Characterization of the Wikipedia Traffic

Antonio J. Reinoso
Libresoft Research Group (URJC)
Fuenlabrada (Spain)
ajreinoso@libresoft.es

Rocío Muñoz-Mansilla
Department of Automation and Computer Science (UNED)
Madrid (Spain)
rmunoz@dia.uned.es

Israel Herraiz
Department of Applied Mathematics and Computing (UPM)
Madrid (Spain)
israel.herraiz@upm.es

Felipe Ortega
Libresoft Research Group (URJC)
Fuenlabrada (Spain)
jfelipe@libresoft.es

***Abstract*—Since its inception, Wikipedia has grown to a solid and stable project and turned into a mass collaboration tool that allows the sharing and distribution of knowledge. The wiki approach that basis this initiative promotes the participation and collaboration of users. In addition to visits for browsing its contents, Wikipedia also receives the contributions of users to improve them. In the past, researchers paid attention to different aspects concerning authoring and quality of contents. However, little effort has been made to study the nature of the visits that Wikipedia receives. We conduct such an study using a sample of users' requests provided by the Wikimedia Foundation in the form of Squid log lines. Our sample contains more that 14,000 million requests from users all around the world and directed to all the projects maintained by the Wikimedia Foundation, including different editions of Wikipedia. This papers describes the work made to characterize the traffic directed to Wikipedia and consisting of the requests sent by its users. Our main aim is to obtain a detailed description of its composition in terms of the percentages corresponding to the different types of requests making part of it. The benefits from our work may range from the prediction of traffic peaks to the determination of the kind of resources most often requested, which can be useful for scalability considerations.**

Keywords-Wikipedia; Traffic characterization.

I. INTRODUCTION

Wikipedia stands as the most successful wiki-based project and provides a vast compilation of contents related to all the knowledge areas. Furthermore, the Wikipedia underlying philosophy promotes the collaboration and participation of users in the production of pieces of knowledge that will remain available for the whole community. This new paradigm for knowledge generation has attracted great attention and has propitiated the consolidation of Wikipedia as a mass collaboration tool. Such acceptation can be regarded just from the continuously increasing number of visits to the different Wikipedia domains that places its web site within the six most visited ones all over the Internet [1].

Regarding its contents, Wikipedia is divided in approximately 270 [2] editions that correspond to the same number of languages. All these editions sum up to 19 million articles, which correspond to encyclopedic entries about particular subjects, dates or people. Wikipedia articles address topics corresponding to traditionally academic disciplines as well as to cultural, sportive or artistic manifestations. In addition, they also deal with highly topical subjects, biographies from live persons or topics related to general public entertainment.

In respect to the audience, Wikipedia editions receive approximately 13,500 million visits a month. This observation can be considered as a good indicator of its acceptance and popularity among users. Such number of visits constitute an absolute challenge in terms of management of requests and content delivery. Concerning this topic, several re-arrangements and re-organizations had to be made in the supporting architecture to meet the scalability demands derived from its rise in popularity and users' participation.

As a result of this relevance, Wikipedia has evolved into a subject of increasing interest for researchers [3]. In this way, different quantitative examinations about its articles, authors, visits or contributions have been undertaken [4], [5]. However, most of the previous research involving Wikipedia deals with the quality and reliability of its contents ([6], [7], [8]) or study its growing tendency and evolution [9], [10]. By contrast, very few studies [11], [12], [13] have been devoted to analyze the manner in which users interact and make use of Wikipedia.

Therefore, this paper aims to analyze the different kind of requests submitted to Wikipedia by its users in a effort to determine both quantitative and qualitative features of such traffic. The major benefits from our study may range from a detailed characterization of the requests sent to the Wikipedia supporting architecture to the forecasting of systems' overload during stress peaks. In addition, proper knowledge about most requested resources may lead to systems improvements concerning the delivery management policy. Finally, we also provide several comparisons amongst the different Wikipedia editions in order to assess differences or trends relative to particular editions. Moreover, we also outline those evolutions that do not fit the general tendency

 ISBN: 978-1-61208-200-4

resulting from the observation of all the received requests.

Our analysis focuses on the largest Wikipedia editions in terms of their number of both articles and requests. In addition, we have analyzed the traffic during a whole year (2009) to avoid temporarily localized events. Our main data source consists in users' requests that are stored by special Squid servers that are deployed by the Wikimedia Foundation to deal with all the incoming traffic to its several projects. In this way, information about each individual request is registered in the form of a log line whose fields are later processed by an ad-hoc Java application. This application filters the requests targeting to Wikipedia contents or services and classifies them for countability purposes.

The rest of the paper is structured as follows: Section II presents the data sources used for this study. Section III explains the filtering process for the data sample and the information that can be extracted out of it. After this, Section IV presents the results, and finally Section V concludes this paper and proposes some ideas for further work.

II. The data source

This section aims to describe the information sources used in our study and constituting the main data feeding to perform our analysis. Visits to Wikipedia, in a similar way to any other Internet site, are issued from users' browsers in the form of URLs. These petitions are registered by the Wikimedia Foundation Squid servers in the form of log lines once the requested contents have been served.

Squid servers are a special kind of servers performing web caching that are used by the Wikimedia Foundation as the first layer in its Content Delivery Network. They manage all the traffic directed to Wikipedia as well as to the rest of wiki-based projects. Squids register every responded petition as log lines and a sample of them is sent to universities and research centers.

Squids commonly work as proxy servers performing web caching. In this way, they cache contents previously browsed to make them locally available in the case that requests for the same contents are issued. This results in a significant decrease of the bandwidth consumption and in a more efficient use of the underlying network. Furthermore, Squid servers may also be used to improve web servers by caching the contents repeatedly requested to them. Squid servers are said to work as reverse proxy servers because they try to answer the incoming requests with the cached contents. When successful, this approach avoid the participation of any other system in the delivery of the requested contents. Particularly, this prevents the operation of database or web servers purportedly placed behind them.

In the case of the Wikimedia Foundation, two layers of Squid servers are placed in front of its Apache and database servers. In this way, most of the requested content is directly served from the Squid subsystem without involving any of the other servers. As the Wikimedia Foundation maintains several wiki-based projects, such as Wikipedia, Wikiversity or Wikiquote, the Squid layers have to deal with all the incoming traffic directed to these projects.

Currently, there are two large Squid server clusters: a primary cluster (located in Tampa, Florida) and another secondary cluster (located in Amsterdam) that only performs web caching. These Squids servers usually run at a hit-rate of approximately 85% for text and 98% for media using CARP (Cache Array Routing Protocol) [14]. Users' requests are firstly routed to one of the Squid clusters using a DNS balancing policy.

However, all the contents requested by users are not cacheable. The pages sent to registered and logged-in users, for example, cannot be cached as they include customized parts as the users' nicknames or, even, personalized options for page displaying such as skins or templates.

Squid systems log information about each served request disregarding whether the answer could have been found in the cache or, on the contrary, it was a tailored page built up by web servers. Every Squid server packages and sends its log lines to a central aggregator host. Here, there is a program in charge of their reception that, in addition, sends them to the set of registered log processors. Basically, a log processor consists either in a file processor, that writes lines to a file, or in a pipe processor, that sends them to a specific command trough a pipe. Both of them use a sampling factor to determine the next line to be written or piped. In turns, another program does the opposite operation and picks the lines to send them through a UDP packet stream. This is how Wikimedia Foundation Squid log lines finally reach our storage systems.

Each log line from a Wikimedia Squid server corresponds to a served user request and constitutes a really valuable data feed because, among several other information, it includes the URLs submitted by the user along with the date at witch the corresponding content was sent in response.

III. Methodology

The analysis presented here is based on a sample of the traffic directed to all the Wikimedia Foundation wiki-based projects during 2009. The sampling factor used for generating our data feed was 1%, which means that we received one in every hundred requests composing the traffic to the several projects maintained by the Wikimedia Foundation. In general terms, more than 14,000 million log lines have been parsed and filtered for this study.

To begin with, we had to separate the requests directed to Wikipedia from the ones targeting to projects like Wikiquote, Wikiversity, etc. In addition, we have only considered consolidated and assiduous Wikipedias in order to focus on highly active editions. Specifically, we have analyzed the requests corresponding to the ten top-most editions regarding their number of, both, articles and visits. These editions

 ISBN: 978-1-61208-200-4

are the German, English, Spanish, French, Italian, Japanese, Dutch, Polish, Portuguese and Russian ones.

The streaming made up of the log lines from the Wikimedia Foundation Squid systems is daily rotated in such a way that lines corresponding to different days are separated in different files. Once stored, log lines are completely available for their processing using an ad-hoc java written application: the *WikiSquilter* tool [15]. The processing consists in a parsing phase devoted to extract the relevant information fields from the log lines. Then, these elements are filtered in order to determine what lines correspond to requests considering of interest according to the directives of the driven analysis. Finally, data related to filtered requests are normalized and stored in a relational database for further examinations.

Log lines received from the Wikimedia Foundation offer a valuable information by themselves though they do not contain specific fields with the necessary data to conduct our analysis. However, these data can be obtained from the URLs submitted by users when they send a request. In this way, URLs have to be parsed to look for the precise elements involved in the characterization process. In particular, there are elements that can be easily extracted from requests such as the following ones:

1) The Wikimedia Foundation project, such us Wikipedia, Wiktionary or Wikiquote, to which the URL is directed.
2) The corresponding language edition of the project.

For the rest of information elements, the parsing process relies on the use of regular expressions to determine the syntactical structure of requests and, consequently, their purported type. In particular, we aim to characterize users' petitions consisting in:

1) Visits, intended as requests for browsing (reading) Wikipedia articles that do not convey any other action.
2) Any action such as previews, edit historical reviews but excluding edits and searches that are treated separately.
3) edits, sent to modify Wikipedia contents that cause the issue of write operations to the database.
4) searches, looking for articles related to a certain topic.
5) api calls, that request any of the built-in functionality offered by the mediawiki software.
6) skin/css requests, that demand customized elements or choices used in the visualization and presentation of Wikipedia contents.
7) media wiki extensions, that are requests for extensions added to provide new functionalities through third-party code ready to be set up together with the mediawiki core.

The filter process consists in assessing whether each analyzed URL is considered significant for our analysis. This is done by checking whether the information elements it contains correspond to the ones in which our work is focusing on. The filter implementation is realized on the basis of a hash structure holding the information elements considered of interest for the analysis as well as their corresponding normalized database codes to be used in the insert operation issued to database management system.

In general terms the application has been designed and developed with strong adherence to the principles of efficiency, robustness and accuracy. However, flexibility and extensibility directives have been also reinforced. Efficiency is gained through several elements such as multithreaded design and filter's O(1) complexity derived of the hash basis. Application's robustness has been achieved by means of the capability of detecting malformed URLs. Flexibility makes the application suitable of being used with whatever log lines with the only requirement of specifying in the corresponding XML file the elements to be parsed and filtered. The software architecture of the application allows to easily include new services that can even involve new data to be processed, so extensibility has been also considered.

IV. Analysis and results

This section provides a quantitative analysis of the traffic composition in the aim of providing an adequate characterization of the requests directed to Wikipedia. This kind of analysis may contribute to describe the way in which Wikipedia is being utilized by its community of users. In addition, our results may serve as an estimation of the operational overload for the systems in charge of supporting the different Wikimedia Foundation wiki-based projects and, particularly, Wikipedia.

Therefore, we present here the characterization of the different types of requests composing the traffic to the Wikipedia editions under study. Furthermore, we are also presenting information related to the general traffic to all the Wikimedia Foundation projects. Traffic information is always computed in terms of number of requests, disregarding, by the moment, considerations about amount of information or transference rates. In addition, we usually present the daily averaged number of requests when larger time units, such as months, are considered in order to avoid the introduction of biased perceptions due to the differences in the number of days. We have to note that technical problems have prevented us from obtaining the traffic information of all the days from 2009. However, we have only failed to receive the traffic of just 4 days, which is an absolute success in terms of the reliability of our receiving infrastructure.

First of all, we consider of interest to determine how the overall traffic to the Wikimedia Foundation is distributed among its different projects during 2009. This is shown in Table I, which provides the percentages of the total traffic corresponding to each particular project. As it is clearly seen, the largest percentage corresponds to the requests

 ISBN: 978-1-61208-200-4

WMF project	Percentage of traffic attracted
Wikipedia	49.47%
Wikiversity	0.03%
Wikibook	0.23%
Wiktionary	0.52%
Wikiquote	0.16%
Species	0.01%
Wikinews	0.06%
Wikisource	0.13%
Commons (images)	1.26%
Uploaded resources	46.72%
Other	1.41%

Table I
TRAFFIC DIRECTED TO EACH WIKIMEDIA FOUNDATION PROJECT AND TO PREVIOUSLY UPLOADED RESOURCES.

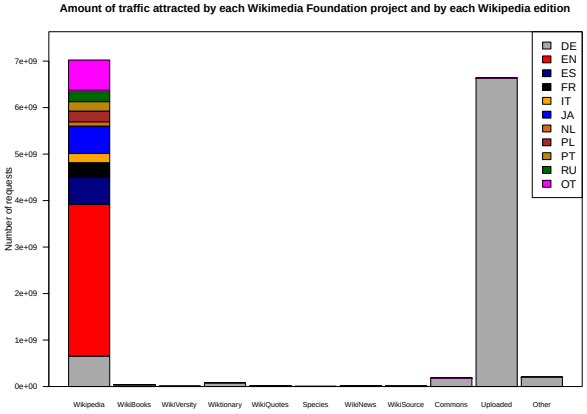

Figure 1. Amount of traffic corresponding to each Wikimedia Foundation project and to each edition of Wikipedia during 2009.

for Wikipedia articles (49.7%). Interestingly, almost the remaining half of requests (46.72%) corresponds to images and other multimedia resources uploaded to the platform to be referenced not only from Wikipedia articles but also from articles belonging to the rest of wiki-based projects. All together, these two types of requests add up to the 96% of all the traffic received by the Wikimedia Foundation servers. Figure 1 shows the relevance of both kind of requests in the traffic and also includes the amount of it corresponding to each Wikipedia edition. As it is shown, the English Wikipedia (EN, in red) attracts much more traffic than any other edition followed by the German (DE, gray), the Spanish (ES, dark blue) and the Japanese (JA, blue) ones.

Figure 2 presents the monthly evolution of the traffic directed to all the Wikimedia Foundation projects during 2009. The vertical axis shows the daily average of requests corresponding to each particular project and to common resources, mainly images, requested by users. In order to adequately examine these figures, it is important to remark that they correspond to the daily average of the sample we receive, which is the 1% of the total traffic, so real ones

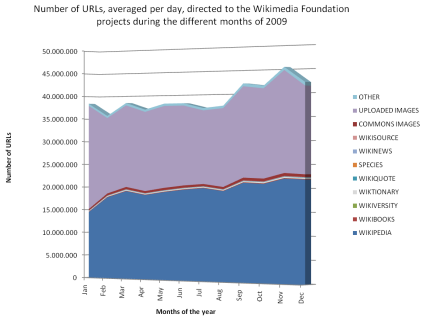

Figure 2. Evolution of the daily averaged number of requests to each Wikimedia Foundation project in every month of 2009.

would be, for example, 30 * 100 times higher in the case of months having 30 days. From Figure 2 we can also compare the monthly evolution of the traffic to Wikipedia in respect to the total traffic to all the supported Wikimedia Foundation projects. As it is shown, though both traffics seem to follow quite similar monthly evolutions, there are some differences such as the tendencies observed in February that indicate that the total traffic decreased whereas the number of requests to the set of Wikipedias increased.

From here on, we aim to characterize only the traffic corresponding to the Wikipedia project. In this way, our first objective is to determine the number of requests directed to each one of its editions and, particularly, to the ones considered in this work. Thus, Figure 3 shows the distribution of the requests to Wikipedia over its different editions in every month of 2009. The English Wikipedia is still the most popular, and receives a volume of traffic much higher than the other editions. Besides this, we considered of interest to aggregate the daily average of the traffic to each Wikipedia edition throughout the entire 2009 and to present their corresponding percentages in respect to the total traffic to the Wikipedia project. Table II presents this information. As we can see, the considered editions attract more than 91% of the total traffic to Wikipedia. This is important in terms of the relevance of the considered set of editions. The evolution of the daily average of requests for each particular edition in the different months of 2009 is presented in Figure 4. As it is shown, not all the Wikipedias follow the same distribution of traffic over time, which can mean different temporal patterns of use.

We can also compare the evolution of the traffic to the different editions of Wikipedia with the progression of their respective sizes. Larger Wikipedias may attract a higher number of requests as a result of their purportedly bigger supporting community. However, this is not always true according to the Figures 5 and 6 which present, respectively, the amount of traffic attracted by each Wikipedia in every month of 2009 and their sizes expressed in number of articles during the same months.

 ISBN: 978-1-61208-200-4

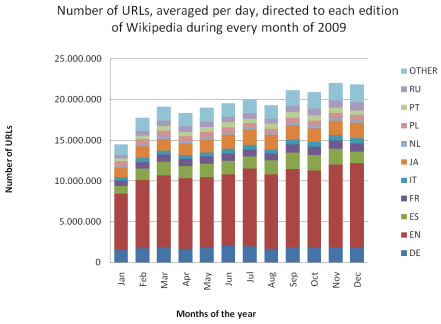

Figure 3. Comparison of the traffic directed to each edition of Wikipedia in every month of 2009.

Wikipedia edition	Daily average of attracted traffic	Percentage
DE	21,767,176.73	9.40%
EN	108,407,534.61	46.45%
ES	19,336,747.61	8.25%
FR	10,622,527.01	4.54%
IT	6,516,987.21	2.79%
JA	19,591,570.27	8.38%
NL	3,128,496.65	1.34%
PL	7,628,743.39	3.30%
PT	6,755,424.08	2.87%
RU	8,269,484.01	3.51%
REST	21,467,547.49	9.17%

Table II
AGGREGATED DAILY AVERAGED NUMBER OF REQUESTS ATTRACTED BY EACH CONSIDERED EDITION OF WIKIPEDIA DURING THE WHOLE 2009. THE TRAFFIC CORRESPONDING TO THE REST OF DISREGARDED EDITIONS IS PRESENTED SUMMARIZED UNDER THE 'REST' ENTRY.

Considering that the English and the Russian Wikipedias are, respectively, the largest and the smallest ones, the same is not valid for the amount of traffic . The case of the Spanish Wikipedia is even more curious because in spite of being situated among the three editions with lesser volumes

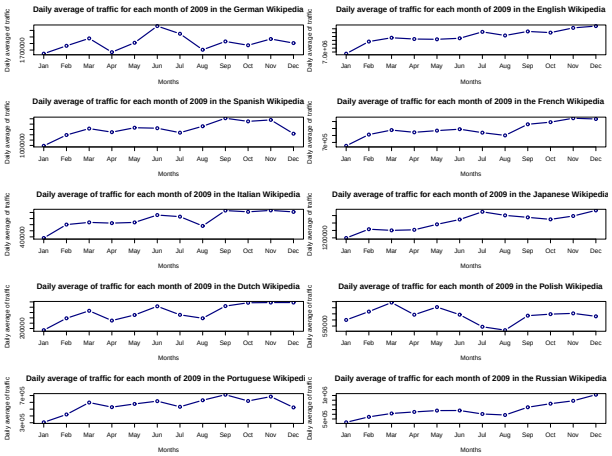

Figure 4. Evolution of the daily averaged traffic directed to each edition of Wikipedia over the different months of 2009.

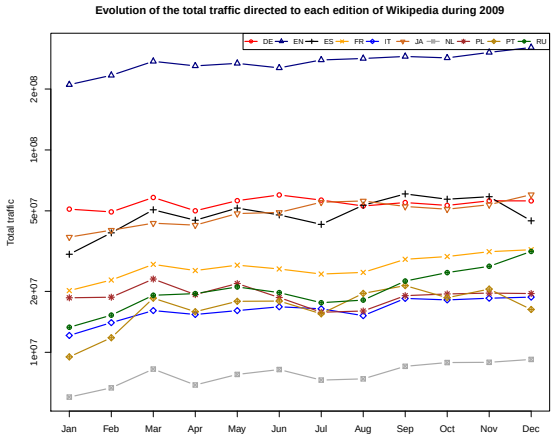

Figure 5. Monthly evolution of the total traffic directed to each edition of Wikipedia throughout 2009.

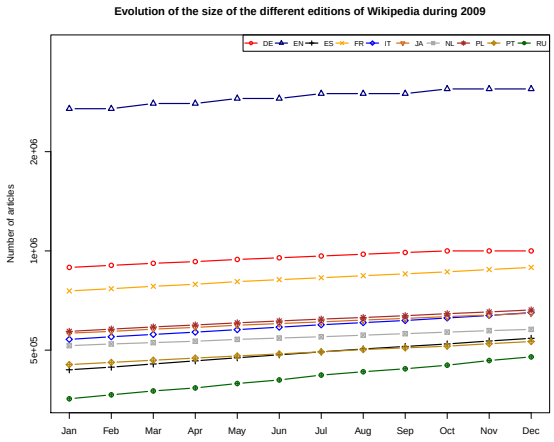

Figure 6. Monthly evolution of the size of the different editions of Wikipedia throughout 2009.

of articles, regarding its traffic, it ranges from the fourth to, even, the second most requested edition. This finding is specially interesting because it proves that resources related to storage and traffic management are not proportional at all, what should be considered particularly in scalability issues.

Surely, it is more interesting to obtain a characterization of the traffic directed to each edition of Wikipedia. This information could be interpreted as an approximation to the use given to each Wikipedia edition by its corresponding community of users. So,Table III shows the percentage of traffic directed to each Wikipedia edition that consist in visits to articles, requests for edit operations, actions such as history reviews or pre-visualizations performed on articles, search operations, css files used to present tailored pages or, even, the Wikipedia icon itself.

From Table III it is interesting to note the high percentage of requests corresponding exclusively to visits as well as to elements related to the presentation and visualization of the

Ed.	Visits to articles	Actions (exc. edit & search op.)	Edit op.	Search op.	Api calls	Skins /css	icons	mw ext.	Undet.
EN	21.51%	22.52%	0.27%	4.75%	6.53%	34.62%	4.38%	3.47%	6.95%
DE	16.54%	20.87%	0.23%	4.09%	7.69%	30.74%	3.46%	14.72%	5.98%
ES	13.58%	33.90%	0.31%	4.12%	6.02%	32.13%	3.68%	3.89%	6.80%
FR	18.24%	23.15%	0.33%	4.00%	6.05%	36.87%	4.42%	4.23%	7.04%
IT	19.80%	21.81%	0.43%	4.44%	5.77%	37.57%	4.49%	3.07%	9.69%
JA	20.69%	25.15%	0.37%	4.22%	3.95%	36.01%	4.19%	2.81%	9.22%

Table III
CHARACTERIZATION OF THE TRAFFIC DIRECTED TO SOME PARTICULAR EDITIONS OF WIKIPEDIA IN TERMS OF THE PERCENTAGES CORRESPONDING TO DIFFERENT TYPES OF REQUESTS.

requested information. It is also noticeable the extremely low percentage of edits (requests to commit any changes over the contents) that is two order of magnitude less than visits.

Regarding the different types of actions, it is shown that requests consisting in calls to the MediaWiki API (Application Programming Interface), search operations and mediaWiki extensions (pieces of code to add particular functionalities to the wiki engine) present relevant percentages. Again, this information may be useful to set and configure the range of resources dealing with these types of requests. Particular interesting observations such as the low percentage of visits in the German Wikipedia together with the impressively high ratio of requests demanding mediaWiki extensions in this edition deserve deeper research. In the same way, the lower percentage of visits corresponding to the Spanish edition and its higher number of requested actions also deserve thorough efforts.

V. CONCLUSION

In this paper we have shown how the Wikipedia traffic can be characterized to obtain its detailed composition. Furthermore, the analysis of the traffic directed to all the projects maintained by the Wikimedia Foundation indicated that it was composed mainly by requests to Wikipedia, on the one hand, and requests for previously uploaded resources, on the other hand.

When comparing the monthly evolutions corresponding to the traffic directed to the whole set of the Wikimedia Foundation's projects and to the one consisting in the, requests, just, to the contents from Wikipedia, it was found that both evolutions are considerably similar thought they present some differences. In particular, the traffic to Wikipedia presents a temporal distribution with less drops and with a slope slightly more tending to increase. This can be interpreted as a non-stopping raise in the attention attracted by the Wikipedia project. In addition, situations when the number of requests to Wikipedia increases though the general traffic falls might be explained, for example, because of a raise in the demands of articles with less images or graphical contents.

Focusing on the requests to Wikipedia, we have determined how the traffic is distributed among its different editions and how the number of received requests is not related to the editions' sizes. This is particularly interesting as it shows that resources arranged for storage and delivery do not scale with the same ratio. Wikimedia Foundation systems staff may take this fact into consideration when planning the allocation of the different kind of resources to be involved in the management and serving of the requests directed to particular language editions.

In respect to the distribution of the requests over the different months, it is found that, as expected, the traffic generally decreases during the summer months surely associated with holiday periods. In the rest of the months the tendency of the traffic does not fluctuate very much and usually tends to increase.

Finally, the percentages corresponding to the different types of considered requests found in the traffic to each edition have been presented. These results show a high number of visits and solicited actions, both near 20%. This is particularly noticeable because visits may be replied using cached contents provided they were issued by non-logged users. However, actions can never been answered in that way so that they need the participation of database servers and specific software systems depending on the nature of the requested actions.

Regarding some of the differences observed in the percentages of the different kinds of actions found in the analyzed editions, it is clear that further research is needed to find out concrete situations. Particularly, the outstanding amount of traffic concerning visualization options deserves a closer examination as it represents, in average, a third the total traffic to each editions. Depending on whether it corresponds to the established displaying options or not, it impact on the overall performance can be really remarkable.

REFERENCES

[1] "Information about wikipedia.org in alexa," accesed March 2012. [Online]. Available: http://www.alexa.com/siteinfo/wikipedia.org

[2] "Wikimedia statistics," accesed March 2012. [Online]. Available: http://stats.wikimedia.org/EN/Sitemap.htm

 ISBN: 978-1-61208-200-4

[3] "Academic studies of wikipedia," accesed March 2012. [Online]. Available: http://en.wikipedia.org/wiki/Wikipedia:Academic_studies_of_Wikipedia

[4] J. Voss, "Measuring wikipedia," in *International Conference of the International Society for Scientometrics and Informetrics : 10th*. ISSI, July 2005.

[5] F. Ortega, J. M. Gonzalez-Barahona, and G. Robles, "The top ten wikipedias: A quantitative analysis using wikixray," in *Proceedings of the 2nd International Conference on Software and Data Technologies (ICSOFT 2007)*, INSTICC. Springer-Verlag, July 2007. [Online]. Available: http://libresoft.es/downloads/C4_159_Ortega.pdf

[6] Korfiatis, Nikolaos, Poulos, Marios, Bokos, and George, "Evaluating authoritative sources using social networks: an insight from wikipedia," *Online Information Review*, vol. 30, no. 3, pp. 252–262, May 2006. [Online]. Available: http://dx.doi.org/10.1108\%2F14684520610675780

[7] J. Giles, "Internet encyclopaedias go head to head," *Nature*, vol. 438, no. 7070, pp. 900–901, December 2005. [Online]. Available: http://dx.doi.org/10.1038\%2F438900a

[8] T. Chesney, "An empirical examination of wikipedia's credibility," *First Monday*, vol. 11, no. 11, November 2006. [Online]. Available: http://firstmonday.org/issues/issue11_11/chesney/index.html

[9] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli, "Preferential attachment in the growth of social networks: the case of wikipedia," Feb 2006. [Online]. Available: http://arxiv.org/abs/physics/0602026

[10] D. Spinellis and P. Louridas, "The collaborative organization of knowledge," *Commun. ACM*, vol. 51, no. 8, pp. 68–73, August 2008. [Online]. Available: http://dx.doi.org/10.1145/1378704.1378720

[11] G. Urdaneta, G. Pierre, and M. van Steen, "A decentralized wiki enginge for collaborative wikipedia hosting," in *Proceedings of the 3rd International Conference on Web Information Systems and Technologies*, March 2007, pp. 156–163. [Online]. Available: http://www.globule.org/publi/DWECWH_draft2006.pdf

[12] A. J. Reinoso, "Temporal and behavioral patterns in the use of wikipedia," Ph.D. dissertation, Universidad Rey Juan Carlos, 2011, http://gsyc.es/ ajreinoso/phdthesis.

[13] A. J. Reinoso, J. M. Gonzalez Barahona, F. Ortega, and G. Robles, "Quantitative analysis and characterization of Wikipedia requests," in *Proceedings of the 4th International Symposium on Wikis*, ser. WikiSym '08. New York, NY, USA: ACM, 2008. [Online]. Available: http://dx.doi.org/10.1145/1822258.1822302

[14] M. Bergsma, "Wikimedia architecture," Tech. Rep., April 2007, accesed March 2012. [Online]. Available: http://www.nedworks.org/~mark/presentations/san/Wikimedia\%20architecture.pdf

[15] "The wikisquilter project," accesed March 2012. [Online]. Available: http://www.alexa.com/siteinfo/wikipedia.org

 ISBN: 978-1-61208-200-4