



## Corrigendum

## Corrigendum to “Wikipedia workload analysis for decentralized hosting” [Computer Networks 53 (11) (2009) 1830–1845]

Guido Urdaneta <sup>\*,1</sup>, Guillaume Pierre, Maarten van Steen

VU University, Department of Computer Science, De Boelelaan 1083, 1081HV Amsterdam, The Netherlands

The authors would like to point out an error in article [1]. The paper states in Section 3 that there were 20.6 billion requests in the studied period, but the correct number is 25.6 billion requests. The difference corresponds to media file (image) requests that were not considered originally. This results in some values reported in Tables 1 and 2 and parts of the text being incorrect. Also note that the read/save ratio column in Table 2 had numbers not corresponding to the studied period, but to a previous data set.

The corrections to the text are as follows:

**Section 3, Paragraph 1**

*Where it says:* and contains 20.6 billion HTTP requests

*It must say:* and contains 25.6 billion HTTP requests

**Section 4, Paragraph 2**

*Where it says:* account for more than 64% of all requests. We can also see that page editions (at 0.03%) are very infrequent

*It must say:* account for more than 71% of all requests. We can also see that page editions (at 0.02%) are very infrequent

**Section 4, Paragraph 3**

*Where it says:* Table 2 shows the distribution of request load

*It must say:* Table 2 shows the distribution of page request load

*Where it says:* almost half of the total traffic is directed to the English Wikipedia, and about 90% of the traffic is concentrated in the 10 most popular wikis

*It must say:* almost half of all page requests are directed to the English Wikipedia, and about 90% of the page requests are concentrated in the 11 most popular wikis

**Section 8, Paragraph 2**

*Where it says:* which represent more than 90% of all requests (95% if we ignore cache maintenance requests)

*It must say:* which represent more than 92% of all requests (96% if we ignore cache maintenance requests)

The corrected tables are as follows:

DOI of original article: [10.1016/j.comnet.2009.02.019](https://doi.org/10.1016/j.comnet.2009.02.019)

\* Corresponding author. Tel.: +31 20 5987754.

E-mail addresses: [g.urdaneta@few.vu.nl](mailto:g.urdaneta@few.vu.nl) (G. Urdaneta), [gpierre@cs.vu.nl](mailto:gpierre@cs.vu.nl) (G. Pierre), [steen@cs.vu.nl](mailto:steen@cs.vu.nl) (M. van Steen).

<sup>1</sup> Supported by the Programme Alban, the European Union Programme of High Level Scholarships for Latin America, scholarship No. E05D052447VE.

**Table 1**

Wikipedia request types, and their frequencies expressed in fractions of the total number of requests (numbers do not add up to 100% due to rounding error).

Description	Frequency (%)
Requests for thumbnails of user-uploaded images	32.82
Requests for static files. These are usually files used in the rendering of wiki pages, such as CSS and Javascript files as well as generic images such as bullets	21.03
Requests for unmodified user-uploaded media files, typically images	17.67
Requests for the current version of a wiki page using the default HTML rendering	10.59
Requests for the current version of a wiki page using a different format, such as printer-friendly versions or raw wikitext	6.84
Requests related to cache maintenance	4.17
Requests in which a page name is specified, but the information retrieved is independent from the page's wikitext. For example, Wikipedia allows obtaining the Javascript or CSS used in the rendering of a specified page without obtaining the page itself	3.62
Requests that produce search results in a standardized format known as OpenSearch	1.05
Requests for special pages other than full text search or upload. Some of these special pages result in the retrieval of wiki pages. However, the names of the wiki pages involved cannot be obtained from the trace	0.81
Keyword search requests handled by the search component of the wiki engine	0.65
Requests directed to a web service API. Most of these requests result in the retrieval of wiki pages, but in many cases it is not possible to determine the names of the pages involved	0.11
Requests for the edition history of a page	0.05
Requests for a specific version of a page. It may be a diff operation, in which case two versions are compared	0.05
Requests that result in a page update or creation	0.02
Requests for possible uploads of media files. They can refer to new files, updates of existing files or retrieval of a form that allows the user to send an actual upload request. It is impossible to determine the name of the uploaded file from our traces or whether it is an actual upload	0.002
Other types of request	0.6

**Table 2**

Distribution of page requests and read/save ratios across different wiki projects.

Wiki	Frequency (%)	HTML Read/Save Ratio
English Wikipedia	43.31	590.1
German Wikipedia	9.14	661.4
Japanese Wikipedia	9.10	1371.6
Spanish Wikipedia	8.06	474.3
Wikimedia Meta	5.84	132.8
French Wikipedia	3.61	256.8
Polish Wikipedia	2.96	525.7
Portuguese Wikipedia	2.40	401.8
Italian Wikipedia	2.12	290.4
Wikimedia Commons	1.66	240.1
Dutch Wikipedia	1.25	266.6
Others (<1% each)	10.55	Unknown

## Reference

- [1] Guido Urdaneta, Guillaume Pierre, Maarten van Steen, Wikipedia workload analysis for decentralized hosting, *Computer Networks* 53 (11) (2009) 1830–1845.