MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY

Google hacking

BAKALÁRSKA PRÁCA

Matúš Chamula

Brno, 2007

Prehlásenie

Prehlasujem, že táto bakalárska práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj.

Vedúci práce: Mgr. Marek Kumpošt

Poďakovanie

Moje poďakovanie patrí predovšetkým vedúcemu práce Mgr. Markovi Kumpoštovi za odborné vedenie a pripomienky k danej téme. Zvláštne poďakovanie patrí mojim priateľom, ktorí mi pri jej vypracovaní pomohli.

Zhrnutie

Cieľom tejto bakalárskej práce je špecifikovať konkrétne možnosti využitia vyhľadávača Google ako hackovacieho nástroja a zhodnotiť hrozby a oblasti, ktorých sa táto problematika najviac dotýka. Jej súčasťou je demonštrácia hackingu pomocou Googlu na konkrétnych príkladoch a takisto aj metód ochrany proti takýmto typom útokov.

Kľúčové slová

hacking, Google, Yahoo, vyhľadávanie, GHDB, API, GScan

Obsah

Kapitola 1

Úvod

Rozvoj internetu a jeho technológií napreduje v poslednej dekáde extrémnym tempom a Google, ako jeden z jeho najvyužívanejších prostriedkov, bude nepochybne patriť k medzníkom svojej doby. Jeho databáza naberá obrovské rozmery a niet pochýb, že sa v nej okrem obecných zdrojov informácií nachádza množstvo citlivých údajov. Dáta, ktoré by mali podliehať ochrane, bývajú často dostupné verejne, takže k mnohým z nich sa dá dopátrať pomerne jednoducho, a to využitím internetového vyhľadávača.

Táto práca ma predstaviť a priblížiť rôznorodosť techník tvorby vyhľadávacích dotazov za účelom čo najefektívnejšieho využitia Googlu pri pátraní po potrebných informáciách. Prihliada sa však na skutočnosť, že vyhľadávač Google sa dá okrem klasických činností, na ktoré je predurčený, zneužiť na mnohé nekalé aktivity. Okruh potrebných informácií sa tak zužuje na osobné a citlivé dáta, ktorých znalosť uľahčuje hackerovi zaútočiť na vybranú obeť.

Jadrom hackingu pomocou vyhľadávača Google je najmä ovládnutie pravidiel používania jeho pokročilých operátorov, ktoré pomáhajú konkretizovať oblasť výsledkov a rovnako aj pokročilejšia znalosť chodu rôznych softwarových produktov. Hacker využívajúci Google tak upriamuje svoju pozornosť na obsah stránok automaticky generovaných príslušnou aplikáciou, ktoré mnohokrát prezrádzajú dôležité údaje a má možnosť touto cestou hľadať potencionálnu obeť. Ako tiež uvidíme, Google môže útočníka nasmerovať na množstvo užitočných stránok - či už ide o proxy servery prípadne stránky s úložiskom exploitov, môže z nich útočník v budúcnosti ťažiť. Špecifickou partiou hackingu Googlom je vyhľadávanie užívateľských mien, hesiel alebo telefónnych čísiel, na ktoré je možné naraziť pri bádaní po stránkach prihlasovacích portálov a viacerých dokumentoch pochádzajúcich z firemného intranetu. Získané informácie spadajúce do tejto kategórie dokážu najviac zúročiť sociotechnici.

V úvodnej časti tejto práce si predvedieme elementárne zásady používania Googlu spolu s booleovskými a pokročilými operátormi. Za nimi nasleduje kapitola zaoberajúca sa princípmi hackingu Googlom, ktorá sa takisto venuje tvorbe účinných dotazov slúžiacich na získanie citlivých dát. Ďalšou rozoberanou tematikou sú zásady efektívnej ochrany pred prípadným útokom hackerov využívajúcich Google a výpočet krokov potrebných na uchovanie dôvernosti svojich dát. Na záver budú predstavené možnosti automatizácie vyhľadávania prostredníctvom Google API a taktiež aplikácia GScan slúžiaca ako nástroj na skenovanie domén pomocou dotazov, ktorých výsledkom sú dáta zneužiteľné útočníkom.

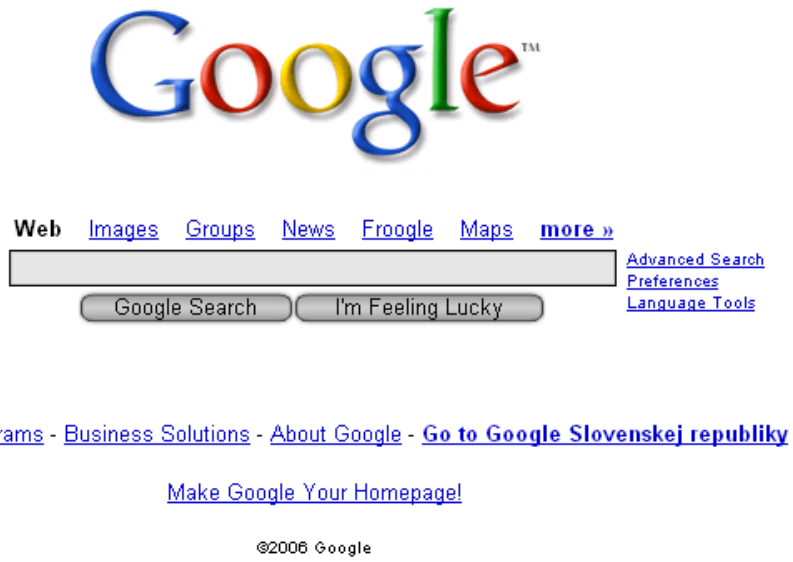
Kapitola 2

Vyhľadávanie na internete

Množstvo údajov zhromaždených na webe začína byť v dnešnej vyspelej dobe neprehľadné a často nejedného užívateľa zmätie. Exponenciálny rast nových dát kladie deň čo deň vyššie nároky na ich uchovávanie a internetová sieť postupom času čoraz viac pripomína knižnicu bez katalógov, ktorá by sa s vylúčením možností a prostriedkov pátrania po informáciách stala nepoužiteľnou a chaotickou. Na dolovanie potrebných poznatkov či dokumentov internetu slúžia vyhľadávače, ktorých je v súčasnosti nemalý počet.

2.1 Google a jeho minulosť

Presne 21. septembra 1999 bol po niekoľkých alfa a beta verziách oficiálne spustený Google, k dnešnému dňu jednotka vo vyhľadávaní [5]. Svojej obľúbenosti vďačí najmä pokročilým možnostiam zadávania dotazov, ktoré sú v kombinácii s mechanizmom generovania výsledkov pilierom úspechu.



Obrázok 2.1: Hlavná stránka Googlu

Samotný názov je odvodený od slova *googol*, ktoré ako prvý použil matematik Milton Sirotta na označenie čísla 10^{100} a Google prevzal tento pojem ako vyjadrenie vôle usporiadať zdanlivo nekonečné množstvo informácií dostupných na internete. Zakladatelia Larry Page a Sergey Brin sa môžu tešiť jeho obrovskej popularite, nakoľko stránky Googlu navštívi v súčasnosti viac ako 380 miliónov unikátnych užívateľov mesačne [5]. Niet sa skrátka čo čudovať, že pojem *google* zdomácnel v bežnej hovorovej reči a dokonca sa začína objavovať aj na stránkach slovníkov.

2.2 Webové rozhranie a služby Google

Hlavnú stránku Googlu zobrazenú na obrázku 2.1 je možné nájsť na <www.google.com>. Už na prvý pohľad pôsobí prívetivo a charakterizuje ju hlavne celková prehľadnosť a jednoduchosť. Nie je teda žiadnym prekvapením, že design Googlu chráni copyright [7].

Prvoradým prostriedkom na interakciu s užívateľom je vstupné textové pole, do ktorého sa vkladajú jednotlivé vyhľadávacie dotazy. Pomocou odkazov viditeľných nad ním je možné dostať sa do ďalších sekcií, ktoré Google ponúka. Ich využitím môžeme vyhľadávať obrázky, správy diskusných skupín či novinové články.

Google postupuje s dobou a svoje možnosti poskytovania služieb rozšíril o populárny *Gmail* s viac ako 2 GB priestoru zdarma a v súčasnosti vytvára konkurenciu na trhu kancelárskych balíkov spustením sady aplikácií *Google Apps for Your Domain*, pričom sú spolu s textovým procesorom *Writely* a tabuľkovým editorom *Spreadsheets* v blízkej budúcnosti pripravované aj ďalšie kancelárske nástroje [6].

2.3 Tvorba vyhľadávacích dotazov

Google stelesňuje veľmi mocný nástroj a v prípade, že je potrebné čokoľvek nájsť, siahne väčšina užívateľov práve po ňom. Na účinné získavanie informácií je však nutné ovládať jeho základnú syntax a v značnej miere sa vyznať v technikách redukcie jednotlivých výsledkov.

2.3.1 Syntax

Irelevantným faktorom pri tvorbe dotazov je tzv. *case-sensitive* vlastnosť, Google totiž veľkosť písmen nerozlišuje. Takže výrazy *hacker*, *HACKER* či *HaCkEr* sú navzájom ekvivalentné. Jedinou výnimkou v tejto oblasti je operátor *OR* [pozri podkapitolu 2.3.2] [7].

Jednou z vymožeností Googlu sú tzv. *wildcards*, čiže zástupné znaky, obvykle chápané ako náhrada jedného, prípadne viacerých symbolov. Google však tento pojem definuje odlišným spôsobom a ich predstaviteľom je hviezdička (*), ktorá vo vyhľadávacom dotaze nahradzuje práve jedno slovo [3].

Pri hľadaní určitej frázy môže byť nápomocný operátor úvodzoviek. Konkrétne slovné spojenie do nich môžeme jednoducho uzavrieť a tým dať Googlu pokyn, aby vypátral celú

frázu v takom poradí slov, v akom sme ich uviedli. Takýto typ vyhľadávania je naviac možné kombinovať s booleovskými a pokročilými operátormi [pozri podkapitoly 2.3.2 a 2.3.3] [7].

2.3.2 Booleovské operátory

Tieto operátory stelesňujú ďalšiu možnosť segmentovania komplikovaných dotazov a dokážu zoskupiť ich elementárne prvky do podoby, ktorá nám prinesie viacero relevantných výsledkov. Okrem toho tvoria výnimku faktu, že dotazy Googlu nie sú *case-sensitive*, nakoľko je nutné operátor *OR* zadávať veľkými písmenami. Z booleovských operátorov rozpozná Google práve tri a síce *AND*, *NOT* a *OR* [3]. Prednosťou každého z nich je možnosť ich kombinácie s pokročilými operátormi [pozri podkapitolu 2.3.3].

Najbežnejším je operátor *AND* a jeho použitím môžeme začleniť do dotazu viacero termínov. Príkladom môže byť výraz:

```
pat AND mat
```

Prostredníctvom nehoho zahŕňame do okruhu vyhľadávania stránky obsahujúce oba parametre. Treba však dodať, že Google považuje operátor *AND* za redundantný a štandardne sa snaží vyhľadávať všetky termíny uvedené v dotaze [3].

Ak však chceme Google donútiť, aby do výsledkov hľadania zaradil konkrétny výraz, máme možnosť využiť operátor *+*, za ktorým musí bez medzery nasledovať požadované slovo [7]. Majme dotaz:

```
"civil war" +spain
```

Ten zabezpečí, že výraz *spain* sa bude nachádzať na každej z výsledných stránok.

Opačnou funkciou disponuje operátor *NOT*, vďaka ktorému je možné časť výsledkov z hľadania odstrániť. Jedným z efektívnych spôsobov aplikácie tohto operátoru je využitie jeho funkčného ekvivalentu v podobe znamienka mínus (-), medzi operátorom a jeho argumentom sa však opäť nesmie vyskytovať medzera [7].

```
windows -xp
```

Vo výsledkoch tohto dotazu sa teda neobjavia stránky obsahujúce výraz *xp*.

Posledným booleovským operátorom je *OR*, ktorý má okrem samotného slova *OR* aj ďalší reprezentant, a to zvislú čiaru (|). Tento operátor napovedá Googlu, že máme v úmysle pátrať buď po jednom alebo druhom slove uvedenom v dotaze. Na výslednej stránke sa tak môže vyskytovať iba jeden z argumentov [9]. Ako jednoduchý príklad uveďme dotaz:

```
"fire fox" | firefox
```

Ten ako výsledok vráti dokumenty obsahujúce frázu *fire fox* alebo výraz *firefox*.

V jednom dotaze je samozrejme možné použiť viacero booleovských operátorov a tým prichádza do úvahy ich priorita. Tá je však prostá, pretože i zložité dotazy chápe Google ako vety čítané zľava doprava. Operátor *AND* ma totiž rovnakú váhu ako *OR* a ten má zas rovnakú váhu ako akýkoľvek pokročilý operátor. Tento faktor ovplyvní jedine poradie výsledkov hľadania, ale nemá žiadny dopad na to, ako Google dotaz spracuje [7].

2.3.3 Pokročilé operátory

Veľmi užitočným nástrojom Googlu pri redukovaní a zužovaní výsledkov jednotlivých dotazov sú pokročilé operátory. Ak sú použité správnym spôsobom, dokážu nás naviesť presne k tým informáciám, ktoré sa pokúšame nájsť. Pokiaľ v dotaze žiadny pokročilý operátor uvedený nie je, bude Google zadaný termín hľadať vo všetkých oblastiach webovej stránky, to znamená v titulku, v texte, v URL a podobne [7].

Operátory tvoria časť vyhľadávacieho dotazu Googlu, avšak na rozdiel od štandardných dotazov pre ne existujú syntaktické pravidlá, ktoré sú nutnosťou na ich korektné používanie. Syntax pokročilých operátorov Googlu má tvar:

```
operátor:argument
```

Medzi operátorom a argumentom sa nesmie vyskytovať medzera, v opačnom prípade Google nemusí pochopiť, že sa jedná o pokročilý operátor a bude ho považovať za ďalší hľadaný termín [7].

Obsahom nasledujúcich riadkov je sumár pokročilých operátorov frekventovane využívaných pri technikách hackingu Googlom spolu s elementárnymi príkladmi ich využitia.

- **`intitle, allintitle`**

  Operátor `intitle` donúti Google vyhľadávať zadaný termín v titulku stránky. Ten sa dá definovať ako text, ktorý obsahuje dokument HTML v tagu <title> a zobrazuje sa v záhlaví prehliadača.

  ```
  intitle:"home page" "Marek Kumpost"
  ```

  Z toho vyplýva, že výsledkom uvedeného dotazu budú stránky obsahujúce frázu *home page* v titulku a *Marek Kumpost* priamo v dokumente. Druhý zmienený `allintitle` zaručuje, že Google vyhľadá v titulku jednotlivo všetky slová a frázy uvedené ako argumenty [3].

- **`inurl, allinurl`**

  Tento operátor obmedzí výsledky na dokumenty obsahujúce zadaný reťazec v URL, ktoré predstavuje samotnú adresu webovej stránky zobrazenú v prehliadači [3]. Na pátranie po slovách *xkumpost* v URL a *GnuPG* kdekoľvek v dokumente slúži nasledovný dotaz:

  ```
  inurl:xkumpost GnuPG
  ```

  Treba však podotknúť, že Google nedokáže v URL účinne vyhľadať komponent protokolu ako napríklad *http://*. S celkovou zložitosťou adresy súvisí aj ďalší problém - URL totiž obsahuje množstvo špeciálnych znakov a akýkoľvek pokus o ich explicitné zadávanie do vyhľadávacieho reťazca môže viesť k nesprávnym výsledkom [7]. Druhým variantom tohto operátoru je `allinurl`, ktorý sa snaží v URL nájsť akékoľvek slovo alebo frázu uvedené ako jeho parametre.

- **`site`**

  Úlohou `site` je umožniť vyhľadávanie výrazov v konkrétne špecifikovanej doméne. Ak teda máme určitú predstavu o informáciách, ku ktorým sa snažíme dopátrať, a zároveň chceme prehľadávať iba stránky Masarykovej a Karlovej univerzity, plne na to postačí dotaz:

  ```
  "prijimaci rizeni" site:muni.cz | site:cuni.cz
  ```

- **`filetype`**

  Tento operátor prikazuje hľadať kľúčové slová v súbore s príslušnou príponou. Okrem štandardných dokumentov vo formáte HTML Google vyhľadáva a indexuje spolu 13 ďalších typov súborov, naviac postupom času pribúdajú ďalšie, ktorých výskyt je však zriedkavejší [7]. Prípony, na ktoré Google upriamuje najväčšiu pozornosť, sú uvedené v tabuľke 2.1.

Adobe Portable Document Format	`pdf`
Adobe PostScript	`ps`
Lotus 1-2-3	`wk1, wk2, wk3, wk4, wk5, wki, wks, wku`
Lotus WordPro	`lwp`
MacWrite	`mw`
Microsoft Excel	`xls`
Microsoft PowerPoint	`ppt`
Microsoft Word	`doc`
Microsoft Works	`wks, wps, wdb`
Microsoft Write	`wri`
Rich Text Format	`rtf`
Text	`ans, txt`

Tabuľka 2.1: Najčastejšie typy súborov indexované Googlom [3]

  Zaradením `filetype` do dotazu máme dokonca možnosť vymedziť typy súborov, ktoré by sa v množine výsledkov vyskytovať nemali [3]. Nasledovný dotaz z výsledkov vylučuje súbory s príponami *pdf* a *ppt*:

  ```
  "google hacking" -filetype:pdf -filetype:ppt
  ```

- **`link`**

  Ako už jeho názov napovedá, sme schopní pomocou `link` nájsť na webe stránky obsahujúce odkazy, ktoré smerujú na zadaný cieľ [3]. Pre syntaktickú správnosť celého výrazu je nutné ako argument uviesť URL:

  ```
  link:www.muni.cz
  ```

- **`inanchor`**

  Tento operátor obmedzí výsledky vyhľadávania na stránky skrývajúce v popise odkazu zadaný výraz. Svojím spôsobom sa ponáša na operátor `link` s tým rozdielom, že nehľadá skutočnú URL, ale textovú reprezentáciu odkazu ukrytú v HTML tagu <a> [3].

- **`numrange`**

  Google má vo svojom repertoári aj menej známy operátor `numrange`, ktorý dokáže vyhľadať čísla v určitom rozsahu, a na rozdiel od predošlých operátorov vyžaduje dva parametre. Prvým je dolná a ďalším horná hranica intervalu čísiel, ktoré sa pokúšame nájsť. Pri zahrnutí `numrange` do dotazu nie je nutné uvádzať jeho názov, oba parametre však musia byť oddelené dvoma bodkami. Pre uľahčenie vyhľadávania Google pri číslach ignoruje okolité znaky, akými sú desatinná čiarka alebo reprezentant národnej meny [7]. Ak teda pátrame po historických faktoch 20. storočia, môže byť nápomocný dotaz:

  ```
  "civil war" 1901..2000
  ```

- **`intext`**

  Operátor `intext` slúži na hľadanie reťazca priamo v texte webovej stránky. Inak povedané, pracuje na rovnakom princípe ako vyhľadávací engine s tým rozdielom, že sa zadaný reťazec nesnaží nájsť v titulku, URL, odkazoch či súboroch so špecifickou príponou [7].

Pokročilé operátory samé o sebe prácu s Googlom značne uľahčujú a poskytujú množstvo ciest ako výsledky dotazov obmedziť na potrebnú mieru. Ich vhodnou kombináciou sa však z Googlu stáva ideálny nástroj na hľadanie citlivých dát, z ktorých môže útočník v priebehu útoku ťažiť. Práve táto skutočnosť je stredobodom celej práce a v ďalších kapitolách bude podrobne rozoberaná.

Kapitola 3

Základy hackingu Googlom

Google sa stal za dobu svojho pôsobenia vo svete internetu štandardom, ktorý v sebe skrýva obrovský potenciál. V tejto kapitole si ukážeme akým spôsobom je vďaka nemu možné dopátrať sa k údajom, ktoré by nemali byť vôbec dostupné. Pri tvorbe tých správnych dotazov totiž Google odhalí často udivujúce výsledky [9]. *Google hacking* ako pojem je však trochu zavádzajúci, nejedná sa totiž o hacking v pravom slova zmysle. *Google hacking* je vlastne synonymom prehľadávania obrovskej databázy Googlu a jeho využitia na nájdenie citlivých dát, potenciálne zraniteľných obetí prípadne zistenie konkrétnejších informácií o potenciálnej obeti útoku [7].

Nasledujúce riadky sa budú venovať technikám tvorby vyhľadávacích reťazcov, ktorých výsledkom sú najmä citlivé a dôverné dáta ako aj ďalšie možnosti ponúkané Googlom a využiteľné hackermi.

3.1 Anonymita s využitím proxy

Najpodstatnejším faktorom každého počítačového hackera je popri všetkej jeho aktivite anonymita a vďaka nekalej činnosti majú hackeri pádny dôvod ukrývať sa pod rúškom utajenia. Odvrátenie vlny podozrenia má teda najvyššiu prioritu a na tento účel je proxy server veľmi účinným prostriedkom.

Proxy stelesňuje akúsi prestupnú stanicu. Dáta, ktoré užívateľ vysiela a prijíma, najskôr putujú k proxy serveru, ktorý sa ako dočasný hostiteľ postará o ich správne presmerovanie. Server, na ktorý sa posielajú príslušné požiadavky, sa teda dozvie jedine IP adresu prestupnej stanice, pričom užívateľ zostáva neidentifikovaný a práve tento fakt sa pre hackera stáva odrazovým mostíkom [7].

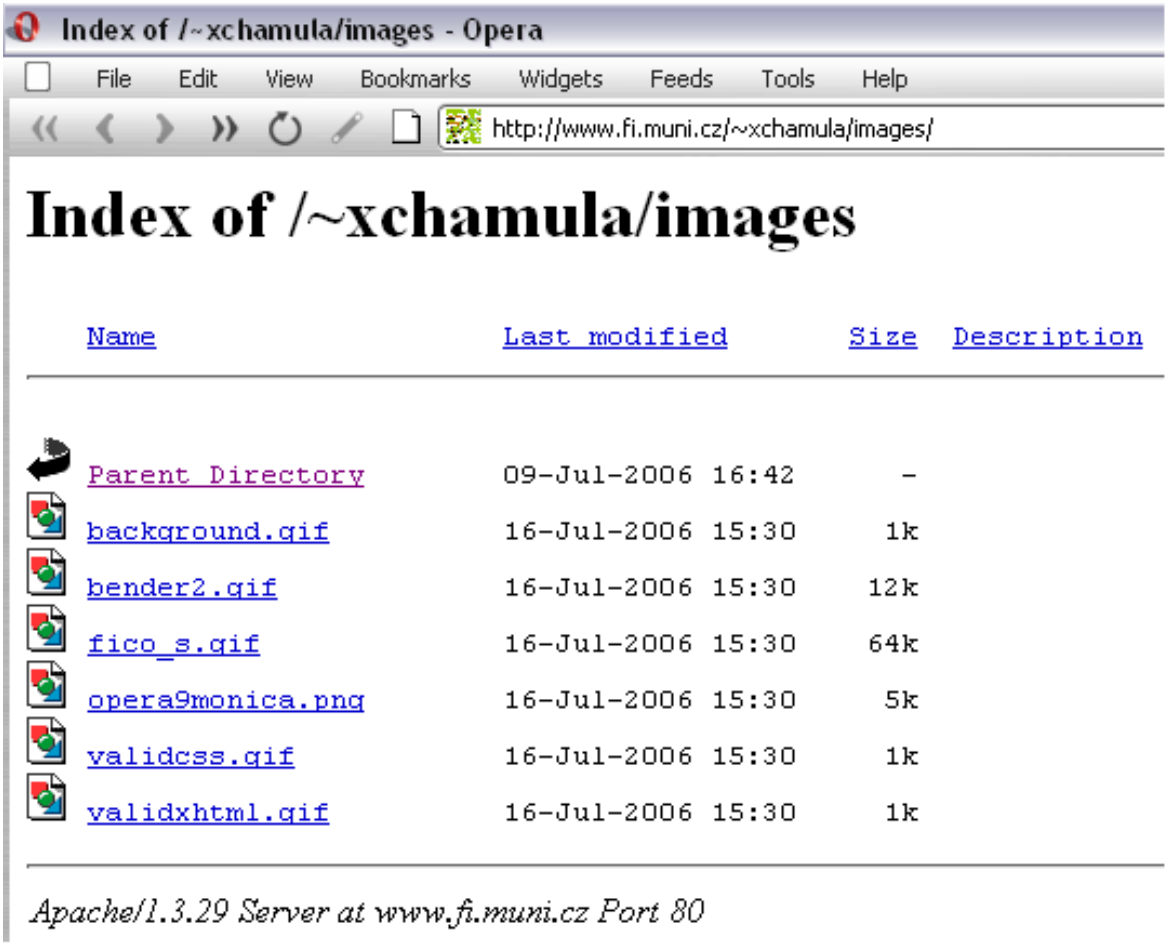
CGI proxy je typom proxy servera vo forme CGI skriptu, ktorý beží na webovom serveri, navonok sa chová ako klasický HTTP proxy a môže poslúžiť ako čiastočná ochrana anonymity. Defaultná stránka *CGI proxy* býva veľmi často uložená v súbore s názvom `nph-proxy.cgi` a obsahuje konkrétny text - *Start browsing through this CGI-based proxy by entering a URL below* [3]. S týmito poznatkami už môžeme vyprodukovať jednoduchý dotaz, ktorý nás odkáže na množstvo *CGI proxy*:

```
inurl:"nph-proxy.cgi" "Start browsing through this CGI-based proxy"
```

Na dosiahnutie konkrétnejších výsledkov využívame fakt, že stránka s *CGI proxy* má názov jej hlavného súboru v URL a tiež znalosť obsahu defaultnej stránky.

3.2 Výpisy adresárov

Výpis adresárov je špeciálnym druhom stránky, na ktorej je zobrazený zoznam súborov a adresárov nachádzajúcich sa na webovom serveri. Typicky obsahujú taktiež titulok nesúci názov aktuálneho adresára a pätu, ktorá vyznačuje jeho koniec [3]. Všetky zmienené prvky sú zobrazené na obrázku 3.1.



Obrázok 3.1: Ukážka výpisu adresára

Výpis adresárov je pre každý webový server špecifický a charakterizuje ho viacero faktorov. Prvým rysom, ktorému treba venovať pozornosť, je titulok obsahujúci frázu *index of*, ktorá figuruje vo všetkých výpisoch. Podľa obrázku 3.1 je takisto badateľné, že súčasťou päty je zobrazenie verzie webového servera. Google dotaz slúžiaci na vypátranie adresárových výpisov web serveru prislúchajúceho k obrázku by mal takúto podobu:

```
"Apache/1.3.29 Server at" intitle:"index of"
```

So znalosťou jednotlivých verzií web serverov možno vypracovať dotazy, vďaka ktorým nás k nim Google dovedie prostredníctvom výpisov adresárov. Tabuľka 3.1 demonštruje príklady týchto dotazov.

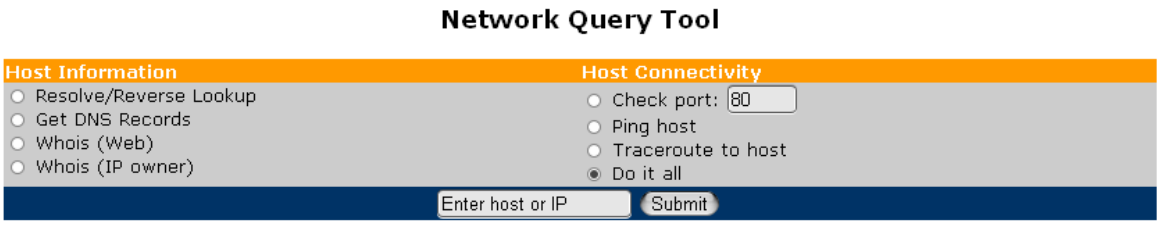
Apache 2.0	`"Apache/2.0 Server at" intitle:index.of`
Ľubovoľná verzia Apache	`"Apache/* Server at" intitle:index.of`
Microsoft IIS 6.0	`"Microsoft-IIS/6.0 Server at" intitle:index.of`
Ľubovoľná verzia Microsoft IIS	`"Microsoft-IIS/* Server at" intitle:index.of`

Tabuľka 3.1: Dotazy pre hľadania výpisov adresárov rôznych WWW serverov [9]

Verzia webového servera je príkladom informácie, ktorá sama o sebe nebezpečná nie je. Pravdepodobnosť výskytu bezpečnostnej chyby u konkrétnej verzie v budúcnosti však rozhodne nie je nulová, pričom útočník môže disponovať exploitom účinným voči danej verzii serveru, a tak sa vďaka tejto technike stáva Google užitočným vyhľadávačom potenciálnych obetí [7].

3.3 Network Query Tool

Google bezpochyby patrí medzi účinné nástroje, ale i s jeho pomocou nie je možné dosiahnuť všetko, a preto je mnohokrát efektívnejšie pri skúmaní poznatkov o potenciálnej obeti využiť služby ďalších prostriedkov. Jedným z nich je aj utilita zvaná *Network Query Tool* zobrazená na obrázku 3.2.



Obrázok 3.2: Aplikácia Network Query Tool

Network Query Tool (*NQT*) je webová aplikácia, takže ju môže využívať každý, kto má prístup ku stránke s *NQT*. Užívatelia webu sú teda schopní získavať záznamy DNS, vykonávať ekvivalenty linuxového príkazu *whois*, overovať aktivitu na konkrétnych portoch alebo dokonca sledovať trasy paketov pomocou *traceroute*. *NQT* okrem iného zaručuje aj vyššiu úroveň anonymity, keďže všetky požiadavky pochádzajú z webu, ktorý túto aplikáciu hostí a webový server defacto maskuje skutočnú adresu užívateľa.

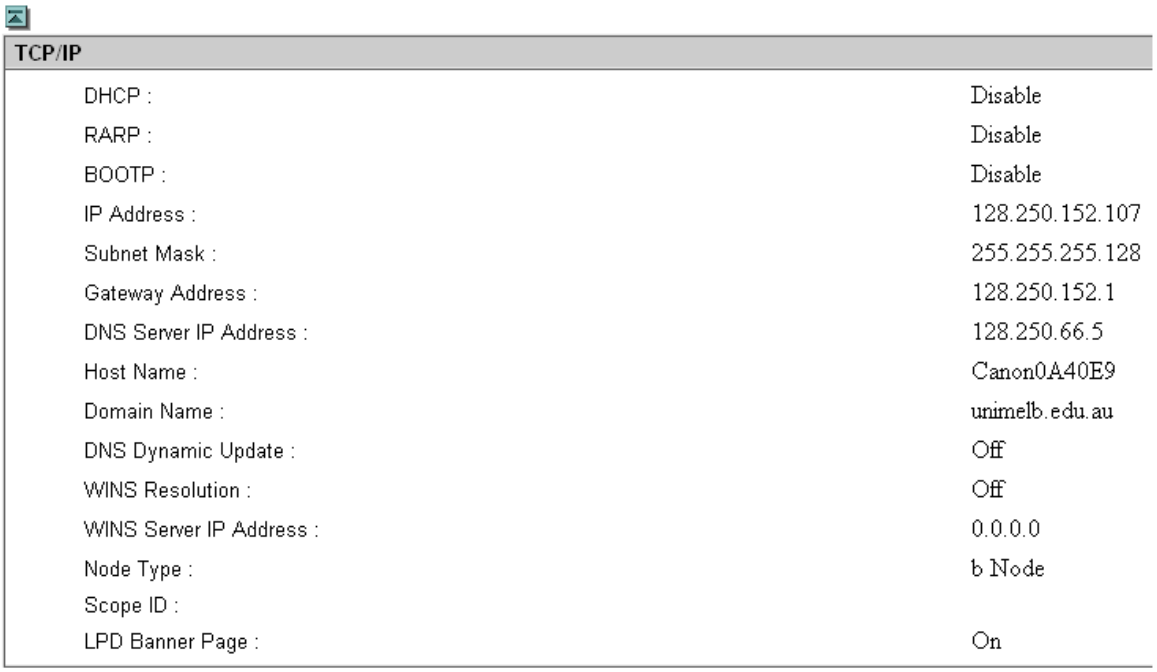
Samotný kód aplikácie *NQT* je napísaný v PHP a býva uložený v súbore `nqt.php`, pričom defaultná *NQT* stránka nesie titulok *Network Query Tool* [7]. Pomocou týchto faktov Google môže pomôcť vystopovať mnohé *NQT*, a to pomocou dotazu:

```
inurl:nqt.php intitle:"Network Query Tool"
```

3.4 Sieťový hardware

Dominantným prostredím na prenos informácií a dát sú počítačové siete a ich využitie značne umocňuje existencia sieťového hardwaru. Zariadenia tohto typu sú najviac využívané vo firemných intranetoch a obvykle disponujú vlastnou webovou stránkou [7]. Ak však na túto stránku existuje odkaz, ktorý sa Googlebotu podarí preliezť, naskytuje sa tak hackerom ďalšia príležitosť dolovania dôležitých údajov.

Aj keď to nie je na prvý pohľad očividné, sieťové zariadenia dokážu prezradiť kvantum užitočných informácií o sieti, v ktorej sa prevádzkujú, a útočníkovi tým značne uľahčia úlohu mapovania celej siete, čo demonštruje obrázok 3.3. Zameriavajú sa teda na ne hlavne hackeri podnikajúci výzvedné akcie proti sieťam [7].



Obrázok 3.3: Stránka tlačiarne Canon ImageReady

Ďalším zaujímavým typom hardwaru, ku ktorému sa možno pomocou Googlu dopátrať, sú webové kamery. Tie síce priame informácie o sieti či iné dáta neposkytujú, no zvyčajne bývajú chápané ako zdroj zábavy pre návštevníka webu nehľadiac na to, že stelesňujú

vynikajúci prostriedok na obhliadku terénu [7]. Útočníkovi tým značne uľahčujú prácu, pokiaľ je jej náplňou fyzická prítomnosť napríklad v budove spoločnosti.

Tabuľka 3.3 je súhrnom dotazov, výsledkom ktorých sú stránky sieťových zariadení.

AXIS Video Live Camera	`intitle:"Live View/ - AXIS"`
Sieťová kamera Canon	`intitle:liveapplet inurl:LvAppl`
webcamXP server	`intitle:"my webcamXP server!" inurl:":8080"`
Sieťové USB disky	`intitle:"Network Storage Link for USB 2.0 Disks" Firmware`
Sieťové VoIP zariadenia	`intitle:"Sipura.SPA.Configuration" -.pdf`
Router Skystream	`intitle:"Skystream Networks Edge Media Router" -securitytracker.com`
Tlačiarne Brother HL	`inurl:"printer/main.html" intext:"settings"`
Canon ImageReady	`intitle:"remote ui:top page"`
Tlačiarne HP	`inurl:hp/device/this.LCDispatcher`

Tabuľka 3.2: Dotazy pátrajúce po sieťovom HW [2][8]

3.5 Exploity a ich využitie

Pri prenikaní do systémov a ich nabúravaní využívajú hackeri celú sieť rôznorodých nástrojov. V arzenáli tých skúsenejších však isto nebudú chýbať časti kódu s názvom exploity, ktoré využívajú chyby v špecifickom softwari, a tak napomáhajú zaútočiť na konkrétny cieľ. Exploitami a škodlivým kódom ako takým sa venuje množstvo stránok a diskusných skupín a vďaka Googlu ich môžeme mať na dosah ruky.

Primárne delenie separuje exploity do dvoch základných skupín - lokálne (local), u ktorých je nutné, aby sa spustili na napadnutom systéme, a vzdialené (remote), ktoré predstavujú druh exploitov spúšťajúcich sa priamo na stroji útočníka, pričom ich úlohou je odoslanie dát do systému obete [7]. Triviálne vyhľadávacie frázy typu *local exploit* či *remote exploit* nás však dovedú k stránkam, ktoré zahŕňajú tematiku exploitov iba z teoretického hľadiska, no archívy kódov obsahujú len zriedka.

Na nájdenie konkrétneho kódu exploitu je potrebné dotazy konkretizovať a tým výsledky hľadania zúžiť. Prevažná väčšina z nich je napísaná v programovacom jazyku *C*, takže stanovenému účelu by poslúžil nasledujúci dotaz:

```
exploit filetype:c
```

Jeho výsledky už budú odkazovať na stránky s požadovaným obsahom.

K pátraniu po kóde exploitov však vedie aj iná cesta založená na znalosti kľúčových slov jazyka *C*, prípadne úryvkov bežných reťazcov zdrojového kódu. V mnohých programoch sa totiž pomocou príkazu *include* pridáva hlavičkový súbor so štandardnou knižnicou pre vstupno-výstupné operácie v tvare `#include<stdio.h>`. Príslušný dotaz Googlu by teda mohol vyzerať takto:

```
exploit "#include<stdio.h>"
```

Jeho výhodou je, že nám poskytne výsledky obsahujúce požadované frázy, pričom sa neobmedzuje len na dokumenty s príponou .c, ale ponúkne nám napríklad aj HTML stránky s úryvkami zdrojového kódu [7].

V prípade, že máme potrebne účinné exploity k dispozícii, zostáva posledným dielom skladačky nájdenie konkrétneho cieľa, prípadne skupiny zraniteľných cieľov pre útok. Jedným z osvedčených ciest je pátranie po cieľoch cez demonštračné stránky, ktoré bývajú generované príslušnou aplikáciou automaticky a obsahujú špecifický text. Tento druh útoku sa snaží využiť bezpečnostné problémy určitej verzie aplikácie a demonštratívne stránky často tento druh informácie poskytujú. S podobnou technikou sme sa stretli pri hľadaní adresárových výpisov [pozri podkapitolu 3.2]. Konkrétnych príkladov hľadania zraniteľných aplikácií existuje obrovské množstvo, jedným z ich solídnych úložíšť je napríklad [2], v tabuľke 3.2 sú uvedené tie najzákladnejšie.

A-CART je zraniteľný voči skriptovaniu cez web	`"Powered by A-CART"`
Na Sugar Suite 3.5.2 a 4.0 je možné útočiť vzdialeným spustením kódu	`"2005 SugarCRM Inc. All Rights Reserved" "Powered By SugarCRM"`
vBulletin 3.0.6 umožňuje vykonávať akýkoľvek kód	`Powered.by:.vBulletin.Version ... 3.0.6`
CubeCart 2.0 je zraniteľný voči SQL injektáži	`"powered by CubeCart 2.0"`
PHP-Fusion verzie 6.00.109 je zraniteľný voči SQL injektáži	`Powered by PHP-Fusion v6.00.109 ©2003-2005. -php-fusion.co.uk`

Tabuľka 3.3: Dotazy zameriavajúce sa na zraniteľné aplikácie [2]

3.6 Užívateľské mená a heslá

Elementárnymi prvkami autentizácie sú užívateľské meno a zodpovedajúce heslo. K prvému uvedenému sa dá pomocou Googlu samozrejme dopátrať omnoho ľahšie, no i tak býva v mnohých prípadoch znalosť užívateľského mena značne podceňovaná. Samotné užívateľské meno síce útočníka pristupovať do systému neoprávňuje, keďže tvorí menej dôležitú polovicu procesu autentizácie, no aj napriek tomu dokáže jeho rozumné využitie priniesť vytúžené ovocie.

Jedným zo spôsobov vypátrania užívateľských mien je vyhľadanie stránok prihlasovacích portálov, ktoré tvoria vstupnú bránu do autentizovaných webov a to pomocou dotazu:

```
login | logon
```

Ich súčasťou okrem očakávaných položiek na zadávaniu údajov pre prihlásenie často býva aj odkaz vedúci na užívateľskú podporu, prípadne dokumentáciu typu *pomôž si sám* [7].

Tieto dokumenty slúžia ako pomoc pri strate či zabudnutí užívateľského mena alebo hesla a môžu útočníkovi poskytnúť záchytné body pri prenikaní do systému.

Ciest, ktoré vedú k dokumentom s užívateľskými menami, je mnoho, často ich môžeme nájsť aj na bežných stránkach. Ďalšou alternatívou ich získania je bádanie vo výsledkoch dotazu ako je napríklad:

```
username | userid | employee.ID | "your username is"
```

Rovnako využiteľným úložiskom cenných dát sú aj firemné intranetové stránky. Pri ich prehľaavaní môžeme naraziť na tzv. technickú podporu alebo *help desk*. Dokumenty z neho získané dokážu byť neobyčajne cenné pri procese zhromažďovania informácií, nakoľko obsahujú kvantum interných údajov a útočník sa z nich môže dozvedieť aj niečo o infraštruktúre firmy [7]. Výraz *help desk* spolu s operátorom `site` umožňujú zúžiť okruh vyhľadávania a zamerať sa na konkrétny cieľ.

Niečo málo o štruktúre firmy sa môžeme dočítať priamo z oficiálnych stránok, tie nám ale poskytnú jedine informácie, o ktorých firma chce, aby sme vedeli. Mnoho spoločností ich však svojim zamestnancom sprístupňuje práve prostredníctvom intranetu. Výraz *intranet* v kombinácii s názvom konkrétneho oddelenia firmy môže vo výsledkoch odhaliť množstvo interných záležitostí týkajúcich sa daného subjektu. Hľadania ako

```
intitle:intranet intext:"human resources"
```

```
intitle:intranet intext:"IT department"
```

útočníkovi okrem užitočných faktov poskytnú mená a kontakty firemných zamestnancov a práve tie sa dajú zužitkovať v neskorších fázach útoku [7].

Pracovníci technickej podpory sú pritom ideálnym terčom pre útok sociotechnikou, kde sa dá znalosť užívateľského mena využiť veľmi efektívne. I v dnešnej dobe totiž platí, že najzraniteľnejším prvkom celého bezpečnostného mechanizmu naďalej zostáva ľudský faktor. Sociotechnika spočíva v ovplyvňovaní ľudí s cieľom oklamať ich natoľko, aby uverili, že sociotechnik je osoba s totožnosťou, ktorú sa snaží predstierať [7]. Pre sociotechnika je preto najdôležitejšie, aby pred začatím útoku získal čo najviac dostupných informácií o pracovníkoch firmy a jej štruktúre. Znalosť užívateľského mena a ďalších konkrétnejších údajov tak útočníkovi ponúka možnosť vydávať sa za osobu, ktorá vo firme pracuje, a tým bezpečnosť prelomiť.

Prvoradým aspektom bezpečnosti je však užívateľské heslo, ktoré by malo spadať do ochrannej zóny s najväčšou prioritou, no aj napriek tomu je možné na pár z nich využitím Googlu naraziť. Heslá objavené týmto spôsobom budú mať s najväčšou pravdepodobnosťou šifrovanú podobu, útočníkovi ale nič nebráni v tom, aby ich pomocou špecializovaného softwaru dešifroval, zobrazil v podobe čistého textu a následne využil pri útoku. Najfrekventovanejšími typmi súborov so zašifrovanými heslami vypátrateľných Googlom sú podporné súbory Microsoft FrontPage, v ktorých sa heslá vyskytujú v zašifrovanej podobe, prípadne exportované registre Windows [7].

Tabuľka 3.4 je výčtom najefektívnejších dotazov pátrajúcich po užívateľských menách a heslách, množstvo ďalších zastrešuje [2].

Podporné súbory MS FrontPage	`"# -FrontPage-" ext:pwd inurl:(service | authors | administrators | users) "# -FrontPage-" inurl:service.pwd`
Stránky s admin prístupom k databázi	`allinurl:admin mdb`
Konfiguračný súbor PHP	`config.php`
Stránky produktov DUware s heslami	`"Powered by Duclassified" -site:duware.com`
Súbory `xls` s užívateľskými menami a heslami	`"login: *" "password: *" filetype:xls`
Prihlasovacie stránky so vzdialeným prístupom	`intitle:"Remote Desktop Web Connection" inurl:tsweb`
Stránky phpMyAdmina	`"phpMyAdmin" "running on" inurl:"main.php"`
Registre MS Windows	`filetype:reg reg HKEY_CURRENT_USER username`

Tabuľka 3.4: Ukážky dotazov pátrajúcich po heslách [2]

Už podľa očakávania možno dedukovať, že hesiel nachádzajúcich sa v archíve Googlu nebude mnoho. No ak sa nám predsa len podarí na nejaké z nich natrafiť, tak už len jeho samotný výskyt prezrádza veľa o úrovni zabezpečenia celého servera. Môžeme si byť teda istí, že sme práve objavili horúceho kandidáta na útok.

3.7 Vyhľadávanie zaujímavých dát

Google toho vo svojom vnútrajšku schováva veľa a jeho využitie siaha ďalej ako len na mapovanie siete či zisťovanie slabín možného terča útoku. Tvorbou správne upravených dotazov môžeme dolovať z indexu Googlu veľmi zaujímavé informácie.

Rozmach P2P sietí nastolil novú éru prenosu dát, vo väčšine prípadov tých nelegálnych. Či už ide o hudbu, prípadne filmy, nelegálnych kópií neustále pribúda. Prečo však miesto P2P sietí nevyužiť obrovskú databázu Googlu, ktorá nám napovie, kde takéto súbory hľadať. Vezmime do úvahu dotazy:

```
"parent directory" DVDRip -xxx -html -htm -php -opendivx -md5 -md5sums
```

```
"parent directory" MP3 -xxx -html -htm -php -opendivx -md5 -md5sums
```

Zmenou názvu adresára docielime, že vo výsledkoch, ktoré nám Google poskytne, bude množstvo užitočných dát na stiahnutie [11].

Rovnako využiteľný môže byť dotaz

```
inurl:microsoft filetype:iso
```

ktorý vypátra tzv. *ISO obrazy* obyčajne obsahujúce inštalačné súbory k softwaru. Argument operátoru `inurl` je samozrejme variabilný [11].

V drvivej väčšine prípadov je k právoplatnému používaniu SW nutné disponovať licenčným kľúčom. Ani to však pre nás vďaka Googlu nepredstavuje prekážku, v jeho archíve sa totiž objavujú i stránky s platnými licenčnými číslami. V prípade operačného systému Windows XP bude istotne užitočný dotaz:

```
"Windows XP Professional" 94FBR
```

Reťazec *94FBR* značne zredukuje množstvo stránok, ktoré platné číslo obsahujú, v licenčných kľúčoch sa totiž vyskytuje veľmi frekventovane. Jeho ďalšími variantmi môžu byť *GC6J3*, *GTQ62*, *FP876*, *D3DX8* [2]. Tento typ vyhľadávania je obmedzený jedine fantáziou užívateľa a pomocou Googlu sa dá nepochybne dopátrať ku kvantu sériových čísiel rôznych produktov.

Rovnako využiteľnou skupinou údajov môžu byť osobné dáta, ktoré nájdu uplatnenie najmä u sociotechnikov. Medzi ich ideálne zdroje patria životopisy, ktoré prezrádzajú adresy, telefónne čisla a ďalšie údaje týkajúce sa konkrétnych osôb. Na ich vyhľadanie postačí zadať Googlu dotaz [2]:

```
"phone * * *" "address *" "e-mail" intitle:"curriculum vitae"
```

3.8 Porovnanie možností Google a Yahoo

Rýchly rozmach Googlu spôsobil jeho vzostup v rebríčku vyhľadávačov až na najvyššiu priečku, pričom jeho pokrok smeruje dopredu míľovými krokmi. Google dokázal svojou stratégiou prenikania na trh prestihnúť aj svojho nedávneho najväčšieho rivala, vyhľadávač Yahoo. Google hacking ako taký ťaží hlavne z jeho efektivity a rozsiahlej kolekcie informácií, ktoré Google zastrešuje. Obsahom tejto kapitoly však bude zhrnutie možností použitia techník hackingu Googlom vo vyhľadávači Yahoo.

Rovnako ako všetky ostatné vyhľadávače, ani Yahoo by sa pri zužovaní výsledkov jednotlivých dotazov nezaobišiel bez booleovských a pokročilých operátorov. Väčšinu z tých, ktoré sú obsiahnuté v repertoári Googlu, Yahoo rozpozná, prípadne má za ne ekvivalentnú náhradu.

Google	Yahoo
`intitle`	`intitle, title`
`inurl`	`inurl`
`site`	`site, domain, hostname`
`filetype`	`originurlextension`
`link`	`link`

Tabuľka 3.5: Pokročilé operátory Google a Yahoo [12]

Pri používaní booleovských operátorov u oboch vyhľadávačov by sme hľadali rozdiely len ťažko, ich syntax a aplikácia je totiž identická. Jediným faktom, ktorým sa Yahoo od Googlu líši, je, že každý booleovský operátor má práve jeden reprezentant, a to znamienko

plus (*+*), znamienko mínus (-) a *OR*, kdežto u Googlu je možné využívať aj ich synonymá *AND*, *NOT* a zvislú čiaru (|) [pozri podkapitolu 2.3.2] [12].

V prípade pokročilých operátorov je už situácia odlišná, niektoré z operátorov Googlu vo vyhľadávači Yahoo nemožno používať. Konkrétne ide o `inanchor`, `numrange` a `intext`. Frekvencia ich využívania nie je síce častá, no aj napriek tomu nám poskytujú možnosti, ktorými sa dokážeme dostať ku konkrétnejším údajom. Tabuľka 3.5 zobrazuje zoznam pokročilých operátorov, ktorými disponujú oba vyhľadávače zároveň.

Kapitola 4

Ochrana pred hackermi využívajúcimi Google

Hacking a jeho praktiky tvoria spletitú pavučinu možností, proti ktorým sa väčšina z nás snaží brániť. Je preto naliehavé oboznámiť sa so základmi tvorby ochranných opatrení či bezpečnostnej politiky a tým získať bližší pohľad aj na druhú stranu mince. Predmetom hackingu Googlom nie je iba využitie možností tohto vyhľadávača na získavanie rôznorodých informácií o možnej obeti útoku, ale i vniknutie do problému ochrany svojich vlastných súborov a údajov spočívajúce v pochopení trikov, ktoré je schopné útočník zužitkovaním svojich vedomostí pomocou Googlu realizovať.

4.1 Základné pravidlá bezpečnosti

Element ľudskej nedbalosti je činiteľ, z ktorého sa snaží ťažiť každý hacker využívajúci Google a s ním nastupuje na scénu termín *Googledork*, ktorý vytvoril Johnny Long. Pôvodne označoval hlúpu osobu odhalenú Googlom, no po prevzatí tohto výrazu médiami nazýva tých, ktorí na internete zverejnili dôverné informácie [3]. Údaje, ktoré by mali podliehať ochrane sa totiž na webe objavujú nevedomky, ale aj z nepozornosti, pričom ich na dosah ruky vystavujú samotní užívatelia.

Zo všetkých faktov však treba dbať na ten najpodstatnejší, a to že webový server je primárne určený na ukladaniu dát pre širokú verejnosť. Presunutím citlivých údajov z verejne dostupného servera na iné dôveryhodnejšie úložisko je zásadným krokom v prevencii proti úniku informácií.

4.2 Blokovanie výpisu adresárov

Obsahom adresárových výpisov bývajú okrem súborov často aj iné položky, z ktorých je útočník schopný ťažiť [pozri podkapitolu 3.2]. Tento výpis sa najčastejšie zobrazuje v prípade, že webový server nebol schopný nájsť v adresárovej štruktúre tzv. indexový súbor, obvykle pomenovaný ako `index.html`, `index.php`, `default.asp` atď., definovaný v konfigurácii serveru a reprezentujúci úvodnú stránku. Pokiaľ nás situácia nenúti povoliť návštevníkom webu prehliadanie adresárov v štýle FTP, mali by byť výpisy adresárov v každom prípade vypnuté [7].

4.3 Obmedzovanie webových robotov

Internetové vyhľadávače usporadúvajú gigantické množstvá dát a popri procese ich indexácie sa pre ne roboti stávajú ideálnymi partnermi. Webový robot podrobne skúma prístupné servery a orientuje sa podľa nájdených hypertextových odkazov. Po indexácii zostavených údajov sa vyhľadávače pokúšajú odhadnúť, kde by sa mohla žiadaná informácia vyskytovať, a tým sa celý priebeh prehľadávania webu automatizuje [7].

Google na tento účel využíva robota s názvom *Googlebot*. Akonáhle *Googlebot* prelezie webový server, odošle príslušné údaje do databázy Googlu, ktorá sa postará o ich ďalšie spracovanie [7]. Ak sa mu to však podarí v okamihu, keď sa na ňom vyskytujú dôverné dáta, dostávajú hackeri využívajúci Google jedinečnú príležitosť na ich získanie.

4.3.1 robots.txt

Pri preliezaní webu sa roboti najskôr obzerajú po súbore `robots.txt`, ktorého obsah a štruktúra podlieha určitým štandardom a jeho súčasťou sú inštrukcie pre jednotlivých robotov. Pomocou nastavení v tomto súbore máme možnosť explicitne určovať, ktoré súčasti webu majú roboti obchádzať [10].

Súbor `robots.txt` sa musí nachádzať priamo v koreňovom adresári webového serveru. Repertoár inštrukcií používaných v tomto súbore pozostáva iba z dvoch príkazov - *User-agent* a *Disallow*. Argument príkazu *User-agent* určuje, ktorému robotu inštrukcie adresujeme, prípadne môžeme použiť zástupný znak * a tým smerovať pokyny na akéhokoľvek existujúceho robota. Príkazom *Disallow* definujeme časti webu, ktoré robot preliezať nemá [10]. Príklad použitia `robots.txt` demonštruje nasledujúci výpis:

```
User-Agent: *
Disallow: /
```

Takto zostavený súbor zabráni, aby roboti indexovali akúkoľvek časť webu (použitie samotného lomítka vyjadruje celú adresárovú štruktúru) [7].

Obsahom položky *Disallow* však môže byť konkrétna cesta k adresáru, prípadne definícia prípon súborov, ktorým sa majú roboti vyhýbať. Pre zakázanie prístupu robota Googlu k súborom v adresári /tmp a rovnako aj ku všetkým PDF súborom stačí implementovať tieto príkazy [10]:

```
User-Agent: Googlebot
Disallow: /tmp
Disallow: /*.pdf$
```

Zástupný znak * v tomto prípade, ako aj u *Disallow* všeobecne, predstavuje ľubovoľnú postupnosť znakov, pričom znak *$* indikuje koniec názvu. V prípade, že stihol Googlebot v minulosti preliezť zložku /tmp spolu so súbormi PDF, prejaví sa odstránenie požadovaného obsahu z indexu Googlu po najbližšom prelezení stránky Googlebotom [7].

Rozsah obmedzení uvedených v súbore `robots.txt` väčšina webových robotov rešpektuje. Niektoré vyhľadávače so zlou povesťou, prípadne samotní hackeri, však jeho obsah rešpektovať nemusia. Jedným z účinných trikov je prvotné vyhľadanie súboru `robots.`

`txt`na základe ktorého získa útočník predstavu o súčastiach webu, ktoré sa administrátor snažil pred očami vyhľadávačov ukryť. Na jeho nájdenie je možné využiť Google, a to pomocou dotazu [7]:

```
inurl:robots.txt filetype:txt
```

Uplatnenie súboru `robots.txt` je teda potrebné zvážiť, v opačnom prípade môže jeho využitie viesť k nemilým následkom.

4.3.2 META tagy

Na niektorých webových serveroch majú právo vytvárať súbor *robots.txt* výlučne administrátori. Za takejto situácie je jedinou možnou cestou obmedzovania robotov využitie nasledujúceho META tagu v hlavičke HTML dokumentu:

```
<meta name="robots" content="noindex, nofollow">
```

Hodnota *noindex* v položke *content* zabezpečí, že obsah stránky nebude indexovaný a *nofollow* dáva robotovi najavo, že pri preliezaní dokumentu nemá sledovať odkazy vyskytujúce sa na stránke [10].

4.4 Google index

Priemernú webovú stránku Googlebot obyčajne prelieza raz za pár týždňov. Zmeny implementované v súbore `robots.txt`, prípadne priamo v dokumente HTML pomocou META tagov sa teda aj napriek snahe neprejavia okamžite. Google však pre túto situáciu ponúka riešenie v podobe *automatic removera* na <http://services.google.com:8882/urlconsole/controller>. Jeho využitím je možné požiadať Google, aby preliezol updatovaný súbor *robots.txt* a takisto umožňuje zadať URL stránky, ktorá obsahuje patričné META tagy [1].

Kapitola 5

Automatizované vyhľadávanie Googlom

V predošlých kapitolách sme sa oboznámili s princípom dolovania citlivých údajov z databázy Googlu a rovnako boli predstavené techniky tvorby vyhľadávacích dotazov, ktoré na tento účel slúžia. Ich najznámejším úložiskom je *Google Hacking Database* (*GHDB*) [pozri podkapitolu 5.2]. Zakladateľom databázy je Johnny Long, uznávaný špecialista v oblasti hackingu a IT security [2].

Situácia, ktorá však vyžaduje, aby sme sa využitím prvkov dostupných v *GHDB* snažili získať čo najviac užitočných informácií z konkrétnej domény, nastoľuje otázku automatizácie celého procesu zadávania dotazov, ktorá by podobné úlohy dokázala značne zefektívniť. Práve na tento účel ponúka Google adekvátnu alternatívu svojho webového rozhrania v podobe *Google API*, ktoré si v tejto kapitole predstavíme.

5.1 Google API

Google API bolo vytvorené pre vývojárov, ktorí chcú použiť Google ako prostriedok vo svojich aplikáciách a potrebujú využívať rozhranie pre online prístup k jeho databáze. Ide o plnohodnotné API, s vlastnou sadou volaní API, pričom software sa k nemu pripája vzdialene [7]. Každý, kto má o služby *Google API* záujem, si však musí u Googlu vytvoriť vývojársky účet, ku ktorému mu bude poskytnutý licenčný kľúč sprevádzajúci všetky dotazy smerované samotnému API. Google každé použitie kľúča sleduje a touto formou využívanie *Google API* značne obmedzuje - s jedným licenčným kľúčom je totiž možné v priebehu 24 hodín odoslať maximálne 1000 dotazov [7]. Rovnako existujú v rámci *Google API* ďalšie obmedzenia zobrazené v tabuľke 5.1.

Max. dĺžka vyhľadávacieho reťazca	2048 B
Max. počet slov v dotaze	10
Max. počet operátorov `site` v dotaze	1
Max. počet vrátených výsledkov na dotaz	10
Max. index vráteného výsledku	1000

Tabuľka 5.1: Obmedzenia dotazov adresovaných *Google API* [4]

Google u svojho API dodržuje štandardy *Simple Object Access Protocol* (*SOAP*) a *Web Services Description Language* (*WSDL*), jeho služby teda nie sú závislé od použitej platfor-

my a vývojári majú pri výbere programovacieho jazyka voľnú ruku [7].

5.2 Google Hacking Database

Techniky tvorby dotazov pre vyhľadávač Google, ktorých výsledkom sú najmä citlivé a dôverné dáta sú rôznorodé a *GHDB* dotazy práve tohto typu združuje do jedného celku. *Google Hacking Database* predstavuje v súčasnej dobe najrozsiahlejší depozitár hackerských techník realizovateľných pomocou Googlu, pričom sa na jej pravidelnom rozširovaní podieľa množstvo nadšencov a v súčasnej dobe zahŕňa takmer 1500 dotazov, ktoré ďalej separuje do viacerých kategórií [2].

- **Informačné správy a zraniteľnosti**

  Dotazy pátrajúce po zraniteľných serveroch prípadne aplikáciách, a to pomocou automaticky generovaných informačných správ.

- **Chybové hlášky**

  Hľadania stránok s chybovými hláškami, ktoré dokážu prezradiť množstvo informácií.

- **Súbory obsahujúce zaujímavé informácie**

  Výsledkom takýchto dotazov síce nie sú heslá ani ďalšie tajné dáta, no aj napriek tomu sa vďaka nim dostaneme k zaujímavým údajom.

- **Súbory obsahujúce heslá**

  Reťazce, ktorých výsledkom sú súbory s heslami.

- **Súbory obsahujúce užívateľské mená**

  Táto kategória zoskupuje rôzne spôsoby získavania užívateľských mien.

- **Footholds**

  Príklady dotazov, ktoré hackerovi pomáhajú získavať oporné body (tzv. *footholds*) na podniknutie útoku.

- **Prihlasovacie portály**

  Dotazy lokalizujúce prihlasovacie portály k autentizovaným webom.

- **Informácie o sieti**

  Patria sem zväčša stránky s log súbormi firewallu a ďalšími užitočnými dátami o sieti.

- **Adresáre s citlivými dátami**

  Zbierka dotazov, ktoré z indexu Googlu dolujú adresáre obsahujúce citlivé a tajné informácie.

- **Informácie o online nákupoch**

  Príklady dotazov, ktoré odhaľujú podrobnosti o nákupoch realizovaných prostredníctvom internetu, ako sú informácie o zákazníkoch či objednávkach, čísla kreditných kariet a iné.

- **Sieťový hardware**

  Táto kategória obsahuje stránky sieťových tlačiarní, webových kamier a ďalších typov zariadení odhalených Googlom.

- **Zraniteľné súbory**

  Dotazy, ktorých výsedkom je množstvo zraniteľných súborov odhalených Googlom.

- **Zraniteľné servery**

  Hľadania tohto typu odhaľujú servery s bezpečnostnými chybami, využívajúc techniku odlišnú od hľadania zraniteľných súborov.

- **Detekcia webových serverov**

  Reťazce tejto kategórie demonštrujú vytváranie profilu webových serverov využitím Googlu.

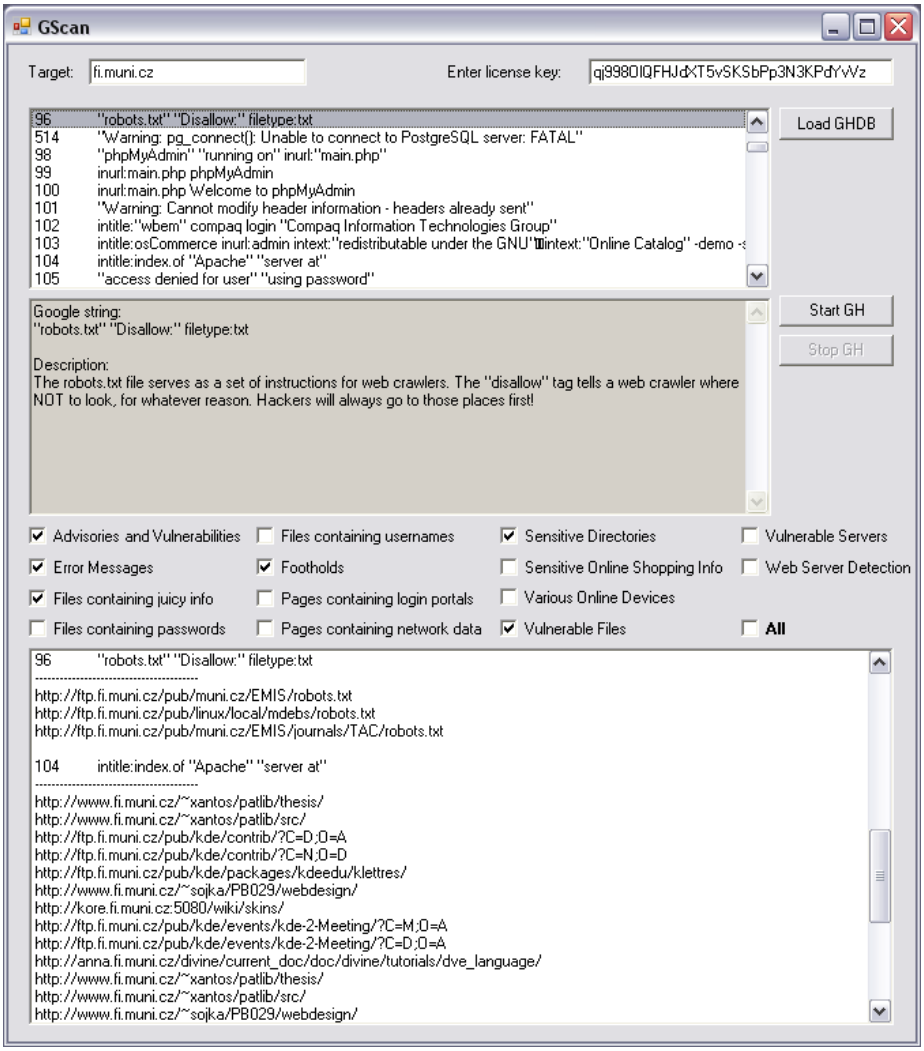
5.3 GScan

GScan predstavuje softwarový nástroj, ktorého primárnou funkciou je automatizácia procesu zadávania dotazov pre vyhľadávač Google využitím *Google API*. Svojimi schopnosťami sa ponáša na aplikáciu Wikto, pomocou ktorej je možné importovať všetky dotazy z *GHDB* a následne skenovať vybranú doménu dostupnými reťazcami nachádzajúcimi sa v databáze. GScan však prináša výhody prameniacie z kategorizácie jednotlivých položiek *GHDB*. Jeho najväčším prínosom je nepochybne filter dotazov obsiahnutých v *GHDB*, vďaka ktorému má užívateľ možnosť začleniť do okruhu reťazcov použitých počas procesu skenovania jedine tie, ktoré sa vyskytujú vo zvolenej kategórii, a tým čiastočne obísť obmedzenie 1000 dotazov na deň súvisiace s pravidlami používania *Google API*.

Zdrojový kód aplikácie GScan je napísaný v jazyku *C#*, ktorý je v súčasnej dobe najpoužívanejším programovacím jazykom pre spoluprácu s *Google API* a na jej spustenie je preto potrebné disponovať nainštalovaným rozhraním *.NET Framework* verzie 2.0 [4]. Samotná aplikácia je koncipovaná ako `exe` súbor spolu s pribalenou databázou vyhľadávacích reťazcov uloženou v súbore s názvom `ghdb.xml`. GScan teda možno rozbehnúť bez nutnosti inštalácie. Všetky jej súčasti spolu so zdrojovým kódom sa nachádzajú na priloženom CD.

Ihneď po spustení sa objaví hlavné okno, ktorého jednotlivé ovládacie prvky zobrazuje obrázok 5.1. Pred samotným inicializovaním skenovania je nutné vyplniť obidve položky užívateľského vstupu v hornej časti okna. Argumentom kolónky **Target** je doména, ktorú má užívateľ v pláne skenovať. Vedľa nej sa nachádza textové pole, do ktorého musí byť

zadaný platný licenčný kľúč potrebný pre využívanie *Google API*. V prípade, že sa tak nestane, obdrží užívateľ chybovú hlášku s výzvou na zadanie funkčného kľúča.



Obrázok 5.1: Hlavné okno aplikácie GScan

Tlačidlo **Load GHDB** slúži na import databázy Google dotazov z priloženého súboru `ghdb.xml`, ktorý je nevyhnutné vyhľadať v adresárovej štruktúre. Akonáhle sa proces importovania *GHDB* skončí, sú všetky dotazy i s podrobným popisom zobrazené v stavových oknách. *GHDB* sa postupom času pochopiteľne rozrastá, a preto je potrebné položky v súbore `ghdb.xml` updatovať. Jeho aktuálnu verziu je možné nájsť na adrese <`http://johnny.ihackstuff.com/xml/ghdb.xml`>

Pod stavovými oknami sa nachádzajú zaškrtávacie polia s názvami jednotlivých kategórií *GHDB*, pomocou ktorých má užívateľ možnosť začleniť do okruhu dotazov iba reťazce

zvolených kategórií. V dobe vzniku tejto práce delila *GHDB* dotazy do 14 kategórií, pričom každá z nich korešponduje práve s jednou položkou filtra. Zaškrtnutie položky *All* spôsobí označenie všetkých kategórií.

Ak sú všetky potrebné údaje vyplnené správne, môže užívateľ začať so skenovaním kliknutím na tlačidlo **Start GH**. GScan začne aplikačnému rozhraniu posielať dotazy spadajúce pod vybrané kategórie a ich výsledky zobrazovať v spodnom stavovom okne. Pomocou tlačidla **Stop GH** je možné celý proces skenovania ukončiť.

Kapitola 6

Záver

Google hacking patrí v oblasti bezpečnosti medzi frekventovane rozoberané tematiky, a to i vďaka tomu, že na poli hackingu predstavuje jednu z najčerstvejších noviniek. Charakterizuje ho najmä jeho jednoduchosť a nenáročnosť - hackerom využívajúcim Google sa totiž môže stať aj bežný internetový užívateľ, keďže úroveň vstupných znalostí je minimálna a obmedzuje sa na pravidlá a spôsoby použitia booleovských a pokročilých operátorov Googlu spolu s určitým stupňom počítačovej gramotnosti. Vďaka dostupnosti jednotlivých vyhľadávacích reťazcov Google Hacking Database sa táto jednoduchosť ešte umocňuje.

Rozrastanie webu a jeho samotného obsahu však so sebou prináša aj dopad na databázu Googlu spôsobujúci jej rapídny nárast. Je pochopiteľne zrejmé, že tento trend si svoje tempo aj naďalej udrží, zraniteľných cieľov a možných spôsobov hľadania citlivých dát bude teda neustále pribúdať. Množstvo dotazov pátrajúcich po nich je prakticky neobmedzené a závisí jedine od fantázie a tvorivosti hackerov. Google hacking má preto skvelé vyhliadky stať sa jednou z najpoužívanejších techník hľadania citlivých dát.

Aplikácia GScan ako súčasť bakalárskej práce svojou schopnosťou kategorizácie položiek *GHDB* oproti ostatným aplikáciám podobného typu zefektívňuje a rozširuje možnosti využitia samotnej databázy. Predstavuje tak ideálny prostriedok na skenovanie vlastného webu, ktorým má užívateľ možnosť predísť potenciálnym rizikám vystaveniu dôverných dát na internete.

Stavebným kameňom hackingu pomocou Googlu je ľudská nedbanlivosť, ktorá spôsobuje výskyt dôverných údajov priamo na internete. Človek, ako najzraniteľnejší bod bezpečnostnej stratégie, je v úvodných fázach útoku najvhodnejším cieľom. Námetom ďalšej práce by teda mohol byť podrobnejší rozbor rozvíjajúcich sa techník hackingu Googlom koncipovaný smerom k získavaniu osobných dát a ich následné zneužitie sociotechnikou.

Literatúra

[1] Calishain, T. a Dornfest, R.: *Google Hacks*, O'Reilly, January 2005, 05-96008-57-0. 4.4

[2] Long, J.: *Google Hacking Database*, <http://johnny.ihackstuff.com/index.php?module=prodreviews> . 3.2, 3.5, 3.3, 3.6, 3.4, 3.7, 5, 5.2

[3] Long, J.: *The Google Hacker's Guide*, 2004, <http://johnny.ihackstuff.com/security/premium/The_Google_Hackers_Guide_v1.0.pdf> . 2.3.1, 2.3.2, 2.3.3, 2.1, 2.3.3, 3.1, 3.2, 4.1

[4] google.com: *Google SOAP Search API*, 2006, <http://code.google.com/apis/soapsearch/index.html> . 5.1, 5.3

[5] google.com: *Google corporate information*, January 2006, <http://www.google.com/corporate/history.html> . 2.1, 2.1

[6] labs.google.com: *Google Labs*, 2006, <http://labs.google.com/> . 2.2

[7] Long, J.: *Google hacking*, Zoner Press, 2005, 80-86815-31-5. 2.2, 2.3.1, 2.3.2, 2.3.3, 2.3.3, 3, 3.1, 3.2, 3.3, 3.4, 3.4, 3.5, 3.6, 4.2, 4.3, 4.3.1, 5.1, 5.1

[8] Macok, M. a Stradal, V.: *Google hacking*, DCIT, s.r.o., 2006, <http://www.dcit.cz/files/bezpecnost/IntSecShow_2005_google.pdf> . 3.2

[9] Piotrowski, M.: *Dangerous Google - Searching for Secrets*, Hakin9 Magazine, April 2005, <http://www.hakin9.org/en/archive/attachments/google_en.pdf> . 2.3.2, 3, 3.1

[10] www.robotstxt.org: *The Web Robots Pages*, <http://www.robotstxt.org/> . 4.3.1, 4.3.2

[11] I-Hacked.com: *Google Hacking*, September 2004, <http://www.i-hacked.com/content/view/23/42> . 3.7

[12] yahoo.com: *Yahoo! help*, 2006, <http://help.yahoo.com/help/us/ysearch/basics/basics-04.html> . 3.5, 3.8