GREEN WEB ENGINEERING
A Set of Principles to Support the Development and Operation of “Green” Websites and their Utilization during a Website’s Life Cycle

Markus Dick, Stefan Naumann and Alexandra Held
University of Applied Sciences Trier, Umwelt-Campus Birkenfeld, Campusallee, 55768 Hoppstädten-Weiersbach, Germany

Keywords: Green Web Engineering, Green IT, Sustainable Development, Caching, Compression, Power Indicator.

Abstract: The power consumption of ICT and Internet is still increasing. To date, it is not clear if the energy savings through ICT overbalance the energy consumption by ICT, or not. In either case, it is suggestive to enforce the energy efficiency of the Web. In our paper, we present a set of 12 principles, which help e.g. to reduce the net load by caching or compressing. In order to classify our suggestions we group them by three main roles in Web Engineering. Additionally, we recommend using data centres which utilize “classic” Green IT.

1 INTRODUCTION

Global warming, greenhouse gas (GHG) effects, climate change, and sustainable development (SD) are key challenges of the 21st century (World Commission, 1991; Gore, 2006). Information and Communication Technology (ICT) takes a double-edged part within these challenges. On the one hand, ICT can optimize material flows and therefore reduce energy consumption (Hilty, 2005). On the other hand, ICT itself is consuming more and more energy: the energy consumption of IT and especially of the Web is still increasing (Erdmann et al. 2004). E.g. the estimated power consumption of data centres in the U.S. was increasing from 28 in 2000 to 61 billion kw/h in the year 2006 (U.S. EPA, 2007) and in the world from 58 billion kw/h in 2000 to 123 billion kw/h in 2005 (Koomey, 2007). Consequently, considering problems like climate change, reducing energy consumption of the Web is necessary.

Our paper discusses some possibilities to reduce especially the data transfer of the net from a technical point of view and suggests a classification through a conceptual framework for existing and prospective solutions. From a theoretical point of view, our approach can be classified into the new research field of "Sustainability Informatics" (Naumann, 2008a).

2 BACKGROUND

Up to now, several publications examine the relationship between the field of sustainability and ICT. They discuss the impact of IT on the environment (Göhring, 2004; Hilty, 2008) or consider the balance between energy savings and energy consumptions by ICT (Coroama & Hilty, 2009). Especially, to date it is not clear whether energy consumption by ICT is greater or smaller than energy savings by ICT, e.g. because of more efficient processes or simulations of scenarios.

Regarding "Green Web Engineering" the field of "Green IT" with suggestions of virtualization, green data centres etc. is covered amongst others by Hird (2008) and Velte (2008). Especially a tool to visualize, whether a website is hosted with renewable energies is described in Naumann (2008b). An approach, which tries to calculate the environmental impact of a website, is the CO2-Stats-Project (http://www.co2stats.com). However, since this approach is valuable, it seems not clear to the public how the measurement is calculated in detail.

Hilty (2008) gives some hints for software developers which can be partly applied to Web developers. Abenius (2009) introduces a model for green software which comprises aspects like speed, energy modelling and existing tools.
3 **GREEN WEB ENGINEERING AND THE LIFE CYCLE OF WEBSITES**

Firstly, we underlay our conceptual framework with a definition: “Green Web Engineering describes the art of developing, designing, maintaining, administering, and using a website in such a manner, that direct and indirect energy consumption within the complete life cycle of a website is reduced.”

The aim of our framework is to reinforce Green Web Engineering. In order to classify our suggestions we distinguish different roles within a website life cycle.

**Designers** are developers, web designers, or authors. Here, we understand *developers* as “classic” software developers, who program server-side applications, JavaScript, AJAX etc., or fulfill other more technical oriented tasks of website developing. *Web designers* are responsible for the screen design, decide on layout and colours, and develop screen flows. **Authors** maintain a website by generating content, deciding about the content structure, giving permissions to other authors etc. **Administrators** maintain a website from a technical point of view, allocate space and supply virtual machines, configure ports etc. Website administrators maintain a dedicated website and are responsible (in conjunction with the authors) for access permissions.

**Users** of a website are all people, who have access to the site and obtain information, communicate with other users etc.

4 **A SET OF 12 GREEN WEB ENGINEERING PRINCIPLES**

In this subsection we present 12 principles how to develop, maintain, and use a green website regarding our definition of “Green Web Engineering”. Our suggestions are structured through our classification of typical user roles within the website life cycle.

4.1 **Suggestions for Administrators**

4.1.1 **Configure HTTPs Caching Support**

Since the configuration of the user’s web browser cannot be affected by web designers or administrators, they have to focus on the server-side configuration aspects of caching. Where approx. 80% of the web users have their browsers configured for caching, however 20% always have an empty cache (Theurer, 2007). The inclusion of caching metadata by the web server will significantly decrease the amount of HTTP requests and HTTP responses. Caching in HTTP/1.1 is designed to reduce

- the need to send requests to servers (“expiration” mechanism) and
- the need to send full responses back to the clients (“validation” mechanism).

The validation mechanism does not reduce the amount of HTTP-requests but it reduces the payload of the HTTP responses that are sent back to the client and thus addresses network bandwidth reduction (Fielding et al., 1999:74).

In order to facilitate the expiration mechanism of HTTP servers, administrators can specify an Expires or Cache-Control header in their response. The Expires header as described with HTTP/1.0 (Berners-Lee, 1996:41) defines the absolute date after which the response is expected to be stale. One minor problem with the Expires header is that it uses explicit date and time strings and thus requires server and client clocks to be synchronized (Crocker, 1982:26; Souders, 2007:22). Beginning with HTTP/1.1 that limitation has been overcome with the Cache-Control header. It uses the max-age directive to define the seconds, which the requested resource may remain in the cache. To stay compatible with older HTTP clients that do not support the HTTP/1.1 protocol, one can define the Expires header alongside with the Cache-Control header. In that case the Cache-Control header overrides the Expires header.

The validation mechanism is utilized by HTTP clients that have a resource entry in their cache that has already expired. In that case the client may send a conditional HTTP request to the server. This conditional HTTP request looks exactly like a normal HTTP request but in addition it carries a so called “validator” that may be used by the server to decide whether the resource requested by the client is still up to date or needs to be refreshed. In case that it needs a refresh the new data is sent to the client, otherwise the server responds with the HTTP status code "304 Not modified". There are two validators that may be used: Last-Modified dates or Entity Tag cache validators. In case of the Last-Modified date a cache entry is considered to be valid if the requested resource has not been modified since the given Last-Modified date. An Entity Tag is a unique identifier for a specific version (entity) of a resource (Fielding et al., 1999:85). The calculation of the Entity Tags needs a refresh the new data is sent to the client, otherwise the server responds with the HTTP status code "304 Not modified". There are two validators that may be used: Last-Modified dates or Entity Tag cache validators.
cache-conditional requests if provided by the server. In addition, clients should also apply a Last-Modified date if one was set (Fielding et al., 1999:86). In order to reduce the total amount of HTTP requests or HTTP payload sizes, we suggest configuring the client cache support properly. This means:

1. Setting far future expiration dates and cache-control headers for resources that infrequently change
2. Setting Last-Modified headers and Entity Tags for all resources that do not need recalculation on subsequent requests (mainly static content)

A simple example configuration fragment for the popular Apache web server may look like this:

```apache
ExpiresActive On
<FilesMatch "\.\.(html|jpg|png|js|css)$">
  ExpiresDefault "access plus 355 days"
  FileETag MTime Size
</FilesMatch>
```

The Apache configuration directive ExpiresDefault handles both, the generation of an Expires header and the generation of a Cache-Control header for the given resource types.

### 4.1.2 Use Compression

Today, many modern web browsers support some kind of compression. Compression reduces not only response size and thus transfer time, but also power consumption as a result of the smaller size and shorter transfer times.

Web browsers usually support the GZIP compression format or DEFLATE. Both formats are especially named in the HTTP/1.1 specification. The Accept-Encoding header is used by web browsers to indicate which content encodings they support. A web server may compress the content using one of the compression methods listed by the browser and must notify the browser within the Content-Encoding response header which compression method is used (Fielding et al., 1999).

Compression is not as simple as it seems, because there are some older browser versions that claim that they support compression but actually do not, because of incompatibilities or bugs. On the other hand, more than 95% (ADTECH, 2008) of all installed and used browsers in Europe are known to support GZIP compression. Therefore, regarding our goal of “Green Web Engineering” it is reasonable to enable compression on the server side.

However, not all content types are suitable for compression, e.g. compressed image file formats, compressed music and video files or PDF documents. Compressing these file types is sometimes even counterproductive. Hence, compression should be used for files that are well compressible like text based files. With a simple example website that was compressed via GZIP compression with an Apache web server, we achieved traffic savings as shown in table 1. The average saving for the whole example site is approx. 60% (assuming PNG-images are not compressed).

<table>
<thead>
<tr>
<th>Example content</th>
<th>Total size (KB)</th>
<th>Compressed size (KB)</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>index.html</td>
<td>5.45</td>
<td>2.44</td>
<td>55.3%</td>
</tr>
<tr>
<td>style.css</td>
<td>2.73</td>
<td>0.68</td>
<td>75.1%</td>
</tr>
<tr>
<td>prototype.js</td>
<td>126.00</td>
<td>29.51</td>
<td>76.6%</td>
</tr>
<tr>
<td>ida-logo.png</td>
<td>24.80</td>
<td>24.86</td>
<td>-0.2%</td>
</tr>
<tr>
<td>ucb-logo.png</td>
<td>9.27</td>
<td>9.28</td>
<td>-0.1%</td>
</tr>
</tbody>
</table>

Here we see a simple example configuration fragment for the Apache web server using the compression module for text based document types:

```apache
AddOutputFilterByType DEFLATE text/html
AddOutputFilterByType DEFLATE text/css
AddOutputFilterByType DEFLATE application/javascript
```

### 4.1.3 Apply “Green IT” Concepts

So far, several web hosts exist that offer web hosting with renewable energy. Additionally, administrators can apply the newest techniques regarding Green IT like virtualization strategies.

### 4.2 Suggestions for Developers, Designers and Authors

#### 4.2.1 Support Caching on the Server Side

As described earlier in section 4.1.1 caching capabilities of HTTP are designed to reduce the amount of HTTP requests and transferred data volume significantly. Up to now we examined only web server configurations that handle static content. Content that is dynamically generated by scripts or programs on the server side is usually not treated to be cacheable and therefore not tagged with caching metadata by default.

However, there may be cases where dynamic content changes less frequently than once for each request. Therefore caching it may be possible without the risk of presenting obsolete data to visitors. In our opinion the following tasks are important:
4.2.2 Minimize JavaScript

JavaScript is used widely on websites to enhance interactivity. JavaScript code fragments are often embedded directly in HTML or in specific JavaScript files. The former increases the response size, whereas the latter increases the amount of HTTP requests especially when more than one JavaScript file is used in a website.

To minimize HTML content size we suggest externalizing any JavaScript code fragments embedded in HTML into special JavaScript files. On the one hand, this results in an extra HTTP file request, but on the other hand it

1. enables browsers to cache these files and to provide them in subsequent requests without calling the origin server
2. reduces the HTML download size especially when the HTML content can not be cached because it must be recalculated for each request.

If JavaScript code resides in dedicated files, we recommend defining distinct Expires header rules in web server configurations, so that these files expire further in the future than the HTML content files do (assuming that HTML content changes more frequently than the technical framework).

If more than one JavaScript file is used in a website, it may save some HTTP requests if these files are merged into one file. Despite increasing the JavaScript file size, the number of transferred bytes is much smaller, because only one HTTP request must be sent and probably revalidated when users are pressing the reload button in their web browser.

The JavaScript file size can be significantly minimized by removing all unnecessary whitespace characters or comments. It can be further reduced by removing all functions, methods or code blocks that are not needed within a particular website. This is true for websites that use only a few functions or methods of huge JavaScript libraries.

Another technique is obfuscation. Most available tools first apply the minimization technique and then shrink the code by substituting variable, method or function identifiers and also method and function parameter names with shorter ones. Self defined API functions referenced within HTML code must remain untouched, of course.

Table 2: Minimization vs. obfuscation vs. compression.

<table>
<thead>
<tr>
<th>Prototype-1.6.0.3.js</th>
<th>Uncompressed (KB)</th>
<th>GZIP compressed (KB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total size</td>
<td>126.70</td>
<td>28.49</td>
</tr>
<tr>
<td>Minimized size</td>
<td>93.09</td>
<td>22.71</td>
</tr>
<tr>
<td>Obfuscated size</td>
<td>87.66</td>
<td>24.94</td>
</tr>
</tbody>
</table>

Table 2 compares the effectiveness of the minimization tool JSMIn (http://javascript.crockford.com/) against that of the obfuscation tool ShrinkSafe (known from the popular Dojo Toolkit...
http://www.dojotoolkit.org/) while minimizing the well known JavaScript library Prototype (http://www.prototypejs.org/). As one may notice, obfuscation is more effective than minimization in the case of special tools used here. This results from the fact that minimization does not substitute any function identifiers with shorter names as the default obfuscation configuration does. In contrast to GZIP compression the overall savings of minimization or obfuscation are less effective (and in that special case of obfuscation even counterproductive, because there are fewer repetitions left in the JavaScript code that may possibly lead to better compression results). Further comparisons with JavaScript libraries lead to similar results (Dojo Toolkit, script.aculo.us).

Against this background we recommend to minimize JavaScript with a tool like JSMin and to compress the HTTP response with GZIP to achieve lower data transfer volumes. We recommend using obfuscation in scenarios where compression can not be used (e.g. if HTTP clients are not fully compatible with compression algorithms).

Additionally, we recommend minimizing or obfuscating the files before they are deployed on the server instead of minimizing or obfuscating them dynamically for each HTTP request to save processor time and thus power consumption.

### 4.2.3 Minimize and Optimize CSS

Minimizing and optimizing CSS files offers many possibilities to reduce HTTP requests. Savings from 40% to 60% are possible, if HTML or JavaScript operation is implemented with CSS (King, 2008: 177). Through a layout, which is designed with CSS, it is possible to save 25% to 50% of HTML filesize (King, 2008: 180).

Table 3 shows the filesize of an exemplary CSS stylesheet after optimization and minimization with different strengths.

<table>
<thead>
<tr>
<th></th>
<th>Filesize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without minimization</td>
<td>2,812 kb</td>
</tr>
<tr>
<td>With standard minimization</td>
<td>2,181 kb</td>
</tr>
<tr>
<td>With high minimization</td>
<td>2,040 kb</td>
</tr>
<tr>
<td>With highest minimization</td>
<td>2,021 kb</td>
</tr>
</tbody>
</table>

These results were obtained with CSSTidy (http://cssstidy.sourceforge.net/). This tool uses some techniques to optimize CSS code in different ways and strengths. In this case, the standard, high and highest minimizations were tested.

Standard minimization means that the compression goal is to reach a balance between size and readability. With this minimization, the file is 631 kb smaller and the compression ratio is 22.4%. One technique the tool uses are abbreviations, for example with colour definitions. The long definition `{color:#ffcc00;}` can be written as `{color:#fc0;}`.

The second minimization is “higher”; the readability should be moderate and the size small. The example file was reduced with this compression by 27.5%. In addition to the other techniques unnecessary whitespaces, strings and blank lines are deleted. This may constrain the readability but reduce filesize.

The third minimization is the “highest”. Its main goal is to achieve the smallest filesize. The readability is not considered. The compression ratio is the highest: 28.1%. In this case, additionally word-wraps are deleted. The whole code therefore is in one line.

There are further techniques, which optimize CSS. Examples are to replace inline styles with type selectors, to remove unnecessary code or to use descendant or contextual selectors instead of inline classes. CSS provides grouping multiple selectors with the same declaration and grouping multiple declarations with the same selector in one rule set. Duplicate element declarations can be avoided through the use of inheritance. Some separated property declarations can be combined in one single declaration. Margin, border, font, background, list-style and outline are special shorthand notations which combine separate single properties. Names of classes or IDs should be short or abbreviated. Long names are only important for developers, not for users. On the other hand, short names should be meaningful for developers in the context, too.

The CSS declarations should generally be saved in an external file and then included in the website documents. This minimizes the filesize of the web document and the CSS declarations will be loaded once and not with every request of the website.

CSS can also be used to replace less efficient code. One example is the table definition in HTML, which can be replaced by CSS. Another point is the usage of embedded JavaScript, as 84.4% of web pages use it. These methods can often be replaced with more efficient CSS code. Examples are dropdown menus and the rollover effect (King, 2008: 177).
4.2.4 Optimize Graphical Elements and Logos

54% of an average website is graphics (King, 2008: 157). Because of this, optimizing images is a promising way to minimize HTTP responses.

A logo, unlike a picture is a graphical object, which symbolizes mostly a company or an organization. The logo is normally included many times in a website. Thus, optimizing the logo can be more efficient than optimizing a picture, which is mostly shown once on a website.

Typically a logo is not a photograph or chart and in comparison to these smaller, so it can be saved as a GIF or PNG file. PNG has a better compression algorithm and especially flat-colour images are 10% to 30% smaller than in GIF (King, 2008:169).

There are some possibilities to optimize a logo. The most efficient ones are: changing the colour mode from RGB to an indexed palette, compressing, saving with special web settings and replacing graphical text with CSS formatted text.

The following example shows how a logo can be optimized. The logo consists of a graphical object, a background with colour gradient and graphical text (see Figure 1). The first step is to reduce the colours from RGB to an indexed palette. In this special case the colours can be reduced to six. Table 4 shows the filesize after optimization.

Table 4: Comparison of PNG fileizes after optimization.

<table>
<thead>
<tr>
<th>Original logo with RGB colours</th>
<th>16,959 bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logo with 16 colours</td>
<td>2,956 bytes</td>
</tr>
<tr>
<td>Logo with 6 colours</td>
<td>1,984 bytes</td>
</tr>
<tr>
<td>Logo with 4 colours without text</td>
<td>604 bytes</td>
</tr>
</tbody>
</table>

The final logo with CSS formatted text is shown in Figure 2. Looking at this example the filesize can be reduced from 16,959 bytes to 604 bytes (96% reduction).

![Software Development Platform Umwelt-Campus Birkenfeld](image1.png)

Figure 2: Final logo with 4 colours and CSS font.

Single-pixel, plain-coloured images, e.g. GIFs, are often used to achieve a certain stretching distance. These images can be replaced with CSS-code, where spacing cells have the same function as the images. This causes the layout to be the same, but through the usage of CSS the download of the images is not necessary anymore.

Images, which are lying relatively close to each other, can be combined into one picture. Additionally, links on the single pictures can be set with an image-map (King, 2008:165). Another possibility to combine images is to use CSS sprites. Sprites can be very effective, if many small images like icons are on a website. The required part of the whole image will be shown with the background-position rule at the designated position on the website (Souders, 2007).

4.2.5 Optimize Photographs

The filesize of a photograph is mostly greater than that of a logo. Consequently, there are more options to reduce the filesize.

Generally spoken, images should be stored with special web settings and the dimensions should be equal to the desired dimensions on the website.

A special technique to reduce the filesize is blurring the background. The blur technique can be used for images, where some parts are not as important as others (King, 2008:169).

![Software Development Platform Umwelt-Campus Birkenfeld](image2.png)

Figure 3: File with 100% quality.

The next step is the transformation from graphical text to CSS text. First, the graphical text will be deleted from the image. Afterwards, the text will be included in the image with CSS code, which defines a layer that has the logo as a background image. The dimensions are equal to the image dimensions. The CSS code for the example is as follows:

```css
logo { margin: 0px; width: 370px; height: 64px; padding: 10px 70px; font:bold 19px Arial, sans-serif; background-image:url(logo.png); background-repeat:no-repeat; }
```

Figure 3 shows the image with 100% JPEG quality and no blur, the filesize is 126,246 bytes. After blurring the background with strength 8, the filesize is
only 18,162 bytes; this image is shown in figure 4. It shows that the person in the foreground is in a satisfying quality whereas the less important background is blurred. But still, the main image information is conveyed.

Figure 4: File with 80% quality and blurred background.

4.2.6 Optimize Videos and Animations

Multimedia files like videos account for 98.6% of the transferred bytes in the Internet (King, 2008:159). Hence, minimizing multimedia files can reduce traffic. Basic techniques are editing with special web settings, using special codecs, reducing the frame rate and using professional equipment.

Other types of multimedia files are Flash animations. These can also be reduced and optimized. Here, non-optimized images and too many animation frames are the problems. The pictures should be optimized with image processing programs before they are imported in Flash. Animations should have a reduced frame rate (King, 2008:170).

4.3 Suggestions for Users

4.3.1 Configure the Web Browser

The former sections focused on the web designers’ and administrators’ options to reduce network traffic and thus power consumption. As was mentioned, administrators have no possibility to influence the browser configuration directly, like e.g. forcing the browser to cache delivered content according to the corresponding HTTP response header fields. Web users can also contribute a large part to the reduction of network traffic and power consumption by configuring the caching capabilities of their web browsers in a way that far future expiry dates can take effect, or by installing web browsers that are fully compatible with GZIP based HTTP compression. Furthermore, there is a lot of advertisement on web pages today that causes additional, but avoidable network traffic.

Thus, we recommend that one should
- configure large caches in Web browsers
- not clear caches during browser shutdown
- use browsers supporting GZIP compression
- use advertisement blockers to block advertisement images and Flash

4.3.2 Visualize Green Power Status of a Website

In order to assess whether a website is hosted with renewable energies, tools like the “Power Indicator” (Naumann, 2008b) can help users by visualizing the “green power state” of a website.

4.3.3 Apply “Green IT” on the Client Side

Of course, users of a website also have the possibility to obtain renewable energies within their companies or households. It is also recommended to utilize the available energy management options of client computers.

5 SUMMARY AND OUTLOOK

The power consumption of ICT and the Internet is still increasing. However, currently it is not clear whether energy savings through ICT overbalance its energy consumption, or not. In either case it is rational to enhance the energy efficiency of the Web.

In our paper we presented a set of 12 principles, which help e.g. to reduce the net load through caching or compressing. In order to classify our suggestions we grouped them by three main roles in Web Engineering: developers, administrators, and users of websites. Additionally, we recommend using data centres, which use renewable energies and follow the principles of classic Green IT.

Further evaluations have to be done that compare the overall traffic and processing time savings of a website if our principles are applied in different combinations or not. Therefore our current studies are focussing on measurement environments and approximations that allow us to draw conclusions on the resulting power savings from a local and probably also global point of view.

Our next steps are to detail and to broaden our model. More details can help to find solutions for more special problems regarding energy efficiency. A generic process model for “Green Web Engineering” is also reasonable and can be integrated in standardized models of Software Engineering. The broadening comprises e.g. aspects like barrier-free
techniques, since the idea of sustainable development contains also aspects of sociability and free access to common goods.

Additionally, it is worth to examine and evaluate our framework in detail and to suggest a ranking of “best practices”.

ACKNOWLEDGEMENTS

This paper evolved from the research project “Green Software Engineering” funded by the German Federal Ministry of Education and Research.

REFERENCES


