

Non-Search Engine Referred Traffic to Wikipedia.org

Mikhail Popov

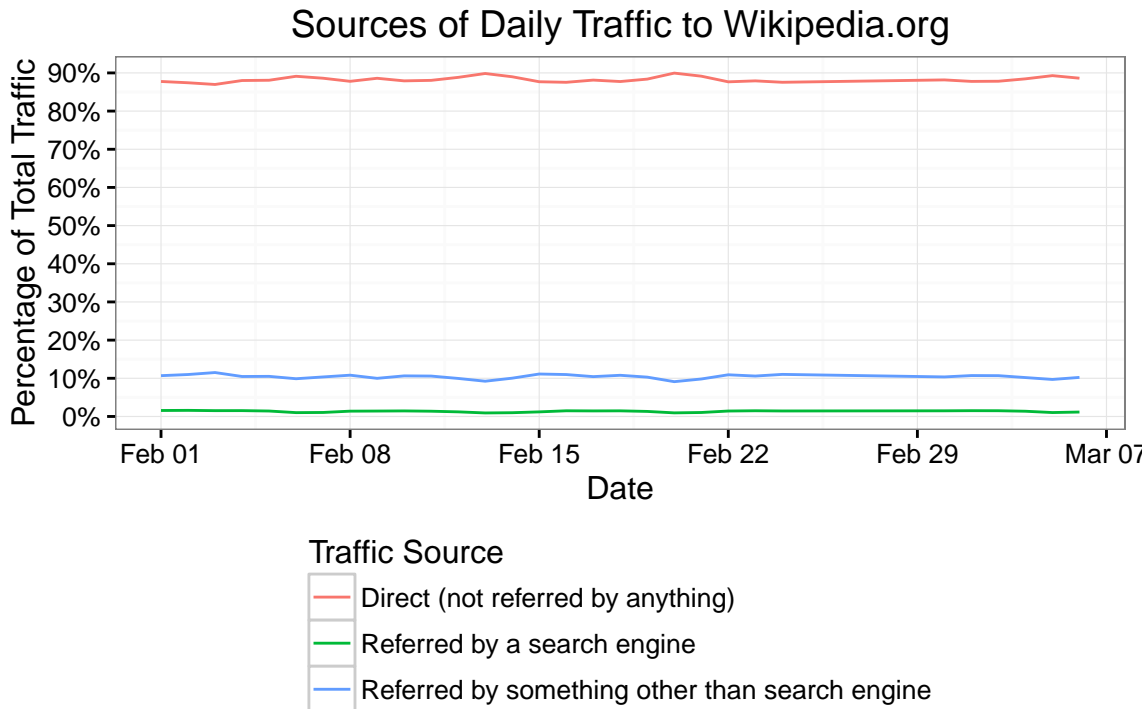
March 16, 2016

Executive Summary

On 10 March 2016 we deployed a change to the External Traffic part of the Portal dashboard that broke down the "non-search engine" traffic into direct traffic and "referred by something other than search engine" traffic. Approximately 10% of all traffic to wikipedia.org is referred by something other than a search engine.

In this report we dove deeper into that 10% traffic to see where the Portal users are coming from:

- About 77% of referred traffic (which accounts for 10% of overall traffic to wikipedia.org) is from Wikimedia projects, most notably wikipedia.org itself, which accounts for nearly all of the internally-referred traffic.
- External traffic (21.6% of referred traffic) is largely accounted by Lightning newtab (a Chrome extension, 23% of external traffic) and AlexaSurfing (Alexa rank booster, 10.3% of external traffic).



Analysis of 10% Traffic

This report looks at web request logs from 8 March 2016 to 15 March 2016, filtered by presence of a referer but specifically not one of the major search engines the refinery recognizes (Baidu, Bing, Google, Yahoo!, Yandex). External traffic accounts for 21.6%, internal for 76.8%, and unknown referrers for 0.5% on average (median).

Date	External	Internal	Unknown
2016-03-08	17.86%	81.82%	0.32%
2016-03-09	16.41%	75.49%	8.11%
2016-03-10	18.67%	80.83%	0.50%
2016-03-11	23.03%	76.52%	0.44%
2016-03-12	22.65%	77.08%	0.28%
2016-03-13	20.56%	78.93%	0.51%
2016-03-14	24.18%	75.19%	0.63%
2016-03-15	24.37%	74.76%	0.88%

We explored some of the unknown and internal referrers, and found the following statistics:

	Requests	% of 10% Traffic
Referred by ourselves (wikipedia.org)	7709525	76.84%
Unknown referrer format	175256	1.75%
Internal referer (non-Portal)	96253	0.96%
Referred by Wikimedia Nederland (wikimedia.nl)	54600	0.54%
Empty scheme	31054	0.31%
Referred from a file	119	<0.002%
Referred by hndUnblock.cgi on Cisco/Linksys routers	7	<0.0001%
Referred by IPv6	6	<0.0001%
Garbage referer metadata	3	<0.0001%
Referred by mailto link	2	<0.0001%

For the analysis of external traffic, we excluded the requests matching the categories above.

Analysis of Internal Traffic

The table below shows the top 11 WMF projects with respect to percentage of internally-referred traffic they accounted for.

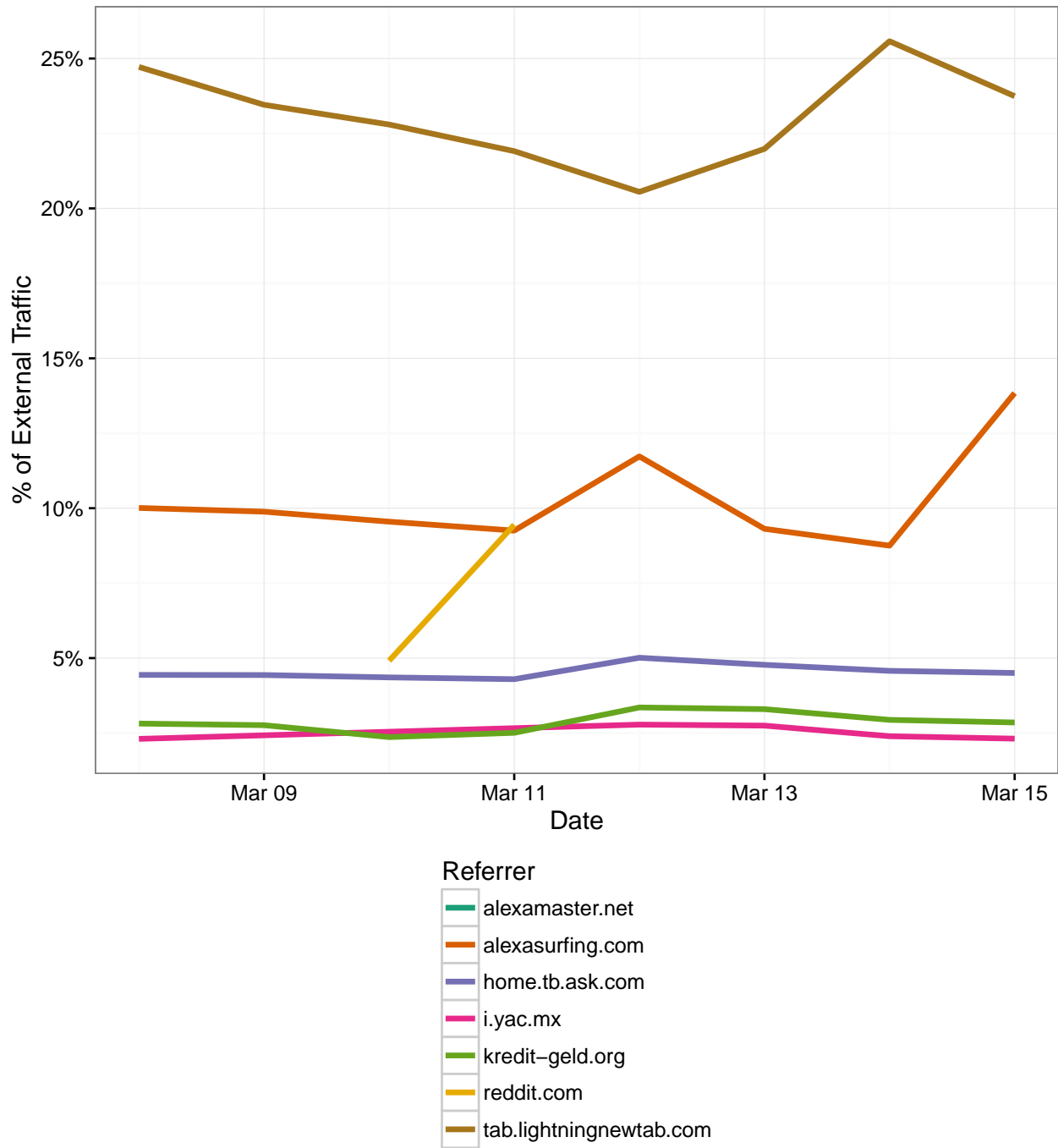
Domain	% of Internal Traffic
wikipedia.org	98.766%
de.wikipedia.org	0.811%
wikivoyage.org	0.083%
wikimedia.org	0.064%
en.m.wikipedia.org	0.048%
es.m.wikipedia.org	0.036%
es.wikipedia.org	0.032%
en.wikipedia.org	0.027%
wikimediafoundation.org	0.018%
wikinews.org	0.018%
wiktionary.org	0.015%
... 202 others...	0.081%

Analysis of External Traffic

The two major sources of traffic to Wikipedia.org appear to be... Lightning newtab (a Chrome extension) and AlexaSurfing: Alexa Rank Booster (an Alexa traffic exchange service).

Source	% of External Traffic
tab.lightningnewtab.com	23.16%
alexasurfing.com	10.29%
home.tb.ask.com	4.53%
kredit-geld.org	2.84%
i.yac.mx	2.39%
reddit.com	2.30%
search.ask.com	1.84%
facebook.com	1.79%
rdr.interstitialadservices.xyz	1.62%
loutrakiblog.gr	1.59%
... 28478 others...	47.65%

Top 5 Sources of External Traffic to Wikipedia.org



There is some fluctuation in traffic from those sources, but we can still get a good picture of the top referral traffic in a week.