THE EXISTENCE PROOF SERVICE OF THE WEB PAGES
New Web Service to Get Grounds of the Existence of the Web Pages

Yuka Obu[1], Masahiro Miyata[1], Juyeon Choe[2], Katsumi Terakado[2] and Tatsuhiro Yonekura[3]

[1]*Graduate School of Science and Engineering, Ibaraki University, Hitachi, Ibaraki, Japan*

[2]*KCS Corporation, Mito, Ibaraki, Japan*

[3]*Department of Computer and Information Sciences, Faculty of Engineering, Ibaraki University, Hitachi, Ibaraki, Japan*

Abstract: Recently, more people exchange the information via the Web. However, because the information on the Web is electronic-base data, the information can be deleted or changed frequently. Therefore, it is difficult to prove to the third party that the content existed or it has not been changed. Hence, we developed the Web service called "The existence proof service of the Web pages" that creates the Web cache and adds the time stamp as the grounds that the cache has not been changed. With this service, the Web page can be proved to have existed in a certain URL at certain time. This can be used for verification of the alteration of the Web page.

1 INTRODUCTION

Recently, because of the wide u se of the Internet, more people exchange the information via the Web and they can see a lot of scenes that the Web technology is used to get information. However, because the information on the Web is electronic-base data, the information can be deleted or changed frequently. There are some cases that we can not find the same information on the same URL afterward. Therefore, it is difficult to prove to the third party that the content existed or it has not been changed. This is one of factors to lower the reliability of the Web information. It is necessary to preserve Web information beforehand to prove the existence of information.

We can find some Web services that can be used for this purpose. However, these Web services only make the Web caches. There are no grounds which prove that the cache has not been changed. Although some Web services create a PDF file from the Web page with the time stamp, peculiar information of the Web, like meta-tag or link, can not be preserved in this kind of cache.

Hence, we released the Web service called "The existence proof service of the Web pages" that creates the Web cache and adds the time stamp as the grounds that the cache has not been changed. It can be proven that the Web page existed on an URL at the certain time and that the content of cache has not been changed. Moreover, in this paper, the purpose of use of the Web service, user's attribute, and the degree of the interest had been analyzed.

2 RELATED WORKS

It is necessary to preserve information on the Web beforehand in order to prove the existence of the content at a certain time. There are some Web services that can be used for this purpose.

"toread" (sidefeed, Inc., 2008) makes the cache of the requested Web page, and sends it to registered e-mail address. The user can confirm the content of the Web page later. "Web Gyotaku" (Affility. Co.,Ltd., 2006) records the cache of the requested Web page on the server. Users can refer the cache on this service instead of the original Web page. "kwout" (HeartRails Inc., 2007) enables the user to clip a part of the Web page as an image and quote it. The clipped data is able to be contributed to "Flickr" (Yahoo! Inc., 2004) or "Tumblr" (Tumblr Inc., 2007). "Wayback Machine" (Internet Archive, 1996) enables user to retrieve and inspect the cache acquired automatically. The requested URL is retrieved and archived contents of the

Web page are displayed in the time series. These services only make the cache of the Web page simply, therefore they can not prove that the cache has not been changed after the creation, by using some evidences. Consequently, even if the deletion and the change of the Web page can be recognized, trust from the third party cannot be obtained.

"HTML2PDF.BIZ" (Asial Corporation, 2007) preserves the Web page as PDF file with the time stamp. When the URL is input, the appearance of the Web page is converted to a PDF file. The PDF file with the time stamp can be verified that it has not been falsified. However, because this service uses PDF file, the preserved information is limited only for the screen shot. Thus, the particular information in the Web page maintained in the HTML though not displayed when referring by the browser, e.g. meta data, link URL, and so on, can not be preserved.

To deal with issues mentioned above, the new service that can prove the existence of the Web page, namely, "The existence proof service of the Web page", has been developed and released.

3 THE EXISTENCE PROOF SERVICE OF THE WEB PAGES

3.1 The Time Stamp

These days, enormous quantity of data came to be exchanged on the Web, even if it is very important. Because the electronic document do not keep the signs of correction, and so on, and they do not deteriorate, it is difficult to trace the history. Hence, the mechanism that proves "When", "Who", and "What data" is important for the evidence of data securing.

In electronic data such as files, the time stamp can be the evidence that ensures the following two effects.

The Proof of the Existence. It can be proven that data existed certainly at the date when the time stamp was given.

The Proof of Completeness. It can be proven that the data has not been falsified since the data when the time stamp was given.

The electronic data with the time stamp is made at the time stamp authority, by uniting time information with the hash value of the data (Adams et al., 2001). In the verification of the data falsification, by comparing former data with the hash value of data with the time stamp, it can be confirmed that the data existed at the time indicated with the time stamp and is not falsified after that.

In Japan, Ministry of Internal Affairs and Communications (MIC) issued the guideline of time business (MIC, 2004a), (MIC, 2004b). And based on th guideline, the accreditation program established to approve if the services of a TA (Time Authority) and a TSA (Time Stamping Authority) meet the required criterions (JADAC, 2005).

3.2 Genaration of the Web Cache with the Time Stamp

The existence proof service of the Web pages (the existence proof service hereafter) had been developed and released (KCS Corp. and Yonekura lab. at Ibaraki University, 2008). This service proves the Web page existed in a certain URL at a certain time and the cache of the Web page has not been falsified after it was generated. The service makes the cache of the Figure 1 is a screenshot of using this service.

Figure 1: Screenshot of getting the cache from the existence proof service of the Web page.
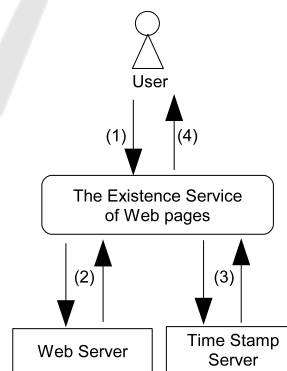

Figure 2: Overview of the existence proof service of the Web pages.

Figure 2 shows the outline of the existence proof service. The steps to use this service are as follow; **(1)** The user sends the URL of the Web page, which he or she wants to get the evidence of the existence, to this service. **(2)** Service acquires information on the Web page from received URL and makes cache. **(3)** The

573

service requests the time stamp server to add the time stamp to the created cache. **(4)** The service replies to the user cache with the time stamp.

The ways to use this service are **(1)** user accesses to the Web site of the existence proof service and inputs the URL of the Web page, **(2)** user uses the toolbar for a Web browser that the existence proof service offers. The toolbars are for Internet Explorer shown in figure 3 and for Firefox.

Figure 3: Screenshot of the the IE toolbar for the existence proof service.

Cache and the time stamp on the Web page are downloaded to the user with the ZIP file by using this service. Figure 4 shows the composition of the downloaded ZIP file. The downloaded ZIP file is decompressed, two files, cahes.zip and caches.zip.tst. HTML, CSS, and Image files are compressed into the caches.zip. Caches.zip.tst is the time stamp file for caches.zip. The HTML file decompressed from caches.zip is permitted to be confirmed. However, when new caches.zip is created again from the decompressed HTML files, it is considered that the files has been falsified by verification process described in 3.3. Therefore, it is necessary to preserve caches.zip with caches.zip.tst.
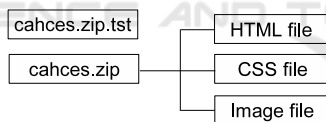

Figure 4: The contents of the downloaded ZIP file.

3.3 Verification

It is verified that the cache file downloaded from the existence proof is not falsified with a time stamp and a free verification tool that PFU Ltd. offers (PFU Limited, 2004). Figure 5 is a screenshot of the verification tool. First of all, the file paths of the time stamp and the cache are registered (Figure 5 ①). Next, the verification button is pushed for the verification execution (Figure 5 ②). Then the result of the verification is displayed (Figure 5 ③).

In the verification tool, the verification is achieved using the hash value contained in the cache and the time stamp. The hash value is calculated using the same function used when the time stamp was generated. If the hash values are the same, it is confirmed that the cache has not been falsified.

Figure 5: Screenshot of the verification execution.

4 THE ANALYSIS OF THE USERS

4.1 Users of the Existence Proof Service

The existence proof service was released to the public on January 21, 2008, in Japan (Nikkei Inc., 2008), (CyberAgent, Inc., 2008), (ITmedia Inc., 2008), (Rakuten, Inc., 2008). Page of the service till October 8, 2008 is 21895.

When the existence proof service toolbar is downloaded, the user answers the questionnaire. The number of valid response is 162 people, and the results are shown in Table 1, Table 2, and Table 3.

Table 1: Result of the questionnaire about age group.

Age	Number	Percentage (%)
10 - 19	28	17.3
20 - 29	60	37.0
30 - 39	39	24.1
40 - 49	25	15.4
50 - 59	9	5.6
60 - 69	0	0
70 -	1	0.6

Table 2: Result of the questionnaire about occupation.

Ocupation	Number	Percentage(%)
Employee	73	45.1
Manager	3	1.9
Part-timer	6	3.7
Self-emplyed	7	4.3
Public service	7	4.3
Student	44	27.2
Without occupation	6	3.7
Other	16	9.9

From Table 1 and Table 2, it is conjectured that students and company employees who are 10 to 49 years old have many interests in this service.

Moreover, about information source of this service, shown in Table 3, 45.7% is news site, 14.8% is someone's Web site, 13.0% is search engine, thus it is understood that the user who knew this service

Table 3: Result of the questionnaire about information source of this service.

Corcumstamce	Number	Percentage(%)
Newspaper	14	8.6
News site	74	45.7
Informed from friend	6	3.7
Search engine	21	13.0
Someone's Web site	24	14.8
Other	23	14.2

from Web site accounts for about 73%. About the usage condition of this service, the frequency of the use from this service's Web site and from the toolbar are 1124 times and 572 times respectively.

Concerning the target for the existence proof, portal site, news site, company's site, weblog site, bulletin boards, shopping site and auction site are targets of this service. It is thought that Web user's interest is high. Especially, about the existence proof by using the toolbar, there are a lot of usage for the site where the update frequency is high, such as the weblog, the news site, auctions, and the shopping sites.

For more practicable service and the expansion of the usage, it is necessary to construct the mechanism that enable to use this service on various sites directly.

4.2 Non Users of the Existence Proof Service

In order to research the potential user's needs for the service that acquires grounds of the existence of the Web page, the questionnaire survey was conducted to people who had not used the existence proof service. Investigation objects are 32 people in their twenties, and all members have the experience of using some Web services but have no technical knowledge.

First, the objects were asked why they use the Web services. The result is shown in Figure 4. The remarkable reasons were "quick", "convenient", and "easy". From this result, it is confirmed that people use the Web service because it is simple and easy to use even if they do not have any technical knowledge.

Secondly, the objects were asked what kind of Web service they use. The result is shown in Table 5. Though The usage of the movie sharing site is remarkable, the other Web services are used in almost the same extent, hence it is considered that each Web service is used on average.

The last question was about the needs of the existence proof service. Result is shown in Table 6. About 25% of the objects wanted to use the service, though about 70% answered that they would not use the service. From this result, the potential users do not think

Table 4: Result of question1 : "Why do you use the Web services?".

Answers	Number
Information is obtained quickly.	9
It is convenient	8
It is easy.	7
Desired information can be acquired.	4
Because of the use of the Weblog or SNS.	4
Information is new.	2
It is easy to obtain information.	1
It is interesting.	1
Killing time.	1
The wanted thing can be obtained cheaply.	1
Information all over the world is available.	1
Other	2

Table 5: Result of question2 : "What kind of the Web service do you use?".

Answers	Number
Weblog	17
SNS	16
Individual homepage	14
Bulletin board	11
Movie sharing site	22
Auction	7
Shopping site	13
News site	13
Other	1

that they have significant need to use the service to proof the existence of the Web page.

Table 6: Result of question3 : "Do you want to use the Existence proof service of the Web pages?".

Answers	Number	Percentage(%)
Yes, I do.	8	25
No, I don't.	19	60
I don't need to use.	2	6
I have no idea.	3	9

From the result above, it can be confirmed that the ordinary users do not have so much consideration about the risk when they use the Web service, and they also do not feel the necessity of our service very much. The objects who want to use the existing proof service mostly use SNS and movie sharing site a lot, and the objects who do not want to use the service mostly use the weblog and movie sharing site. And there were some opinions that it is both-

eresome to use the Web service while considering the risk and the way to use the existence proof service is too complicated. Consequently, the existence proof service is supposed to have the mechanism that is easily available on the SNS, the weblog, and the movie sharing site. For instance, the mechanism to acquire the grounds of the existence of the Web page that the service periodically checks the requestedWeb site automatically and provides the notification if there is a change is considered.

5 CONCLUSIONS

"The existence proof service on the Web pages" was released to the public. This service generates the cache of the Web page and adds the time stamp to the cache. By using the obtained data, it can be proven that theWeb page existed in a certain URL at a certain time and the cache has not been falsified. Moreover, from the result of the questionnaire and the service operation results, it is inferable that the users have high interest about this service.

In this service, the cache includes HTML, CSS, and image files. The cache can be preserved including information that cannot be confirmed by the displayed appearance on a browser, such as the link information, meta information, and so on. Therefore, this data can be used for the perpetuation of evidence for the misrepresentation of the link. Moreover, because the difference of the appearance can be confirmed by displaying the cache on various browsers, information that is more various than only the preservation of the screenshot can be maintained.

In the future, the reliability securing on the Web page, the compliance measures of the enterprise, and the user protection are supposed to be achieved, by developing the protocol of proof of sessions and API of this service.

Concretely, the existence proof service can be set up on various shopping sites, corporate sites, and so on, by developing API of present service, and it will be possible to contribute to the improvement of the reliability maintenance of the Web sites. For the enterprise that sells the product and service, it is expected that using this service positively becomes an appeal for the reliability of the site. Moreover, for users, they can devise acountermeasure to the problem that occurs while using the Web services.

In order to improve the reliability of the existence proof service, the establishment of the standpoint as the third-party institution by the cooperation of the municipality and the official organnization is one of the future tasks. And for the stability of the service,

consideration about the system performance and the load tolerance is necessary.

REFERENCES

Adams, C., Cain, P., Pinkas, D., and Zuccherato, R. (2001). Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). RFC 3161 (Proposed Standard).

Affility. Co.,Ltd. (2006). Web Gyotaku. http://megalodon.jp/.

Asial Corporation (2007). HTML2PDF.BIZ. http://www.html2pdf.biz/.

CyberAgent, Inc. (2008). Nikkan AmebaNews. http://news.ameba.jp/domestic/2008/01/10463.html.

HeartRails Inc. (2007). kwout. http://kwout.com/.

Internet Archive (1996). Wayback Machine. http://www.archive.org/web/web.php.

ITmedia Inc. (2008). ITmedia News. http://www.itmedia.co.jp/news/articles/0801/21/news041.html.

JADAC (2005). Accreditation program for time-stamping service. http://www.dekyo.or.jp/tb/english/index.html. Japan Data Communications Association.

KCS Corp. and Yonekura lab. at Ibaraki University (2008). The Existence proof service of the Web pages. http://www.existingproof.jp.

MIC (2004a). About the guideline of time business, in japanese. http://www.soumu.go.jp/s-news/2004/041105_3.html. Ministry of Internal Affairs and Communications.

MIC (2004b). The guideline of time business, in japanese. http://www.soumu.go.jp/s-news/2004/pdf/041105_3_01.pdf. Ministry of Internal Affairs and Communications.

Nikkei Inc. (2008). NIKKEI.net Interactive. http://www.nikkei.co.jp/.

PFU Limited (2004). PFU TimeStamp Service download client tool for verification. http://www.pfu.fujitsu.com/tsa/download/client.html.

Rakuten, Inc. (2008). infoseek News. http://news.www.infoseek.co.jp/.

sidefeed, Inc. (2008). toread. http://toread.cc/.

Tumblr Inc. (2007). Tumblr. http://www.tumblr.com/.

Yahoo! Inc. (2004). Flickr. http://www.flickr.com/.