

# OSINT 검색 서비스

## OSINT 정의

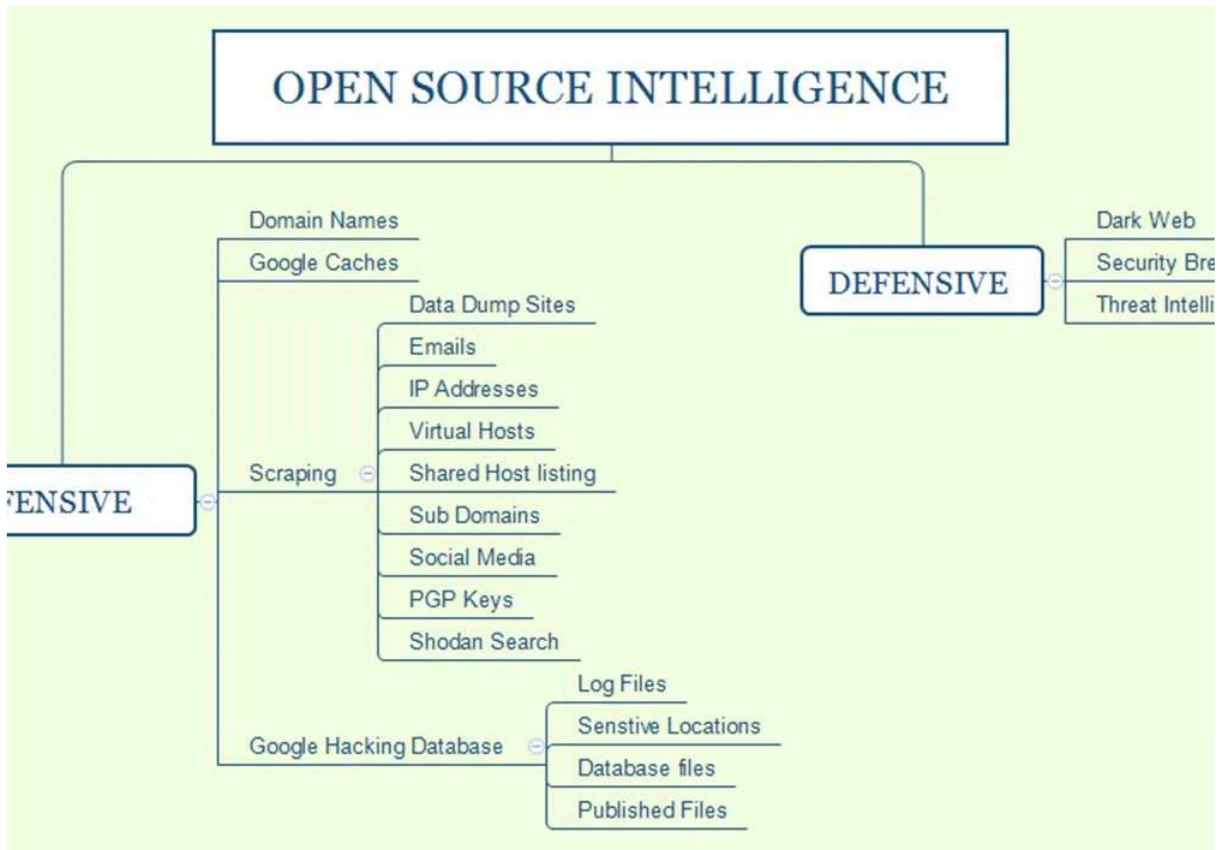
---

- OSINT는 Open Source Intelligence의 약자
- 공개적으로 사용 가능한 소스에서 수집한 인텔리전스
- 오픈 소스 소프트웨어나 공공(Public) 정보와는 관련이 없으며, 다양한 공개된 정보 소스에서 수집한 인텔리전스
  - 정보(Information) : 다양한 출처에서 수집된 데이터나 사실
  - 인텔리전스(Intelligence) : 특정 목적을 위해 수집되고, 분석 및 해석된 정보, 정보에 의미를 부여하고, 특정 맥락에서 유용한 통찰을 제공
- 대표적인 정보 수집 출처로는 크리미널아이피(Criminal IP), 쇼단(Shodan), **구글(Google)**, 야후(Yahoo), 트위터(Twitter) 등
- 프라이빗 정보와 다크웹 정보도 수집하는 것이 OSINT와 연관이 있지만, Public하게 공개되어 있지는 않음!!!!

## OSINT 공격자 및 방어자 관점

---

- 공개적으로 사용 가능한 소스는 담당자(방어자)와 범죄자(공격자) 모두 접근 가능
- 방어자와 공격자 모두에게 기회를 제공하기 때문에, 회사의 취약점을 학습하고 조치할 수 있는 동시에 공격자는 취약점을 악용할 수 있음
  - **방어자**는 OSINT를 활용하여 회사의 보안 취약점을 사전에 발견하고, 이를 개선함으로써 보안을 강화. 예를 들어, 쇼단(Shodan)을 이용해 인터넷에 노출된 장치를 찾아내고 보안 설정을 강화.
  - **공격자**는 OSINT를 이용하여 공격 대상을 조사하고, **취약점을 발견하여 악용**. 예를 들어, 소셜 미디어나 데이터 덤프 사이트에서 수집한 정보를 이용해 피싱 공격 계획.



Reference: mastering kali linux for Advanced Penetration Testing

## OSINT 사례 - 트위터 (X)

신기사

### 드? 외계인용 메시지?...미 전략사령부

21-03-30 06:31

나라 기자

단 페이지

이담 섞인 억측 날다 금세 삭제...사과 트윗도 지워져

합뉴스) 백나리 특파원 = 미군 전략사령부 트위터 계정에 정체를 알 수 없는 소동이 발생했다.

시간) 미 정치전문매체 더힐에 따르면 미 전략사령부 트위터 계정에 올랐다가 사라졌다.

설명 없이 암호처럼 ;l;;gmlxzssaw라고만 적힌 트윗이었다.

용자들 사이에서는 당장 농담 섞인 억측이 시작됐다. 실수로 핵무기 발사 명령을 내린 것이라든가, 컴퓨터 자판에 올라간 것이란 댓글 등이 줄을



US Strategic Command  
@US\_Stratcom

;l;;gmlxzssaw

19:48 · 3/28/21 · Twitter Web App

2,357 Retweets 2,075 Quote Tweets 5,497 Likes



Canadian Forces in Canada  
Replying to @US\_Stratcom

These things happen.  
This might even happen to you one day

It's okay, folks.

142

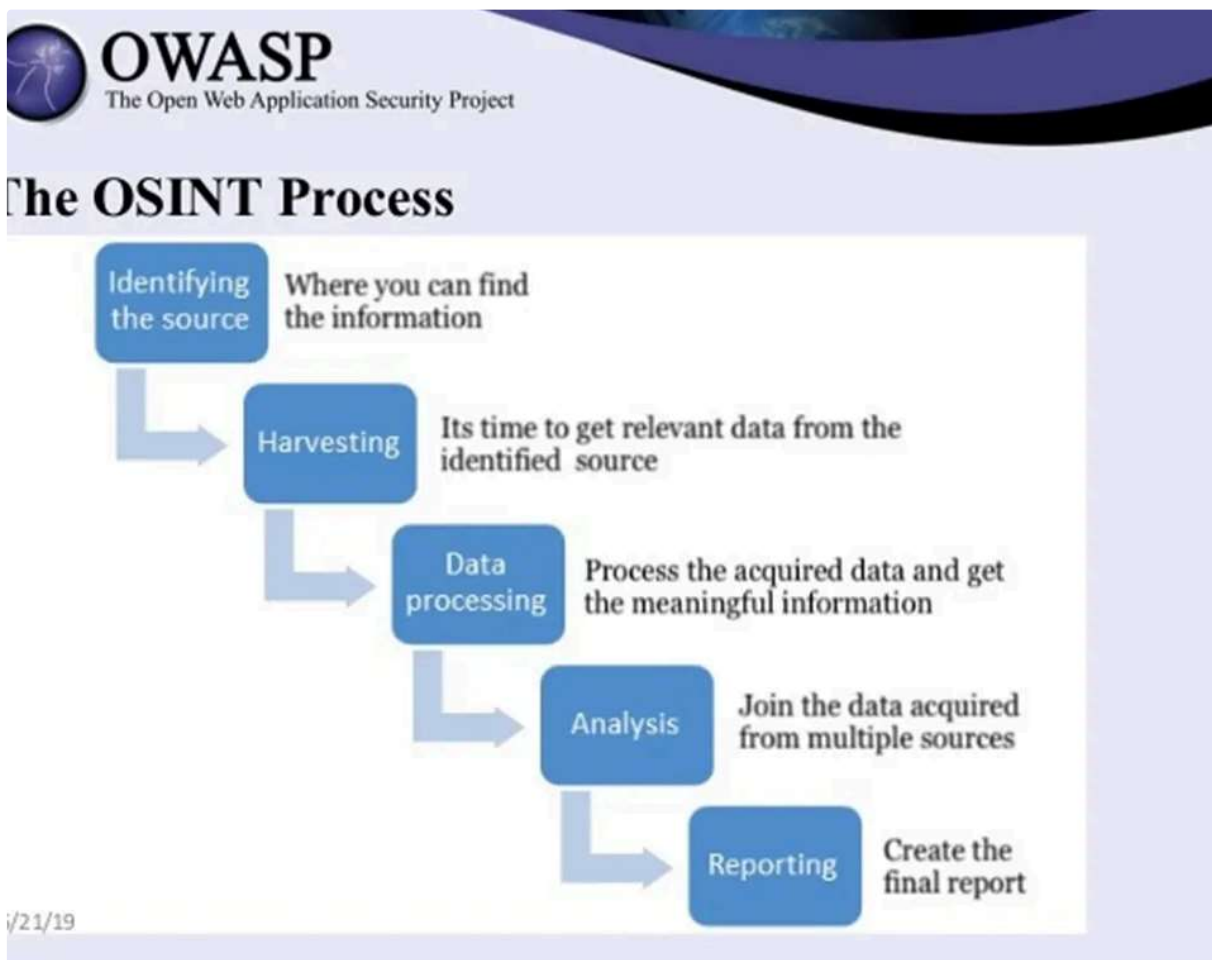
283

3,167



## OSINT Process

- OSINT Process는 정확하게 정해진 것이 없어 조직에 따라 요구사항에 맞춰 제작하는 것이 중요
- Identifying the source(소스 식별)  
정보를 찾을 수 있는 곳과 어떤 정보를 얻어야 하는지 식별하는 단계
- Harvesting(수확)  
식별된 소스에서 관련된 데이터를 가져오는 단계  
쇼단, 구글, 트위터 등에서 정보 수집
- Data Processing(데이터 처리)  
획득한 데이터를 처리하고 의미있는 정보를 얻는 단계
- Analysis(분석)  
여러 소스에서 수집한 데이터를 결합하는 단계
- Reporting(보고)  
앞서 진행한 단계를 종합하여 최종 보고서 작성



## OWASP\_OSINT\_Presentation

## 서브 도메인 정보 수집

### 서브 도메인

- 도메인(Domain)
  - 도메인은 숫자로 이루어진 IP 주소를 쉽게 기억하고 사용할 수 있도록 영문으로 표현한 것.
  - 예를 들어, `123.456.789.0` 이라는 IP 주소 대신 `example.com` 과 같이 사용
- 서브 도메인(Subdomain)
  - 서브 도메인은 기본 도메인에 추가되는 확장 부분으로, 보조 도메인 또는 2차 도메인으로도 불림.
  - 예를 들어, `blog.example.com` 에서 `blog` 가 서브 도메인으로, 서브 도메인은 특정 웹 서비스나 섹션을 분리하여 관리하기 위해 사용
- 보안 중요성
  - 일반적으로 알려진 도메인은 지속적으로 **보안 진단을 받아 안전할 가능성이 높음**
  - 그러나 보안 담당자나 인프라 담당자가 모르는 서브 도메인이 존재할 수 있으며, 이는 보안상의 허점이 될 수 있음 (recruit, test...)
  - 따라서 서브 도메인의 존재 여부를 확인하고, 그에 대한 보안 점검을 수행하는 것이 필수적

### Fierce 도구

- 서브 도메인 식별을 위해 사용
- `/usr/share/amass/wordlists/fierce_hostlist.txt` 사전 파일

```
sudo apt update sudo apt install fierce sudo fierce --domain google.com
```

### Google and Alphabet Vulnerability Reward Program (V...

Services in scope Qualifying vulnerabilities Non-qualifying vulnerabilities Reward amounts for security vulnerabilities

 <https://bughunters.google.com/about/rules/google-friend...>



Google Bug Hunters

# 한글 폰트 설치 `sudo apt install fonts-nanum*`

`sudo mousepad /usr/share/amass/wordlists/fierce_hostlist.txt`

## Netcraft 서비스

- **Netcraft**는 영국에 본사를 둔 인터넷 서비스 회사로, 사이버 범죄 중단, 애플리케이션 보안 테스트 및 자동화된 취약점 스캔을 포함한 다양한 인터넷 보안 서비스를 제공하는 회사
- **DNS 정보를 검색할 수 있는 무료 온라인 웹 서비스**
  - Netcraft는 DNS 정보뿐만 아니라 웹 사이트의 운영체제, 웹 서버 소프트웨어 버전, 서브 도메인 정보, SSL 인증서 정보, 서버의 생성 날짜와 가동 시간 등 다양한 정보를 제공하는 무료 온라인 웹 서비스를 제공

[searchdns.netcraft.com](https://searchdns.netcraft.com/)

 <https://searchdns.netcraft.com/>

## DNS Dump 사이트

### DNSDumpster - Find & lookup dns records for recon ...

Free domain research tool to discover hosts related to a domain. Find visible hosts from the attackers perspective for Red and Blue

 <https://dnsdumpster.com/>



## 구글 옵션을 이용해서도 서브 도메인 확인 가능



site:naver.com -site:www.naver.com



쇼핑 이미지 동영상 짧은 동영상 도서 웹 : 더보기

네이버 프리미엄콘텐츠

<https://contents.premium.naver.com>

## 버 프리미엄콘텐츠

인기 프리미엄 채널 · 고배당주 연구소 · 라오어와 미국주식 함께 걷기 1k · 재테크농부 1k · 수급단타왕x  
식스터디 1k · 도키와미국주식 1k.

네이버 개발자 센터

<https://developers.naver.com>

## 버 개발자 센터 - NAVER Developers

| 오픈 API들을 활용해 개발자들이 다양한 애플리케이션을 개발할 수 있도록 **API 가이드**와 **SDK**를 제공함  
제공중인 오픈 API에는 네이버 로그인, 검색, ...

## Internet Archive

- 수백만 권의 무료 도서, 영화, 소프트웨어, 음악, 웹 사이트 등의 비영리 도서관
- 전세계의 모든 웹 서비스에 대하여, 과거부터 현재 까지 웹 서비스 기록을 수집하여 보관하는 웹 서비스

Internet Archive: Digital Library of Free & Borrowable Books, Movies, Music & Wayback Mac...

Internet Archive is a non-profit digital library offering free universal access to books, movies & music, as well as 624 billion archived web pages.


<https://archive.org/>

## 구글 해킹 (Google Dork , Google Hacking) (별 5개)

- 구글 검색 및 기타 구글 애플리케이션 서비스를 활용한 정보 수집
- 구글에서 제공하는 다양한 검색 옵션을 활용

- **검색 옵션을 악의적인 목적으로 이용**하여 “구글 해킹” 용어 생김
- 구글봇이 수집하는 데이터를 서버에 캐시상태로 저장하기 때문에, 해당 사이트가 삭제되거나 한 후에도 오랜 시간이 지나기 전엔 검색결과에 노출되기 때문에 이전 페이지가 그대로 노출 될 수 있으며, 이 데이터를 모으면 손쉽게 취약점을 찾을 수 있다.

 googledork.pdf 2740.5KB

```
site:naver.com inurl:admin filetype:xlsx # 관리자 페이지에서 xlsx 노출된 것  
찾아!! site:naver.com filetype:log site:naver.com inurl:admin  
intext:password
```

## 구글 해킹 사례

[단독] 구글神 때문에...공무원 개인정보 및 관리자 페이...

지난 3월 3일 본지는 구글 검색에 의한 개인정보 노출의 심각성에 대해 문제를 제기한 바 있다. 너무나 열심히 일하는(?) 구글 봇 때문

 <https://www.boanews.com/media/view.asp?idx=53701>

[illegible]

내 결제정보의 구글 노출 여부, 어떻게 확인하나

최근 에스아이알소프트(SIR)가 제공하는 무료 웹 게시판 '그누보드 5'와 쇼핑몰 구축 솔루션 '영카트5'의 결제정보 노출 취약점이 발견

 <https://www.boanews.com/media/view.asp?idx=76120>

```
116.120.58.37 > lgxpay > lgdacom > log
```

```
... Response (LGD_RECEIVER, 0) = 2015-08-21 10:58:12 [DEBUG] ..... Response (LGD_BUYER, 0) = 2015-08-21 15:40:45 [DEBUG] ...
```

2015-06-17 02:02:58 [INFO] [] XPayClient



## 검색 옵션

다양한 검색 옵션으로 사용자가 원하는 결과를 보다 정확하게 검색

다 검색옵션  
intitle:  
inurl:  
:  
erange:  
ne:  
nebook:  
ited:  
ings”

순위	Port 번호	이벤트수	패킷수
1	80	1093	9183
2	25	575	183062
3	80,139,1025,2745,612 ...	351	1939
4	4899	300	8077
5	445	292	5409
6	21	152	1904
7	901	113	2426
8	1433	90	9157
9	1080	86	364
10	9898	83	829

## 구글 해킹 옵션


옵션	내용
inurl	-URL에서 특정 문자열을 검색 -URL에서 특정 웹 페이지나 디렉토리를 찾을 때 주로 사용
intitle	-title(타이틀)에 특정 단어나 구문이 제목에 포함된 페이지를 검색 -특정 주제와 관련된 웹 페이지를 찾을 때 주로 사용
filetype	-특정 파일 형식을 검색 -특정 파일을 찾거나, 특정 파일 형식으로 작성된 문서를 찾을 때 주로 사용
site	-특정 도메인 또는 하위 도메인에서 검색 -특정 웹 사이트에서 정보를 찾을 때 주로 사용
intext	-페이지의 본문에서 특정 단어나 구문을 검색 -특정 주제와 관련된 웹 페이지를 찾을 때 주로 사용



옵 션	내 용
쌍따옴표(" ")	-특정 단어나 구문이 정확히 일치하는 검색 결과를 표시하는 데 사용 -"아이폰 갤럭시"를 검색하면 아이폰과와 갤럭시라는 단어가 포함된 결과가 표시되지만, "아이폰"과 "갤럭시"가 함께 포함된 결과는 표시되지 않음
더하기 기호(+)	-검색 결과에 반드시 포함해야 하는 단어나 구문을 지정하는 데 사용 -"아이폰 +갤럭시"를 검색하면 아이폰에 대한 검색 결과에서 반드시 갤럭시라는 단어가 포함된 결과만 표시
빼기 기호(-)	-검색 결과에서 특정 단어나 구문을 제외하는 데 사용 -"아이폰 -갤럭시"를 검색하면 아이폰에 대한 검색 결과에서 갤럭시와 관련된 결과는 표시되지 않음
와일드카드(*)	-검색어 구문의 일부분을 대체하는 데 사용 -"apple * iphone"를 검색하면 "apple Korea iphone"나 "apple Store iphone" 등과 같은 다양한 검색 결과가 표시

이를 대응하기 위해서 robots.txt 파일을 루트 디렉터리에 추가해서 설정.. 하지만..

 **cjlandkids.kr**  
https://cjlandkids.kr  
https://cjlandkids.kr/  
페이지에 관한 정보가 없습니다.  
주 알아보기

← → G  cjlandkids.kr/robots.txt

```
# robots.txt
User-agent: *
Disallow: /
```

```
User-agent: * Disallow: /
```

User-agent: \* Disallow: /admin/ #관리자 페이지 접근 제어 미흡 (인증 미흡) Disallow: /documents/ # 중요 버전 정보 노출 Disallow: /images/ # 중요 버전 정보 노출 , index of 취약점(?)! Disallow: /passwords/ # index of 취약점, 중요 버전 정보 노출

```
import requests from bs4 import BeautifulSoup def get_robots_txt(url):
robots_url = url.rstrip('/') + "/robots.txt" try: response =
requests.get(robots_url, timeout=5) if response.status_code == 200:
print("[+] robots.txt 파일 발견!") print(response.text) return
response.text else: print("[-] robots.txt 파일이 없습니다.") return None
except requests.RequestException as e: print(f"[!] 요청 중 오류 발생: {e}")
return None def get_disallowed_paths(robots_txt): disallowed_paths = []
```