"I know Google-Fu!" ehk hakkame asjadeotsijateks

Kaido Kikkas

(Algselt avaldatud ajakirjas "30 pluss" veebruaris 2016, siin ilmub kokkuleppel toimetusega)

"Kullakänkarid ja jaanalinnusuled, surnud rotid ja paukkompvekid, ja veel tillukesed, tillukesed mutrid ja muu sihuke kraam." - Pipi Pikksukk

Seekordne jutt räägib otseselt küll Internetist info otsimisest, ent siit ei puudu ka meie sarja läbiv turvalisuse aspekt. Vanasõna ütleb: "Kes otsib, see leiab" - kuid küsimus on, kas kõik leitu pidi üldse leitav olema? Nii mõnelgi alljärgnevate lihtsate näidete abil võrgust leitud failide omanikul ei ole halli aimugi, et nende "tillukesed mutrid" mõne hakkaja Pipi näppu on juhtunud.

Ja võiks veel tsiteerida üht tuntud ameerika ütlust "Ühe inimese rämps on teise inimese varandus" - ehk isegi äravisatud kraam (olgu siis tegu arvutiekraanil prügikasti lohistatud failide või füüsilisse paberikorvi visatud paberdokumentidega) võib osutuda mõnele vägagi huvipakkuvaks. Erandiks ei ole ka Internetis leiduv materjal.

Astrid Lindgreni kuulus raamatukangelane Pipi Pikksukk otsustas ühes loos hakata asjadeotsijaks. Alguses tunti suurt rõõmu pelgalt purkidestki, siis aga jäi ette ka üks maas põõnutanud onu - ning vaid Tommy ja Annika sekkumine päästis tolle üleskorjamisest ja küülikupuuri sattumisest. Samasuguseid agaraid asjadeotsijaid on aga täis ka Internet.

Google-Fu algtõed

Internetis asjadeotsimiseks on olemas kunst, mida võiks nimetada Google-Fu'ks. Selles kunstis ei anta küll värvilisi vöid, kuid omad tasemed on olemas siingi. Lihtsaimat taset valdame me ilmselt kõik - avame Google'i otsingu ja kirjutame lahtrisse soovitava sõna või fraasi. Ent Google võimaldab tegelikult palju enamat.

Esimene samm edasi on õppida välistama osasid vastusevariante. Näiteks on eesti keeles sõnal "tee" kolm tähendust ning kui otsime jooki, mitte rada, võime kirjutada otsingusse näiteks "tee -tänav -maantee". Nii saame vastustest sageli palju ülevaatlikuma pildi. Samamoodi võime määrata ära sõna, mis igal juhul peab otsitaval lehel esinema - näiteks "tee +roheline" jätab välja üksnes mustast teest rääkivad lehed.

Suurem osa Google-Fu'st aga toimub täpsustavate parameetrite abil (üldkujul "parameeter.väärtus"), mis kirjutatakse samamoodi otse päringulahtrisse.Veidi tõsisema asjadeotsimise jaoks lisame esmalt domeenitäpsustuse. See võib olla näiteks kasvõi kodumaine .ee - otsinguga "kummipuu site:.ee" näitab meile Eesti veebis asuvaid lehti, kus on juttu kummipuust. Sellega aga annab minna palju täpsemaks - näiteks "site:kontor.minufirma.ee/juhan".

Kui tavaline Google'i otsing keskendub veebilehtedele, siis asjadeotsija üheks heaks töövahendiks on "filetype"-otsing kindlat tüüpi failide seast (näiteks .pdf või .doc). Eriti aga võib huvitavaid tulemusi anda täpsustatud domeeni ja failitüübi kooskasutus - näiteks võib otsinguga "site:minukonkurent.ee filetype:docx" saada jälile nii mõnelegi huvitavale dokumendile, mille sealse ettevõtte hooletud asjapulgad on kogemata kuhugi veebikataloogi salvestanud. Otsida võib ka näiteks

- xls ja xlsx - Exceli tabelid; firma kontekstis tähendab see enamasti arvandmeid, sh raha liikumise kohta.
- ppt ja pptx - PowerPointi esitlused; võib näiteks saada teada, mida rääkis firma tegevjuht hiljutisel aktsionäride üldkoosolekul....
- pdf - ilmselt kõige universaalsem dokumendiformaat, võib sisaldada mida iganes.
- zip, rar, tar.gz.... - pakitud failid. Sageli pakitakse mingi hetkel mittevajalik suur infokogum küll kokku, ent pakk unustatakse sinnasamasse veebikataloogi.
- exe - Windowsi-maailma programmifailid. Vahel võib olla tegu ka ebaseadusliku tarkvaraga, millest teadasaaja võib kiusu pärast teatada "kuhu vaja" (nagu vanasti öeldi). Sama võib juhtuda muusika- või filmikoguga, ka mõnda sorti pildid ja videod võivad olla üsna plahvatusohtlikud.

Kui nüüd keegi ära ehmatas ja küsis: "Kas ikka nii tohib?", siis võib täie rahuga vastata, et kõik siinkirjeldatu on täiesti seaduse piires. **Ainus mõjus vastuabinõu on koristada veebist ligipääsetavatest kataloogidest (htdocs, public_html vmm) ära KÕIK failid, mida võõrad nägema ei peaks.** "Ega keegi ju seda üles ei leia" ei ole argument.

NB! Siin lühidalt kirjeldatud Google-Fu võtted olid mõeldud vaid tavakasutajatele põhiliste ohtude teadvustamiseks. "Kõrgemate vööde" programmis on aga küllaldaselt tehnikaid ka IT-spetside tegevusväljalt (Google-Fu abil leitakse tänini avaveebist näiteks logi- või konfiguratsioonifaile ja SQL-andmebaaside tõmmiseid, ent ka tuntud veebitarkvarade ebaturvalisi paigaldusi jpm) - nii et meistri käes ei ole see kunst vähem ohtlik kui mõni füüsilise maailma võitlusviis. Tõsisematele huvilistele võib soovitada näiteks Patrick Engelbretsoni raamatut "The Basics of Hacking and Penetration Testing" ja Johnny Longi "Google Hacking for Penetration Testers". Mainigem aga, et see valdkond areneb kiiresti ning mõned isegi paari aasta vanuses raamatus ära toodud võtted ei pruugi täna enam toimida, nende asemele aga tulevad uued.

Kui kusagilt ei leia, siis kusagilt ikka leiab

Tuletame meelde üht punkti Facebooki teemas, mis keelas seal toimetada mingi aine mõju all olles. See kehtib tegelikult laiemalt kogu võrgumaailma kohta ning üheks peapõhjuseks on just info säilimine.

Kodanik, kes on purjus peaga laamendades saatnud veebis pimedasse paika kõik alates peaministrist ja lõpetades naabri koeraga, võib küll kaineks saades enda rumalust mõista ja kogu tekitatud läbu kiiresti ära kustutada - kuid juba võib olla hilja, kuna Google'i otsirobot on joogise pea sünnitise üles leidnud ja puhvrisse talletanud. Näiteks kui tekst pandi ajaveebi http://minublogi.blogspot.com, tuleks varem seal olnut "guugeldada" umbes sellise käsuga: "on üks igavene cache:minublogi.blogspot.com". Tähelepanuks: selline lihtne puhvriotsing ei laiene lehel olevatele linkidele - neile klõpsates kuvatakse tulemus juba otse veebist, mitte puhvrist!

Teine, veelgi tõsisem infosäilitaja on Interneti Arhiiv (*Internet Archive*; http://www.archive.org). See on tegelikult täiesti soliidne mäluasutus, mis sarnaselt päris arhiivide ja raamatukogudega hoolitseb inimkonna teadmiste säilitamise eest. Sealt leiab mitmete huvipakkuvate veebilehtede info ka päris kaugest (arvutiajastu mõistes) minevikust - näiteks on seal tänini talletatud veebileht, mille lõi siinkirjutaja koos kolleegidega TTÜ-s aastal 1996 ja mis kadus veebist 2002. aastal.

Interneti Arhiiv on talletamisel küll mõnevõrra valivam kui Google'i puhver ning seetõttu iga joogise peaga tehtud röögatust tulevastele põlvedele ei jäta, kuid pikemaealised kahtlase väärtusega asjad (kasvõi igasugused "Ajage X ahju!" stiilis lehed) võivad sinna juba sattuda küll. Päris kindlasti on seal aga materjali, mida kasvõi erinevad poliitikud ja "arvamusliidrid" hiljem

häbenevad ja mis väga tõenäoliselt tulebki kunagi tulevikus veel neid kummitama. Asjadeotsijatele on see igal juhul tänuväärne tööpõld. Arhiivi kasutamiseks tuleb minna archive.org -lehele, sisestada sinna soovitav veebiaadress ning valida salvestiste seast sobiva ajahetke oma (võib aga muidugi ka juhtuda, et Arhiiv ei ole soovitud lehte piisavalt tähtsaks pidanud ning seda ei ole säilitatud).

Ja viimaks on tõsistele asjadeotsijatele olemas ka tumeveeb (*Dark Web*) ehk Interneti "veealune" osa - aga sellest tuleks rääkida kunagi hiljem omaette loos.

Lõpetuseks

Tänasest loost võiks jääda kõlama kaks mõtet. Esmalt on asjadeotsimine Internetis põnev ja teatud piirini täiesti seaduslik tegevus, mida võibki harrastada nagu raadioamatörismi või metalliotsijaga põllul jalutamist. Teisalt aga tuleb arvestada, et asjadeotsijaid on palju ning kaugeltki mitte kõik neist ei ole heatahtlikud. Seetõttu tuleks enda veebivarandus aeg-ajalt üle vaadata ning kõik, mis ei kuulu kaasinimeste pilgu alla, teadlikult veebist eemaldada.