Lab 13

Weblog Pattern Analysis

Lab Objectives

- Manually parse through Web Logs via Opensearch
- Query the LLM to parse through the Web Logs Index
- Limitations of a small embedding model

Bog Standard Web Logs

Let’s open up the Discover page in Opensearch, and open up our “weblogs_raw*” index pattern. Here we will have a heavily censored export of web logs from a production environment. Due to how these logs had to be censored and imported- our fields are not parsed fully. This, while a curse when filtering through manually within Dashboards, will also show another way to incorporate a LLM into your workflows.

[CENSORED] as far as the eye can see

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 157.90.177.215 157.90.177.215 [CENSORED] - - [02/May/2022:00:02:44 -0700] "GET /signup/?langchoice=no&smid=104188011 HTTP/1.0" 302 - "-" "Mozilla/5.0 (compatible; BLEXBot/1.0; +http://[CENSORED]/)" "-" "-" - - - - "-" _id: bNwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 216.127.52.162 216.127.52.162 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 218 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: bdwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 216.127.52.213 216.127.52.213 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 3764 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: btwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 3764 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: b9wNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 14880 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: cNwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 212.104.237.143 212.104.237.143 [CENSORED] - - [02/May/2022:00:02:45 -0700] "GET /favicon.ico HTTP/1.0" 200 894 "https://[CENSORED]/landing/mobile-landers/mg-2column-timeline/?v=new-style-with-canned-header&vidChoice=ji&AFNO=1-5550&UHNSMTY=437&stno=2-630-0-7909-0-0-2050-5666" "Mozilla/5.0 (Linux; Android 9; SAMSUNG SM-G611F Build/PPR1.180610.011) AppleWebKit/537.36 (KHTML, like Gecko) SamsungBrowser/9.0 Chrome/67.0.3396.87 Mobile Safari/537.36" "-" "en-GB,en-US;q=0.9,en;q=0.8" - - - - "-" _id: cdwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 216.127.52.164 216.127.52.164 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 3764 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: ctwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 216.127.52.207 216.127.52.207 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 3686 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: c9wNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:05.843 @timestamp: Aug 5, 2024 @ 12:39:05.843 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:02:45 -0700] "POST /SMLive/SMLResult.xml HTTP/1.0" 200 3686 "-" "curl/7.21.0 (x86_64-pc-linux-gnu) libcurl/7.21.0 OpenSSL/0.9.8o zlib/1.2.3.4 libidn/1.15 libssh2/1.2.5" "-" "-" - - - - "-" _id: dNwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 198.23.167.27 198.23.167.27 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /?sssjson=1&pagenum=4 HTTP/1.0" 200 57306 "-" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.113 Safari/537.36" "-" "-" - - - - "-" _id: YtwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 50.68.167.85 50.68.167.85 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /?v=2 HTTP/1.0" 302 - "-" "Mozilla/5.0 (iPad; CPU OS 15_4 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) CriOS/101.0.4951.44 Mobile/15E148 Safari/604.1" "-" "en-US,en;q=0.9" - - - - "-" _id: Y9wNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 59.153.246.254 59.153.246.254 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /landing/mobile-landers/mg-2column-timeline-more/?vidChoice=montage&AFNO=1-54986&stno=2-1563-0-7789-0-0-3190-5670 HTTP/1.0" 200 7347 "https://[CENSORED]/" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "en-GB,en;q=0.9" - - - - "-" _id: ZNwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 157.90.181.147 157.90.181.147 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /cam/[CENSORED]/?langchoice=pt HTTP/1.0" 200 11558 "-" "Mozilla/5.0 (compatible; BLEXBot/1.0; +http://[CENSORED]/)" "-" "-" - - - - "-" _id: ZdwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 127.0.0.1 127.0.0.1 localhost - - [02/May/2022:00:02:21 -0700] "GET /server-status/?auto HTTP/1.0" 200 233 "-" "Go-http-client/1.1" "-" "-" - - - - "-" _id: ZtwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

> Aug 5, 2024 @ 12:39:03.152 @timestamp: Aug 5, 2024 @ 12:39:03.152 log: 157.90.177.228 157.90.177.228 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /cam/[CENSORED]/ HTTP/1.0" 200 8792 "-" "Mozilla/5.0 (compatible; BLEXBot/1.0; +http://[CENSORED]/)" "-" "-" - - - - "-" _id: Z9wNJJEBfzbu3jp_ZX6P _type:

User Agent Strings

User Agent Strings are the unique identifier for each web browser, or scraping tool or similar, associated with each web request. In these logs, There are a variety of mobile, desktop, and scraping programs. If you are familiar with UserAgentStrings, try to query manually for one of each type. The next slides will have example queries and results for these.

Mobile UserAgentString

`Log:android` OR `log:iPhone` OR `log:iPad`

@timestamp per week

Time	_source
Aug 5, 2024 @ 12:39:05.843	@timestamp: Aug 5, 2024 @ 12:39:05.843 log: 212.104.237.143 212.104.237.143 [CENSORED] - - [02/May/2022:00:02:45 -0700] "GET /favicon.ico HTTP/1.0" 200 894 "https://[CENSORED]/landing/mobile-landers/mg-2column-timeline/?v=new-style-with-canned-header&vidChoice=ji&AFNO=1-5550&UHNSMTY=437&stno=2-630-0-7909-0-0-2050-5666" "Mozilla/5.0 (Linux; Android 9; SAMSUNG SM-G611F Build/PPR1.180610.011) AppleWebKit/537.36 (KHTML, like Gecko) SamsungBrowser/9.0 Chrome/67.0.3396.87 Mobile Safari/537.36" "-" "en-GB,en-US;q=0.9,en;q=0.8" - - - - "-" _id: cdwNJJEBfzbu3jp_cX5H _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:59.571	@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:01:47 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 2025 "-" "Mozilla/5.0 (Linux; Android 12; SM-G988U) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.127 Mobile Safari/537.36" "-" "-" - - - - "-" _id: YNwNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:55.300	@timestamp: Aug 5, 2024 @ 12:38:55.300 log: 49.37.148.71 49.37.148.71 [CENSORED] - - [02/May/2022:00:01:12 -0700] "GET /u/thumbs/4d2222b750f4d0494_SCENE_004_MP4-1800.mp4/4d2222b750f4d0494_SCENE_004_MP4-1800.mp4-4.jpg HTTP/1.0" 302 - "http://[CENSORED]/video/faith-and-austin-get-their-clams-crammed/" "Dalvik/2.1.0 (Linux; U; Android 7.1.1; MI MAX 2 MIUI/V11.0.2.0.NDDMIXM)" "-" "zh-CN" - - - - "-" _id: UNwNJJEBfzbu3jp_Rn5U _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:55.300	@timestamp: Aug 5, 2024 @ 12:38:55.300 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:01:12 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 2009 "-" "Mozilla/5.0 (Linux; Android 9; Moto Z (2) Build/PPX29.159-24) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0 Chrome/100.0.4896.127 Mobile Safari/537.36" "-" "-" - - - - "-" _id: V9wNJJEBfzbu3jp_Rn5U _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:51.337	@timestamp: Aug 5, 2024 @ 12:38:51.337 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:00:35 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 1994 "-" "Mozilla/5.0 (Linux; Android 11; SM-A215U) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.127 Mobile Safari/537.36" "-" "-" - - - - "-" _id: RNwNJJEBfzbu3jp_Nn61 _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:51.337	@timestamp: Aug 5, 2024 @ 12:38:51.337 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:00:36 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 2012 "-" "Mozilla/5.0 (Linux; Android 11; KB2007) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.127 Mobile Safari/537.36" "-" "-" - - - - "-" _id: R9wNJJEBfzbu3jp_Nn61 _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:51.337	@timestamp: Aug 5, 2024 @ 12:38:51.337 log: 54.238.21.82 54.238.21.82 [CENSORED] - - [02/May/2022:00:00:36 -0700] "GET /static-file-streaming/videos/hero-join-496x100-girls.mp4?Xld_rct=1 HTTP/1.0" 302 - "-" "Mozilla/5.0 (Linux; Android 9; V1809A) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.110 Mobile Safari/537.36" "-" "-" - - - - "-" _id: SdwNJJEBfzbu3jp_Nn61 _type: - _index: weblogs_raw _score: -

Time	_source
Aug 5, 2024 @ 12:39:03.152	@timestamp: Aug 5, 2024 @ 12:39:03.152 log: 59.153.246.254 59.153.246.254 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /landing/mobile-landers/mg-2column-timeline-more/?vidChoice=montage&AFNO=1-54986&stno=2-1563-0-7789-0-0-3190-5670 HTTP/1.0" 200 7347 "https://[CENSORED]/" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "en-GB,en;q=0.9" - - - - "-" _id: ZNwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:59.571	@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 223.231.179.26 223.231.179.26 [CENSORED] - - [02/May/2022:00:01:47 -0700] "GET /48xai/video/indian+blowjob HTTP/1.0" 302 - "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 12_5 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) GSA/160.0.373863126 Mobile/15E148 Safari/604.1" "-" "en-us" - - - - "-" _id: WdwNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:59.571	@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:01:47 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 2019 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "-" - - - - "-" _id: W9wNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:59.571	@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 93.72.230.215 93.72.230.215 [CENSORED] - - [02/May/2022:00:01:47 -0700] "GET /landing/mobile-landers/top-cats-v2/?AFNO=1-555063&UHNSMTY=437&stno=2-630-0-7909-0-0-2050-4839 HTTP/1.0" 200 6397 "https://[CENSORED]/landing/mobile-landers/top-cats-v2/?AFNO=1-555063&UHNSMTY=437&stno=2-630-0-7909-0-0-2050-4839" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "ru" - - - - "-" _id: XNwNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:59.571	@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 99.5.91.199 99.5.91.199 [CENSORED] - - [02/May/2022:00:01:47 -0700] "GET /lander/v2/mg-mobile/mg-top-cats-v2-geo/?nav-v2=true&setp1p=PIP_SM&AFNO=1-537289&UHNSMTY=437&stno=2-630-0-5767-0-0-3192-5861 HTTP/1.0" 302 - "https://[CENSORED]/" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "en-US,en;q=0.9" - - - - "-" _id: X9wNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:55.300	@timestamp: Aug 5, 2024 @ 12:38:55.300 log: 127.0.0.1 127.0.0.1 [CENSORED] - - [02/May/2022:00:01:12 -0700] "POST /ajax/config/index.php?ajax=1 HTTP/1.0" 200 2023 "-" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "-" - - - - "-" _id: UtwNJJEBfzbu3jp_Rn5U _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:55.300	@timestamp: Aug 5, 2024 @ 12:38:55.300 log: 207.66.133.63 207.66.133.63 [CENSORED] - - [02/May/2022:00:01:12 -0700] "GET /landing/mobile-landers/top-cats-v2/?AFNO=1-5466&UHNSMTY=447&stno=2-643-0-7698-0-0-3179-5738&Xld_rct=1 HTTP/1.1" 200 6459 "https://[CENSORED]/" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.4 Mobile/15E148 Safari/604.1" "-" "en-US,en;q=0.9" - - - - "-" _id: U9wNJJEBfzbu3jp_Rn5U _type: - _index: weblogs_raw _score: -
Aug 5, 2024 @ 12:38:55.300	@timestamp: Aug 5, 2024 @ 12:38:55.300 log: 92.184.124.64 92.184.124.64 [CENSORED] - - [02/May/2022:00:01:12 -0700] "GET /landing/mg-mobile/mg-scrolling-feed-sarah/?v=dark02&AFNO=1-559837&UHNSMTY=437&stno=2-630-0-8096-0-0-3247-5136 HTTP/1.0" 200 7392 "https://[CENSORED]/" "Mozilla/5.0 (iPhone; CPU iPhone OS 15_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) FxiOS/99.3 Mobile/15E148 Safari/605.1.15" "-" "en-GB,en;q=0.9" - - - - "-" _id: VtwNJJEBfzbu3jp_Rn5U _type: - _index: weblogs_raw _score: -

Desktop UserAgentString

log:Windows

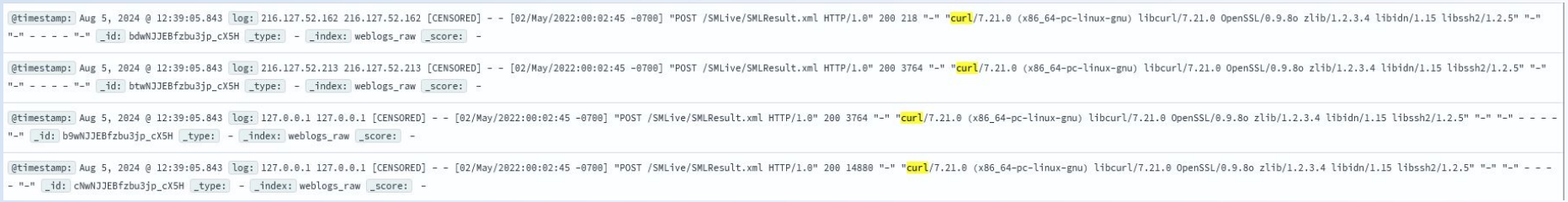
@timestamp: Aug 5, 2024 @ 12:39:03.152 log: 198.23.167.27 198.23.167.27 [CENSORED] - - [02/May/2022:00:02:21 -0700] "GET /?sssjson=1&pagenum=4 HTTP/1.0" 200 57306 "-" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.113 Safari/537.36" "-" "-" - - - - "-" _id: YtwNJJEBfzbu3jp_ZX6P _type: - _index: weblogs_raw _score: -

@timestamp: Aug 5, 2024 @ 12:38:59.571 log: 193.235.141.17 193.235.141.17 [CENSORED] - - [02/May/2022:00:01:47 -0700] "GET /cam/[CENSORED]/ HTTP/1.0" 200 14482 "-" "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/94.0.4606.81 Safari/537.36" "-" "en-US" - - - - "-" _id: YdwNJJEBfzbu3jp_WX7X _type: - _index: weblogs_raw _score: -

@timestamp: Aug 5, 2024 @ 12:38:51.337 log: 154.202.119.143 154.202.119.143 [CENSORED] - - [02/May/2022:00:00:36 -0700] "POST /login.php HTTP/1.0" 200 9121 "https://[CENSORED]/login.php?langchoice=en" "Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36" "-" "-" - - - - "[CENSORED]@[CENSORED].COM"'154.202.114.26 154.202.114.26 [CENSORED] - - [02/May/2022:00:00:36 -0700] "POST /login.php HTTP/1.0" 200 9123 "https://[CENSORED]/login.php?langchoice=en" "Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36" "-" "-" - - - - "[CENSORED]@[CENSORED].COM"'157.90.177.212 157.90.177.212 [CENSORED] - - [02/May/2022:00:00:36 -0700] "GET /cam/[CENSORED]/?langchoice=fr HTTP/1.0" 200 9171 "-" "Mozilla/5.0 (compatible; BLEXBot/1.0; +http://[CENSORED]/)" "-" "-" - - - - "-" _id: StwNJJEBfzbu3jp_Nn61 _type: - _index: weblogs_raw _score: -

Scraper UserAgentString

`NOT log:android AND NOT log:iphone` OR `log:blexbot` OR `log:curl`



IP Addresses

Given that there are a multitude of various IP Addresses, simply pick any random one that suits your fancy. Make sure to save some for later. Here is one for a scraper:

`log:157.90.177.215`

LLM IP

Now that we've manually queried for some information, let's see if the embedding model can keep up with this dataset. Ask the LLM for one of the IPs you found. Try and do this a few times to see if the embedding model is able to access all of the data. Below is an example query. Note any inaccuracies you may find.

```
POST /_plugins/_ml/agents/ d4LuXpgBlaNTCsEIC-kz /_execute
{
  "parameters": {
    "question": "Does the IP 157.90.177.215 exist in
this dataset?"
  }
}
```

LLM User Agent Strings

Let’s try to repeat the previous experiment, but this time around User Agent Strings. Here are some example queries:

```
Are there any mobile User Agent Strings in the Dataset?
Are there any scraper User agent Strings in the Dataset?
What Mobile Devices appear in the Dataset?
```

Try to make some of your own as well! Note any inaccuracies.

Lab End