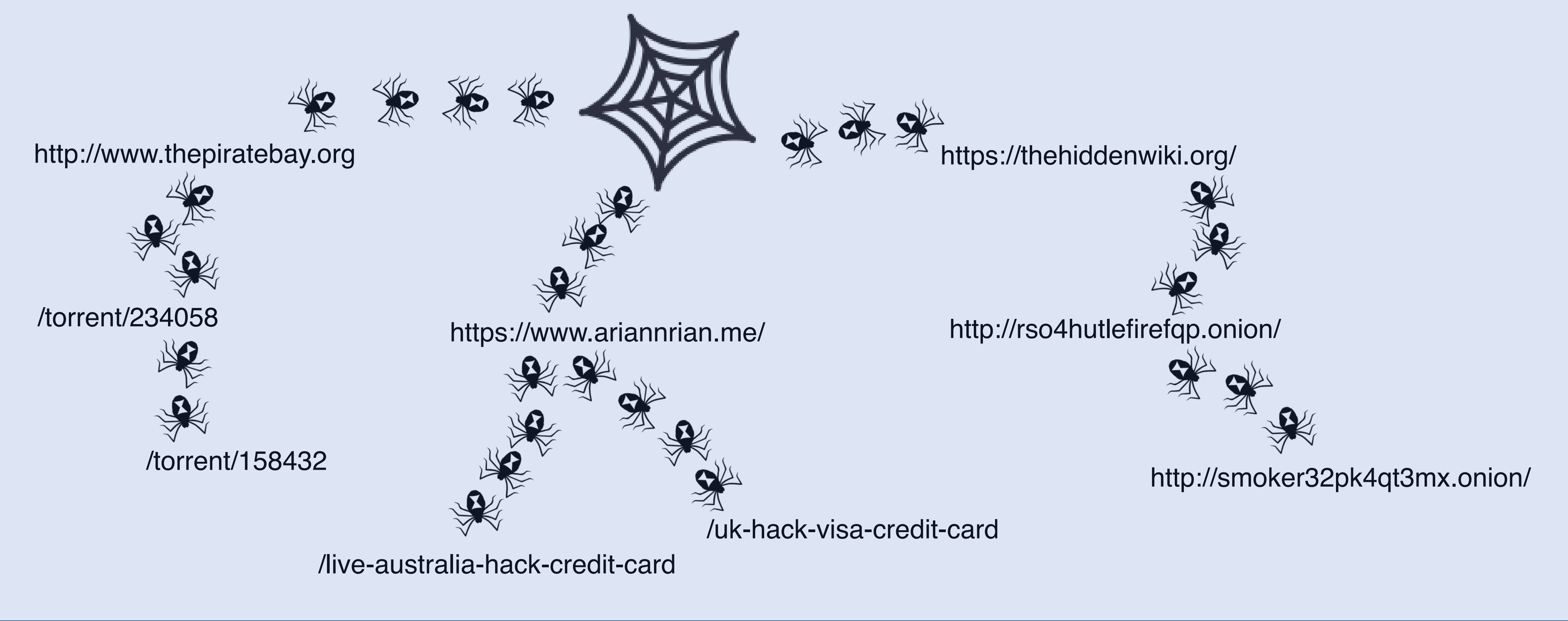


Automatizovaná analýza a archivace dat z webu

Tomáš Kocman, xkocma04@stud.fit.vutbr.cz, ID: 5

Pavouci prochází všechny stránky na webu dle definovaných strategií. Systém je plně konfigurovatelný, lze tedy například nastavit jaké weby a na nich jaké stránky se budou navštěvovat, blacklist na specifické domény nebo kontrolovat s jakou intenzitou je vzdálený webový server navštěvován, aby se předešlo DOS.



Analýza dat nalezených na webových stránkách probíhá použitím regulárního výrazu, který v textu HTML stránky hledá vyhovující řetězec.

Powerwolf - The Sacrament of Sin [Deluxe Edition] (2018)-MFA
Uploaded 07-22 2018, Size 196.75 MiB, ULed by MFA2018
Powerwolf - Complete Discography (2005-2013) [7 Cd's] 320k
Uploaded 07-19 2013, Size 731.4 MiB, ULed by pr369
Powerwolf - The Sacrament of Sin (Deluxe Version) (2018)
Uploaded 07-19 2018, Size 192.66 MiB, ULed by Freak0370
Powerwolf - Discography (2005-2015) [FLAC]
Uploaded 01-15 2017, Size 2.22 GiB, ULed by Anonymous
Powerwolf - Metallum Nostrum 2019
Uploaded 01-12 11:32, Size 120.81 MiB, ULed by RoffeTheCoolMan

Hledáme nelegální BitTorrent magnet odkazy.
Regex: \"magnet\":\\?.*(xl|dn|xt|as|xs|kt|mt|tr).*\"

Marketplace Drugs
http://rso4hutlefirefqp.onion/ - EuCanna - Medical Grade Cannabis Buds, Rick Simpson Oil, Ointments and Creams
http://newpdsuslmzqazvr.onion/ - Peoples Drug Store - The Darkweb's Best Online Drug Supplier!
http://smoker32pk4qt3mx.onion/ - Smokeables - Finest Organic Cannabis shipped from the USA
http://fzqnrlcvhkgbdwx5.onion/ - CannabisUK - UK Wholesale Cannabis Supplier

Hledáme URL adresy do sítě Tor.
Regex: (?:https?:/)?(?:www)?\\.?(\\S{16}|\\S{56})\\.onion\\b

5217295282912579 | 12 | 2021 | 327 | AUSTRALIA | PEI-SHU LIN | 3 Darkana Road | BALHANNAH | 南澳大利
亞 | 5242 | 0431834569 | shane198908@gmail.com |

TAG

Fresh hack Mastercard with CVV 2021 Expiration, Hack Credit Card Number with CVV, Hack Credit Card number with CVV - 2021 Expiration, Hack Mastercard number with CVV

Hledáme čísla odcizených bankovních karet.
Regex: \\d{12,19}

V případě, že na webové stránce existuje nějaký řetězec vyhovující regulárnímu výrazu, celá webová stránka je archivována.

thepiratebay.org	thepiratebay.org/torrent/234058	ariannrian.me	thehiddenwiki.org	rso4hutlefirefqp.onion
- index_files/	- index_files/	- index_files/	- index_files/	- index_files/
- index.htm	- index.htm	- index.htm	- index.htm	- index.htm
- index.rdf	- index.rdf	- index.rdf	- index.rdf	- index.rdf
thepiratebay.org/torrent/158432	ariannrian.me/uk-hack-visa-credit-card	smoker32pk4qt3mx.onion		
- index_files/	- index_files/	- index_files/		
- index.htm	- index.htm	- index.htm		
- index.rdf	- index.rdf	- index.rdf		