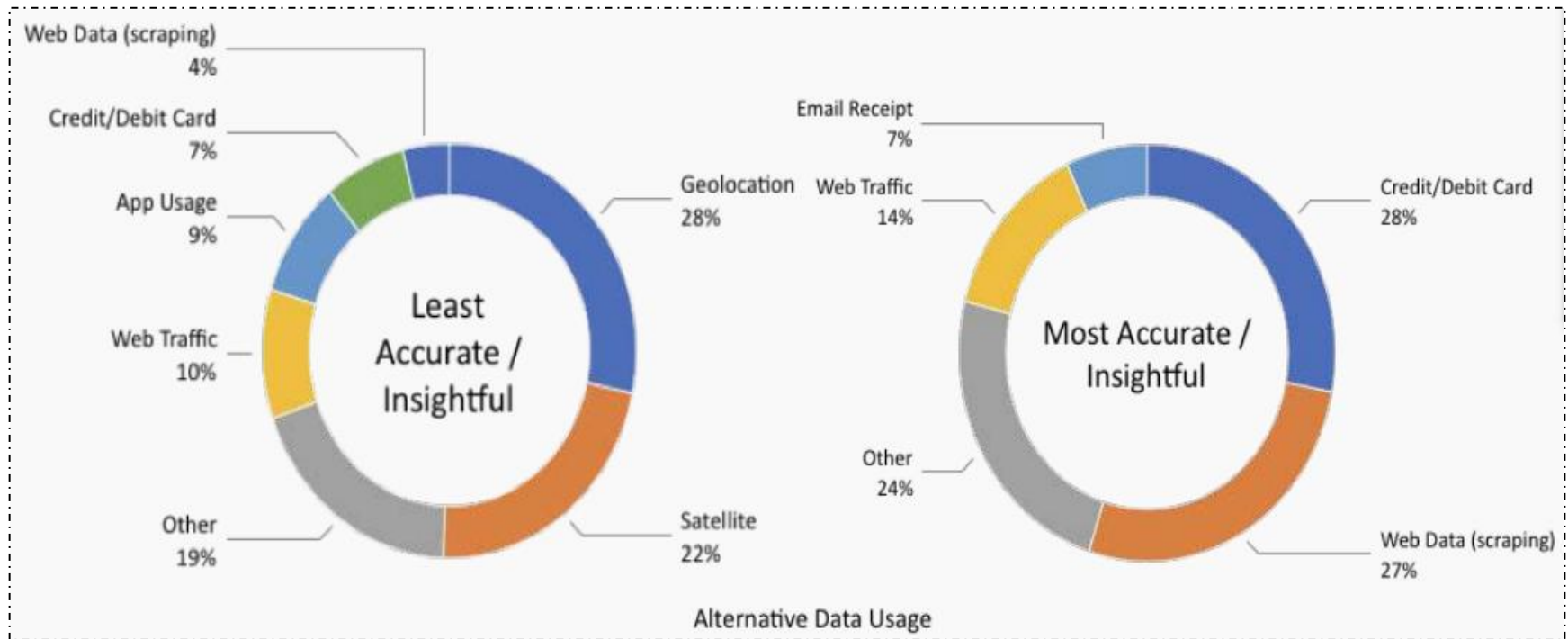
A survey by Ernst & Young shows significant adoption of alternative data in 2017:





## 4.The Market for Alternative Data

- Based on the experience so far, fund managers considered scraped web data and credit card data to be most insightful
- Geolocation and satellite data are considered to be less informative:



## 4.The Market for Alternative Data

### Data providers and use cases

- Social sentiment data
  - Dataminr
  - StockTwits
  - RavenPack
- Geolocation data
  - Advan
- Satellite data
  - RS Metrics
- Email receipt data
  - Eagle Alpha

<https://alternativedata.org/data-providers/>

## 5. Working with Alternative Data

### **Scraping OpenTable data**

- Parsing data from HTML with Requests and BeautifulSoup
- Introducing Selenium – using browser automation
- Building a dataset of restaurant bookings and ratings
- Taking automation one step further with Scrapy and Splash

### **Scraping and parsing earnings call transcripts**

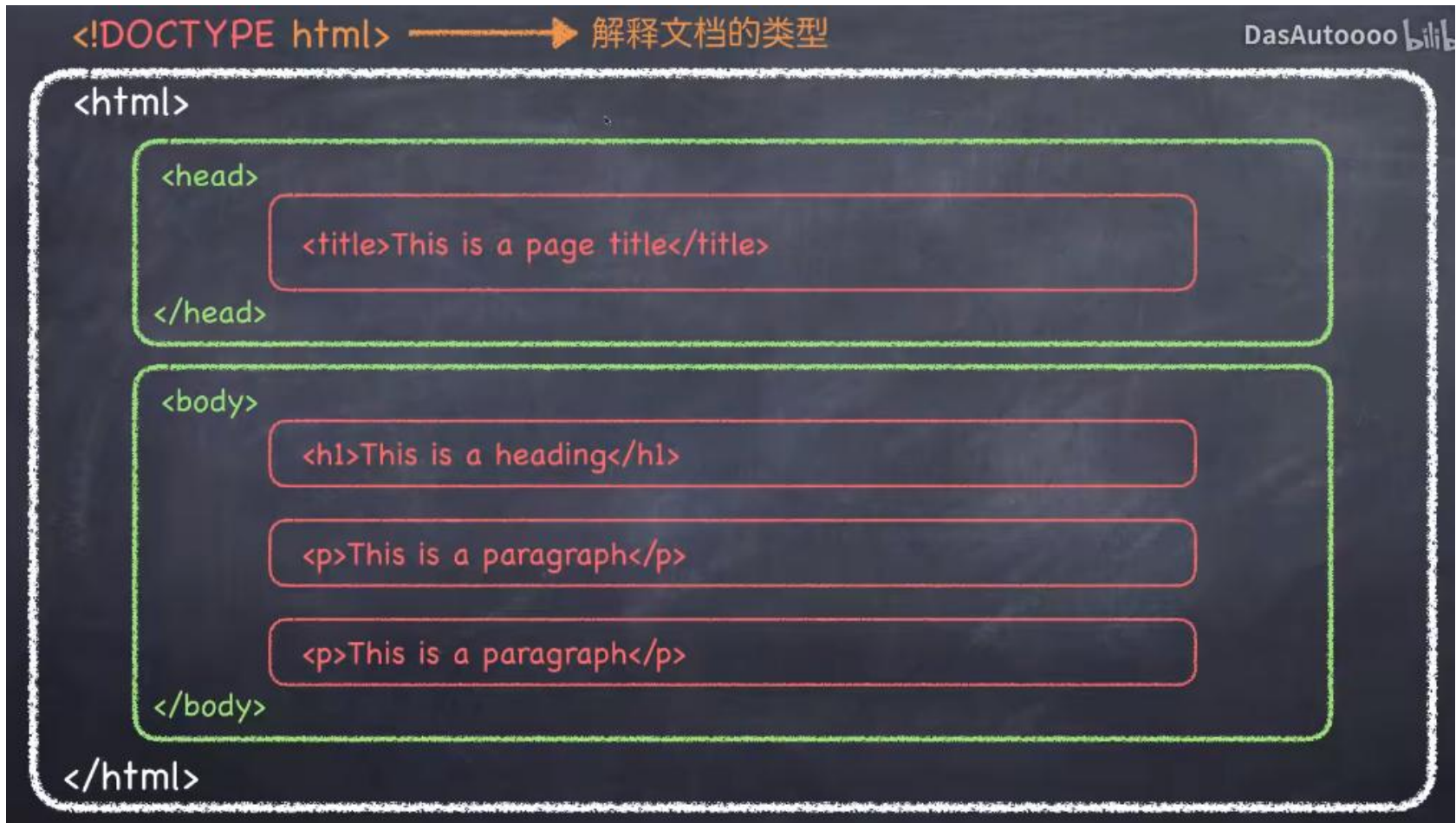
## 5. Working with Alternative Data

### **Hyper Text Transfer Protocol (HTTP)**

- Client(browser) & Server
- Request & Response
- get & post user-agent cookie referer ip
- html

# 5. Working with Alternative Data

## Hyper Text Markup Language (HTML)



## 5. Working with Alternative Data

### Hyper Text Markup Language (HTML)

```
<p class="paragraph" id="para1" >Lorem</p>
```

### Cascading Style Sheets(CSS)

```
.paragraph {  
    color : red ;  
}
```

## 5. Working with Alternative Data

### Hyper Text Markup Language (HTML)

### JavaScript(JS)

- 内部的 JavaScript

```
<script>
```

```
// Your JavaScript
```

```
</script>
```

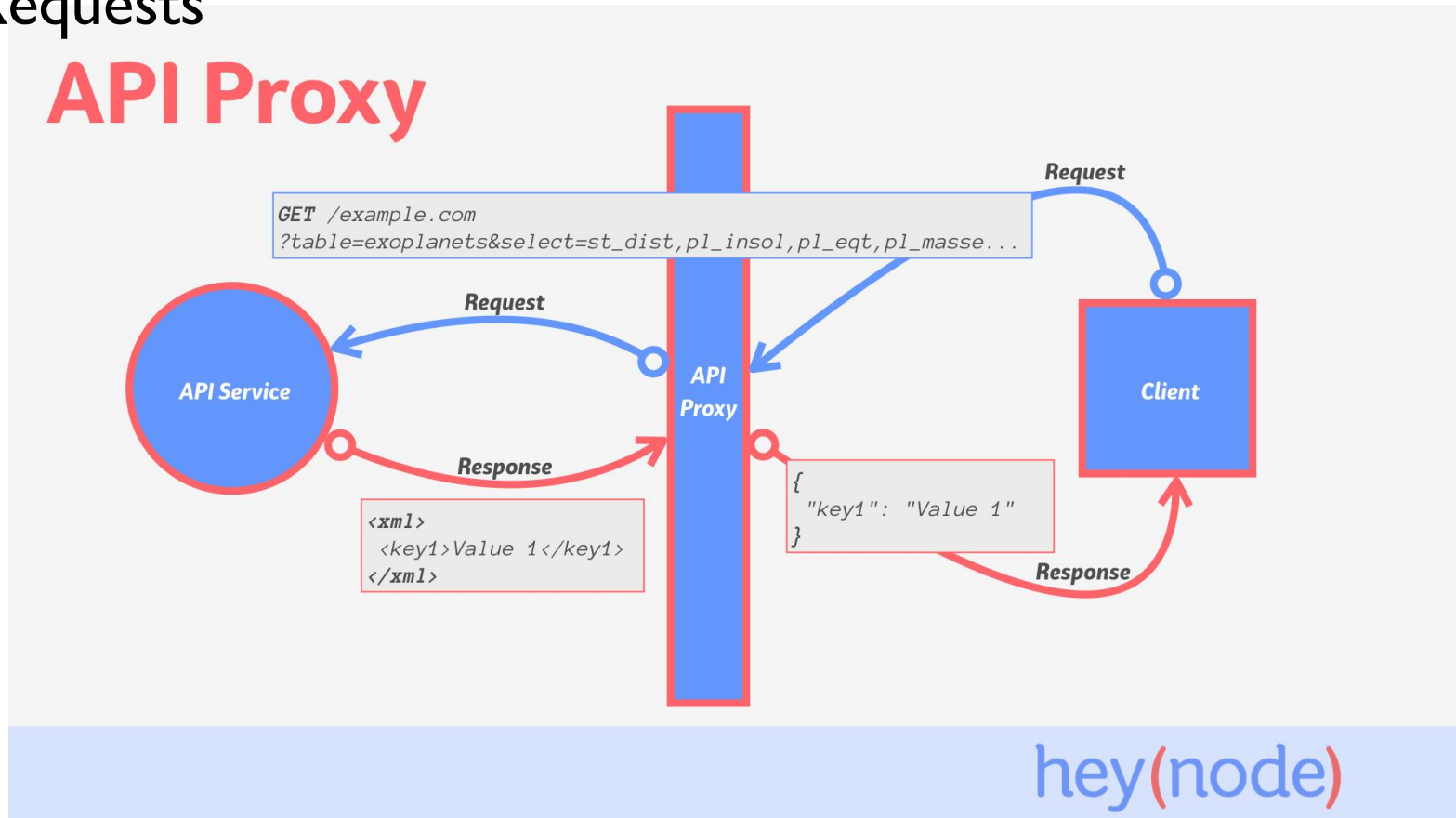
- 外部的 JavaScript

```
<script src="script.js"></script>
```

# 5. Working with Alternative Data

Talk is cheap, show me the code. -- Linus Torvalds

- Requests

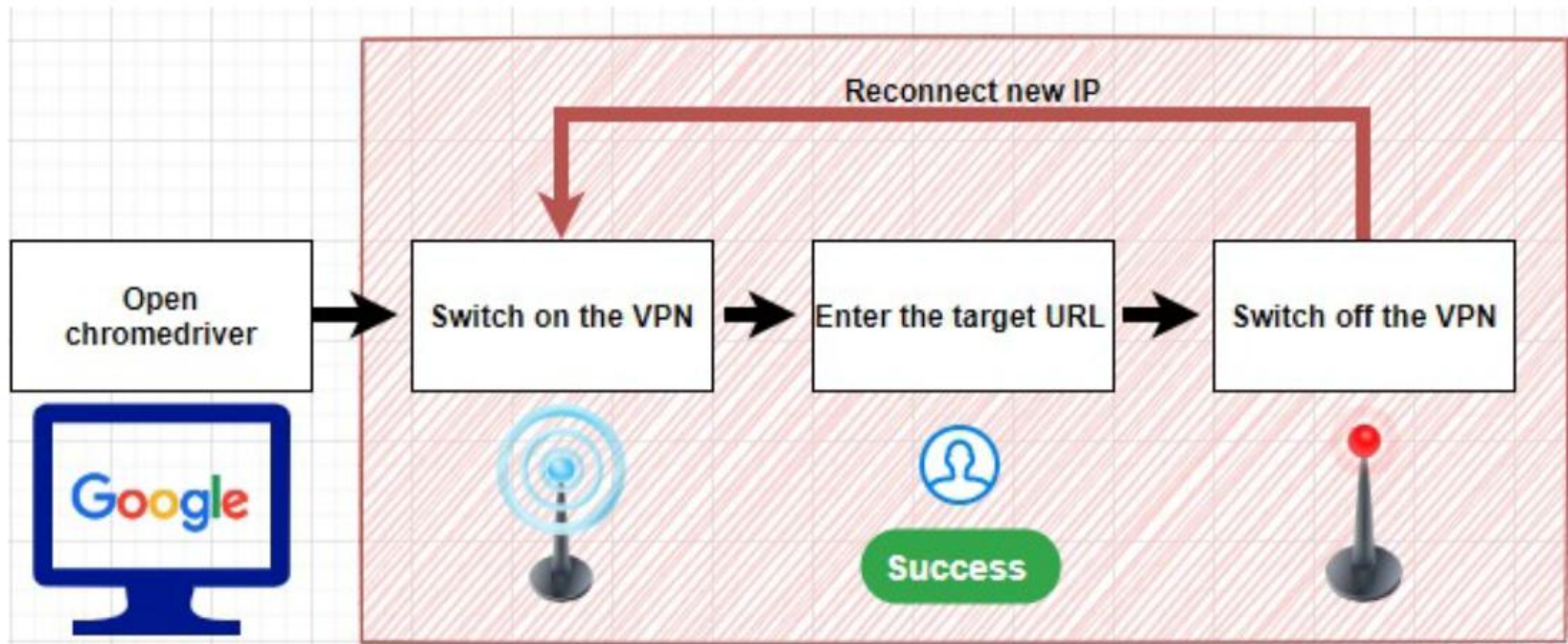




# 5. Working with Alternative Data

Talk is cheap, show me the code. -- Linus Torvalds

- Selenium



# 5. Working with Alternative Data

Talk is cheap, show me the code. -- Linus Torvalds

- Scrapy

