

Based on Web Application Front-end performance optimization

Peipei Ling

School of computer and science technology
Shanghai University
Shanghai, China
radeon_ling@hotmail.com

Abstract—With the development of the Internet, Web applications have become preferred platform of most systems. The whole website has been covered by different regions and more and more pictures, very slow to be opened. At this moment, the website should be optimized, and in all optimizations, front-end optimization is the first step.

Keywords- Front-end, Optimization, Performance, Website, Pre-loading, Delay loading, Web application

I. INTRODUCTION

With the development of the Internet, Web application is an important part of life. According to the CNNIC's statistics, the websites has increased more than 3 million by June 30th 2009. The number of 2009 is 6% than 2008. Throughout those websites performance, many Web is loaded very slowly so as to Access denied. Steve Souder, who is a performance optimization expert in Google, considered that it is only 10% to 20% of the end user response time spent on downloading HTML document. The remaining 80% to 90% of the time is spent in the download page all component. In other words, 80 to 90 percent of users waiting time are from the front page loading of the website. How to effectively improve your response speed is one of the biggest considerations of all in current network conditions, hardware and software. [1] Chen Zhipeng proposed based on CDN approach to web service optimization in his article. His methods could make website and web service increase performance. And Pankaj Solanki considered the cache is also optimized for website in his thesis. But those articles didn't have an overall performance of front-end in the web.

So the thesis will put forward a series of optimization. For instance, transmission optimization, quantity of transmission, Brower cache, pre-load & delay loading, optimization on client.

II. TRANSMISSION OPTIMIZATION

There are 4 items for transmission optimization, including transmission speed, transmission quantity, browser cache, and pre-loading and delayed loading. These 4 items are related to transmission. If visitors open a page of the website, the page should be transmitted from server to client and the time needed is determined by the transmission. The faster the transmission is, the faster the page is opened. As for the transmission speed, how to get it optimized?

The first is the problem of CDN. In case of development of an enterprise website, this problem may not be considered. Because it is just a system used inside enterprises, B/S

system won't be published to the internet. Even if it is published to the internet, the number of users is limited, and also the performance requirement to the system is not very high and strict. But in case of an internet system, this kind of problem should be considered.

The second is the problem of byte optimization. It is in the purpose of minimization of the file to be transmitted as possible, such as pictures, words and etc., but with a premise of ensuring the effect unchanged on web and never reducing users' experience.

A. CDN

In China, there is a problem between network operators, but possible to be changed with the realization of the unification of three nets. If Server A is located in the computer room of Telecom, its client who leases the line of China Netcom will feel slower in visiting the Server A, comparing with visiting the server in the computer room of China Netcom, and then the solution of CDN is supposed to be taken. It means that such a server should be deployed in the computer room of network operator or corresponding servers in different regions. For example, some clients located in Beijing can access the local servers via special policies, or the clients in Shanghai can access Shanghai servers automatically. [2] Thus, performance is promoted. In fact, in this way, server is not optimized in performance, but the transportation between client and server correspondingly changed, the clients to access internet are provided with the shortest path to download the web resources needed to be downloaded, such as HTML, CSS, JS, GIF, JPEG and etc.

B. Byte optimization

CDN is suitable for the website with very huge PV. The byte optimization is suitable for some websites which have fewer PV.

The byte optimization is a method to remove all the meaningless symbols of HTML code such as carriage returns, space and etc. There are just 4 lines of HTML codes in the Microsoft website (Figure 1). It is not different from any other website. It is loaded very faster thought its server is in the USA.

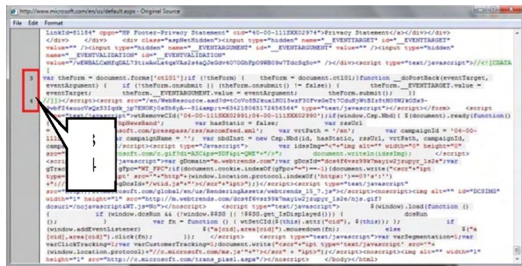


Figure 1. HTML codes of the Microsoft website

It can be seen from figure 1 that the Microsoft's site is optimized in byte. He removed all the spaces and carriage returns from their HTML codes, with the purpose to put the codes which can be placed in a line into a line as possible. The file's size becomes 75.8KB in 3 lines after optimization. (Figure 2)

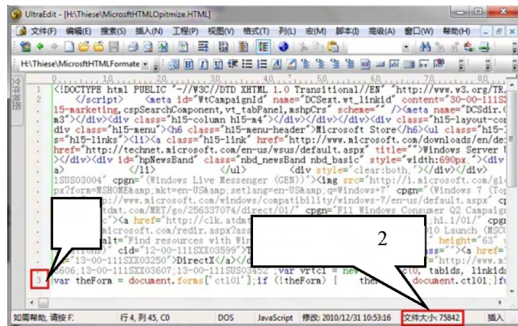


Figure 2. Optimized codes

It is very hard to maintain these optimized codes by developers, and that readability is very poor. But in development stage, it is possible to get a very strong readability form after formatting these codes. After codes formatted (the codes before optimization), codes' size is 303.2 KB (Figure 3), 200 k larger than the optimized codes' size, and no any change in between, except for meaningless spaces and carriage returns, which have been removed, so 200 KB removed. It has no impact on some small websites, but to some large websites, in case of PV increased, about 200 KB will be reduced when only a visitor opens this page of the site, so the throughput is very considerable to server. Therefore, it's very necessary and helpful to optimize byte in the HTML. In a similar case, JavaScript is also compressible, same method as in HTML byte compression. [3]

III. QUANTITY OF TRANSMISSION

A. Reduce the quantity of the request

When a website is published on the internet, the speed opened by visitors is very important. For developers, it is often very important in front-end optimization. If there is only HTML without any other elements like CSS, JavaScript, pictures, flash in the pages of all websites, except for the pure HTML, then the page size is very small, able to be loaded at once. The speed is so fast. But now, the websites' content is so rich including JavaScript, CSS, pictures and

flash, as well as the latest Silverlight and etc. When so rich content include in the page, how to reduce quantity of request? As a matter of fact, many quantity of page request are basically generated by pictures instead of JavaScript, CSS and etc. as they could be cached by cache mechanism. In case of hundreds or thousands small icons in a background management system, then when such a page is opened, it is needed to have the requests of dozens or even hundreds of pictures. At the time, visitors will feel so slow in page loading and performance be affected due to too many pictures loaded (The more request, the less performance). Furthermore, another characteristic is that browser gives subsequent request quantity a limitation, just like in IE, Firefox. The limit gives a maximum quantity, instead of establishment of interlinking depending upon the document number on the webpage, no matter how many. If there are 3 requests of maximum concurrent to the server, at the moment, the quantity of simultaneously downloading and requesting image resources are three links. At that time, the more the requests are, the more the amount of the reads in the queue will be. And that its performance will get down.

B. Appropriately allocate domain

In visiting the video sites such as youku.com, tudou.com and etc., it can be seen that the file e.g. pictures on the websites are saved on a server specially used for saving pictures, with video on specific video server, so it has been realized that different resources are saved on different servers, so as to reduce the request to individual domain. In the case of the same website or domain, in visiting, the quantity of queue request will be so many. So, its advantage is to reduce the access request of a same domain, so as to reach the purpose of acceleration.

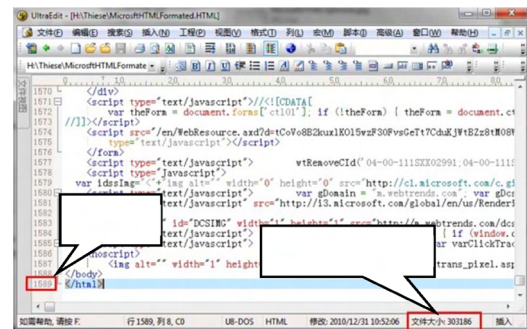


Figure 3. The codes before optimization

C. Avoid the error 404 & redirect

To developers, the error 404 is a neglected issue some time. Redirect is that a page jumps to another page, and then again jumps to another page according to some condition judgment. Sometimes if the logic is very complex, then it likely leads to endless loop so that the browser refreshes constantly. Doing like this will make the browser occupy a link for long, so that the time of request is prolonged and performance reduced. Because that the error 404 is caused by requesting a non-existent page, it itself occupies one link.

D. Use a picture file integrated with multi-image

When a website has many small icons, a good solution is to make those small icons or pictures merged into a picture file before loading its specific part of small icons by means of CSS and JavaScript, and then show the positions and contents of the small icons on the corresponding positions of the webpage where it is. If there are many small icons on the webpage, most time will be wasted in the linking and waiting.

IV. BROWER CACHE

A. Request in GET method facilitates cache s

There is just a Form on webpage. When submitting form, it is easy to cache using the GET method, because it can directly write content into the cache by a link. However, in case of using POST method, the form will be expired. There are no way for developers to cache directly, but it is available to cache by adopting some other ways, e.g. JavaScript, which is just to simulate browser cache, not make use of browser cache. So only the GET method can use browser cache. [7]

B. CSS, JavaScript adopt the external and separated method

When writing HTML code for developing, avoid CCS as possible, but put CCS independently into a file, loading it through the method of Link. Same is for JavaScript. The advantage is that it can load the head of HTML of CCS and JavaScript one-off when the webpage of a website is visited. When clicking a link and turning to another page, there is no need to load the same CCS and JavaScript, but read it directly from cache.

V. PRE-LOAD & DELAY LOADING

Pre-loaded and delayed loading is a networking application for JavaScript. Visitor feels quicker in speed in some respects. But, in terms of time, the time for loading all add-ins may be very long. They are the effects reachable by means of the pre-loaded and delayed loading.

A. Pre-loaded

The JavaScript and CSS displayed immediately to users or required to work immediately when the page is opened is considered as the pre-load.

B. Delayed loading

If there are some information to be displayed after the page is opened with some action from the visitors or in some time, the pre-loaded or delayed loading can be applied to display. There is another situation is to estimate the resources that visit may be visit.

VI. OPTIMIZATION ON CLIENT

A. CSS optimization

1) *Replace internal with external link*: It's better to replace writing the CSS in the HTML with external link way, but just suitable for PC client instead of mobile devices like mobile phone and etc. [4]

2) *Avoid IE filter and CSS expression*: It's better to avoid using IE filter and CSS expression. They make the CPU use ratio enhanced instantly, and are possible to make the CPU overloaded. The use of IE filter should be avoided, but depend upon condition, instead of lump together.

3) *Better to be put to the head in the HTML*: All the browsers such as IE, Firefox, and Chrome optimize CSS. For example, if the tag like `<style></style>` style sheet is put on the head of the HTML, It is very high-efficiency to render once all the elements of the HTML, when the CSS code is being loaded. But if the program finds the CSS again after all the contents in the webpage has been loaded to client, then it renders HTML recursively from cover to cover, making operation efficiency reduced greatly. So it's better to put the CSS to head tag in the HTML during the development phase, even if not on the head, it should be put in front of application element, not behind it, in order to speed up the loading of client-side.

4) *Avoid implementing the background with small picture*: Usually, a back-ground image is needed in making website, which is long in shape, cut into 1px, before realizing a complete background image by use of background-repeat attribute. But the method wants to increase the CPU usage. Only when the long strip used as back-ground makes the time of repeat less, the CPU resource consumption is less. In other words, the background should be reasonably enlarged in width, but not too big, otherwise, it affects to a certain extent the download and reading of documents. E.g. in case of about 100px, it can be used by the format of gif, jpeg, in short, picture should be as small as possible. Thus, even of the picture is repeated, for the width of 900px defined in general sites, only by 9 times of loops, the background can be achieved.

B. JavaScript Optimization

1) *Algorithm optimization (Especially in loop)*: Algorithm optimization is necessary in any programming language. JavaScript is used for interpretive execution, so it is not so high itself in performance, a large part of its efficiencies should be supported by CPU at client-side. In dealing with some loops by use of JavaScript, it will be found that performance is poor and speed low. Normally, this kind of problem is caused from loop. For example, if it is wanted to get a length of some DOM in HTML, it is necessary to loop according to its length.

For example: the efficiency like `int i = 0; i < obj.length; i++` will be slower, as all the attributes in DOM of HTML are achieved by use of method. So, applying the attribute of `obj.length` to get its length is just equal to visiting a method to get the length. In the loop of this section, if it is less than `obj.length` each time, program tends to calculate out the length of DOM elements each time, making the loop slower. Its performance can be improved if only the code of this section is slightly modified.

```
var objLength = obj.length;
```

```
int i=0; i < objLength; i++;
```

Thus, the length of the DOM object can be obtained by one-time calculation, so that the cycle performance of JavaScript can be greatly optimized.

2) *JavaScript doesn't support long integer*: Only integer is supported instead of long integer in JavaScript. Due to its use of the process methods of 32bit, it got restricted to a certain extent. It's necessary to consider the problem when handling huge number on client. It should be achieved by use of some flexible ways. One is to change the number to string; the other is to use AJAX to submit the data to be processed to server, and then result is put to front-end by the form of string. If the long-integer number must be set to a variant, then the variant will be cut off by JavaScript.

3) *There is no real multi-threading in the JavaScript*: JavaScript is not a computer language supporting multi-threading. The multi-threading can be achieved by simulation when it is needed.

4) *Asynchronous is the JavaScript not necessary in page initializing*: The advantage doing this is to ensure the number of files being loaded minimum when the page is initialized. When the page has been initialized before the remaining JavaScript codes is loaded on the background, so as to achieve the effect of optimizing performance.

C. HTML optimization

1) *Try not to use table for layout, especially for external layout*: In constructing a web, the "DIV+CSS" flow-style layout is superior to table structured layout pattern, because the performance of rendering is quite different between flow-style layout and structured HTML on client. The advantage of flow-style layout is that explorer shows data immediately once data is downloaded, but table works like stack. When reading data, it should finish reading all outmost data before showing complete information on explorer. In case that the table nested with table is used in the whole website, with a very big data size, this time, when user inputs address in URL, page will be blank, then shows user total contents at one blow after a while, without a process.

2) *Try best to use the flow-style tag such as div, p and etc.*

3) *Try best to use semantic tag*: Like ``, ``, there has logs of same type elements in HTML tag, such as `` and ` `, `<I></I>` and ``. In HTML, `` represents bold type, without any other meaning, ` ` not only represents bold type, but emphasis. In HTML, `<I></I>` is

italic, without any other meaning, but `` represents this part of contents has been referenced. Usually, in many papers, the emphasized uses bold type, the referenced, italic type. Here, the semantic tag is used. Another advantage of the semantic tag is to allow the disables surf on-line also. If there is no such semantic tag on webpage, it will not be recognized by readable browser, and then they get no way to know by listening if this part of content is emphasized or quoted from somebody.

4) *Reduce the number of DOM, reduce the use of ifram*: Reducing the number of DOM is to reduce the number of tags. It is required to use as little as possible HTML tags to achieve the same effect, not to nest too many HTML tags so that the browser's rendering and CPU consumption can be reduced.

5) *Consider the content in the part of asynchronous loading*: Asynchronous loading is generally achieved by use of AJAX. [5][6] Like the tree menu of explorer, in reading menu, the more the menus, the lower the performance. Because that in rendering the tree, all nodes should be finished reading before loaded onto page. If there are more than ten thousand or millions nodes in the menu, the performance will be seriously declined. So solution is to use asynchronous loading method. When an add symbol in the front of a node is clicked, the child nodes of the node will be loaded. Nowadays, AJAX is used widely in the web, so pay attention to several points as follows:

a) *Avoid concurrency*: At the same time when using AJAX to give out the first request, the second request is submitted. If there is no any limit given to the request, and the response time is longer than the second one, in that way, the result returned from the page is the result given out by the first request. Try better to use the GET method request.

b) *Use GET method for request as possible*: As previously mentioned, GET method is very beneficial to cache. But in the case that it is a must to submit many data on the page, only POST can be used, instead of GET. The reason is that the GET method is submitted by URL, and URL will be limited in length by browser..

REFERENCES

- [1] Nie Yinggao, "Analysis on Frontend Performance Optimization of Website Based on Page Speed," Library of Xian Ning University, 2007.
- [2] Wu Zhipeng, "Web service performance optimization," Xia Meng University, 2007.
- [3] Andrew B. King, Website optimization., vol. 2. O'Reilly, 2008.
- [4] Hyun Joo Lee, "Optimizing Web pages," The University of Texas at Austin Graduate School of Information, February 20, 2003..
- [5] Pankaj Solanki, "Web Caching and Response Time Optimization in Application with Web Browser," Department of Computer Science and Engineering Indian Institute of Technology, Bombay, April, 2009.