Greenalytics: a tool for mash-up life cycle assessment of websites.

Jorge L. Zapico, Marko Turpeinen, Nils Brandt

KTH, The Royal Institute of Technology, Centre for Sustainable Communications

Lindstedtsvägen 3, 10044 Stockholm, Sweden

zapico@kth.se

turpeinen@kth.se

nilsb@kth.se

Abstract

The environmental impact of internet is growing, reaching an estimated 1.4% of world greenhouse emissions. This impact is hidden for both users and web developers. Understanding and analyzing the environmental footprint of a website is not an easy task. The impacts are distributed through multiple hardware networks and a global user base, making the individual impacts difficult to allocate. This article presents the development of a functional application for generating automatic life cycle assessments for web sites based on mashing-up information. This application has the aim of making the impact of websites visible, allowing the instant analysis of their carbon footprint using existing analytics data and presenting it in an understandable and transparent way. The development process is presented with detailed information about how the calculations are performed. The results are discussed around two different cases, focusing on the challenges of calculating the server side impact and the possibilities for improvement.

1. Introduction

Internet usage has grown exponentially during the last decade, and its environmental impact is gaining relevance. Even if individual technologies have improved efficiency, the total impact of internet has continued growing, accounting for 1.4% to 2% of world greenhouse gases emissions (Malmodin, 2010; Climate Group, 2008; WWF and HP, 2008). Due to the virtual nature of the medium, this impact is mostly hidden for both users and developers. Understanding and analyzing the environmental impact of a website is not an easy task. There is a growing amount of works looking at the life cycle of internet services using methodologies as Life Cycle Assessment (Bauman and Tillman, 2004). These are however few and usually based on a global scale, not applicable only for a single website. The impact of a site is distributed through multiple hardware networks with a global user base, making the individual impact difficult to pinpoint.

This article presents the development of a functional application for generating automatic life cycle assessments for web sites based on mashing-up information. This application has the aim of making the impact of websites visible, allowing the calculation of carbon footprint using existing analytics data and presenting it in an understandable and transparent way.

Most websites log extensive and detailed information about their usage. Our application combines this information with environmental impact data to automatically generate a calculation of the carbon impact of the website. This make possible for anyone to login and calculate the impact of their sites. The web tool was developed from scratch using Ruby on Rails[1] and released as a public beta version[2]. The development followed an iterative design approach with an agile style, releasing early working versions for getting

---

[1] More info at http://rubyonrails.org

[2] Available at http://greenalytics.org

feedback. The application, even if functional, is still under development. This article presents the design process, the ideas behind the project and what we have learned until now.

2. The climate impact of internet

The impact of internet in climate change is a topic of increasing interest. The most comprehensive studies have taken a top-down perspective, looking at the whole IT sector and making analysis of its total energy use and climate impact. Malmodin et al. (2010) have looked at the whole ICT sector, calculating a total of 630 Mtons CO_2 (eq) that accounts for 1.4% of the global emissions. This is in the same order of scale as other studies from the Climate Group (2008) and from WWF and HP (2008).

The two most important constituents of this impact are individual users (computers and screens) with 40% of the total impact, and servers with 23% (ITU, 2007). Malmodin et al. (2010) provide similar results, grouping data networks, servers and infrastructure to a total of 170 Mtons CO_2 (eq). More information about the servers energy use can be found in the reports from Koomey (2007a; 2007b).

Moving from the general numbers to more individual scales Taylor and Koomey (2008) and Weber et. al (2009) have mapped the total network and server impact with internet traffic information available, providing measurements of the energy required per data unit. This approach of allocating the total impact to individual traffic is the one we followed in our application.

The impact of internet is not only discussed in research. The media and the public also show interest in this topic. A couple of examples that were widely discussed are:

- Google search compared to a cup of tea: This was a story published at The Times (2009) using data from Alex Wissner-Gross. It approximated the carbon impact of a Google search to 7 grams (comparing two searches with the emissions of boiling a kettle). Google answered[3] and lowered the estimation to 0.2 grams per search. This story was polemic and widely discussed, see for instance by Mills and Koomey (2009).
- Avatars in Second Life as Brazilians: Nicholas Carr (2006) wrote an article arguing that a virtual player in Second Life[4] consumed as much electricity as an average Brazilian. The calculation has been put into question by the owners of Second Life, arguing that it is wrong by several multiples (Tomlinson, 2010).

These two examples show that there is a public interest in this kind of information. These two stories were widely published and generated a lot of conversations. But they also show that the results are controversial. The data contains a lot of uncertainties and when presenting results with exact numbers a lot of backfire was generated with corrections and arguments. This problem has been analyzed by Koomey (2002).

3. Development

Starting from the information available we started designing a tool to calculate the environmental impact of websites in an automatic way. Websites log extensive and detailed information about their usage, as this information is key for webmasters in the development and optimization of the sites. The most com-

[3] Google response at: http://googleblog.blogspot.com/2009/01/powering-google-search.html
[4] Second Life is an online virtual world accessible at: http://secondlife.com/

mon way is to use third-party statistics tools. For our application we chose to work with Google Analytics[5] (GA), the most used web statistics service worldwide (Metalytics, 2009) provided free by Google. This tool is widely used, ranging from big commercial services to personal websites. GA provide detailed information relevant to calculate the impact of the website: the number of visitors to the website, how much time did they spent, from which country, which pages they visited, etc. This information is available for developers using an API[6].

Our idea was to combine the information from GA with environmental impact data to automatically generate a carbon impact of the life cycle of the website, making possible for anyone with a GA account to login and see the impact of their sites. We developed a web tool from scratch using Ruby on Rails and released it as a public beta version.

Figure 1. Screenshot from the application

3.1 Boundaries

- Only greenhouse gases impact is calculated, providing the result in CO_2.
- Only electricity use in the use phase from the servers, infrastructure and visitors' computers is included.
- Hardware production, transport and disposal are not included in this first version.
- The production of the website (the impact related to when the web developer programmed the website) is not included.

[5] Google Analytics can be accessed at http://google.com/analytics
[6] API refers to Application Programming Interface. More info at http://code.google.com/apis/analytics/

3.2 Functional Unit

Greenalytics provides two different functional units: CO_2 per visit, but also the total CO_2 per month/site. The choice of two units was made for enabling both a regular efficiency value per user but also an overview of the total climate impact of the website.

The total amount is also translated into other units (for instance kilometers by car) to help understanding the scale of the emissions. This is performed using carbon.to[7] API.

3.3 Calculations

The impact of the website is divided in two parts: the impact of the website from the users and the impact from the server and internet infrastructure. These two are the biggest impacts of internet as presented in the previous section.

3.3.1 Users impact

The user impact of the website is the potential environmental impact generated by the users while browsing the site. In the current version this includes only to the electricity consumed by their hardware. The application makes use of GA detailed information about the users: their exact number, location and duration of visit. The total CO_2 is calculated by aggregating the impact per country, which is the total time visitors from that country spent in the site multiplied by how much electricity computer consume multiplied by the electricity factor (how much carbon dioxide is emitted per electricity unit) of that country:

$$\sum country\left(\sum time \times kW \times \frac{gCO_2}{kWh}\right)$$

Assumptions:

- We assume that the users' time is spent exclusively browsing the website, so all the electricity use for that time is allocated to the site.
- The type of computer and screen size is not possible to know. We assume a mix between laptops and desktops: 45 per cent using laptops consuming 19W and 55 per cent using desktops consuming 60W. Electricity consumption information is taken from IVF (2007) and the 45/55 distribution from IDC (2008).
- The carbon factors for the electricity are taken from Carma[8] through their API. These factors are calculated for a whole year and do not reflect the specific energy mix when the users visited the site. When the country is unknown, the global average is used.

3.3.2 Server and infrastructure

The server and infrastructure part is calculated using the total data traffic generated by the site and an approximation of the energy used by internet per data unit. GA doesn't provide a direct way of getting the

[7] See http://carbon.to
[8] http://carma.org

total data traffic, so an approximation is calculated by aggregating the total traffic per page (the size of the page per the number of visits it has), multiplying with an energy use per data unit factor (7 kWh/GB, taken from Taylor and Koomey (2008) and Weber et. al (2009)) and finally multiplying with an electricity factor for getting the CO_2 value:

$$\sum page(size \times pagevisits) \times \frac{kWh}{Gb} \times \frac{gCO_2}{kWh}$$

Assumptions:

- Right now a global carbon factor for electricity is used for the servers and infrastructure. This is because there is no straightforward way of getting the location of the server from GA, and also because the infrastructure cannot be pointed to a single country.
- The energy factor of 7 kWh/GB is generic and contains itself many assumptions. Weber et. al (2009) argues that this amount is dropping about 30% per year. This would mean that for 2010 the energy use per gigabyte would be closer to 3 or 4 kWh/GB than to the 2008 baseline. However we have not any fact about this decrease and the application preliminarily uses 7 kWh/GB.
- The data traffic is calculated by the size of the pages, and it doesn't take into account other factors such as caching. Dynamic content as streaming media or AJAX based interactions are not included.

4. Discussion

We have tested Greenalytics with several websites of different sizes. We will present shortly two cases that provide insight in how the results look like. The calculations are for a period of 30 days.

Figure 2. Case 1 (left) and Case 2 (right)

Case 1 is a small/medium website with around 1000 visitors per month. The analysis result was around 2 kg (1.96 gram per visitor). Of which visitors impact was 0.69 kg and the server impact 1.32 kg.
The users impact is lower than normal as most visitors come from Sweden which have a low electricity carbon factor of 37 grams CO_2/kWh.

Case 2 is a site with high monthly traffic around 700.000 visitors per month. The application gives a result of around 1195kg (1.73 gram per visitor). Of which the server accounts for 714kg and the users for 481kg. There was also a big percentage of visitors from Sweden that reduced the percentage of the user impact, but not as much as in Case 1.

4.1 Problems

The main identified problem is the carbon footprint calculation of the server and infrastructure. The actual number is still a rough estimate using a general factor for internet traffic's energy intensity. This approach goes against our ideal of real time and high granularity.

The lack of localization for the server using a global factor for translating electricity to carbon dioxide is the first weak point. We can see clearly that in the test examples. In case one the server is located in the US, where the electricity has a factor of 1345 grams CO_2 per kWh, more than double than the global average. In case two the server is situated in Sweden where the electricity factor is only 37 grams CO_2 per kWh[9]. As most of the checked sites during the development process where situated in US servers we could say that using a global average gave conservative numbers, but for the concrete cases where the server is situated in a country with low carbon intensity electricity as Sweden, the results can be out of scale.

The internet electricity intensity is the other weak point of the calculation. As presented before we use a calculation from Weber et. al (2009). The factor is calculated for 2008 and they argue that the impact is dropping about 30% per year. Rapid technological change adds uncertainties to this calculation. As we can see in figure 5, the use of older numbers for calculating actual servers impact can be misleading, as efficiency changes almost exponentially. For example the increase in virtualization and cloud computing happening now is probably reducing the energy use per computing unit (with cost as a driving force). Our suggestion for this problem is to move towards bottom-up data coming out from the servers and infrastructure instead of basing the calculations in a top-down approach.

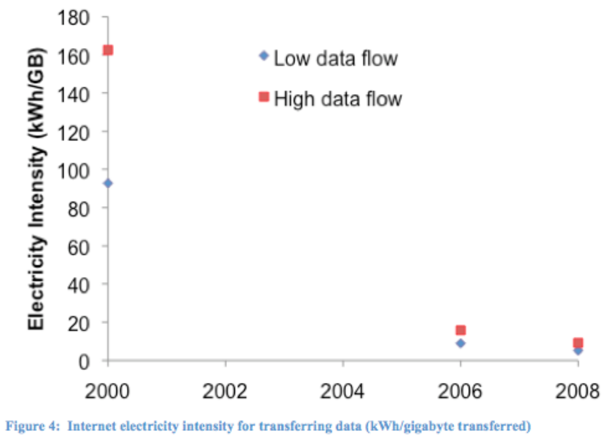

Figure 5. Internet electricity intensity from Weber et. al (2009)

[9] Factors taken from http://carma.org

4.2 Advantages

The data available in internet for many products and user behaviors is growing, and with an increasing use of concepts as APIs, Linked Data or Open Data[10], the creation of mash-ups for calculating environmental impact is realistic. In our case, the use of a mash-up approach had the following benefits:

- Real-time: The results are based on immediate data, the usage data from GA includes visits up to the same day. The carbon factors for the electricity are not static either but change with time as information is updated each year.
- Better data granularity: The use of a programmatic approach allows the gathering and use of bigger data sets as they do not have to be processed by hand. This provides the opportunity to use very detailed information. An example is the user impact calculation where the exact time in seconds is taken from each visit and then aggregated by country (See figure 4). This calculation is based on high granularity real use data instead of in assumptions about the usage.

Argentina 46.9 min 18.87 grams CO2. With a factor of 678.69
Australia 73.5 min 85.63 grams CO2. With a factor of 1965.95
Belgium 29.1 min 12.05 grams CO2. With a factor of 698.78
Brazil 0.2 min 0.01 grams CO2. With a factor of 110.74
Canada 5.2 min 1.45 grams CO2. With a factor of 471.06
Chile 2.2 min 1.2 grams CO2. With a factor of 917.15
China 6.2 min 7.02 grams CO2. With a factor of 1910.02
Costa Rica 12.6 min 0.12 grams CO2. With a factor of 15.66
Croatia 34.2 min 14.23 grams CO2. With a factor of 702.33
Czech Republic 3.9 min 3.78 grams CO2. With a factor of 1637.32
Denmark 43.6 min 21.34 grams CO2. With a factor of 825.76
Egypt 3.8 min 1.85 grams CO2. With a factor of 825.22
Finland 45.7 min 5.36 grams CO2. With a factor of 197.97
France 59.6 min 6.83 grams CO2. With a factor of 193.47
Germany 95.2 min 76.19 grams CO2. With a factor of 1350.79
Greece 43.9 min 45.74 grams CO2. With a factor of 1757.29
Hungary 0.8 min 0.31 grams CO2. With a factor of 657.24
India 58.4 min 61.36 grams CO2. With a factor of 1772.7
Indonesia 9.3 min 7.99 grams CO2. With a factor of 1458.33
Iran 4.4 min 2.1 grams CO2. With a factor of 799.01
Ireland 1.7 min 1.34 grams CO2. With a factor of 1308.59
Italy 16.8 min 9.41 grams CO2. With a factor of 947.15
Japan 0.2 min 0.08 grams CO2. With a factor of 802.57

Figure 4. Screenshot from Greenalytics user impact calculation

- Possible for users to analyze their own sites: Instead of a traditional approach where every site would be calculated individually, a mash-up approach allows any user to connect their existing data and analyze their sites without needing external expertise.

4.3 Future work

The application presented in this article is still an early prototype. The reception has been positive and it's ongoing heavy development for next iterations. Some of the features to be included and future work are:

- Localizing the server so a country specific electricity factor can be used. This could be done either by mapping the server's IP address to a geographical location or by having the users fill the information themselves (either the country or the hosting company). The part of the calculation corresponding to the infrastructure would still be global as the visitors usually come from many different countries.

[10] See http://www.opendatacommons.org/ and http://linkeddata.org/

- Separating the server impact from the infrastructure. This can provide greater granularity both for the calculation and for the visualization of the result. It can be extra important if the users can specify the hosting company. The electricity consumption could be tailored to the specific energy efficiency each hosting could provide.
- Including life cycle data from computer production both for the server and the users equipment. Those emissions are difficult to allocate to a particular usage, but they are important to include. In many cases a computer can have the same amount of embedded greenhouse gases emissions from its production as the total emissions from its use phase.
- Providing actions for the webmasters. Providing information can help increasing the awareness about the carbon emissions of websites. The second step is to use that information for changing behavior. Suggesting actions to decrease the emissions is the functionality that testers asked for the most. Examples could be: to provide links to organizations where the total emissions can be offset; to have best practice guidelines can have examples about how to optimize the server; having a ranking with hosting companies with good energy efficiency.

5. Conclusions

This article presented Greenalytics, an application for the analysis of the climate impact of internet at an individual web scale using a mash-up approach. Even if the results are still an approximation of the real environmental impact (focused on the carbon impact), they give a sense of scale and make it visible. Using a mash-up approach for making the analysis had the benefits of accessing real time data, with a granularity that can surpass traditional life cycle assessments. It also allows users to analyze their own data to check the impact of their website, building a bottom-up measure of the web energy use. As the concern about the energy use and environmental impact of internet is growing, tools as Greenalytics can be key in increasing the visibility and communicating the potential environmental impacts, and could be used as a base for actions towards minimizing them.

From the preliminary results that Greenalytics provide and the discussion points, we can draw three main conclusions:

- The impact of visiting an individual website is low, in the scale of 1 gram, but is not nonexistent. On the other hand, the aggregated impact of a website over time seems to be considerable.
- There is little up-to-date information about the electricity consumption of servers and internet infrastructure. Most of the information, including the one used by Greenalytics, is based on many assumptions and outdated data. We have to be careful when presenting such information and make sure the uncertainties of the calculations are transparent.
- Making dynamic environmental analysis in real time using a mash-up approach is possible and provides several advantages. In our application websites provide an early use-case as systems where usage information is already available an open through APIs.

The impact of IT is of concern and more knowledge is needed. We think that increasing the awareness and gathering data at the individual website level with applications as Greenalytics is a good start.

6. Literature

Bauman, H. Tillman, A.M. 2004. The Hitch Hiker's Guide to LCA. Studentlitteratur, Sweden.

Carr, N. 2006. Avatars consume as much electricity as Brazilians. Retrieved 06/06/2010 at: http://www.roughtype.com/archives/2006/12/avatars_consume.php

IDC (2008). Worldwide Quarterly PC Tracker. www.idc.com/getdoc.jsp?containerId=IDC_P20

ITU. 2007. ICTs and climate change. Retrieved 06/06/2010 at: http://www.itu.int/themes/climate/docs/report/02_ICTandClimateChange.html

IVF Industrial Research and Development Corporation. 2007. European Commission DG TREN Preparatory studies for Eco-design Requirements of EuPs. Personal computers (desktops and laptops) and Computer Monitors. Final Report (Task 1-8) http://extra.ivf.se/ecocomputer/downloads/Eup%20Lot%203%20Final%20Report%20070913%20published.pdf

Koomey, J. 2002. Sorry, wrong number: The use and misuse of numerical facts in analysis and media reporting of energy issues. Annual Review of Energy and the Environment. vol. 27,. pp. 119-158

Koomey, J.G. 2007a. Estimating total power consumption by servers in the U.S. and the World. Lawrence Berkeley National Laboratory and Stanford University.

Koomey, J.G. 2007b. Estimating regional power consumption by servers: a technical note. Lawrence Berkeley National Laboratory and Stanford University.

Malmodin, J. Moberg, Å. Lundén, D. Finnveden, G. Lövehagen, N. 2010. Greenhouse gas emissions and operational electricity use in the ICT and entertainment & media sectors. Journal of Industrial Ecology. Under publishing.

Metalytics. 2009. Retrieved 06/06/2010 at: http://dan.bravender.us/2009/12/21/Metalytics_-_The_Analytics_of_Analytics.html

Mills, E. Koomey. J. 2009. Tempest in a tea-kettle. Retrieved 06/06/2010 at: http://evanmills.lbl.gov/commentary/tempest.html

Mingay, S. 2007. Green IT: A new industry shock wave. Retrieved 06/06/2010 from Gartner: www.netdesign.dk/manedens-tema/telepresence/green-it-the-new-industry.pdf

Roth, K.W. McKeeney, K. 2007. Energy consumption by consumer electronics in U.S. Residences. TIAX LLC.

Taylor, C. Koomey, J. 2008. Estimating Energy Use and Greenhouse Gas Emissions of Internet Advertising. Working paper.

The Climate group. 2008. Smart 2020: enabling the low carbon economy in the information age. A report of the Global eSustainability Initiative (GeSI)

The Times. 2009. Revealed: the environmental impact of Google searches. Retrieved 06/06/2010 at: http://technology.timesonline.co.uk/tol/news/tech_and_web/article5489134.ece

Tomlinson, B. 2010. Greening through IT. Information Technology for Environmental Sustainability. MIT Press, Cambridge.

US EPA 2007. Report to Congress on server and data center energy efficiency. Public Law 109-431. EPA Energy Star program. www.energystar.gov

Weber, C.L. Koomey, J. Matthews, H.S. 2009. The energy and climate change impacts of different music delivery methods. Report to Microsoft Corporation and Intel Corporation.

WWF and HP. 2008. The potential global CO2 reductions from ICT use. WWF.