OTG-INFO-001

استفاده از موتور جستجو در جهت کشف و شناسایی نشت اطلاعات

۱- خلاصه:

موتورهای جستجو برای یافت نتایج، به ترتیب کارهایی را انجام میدهند. ابتدا برنامه های کامپیوتری (یا ربات ها) به صورت مرتب داده ها را جمع آوری می نمایند (کراول کردن میلیون ها صفحه وب). این برنامه ها صفحات وب را بر اساس لینک های موجود در صفحات دیگر یا سایت مپ (نقشه سامانه) پیدا می نمایند. فایل مخصوصی به نام robots.txt وجود دارد که شامل لیستی از صفحات است که وب سایت نمی خواهد این صفحات توسط موتورهای جستجو جمع آوری شوند، پس اگر وب سایت از فایل robots.txt استفاده نماید، صفحات لیست شده در این فایل، توسط موتورهای جستجو نادیده گرفته می شوند. این یک دیدگاه ابتدایی و پایه است – گوگل توضیحات عمیق تری در خصوص چگونگی کارکرد موتورهای جستجو ارائه داده است.

تست کننده ها می توانند از موتورهای جستجو جهت اجرای فرآیند شناسایی در وب سایت ها و نرم افزار های تحت وب استفاده نمایند. دو روش **مستقیم** و **غیرمستقیم** برای کشف و شناسایی توسط موتوهای جستجو وجود دارد: روش **مستقیم** مربوط به جستجوی ایندکس ها و ارتباط بین محتوای کش شده می باشد در حالی که روش غیر مستقیم شامل فهمیدن طرح های حساس و پیکربندی اطلاعات به وسیله جستجوی فروم ها، گروه های خبری و سایت های مناقصه می باشد.

وقتی یک ربات موتور جستجو، کاوش خود را کامل می کند، شروع به ایندکس کردن صفحات وب بر پایه ی تگ ها و خصیصه های مرتبط می نماید (مانند تگ <Title>) برای اینکه بتواند نتایج جستجوهای مرتبط را بازگرداند. اگر فایل robots.txt در طول مدت ارائه خدمات سامانه، بروزرسانی نگردد، و از تگ های داخلی meta که ربات های جستجو را راهنمایی می کنند که کدام صفحات نیازی به ایندکس نداشته، استفاده نگردد، آنگاه ممکن است ایندکس ها شامل محتوایی باشند که مالک سایت علاقه ای به انشار آنها ندارد.

مالکان وبسایت ها ممکن است از نکات قبلی در خصوص فایل های robots.txt، متا تگ های html، احراز هویت، و ابزار های ارائه شده توسط موتورهای جستجو برای حذف برخی از محتواهای مهم استفاده نمایند.

Mehran.python@gmail.com

۲- تست عملی

لازم است تا متوجه شویم چه اطلاعات حساس و تنظیمات مهمی از نرم افزار، سامانه یا سازمان افشا گردیده است که این کار با استفاده از هر دو روش **مستقیم**(سایت خود سامانه یا شرکت) و **غیرمستقیم**(وب سایت های ثالث) انجام می گیرد.

۳- چگونه تست کنیم؟

از یک موتور جستجو، جهت جستجوی موارد زیر استفاده می نماییم:

- نمودار شبکه و تنظیمات آن
- پست ها و ایمیل های آرشیو شده توسط ادمین یا دیگر کارمندان کلیدی سازمان
- مراحل ورود به سامانه و فرمت و قالب نام های کاربری
- نام های کاربری و پسوردها
- محتوای صفحات خطا
- نسخه های تست، درحال توسعه و آزمایشی وبسایت

۴-موتورهای جستجو:

هرگز تست خود را محدود به یک موتور جستجو ننمایید، زیرا موتورهای جستجوی مختلف، بر اساس کاوش گرهای خود و الگوریتم های جستجو مختص به خود، نتایج مختلفی را ارائه می دهند. استفاده از موتورهای جستجوی زیر را در نظر داشته باشید: (مرتب شده بر اساس حروف الفبای انگلیسی)

- Baidu

یکی از محبوب ترین موتورهای جستجوی چینی

- Bing

یکی از موتور های جستجو متعلق به مایکروسافت، دومین موتور جستجوی محبوب در جهان که از جستجوی پیشرفته پشتیبانی می نماید.

- BinSearch.info

یک موتور جستجو برای یافتن گروه های خبری (دودویی)

- DuckDuckGo

یک موتور جستجو با تمرکز بر حریم خصوصی که نتایج را از منابع بسیار متفاوت جمع آوری می نماید که از قواعد جستجو (جستجوی پیشرفته) نیز پشتبانی می نماید.

- Google

در حقیقت محبوب ترین موتور جستجوی جهان، که از سیستم رتبه دهی برای جمع آوری بهترین نتایج مشابه استفاده می نماید که از عملگرهای جستجو (جستجوی پیشرفته) نیز پشتیبانی می نماید.

- Startpage

یک موتور جستجو که عینا از نتایج گوگل استفاده می نماید اما اطلاعات خصوصی، سایت های جستجو شده و لاگ ها را جمع آوری نمی نماید (حریم خصوصی) و همچنین از عملگرهای جستجو (جستجوی پیشرفته) نیز پشتیبانی می نماید.

- Shodan

یک سرویس جستجوگر برای یافتن دستگاه ها و سرویس های متصل به اینترنت، گزینه های جستجو در نسخه رایگان محدودتر از نسخه خریداری شده می باشد.

دو موتور جستجوگر DuckDuckGo و Startpage به دلیل جمع آوری نکردن لاگ ها و تاریخچه جستجو ها، حریم خصوصی بسیار زیادی را برای تست کننده فراهم می کنند.

۵- عملگرهای جستجو:

عملگر جستجو به یک کلمه خاص گفته می شود که توانایی جستجوی منظم را افزایش می دهد و می تواند منجر به یافتن نتایج دقیق تر گردد. به صورت عمومی عملگرها به این شکل نوشته می شوند Operator:query که Operator عملگر مربوطه و query جستجوی ما می باشد. چند عملگر پر استفاده به شرح زیر می باشد:

- Site:
  نتایج جستجو را محدود به یک سایت خاص می نماید.
- Inurl:
  کلمه مربوطه را تنها در URL سامانه ها جستجو می نماید.
- Intitle:
  کلمه مربوطه را تنها در Title سامانه ها جستجو می نماید.
- Intext: یا inbody:
  کلمه مورد نظر را تنها در بدنه صفحات سایت جستجو می نماید.
- Filetype:
  تنها به دنبال نوع خاصی از فایل ها می گردد برای مثال PNG یا PHP

به عنوان مثال، برای یافتن نتایج مروبط به سامانه owasp.org از دستور زیر استفاده می نماییم:

```
site:owasp.org
```

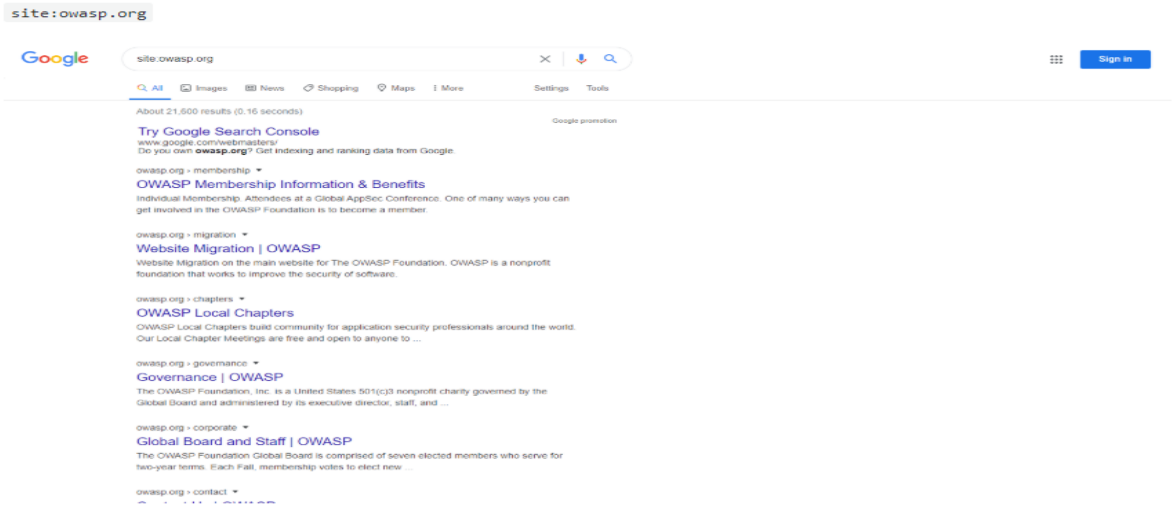

Figure 4.1.1-1: Google Site Operation Search Result Example

۶- مشاهده محتویات کش شده

برای جستجوی نتایجی که قبلا ایندکس شده اند، باید از عملگر :cache استفاده نمایید. این مورد می تواند به شما کمک کند تا محتوایاتی که با گذشت زمان تغییر کرده اند یا دیگر وجود ندارند را پیدا کنید. تمامی موتورهای جستجو، قابلیت جستجو در کش را نداشته و تنها منبع برای این مورد، موتورجستجو گوگل می باشد.

برای دیدن محتوای کش شده سامانه owasp.org از دستور زیر استفاده می نماییم:

```
cache:owasp.org
```

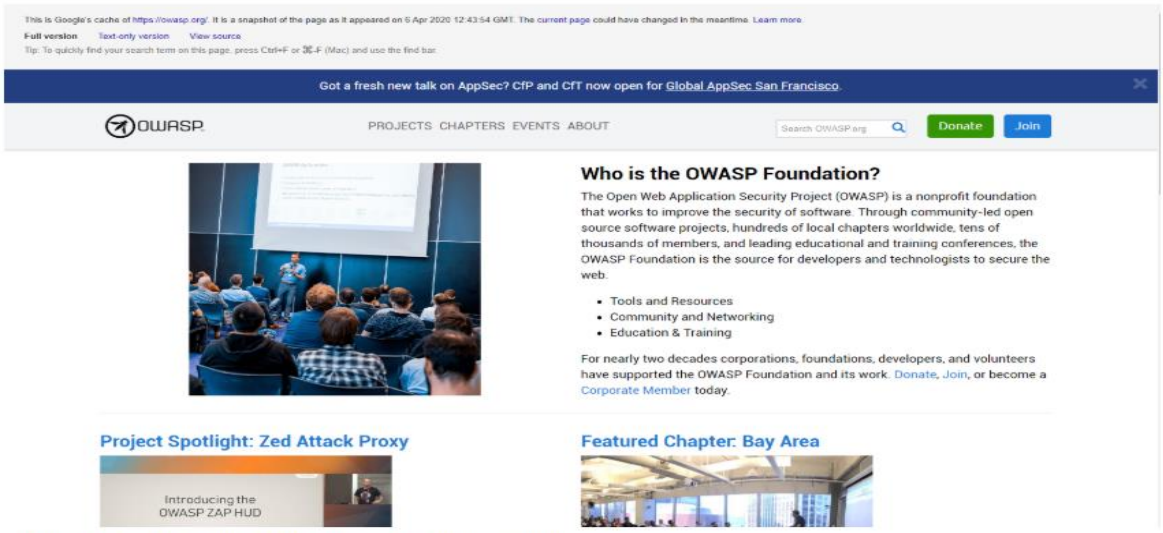

Figure 4.1.1-2: Google Cache Operation Search Result Example

۷- پایگاه داده گوگل هکینگ

جستجو بر اساس عملگرها وقتی با خلاقیت تست کننده ترکیب شود، می تواند نتایج بسیار تاثیرگذاری را بدست آورد. عملگرها را می توانند به نتایج جستجو و اطلاعات و داده های حساس ربط دهند. این تکنیک گوگل هکینگ یا گوگل دورک نام دارد و می توان توسط تمامی موتورهای جستجویی که از عملگرها استفاده می کنند استفاده شود.

یک پایگاه داده از دورک ها، مانند پایگاه داده ی گوگل هکینگ میتواند در پیدا کردن اطلاعات خاص بسیار مفید باشد. بعضی از دسته های دورک ها در پایگاه داده به شرح زیر می باشد:

- پایگاه (Footholds)
- فایل هایی شامل نام های کاربری
- دایرکتوری (مسیر) های حساس
- شناسایی وب سرور
- فایل های آسیب پذیر
- سرویس های آسیب پذیر
- پیام های خطا
- فایل هایی شامل اطلاعات مفید
- فایل هایی شامل پسورد ها
- اطلاعات حساس خرید آنلاین

پایگاه داده موتورهای جستجوی دیگر مانند بینگ و شودان نیز به منابع Bishop Fox's در پروژه ی Google Hacking Diggity Project دسترسی دارند

۸- راه های جلوگیری:

قبل از پست آنلاین، با دقت فراوان حساسیت طرح ها و پیکربندی های اطلاعات بررسی گردد.

به صورت دوره ای و مرتبط، حساسیت طرح ها و پیکربندی های اطلاعات بررسی گردد.

Mehran.python@gmail.com