Saving safely

Secure documentation of your web findings

Albrecht Ude

EIJC / Dataharvest – Mechelen – 03.06.2023, 09:30 - 10:45 h, C2.05

What should be saved and proven at all?

When you save content from the WWW (web pages), you usually want to prove that a **certain content** was available on the web on a **certain date** at a **certain address** (or was *not available at that time*).

So you want to be able to prove

- address

- Timestamp (date and time) and

- Content of a certain web page

You Try to Save a „Web Page" - What is that?

What is called a "web page" is

- created when a web server gets a request
- sends data to a web browser
- via an Internet connection
- which the browser then displays on the screen.

There are a few things to keep in mind.

Most web pages are conglomerates of files which composition changes.

Problem Situation

1) Servers do not always deliver the same data (even in response to identical requests):
- displayed advertisements are exchanged,
- servers react to the data transmitted by the browser with the request:
  - cookies
  - referrer
  - IP address
  - browser type and version   etc.

1 b) Many servers react above all to which end device they are requested from: desktop, tablet, mobile device

2) The connection from the server to the browser can be manipulated (government censorship, intervention of the Internet Service Provider).

Problem Situation

3) You have to limit yourself to saving what your browser displays.

In doing so, the **files** are usually **modifed by the saving** process.

It is clear that any digitally stored document can be manipulated.

What do you want to save?

If you are interested in documenting the **specific combination** of a web page e.g. with an advertisement,

    - make screenshots
        - of the entire screen „Print"-Button on keyboard
        - within the browser

Think of taking a photograph of your entire device.

    in addition to the following:

Best Way: Saving a Web Page in External Web Archives

WayBackMachine
https://archive.org/

Archive today
https://archive.is/

Other Archives

http://timetravel.mementoweb.org/about/

https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives

Saving a Web Page in External Web Archives

Problem
Archiving social media pages (needs account normally)

https://www.bellingcat.com/resources/how-tos/2018/02/22/archive-open-source-materials/

Saving a Web Page with Other External Tools (1)

Search Engine Caches

e.g.
https://pagecached.com/

Saving a Web Page with Other External Tools (2)

Web Archiving Tools (sometimes $)

e.g.
https://www.pagefreezer.com/website-archiving/

https://www.stillio.com/how-to-archive-website

https://conifer.rhizome.org/

Saving a Web Page with Other External Tools (3)

Online PDF-Converters

e.g.
https://tools.pdf24.org/en/webpage-to-pdf

https://webtopdf.com/

https://www.freeconvert.com/webpage-to-pdf

Saving a Web Page locally

Use as much browsers as you use!

- Save option "Web page, complete"
(works fine, saves all pictures & files, modifies HTML!)

- Save option "Web Archive (One File)"
(works fine, modifies HTML!)

- Save option "Web page, HTML only"
(saves rendered HTML)

Saving a Web Page locally

Printout on paper

- has a high "suggestive power" as evidence
- may differ significantly from the screen representation
(different format, possibly a different formatting file (CSS)).

Saving a Web Page locally

Human witnesses

- ask colleagues to save the web pages in question and to be prepared to testify if necessary.

- probably easier to impress a judge with witnesses than to give him detailed technical knowledge.

Thank you for your patience!

Albrecht Ude

albrecht@ude.de

www.ude.de