

Ranke.2 Teaching Aid

The Interface of the Wayback Machine



Compiled by Jordan Ricker, revised by Stefania Scagliola with
assistance from Valérie Schafer
Last modified: 17 December 2020



This teaching aid is meant to guide you through the interface and the various features offered by the Waybackmachine to retrieve archived websites and webpages. It is part of the lesson ‘Websites as historical sources’, offered on the Ranke.2 teaching platform for digital source criticism. See: <https://ranke2.uni.lu>

We hope it can encourage you to use the WBM.

The present Ranke2 team: Stefania Scagliola, Jordan Ricker, Ilaria Comes and Emmanuel Decker

The **Internet Archive** (IA) is the name of a non-profit organization whose goal is to provide free access to anyone in the world to a wide range of books, movies, audio files, and web pages, among others.

The **Wayback Machine** (WBM) is a searchable repository of snapshots of past web pages that have been collected by the IA and by other actors* that have a policy of preserving websites.

The Homepage of the Wayback Machine

URL: <https://archive.org/web/>

The screenshot shows the homepage of the Wayback Machine. At the top, there's a navigation bar with the "INTERNET ARCHIVE" logo, followed by links for ABOUT, BLOG, PROJECTS, HELP, DONATE, CONTACT, JOBS, VOLUNTEER, and PEOPLE. On the right side of the header are "SIGN UP | LOG IN" and "UPLOAD" buttons. Below the header, the "INTERNET ARCHIVE WayBackMachine" logo is displayed, with the tagline "Explore more than 486 billion web pages saved over time". A search bar with the placeholder "http://" and a "BROWSE HISTORY" button are located below the logo. A "DONATE" button is also present. The main content area features a grid of thumbnail images representing different web archive collections. Below this, there are three sections: "Tools" (with links to Wayback Machine Availability API, WordPress Broken Link Checker, and 404 Handler for Webmasters), "Subscription Service" (describing Archive-It and linking to its website), and "Save Page Now" (with a "http://" input field and a "SAVE PAGE" button). At the bottom, there are links for FAQ, Contact Us, and Terms of Service (Dec 31, 2014).

*How the different people, organizations, and bots contribute snapshots to the WM is a complex process.

For more detailed information, see; '[The Internet Archive and the socio-technical construction of historical facts](#)' (Ben-David and Amram, 2018)

Various other media branches of the Internet Archive

URL: <https://archive.org/web/>

This is the search bar to access the web archives of the Wayback Machine

You can search either **with keywords or with a specific URL**

The keyword function is quite limited and does not work like most search engines

INTERNET ARCHIVE

ABOUT BLOG PROJECTS HELP DONATE CONTACT JOBS VOLUNTEER PEOPLE

INTERNET ARCHIVE Explore more than 486 billion web pages saved over time

WayBackMachine

http:// BROWSE HISTORY

Find the Wayback Machine useful? DONATE

Tools

Wayback Machine Availability API
Build your own tools.

WordPress Broken Link Checker
Banish broken links from your blog.

404 Handler for Webmasters
Help users get where they were going.

Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit [Archive-It to build and browse the collections.](#)

Save Page Now

http:// SAVE PAGE

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.

FAQ | Contact Us | Terms of Service (Dec 31, 2014)

This feature lets you enter a URL that you would like saved and added to the Wayback Machine's archive

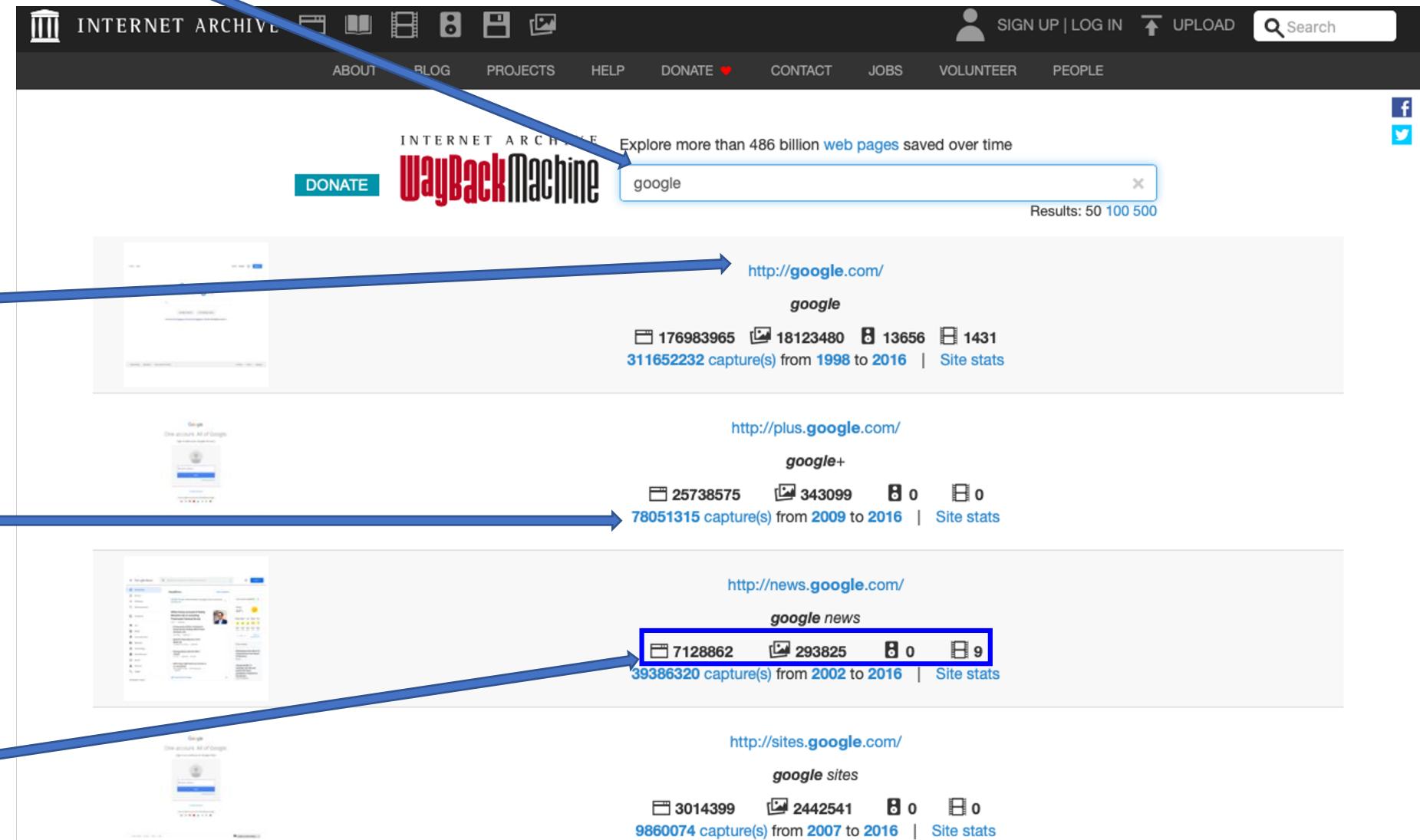
If you search for a keyword like 'google', you will get a list of results in which the keyword appears in the URLs collected by the WBM

Clicking on one of the URLs will bring you to the most recent snapshot of that URL

Clicking on the number of captures will take you to the overall **Calendar** view of that

The different icons tell you how many different types of media (or formats) have been archived by the IA

URL: https://web.archive.org/web/2019*/google



Sometimes when you try to search for a URL or click on a hyperlink, the WBM servers can be quite busy and you will get an endless ‘loading’ icon (lasting longer than 30 seconds).

INTERNET ARCHIVE



Explore more than 486 billion web pages saved over time

DONATE

google.com

Results: 50 100 500

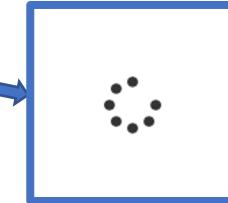
Calendar

Collections beta

Changes beta

Summary

Site Map



[FAQ](#) | [Contact Us](#) | [Terms of Service \(Dec 31, 2014\)](#)



INTERNET ARCHIVE Explore more than 486 billion web pages saved over time

DONATE

google.com

Results: 50 100 500

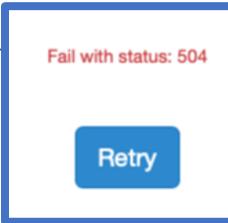
Calendar

Collections beta

Changes beta

Summary

Site Map



[FAQ](#) | [Contact Us](#) | [Terms of Service \(Dec 31, 2014\)](#)

Other times, you may get an error message instead

If either of these happen, be patient. The WBM has nearly half a *trillion* web pages in its servers, so either refresh the page or try again later

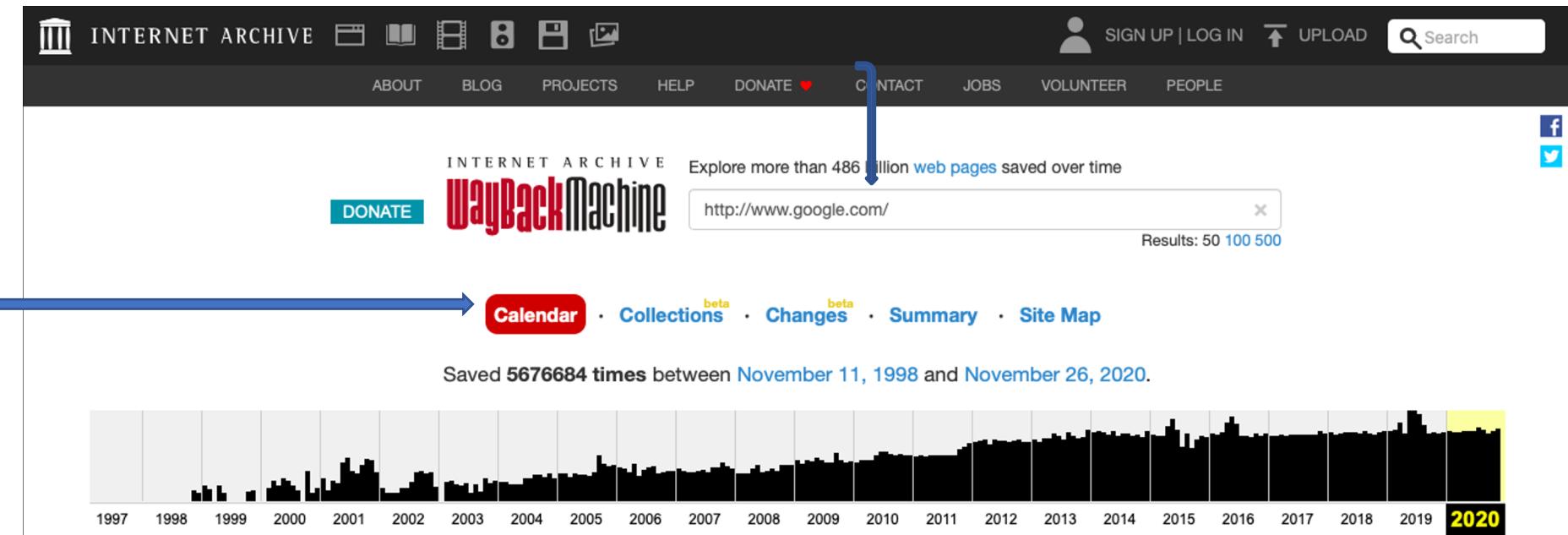


The Wayback Machine is an initiative of the [Internet Archive](#), a 501(c)(3) non-profit, building a digital library of Internet sites and other cultural artifacts in digital form. Other projects include [Open Library](#) & [archive-it.org](#).

Your use of the Wayback Machine is subject to the Internet Archive's [Terms of Use](#).

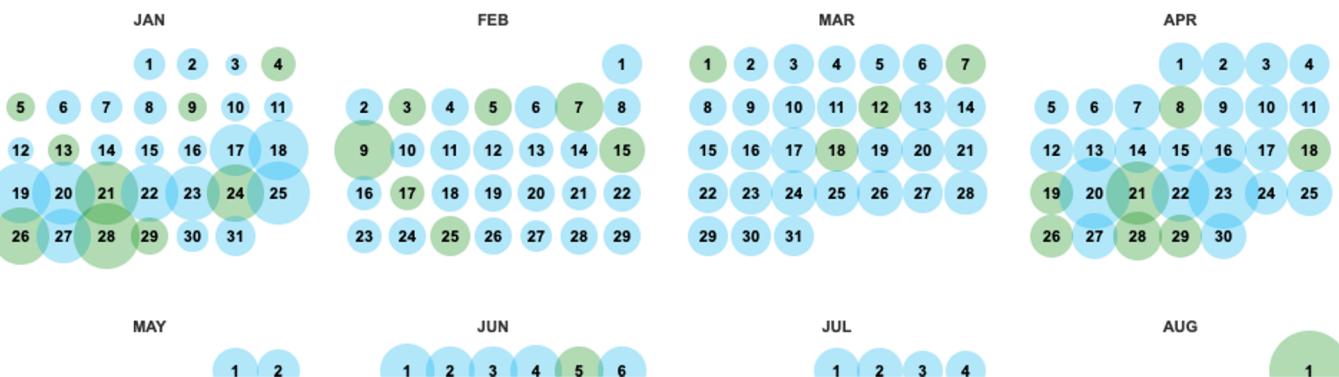
URL: https://web.archive.org/web/*/https://google.com

If you search for a URL like '<https://google.com>', or select the number of captures from a keyword search, a menu will appear in which you are automatically directed to the default feature of the WBM: the **Calendar** feature



The other features are:

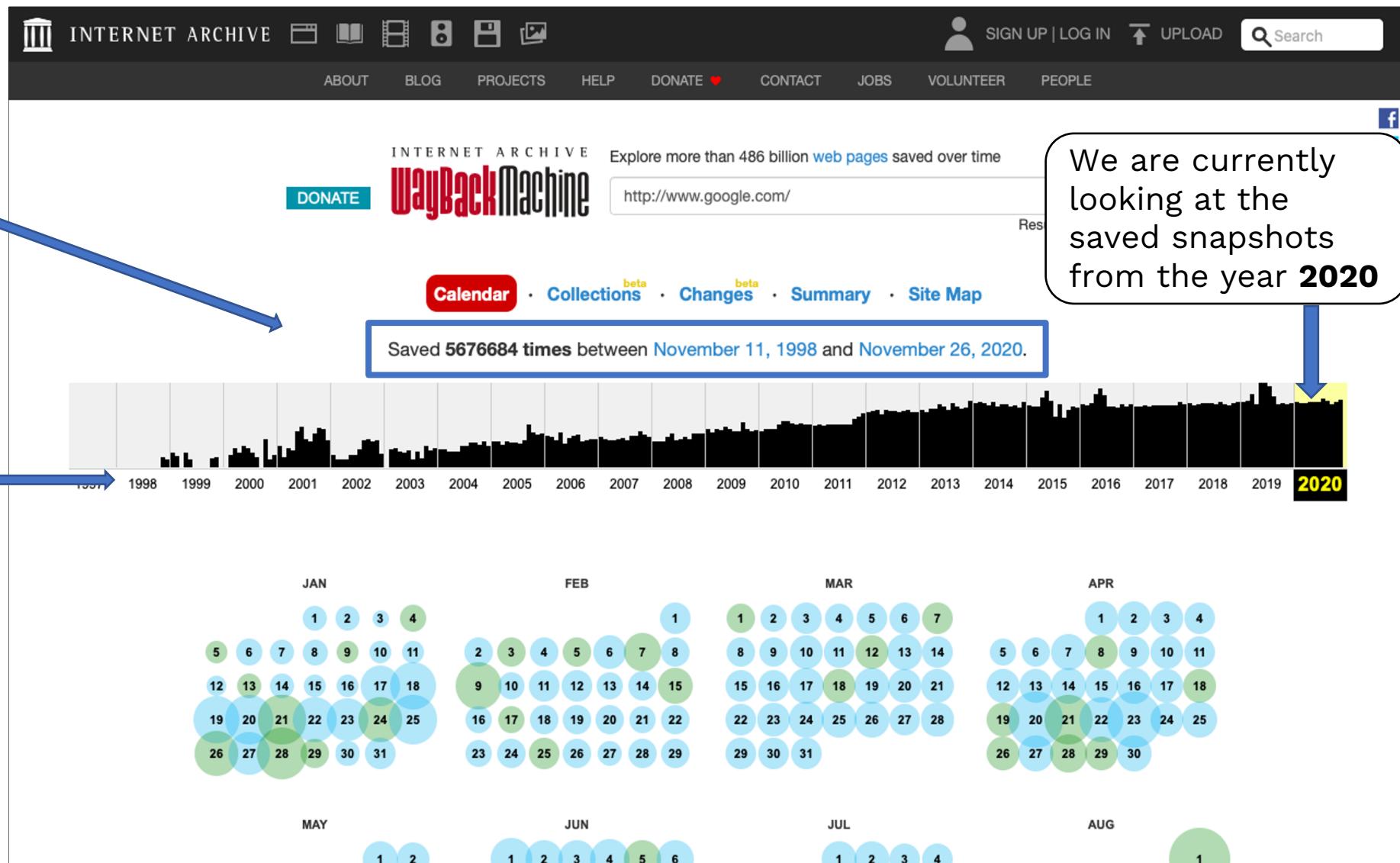
- [Collections](#)
- [Changes](#)
- [Summary](#)
- [Site Map](#)



Calendar

URL: https://web.archive.org/web/*/https://google.com

This indicates the number of times the searched URL has been saved by the WBM and the dates of the oldest and newest snapshot

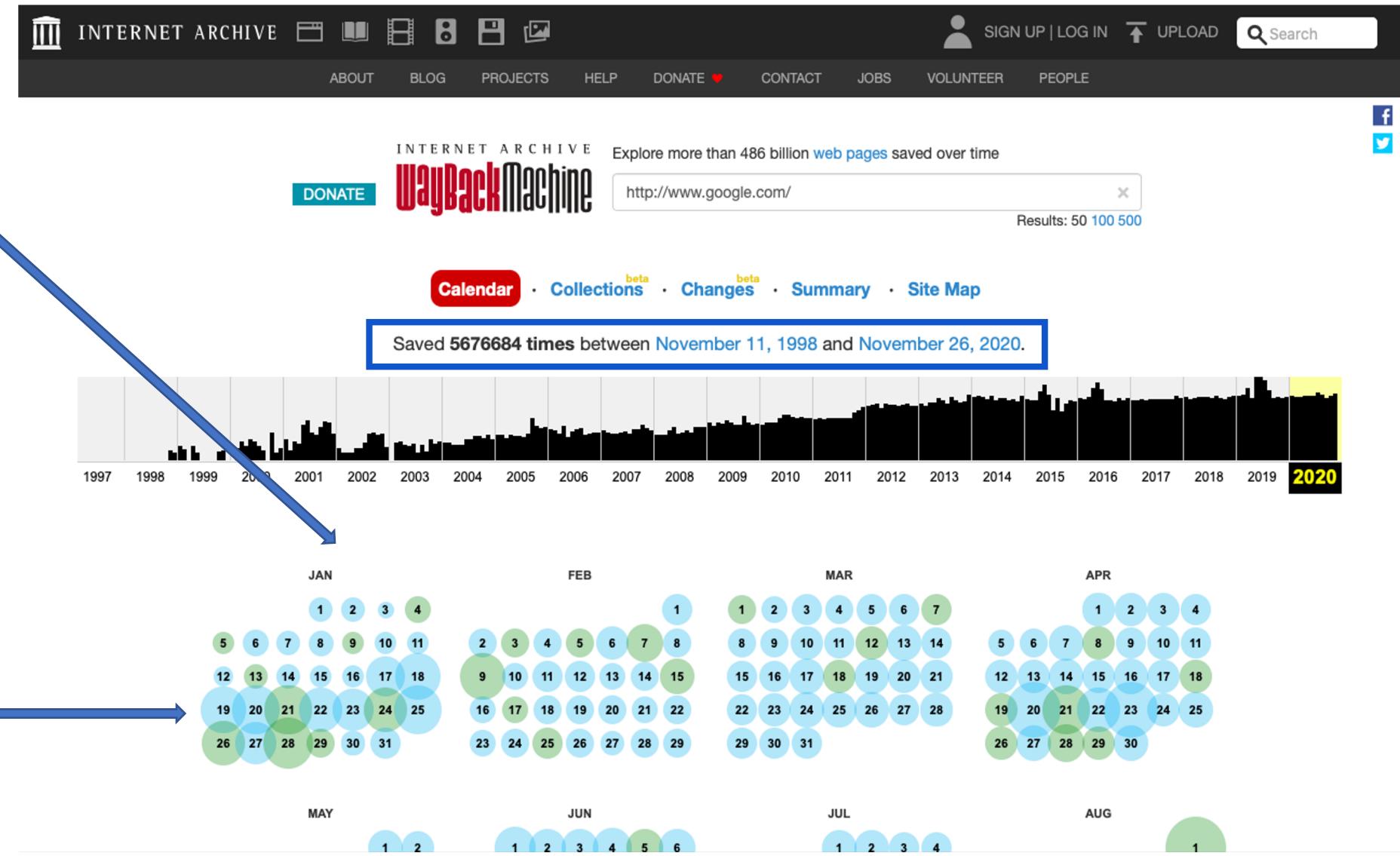


The **Calendar** feature of the Wayback Machine shows you **a graph of the coverage of the website** throughout the years from the first to the last snapshot

URL: https://web.archive.org/web/*/https://google.com

Below, you can see a traditional calendar view with the specific dates and times a snapshot has been saved in a particular year

If a link or dot is **blue** it means that the web server result code the crawler got for the related capture was good. **Green** means the crawlers got a redirect, **orange** means the crawler got a client error, and **red** means the crawler saw a server error. Most of the time you will probably want to select the **blue** options



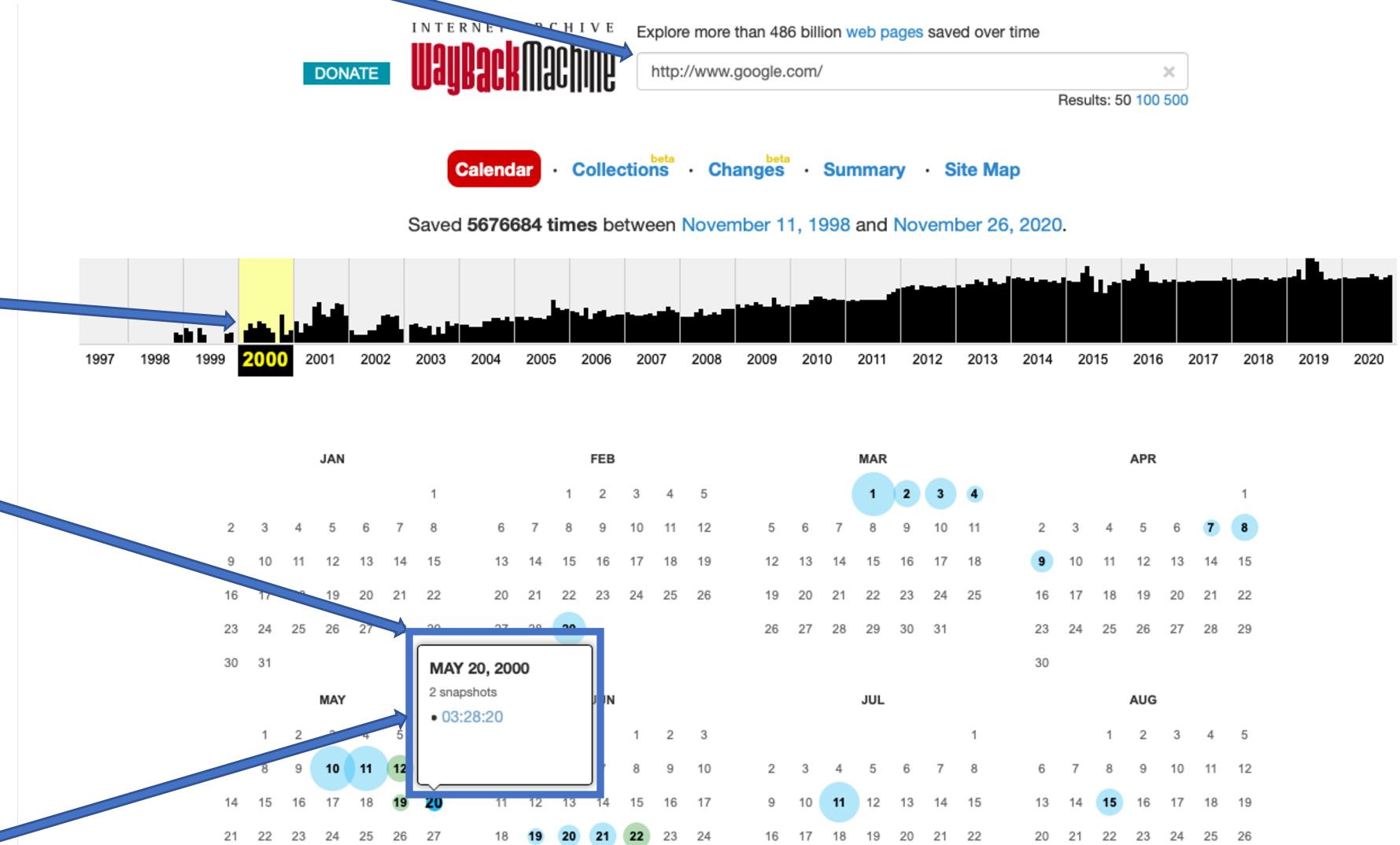
Location of the searched URL:
[URL: http://www.google.com](http://www.google.com)

URL: https://web.archive.org/web/20000901000000*/https://google.com

Within **Calendar** we are now looking at snapshots saved from the **year 2000**. You will notice there are far fewer saved snapshots in this year.

Hovering over a particular **date** (in this case, May 20, 2000) makes a small pop-up appear with a detailed listing of the date and times that a snapshot has been saved

Clicking on the specific **time** will bring you to that saved snapshot.
(see next page)



This is a snapshot of the URL ‘google.com’ from May 20, 2000 at 03:28:20 GMT.

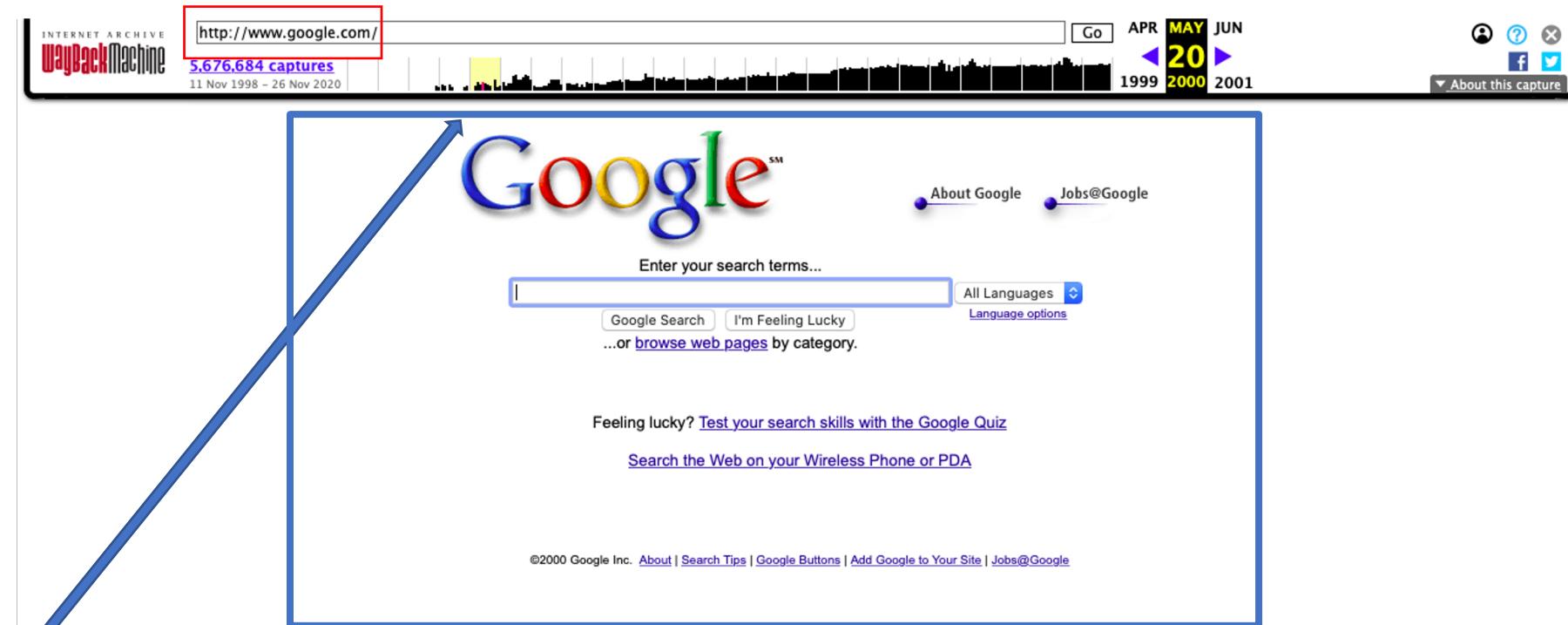
The date and time of the capture are shown in the URL: the date is listed first (2000) followed by the month (05) and the day (20). The time—always in GMT—is then listed, with hours (03) followed by minutes (28) then seconds (20).

You can see a mini version of the **Calendar** view and the relative location of your snapshot here in yellow, along with searched URL.

URL:<https://web.archive.org/web/20000520032820/http://www.google.com/>

Date

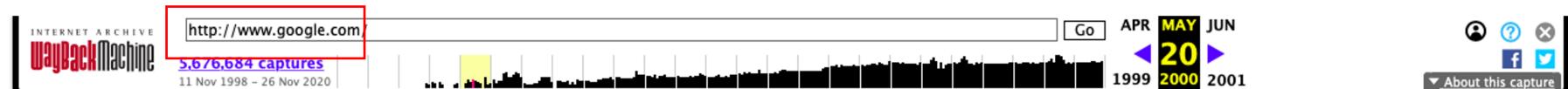
Time



URL:<https://web.archive.org/web/20000520032820/http://www.google.com/>

Date

Time



This is what the homepage of Google looked like on May 20, 2000



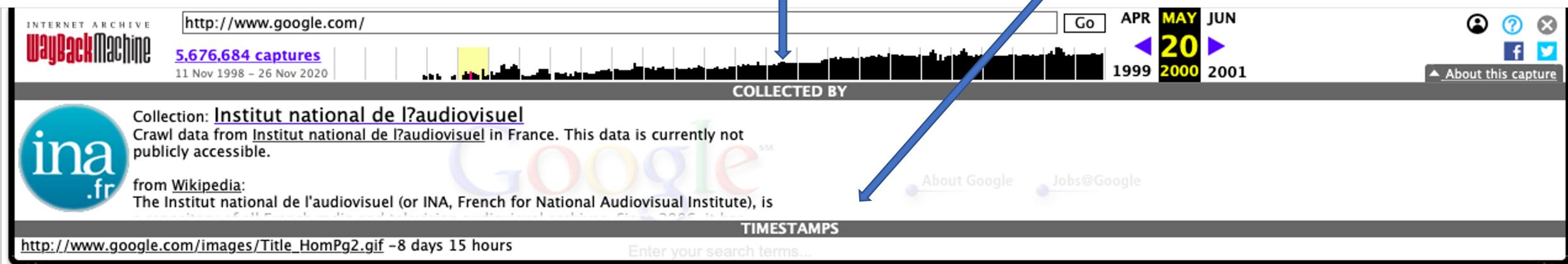
Sometimes, some of the hyperlinks from the snapshot are also archived. See for the result page 15

You can click on the button '**About this capture**' to get more information about this snapshot (see next page)

If you click on '**About this capture**', a field opens below the mini-calendar with information about **who** collected the snapshot of the webpage and at **what time**

Under the **COLLECTED BY** heading, you will see information about the organisation(s) that collected the snapshot

Under the **TIMESTAMPS** heading, you will see when different aspects of the page were crawled (collected) relative to the date of the snapshot itself



We can see in this case that the crawl was performed by the **INA (Institut national de l'audiovisuel)**, a French archive organisation that focuses on audio-visual information

If you click on the hyperlink indicated on page 11, it takes you to a snapshot of a completely different URL:
<http://www.google.com/googlequiz.html>



However, it is important to note that this snapshot is not from the same date as the previous one, May 20, 2000 but instead from **June 19, 2020**, almost an entire month later

You can click on all the hyperlinks from a snapshot in the WBM, **but the archived pages they will take you too will not all be preserved at the same time or even the same date as the original one you found** (and sometimes they're not archived at all)

The screenshot shows the Google Quiz page from June 19, 2020. The title is 'Google Quiz' with a subtitle 'Take the Google Father's Day Quiz'. Below the title is a paragraph about the quiz: 'Honor your father and father figures everywhere by taking our special Father's Day Quiz. Knowledge may be its own reward, but completing the quiz also gives you a chance to win a Google t-shirt for yourself or the dad-like person of your choice. If you do win a shirt and would like to pass it on to someone else, enter their address instead of your own and check the "gift" box on the form. We'll include a nice note telling the recipient how they came to receive official Google garb.' There is a 'Google Search' button above a list of questions. A blue arrow points from the 'About this capture' link in the Wayback Machine interface to this Google Quiz page.

Google Quiz

Take the Google Father's Day Quiz

Honor your father and father figures everywhere by taking our special Father's Day Quiz. Knowledge may be its own reward, but completing the quiz also gives you a chance to win a Google t-shirt for yourself or the dad-like person of your choice. If you do win a shirt and would like to pass it on to someone else, enter their address instead of your own and check the "gift" box on the form. We'll include a nice note telling the recipient how they came to receive official Google garb.

When you take the quiz, you'll find some answers in the results of your first search. Others may require you to do a second search using terms revealed in the first set of results. Don't despair. As Dad always says, "If something's worth doing, it's worth doing right."

How it works:

1. First, you must have cookies enabled on your browser. If you haven't intentionally turned them off, they should be on by default and you don't need to do anything.
2. There are three categories in our quiz and five questions in each. If a question stumps you, you can choose to skip over it and move on to the next one.
3. The questions will appear in the bottom of your screen, with a Google search box above them. Use this search box to find the answer. If you get lost, use the back button on your browser to return to the Google search box page.
4. If you must tear yourself away in the middle of the quiz, fear not. You can pick up where you left off by going back to the quiz category in which you began. You can choose to start the quiz over or continue from the last question you answered
5. Your score will appear at the end of the quiz.

And now, [on to the Google Quiz...](#)

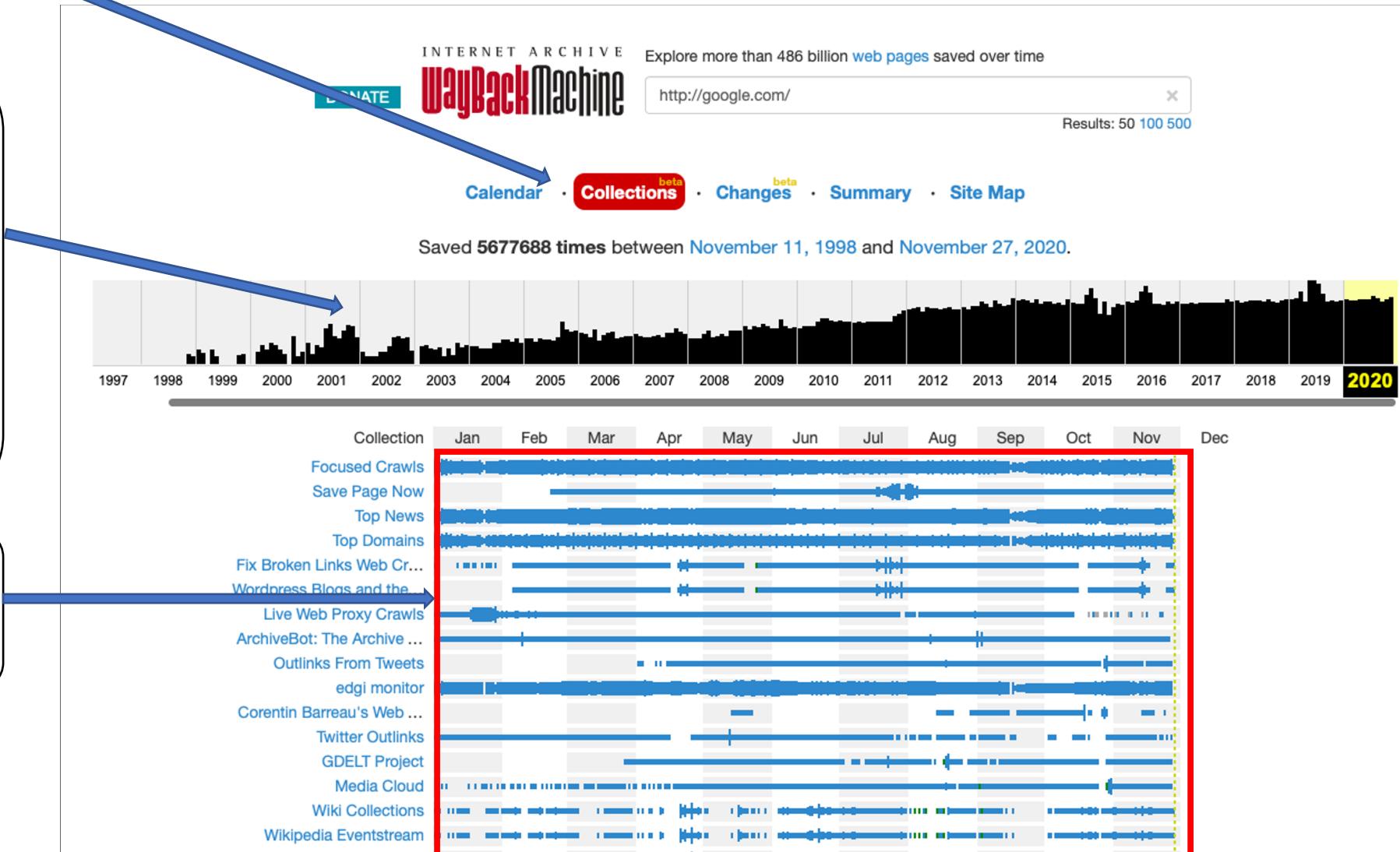
Collections^{beta}

This is the **Collections**
(beta) feature of the
Wayback Machine

Arranged by date like the Calendar feature, **Collections** shows the frequency with which this URL has been saved in various different collections over a selected year. The default view is the current year.

The more marks, the more it was crawled and archived.

URL: https://web.archive.org/web/collections/2020*/http://google.com/

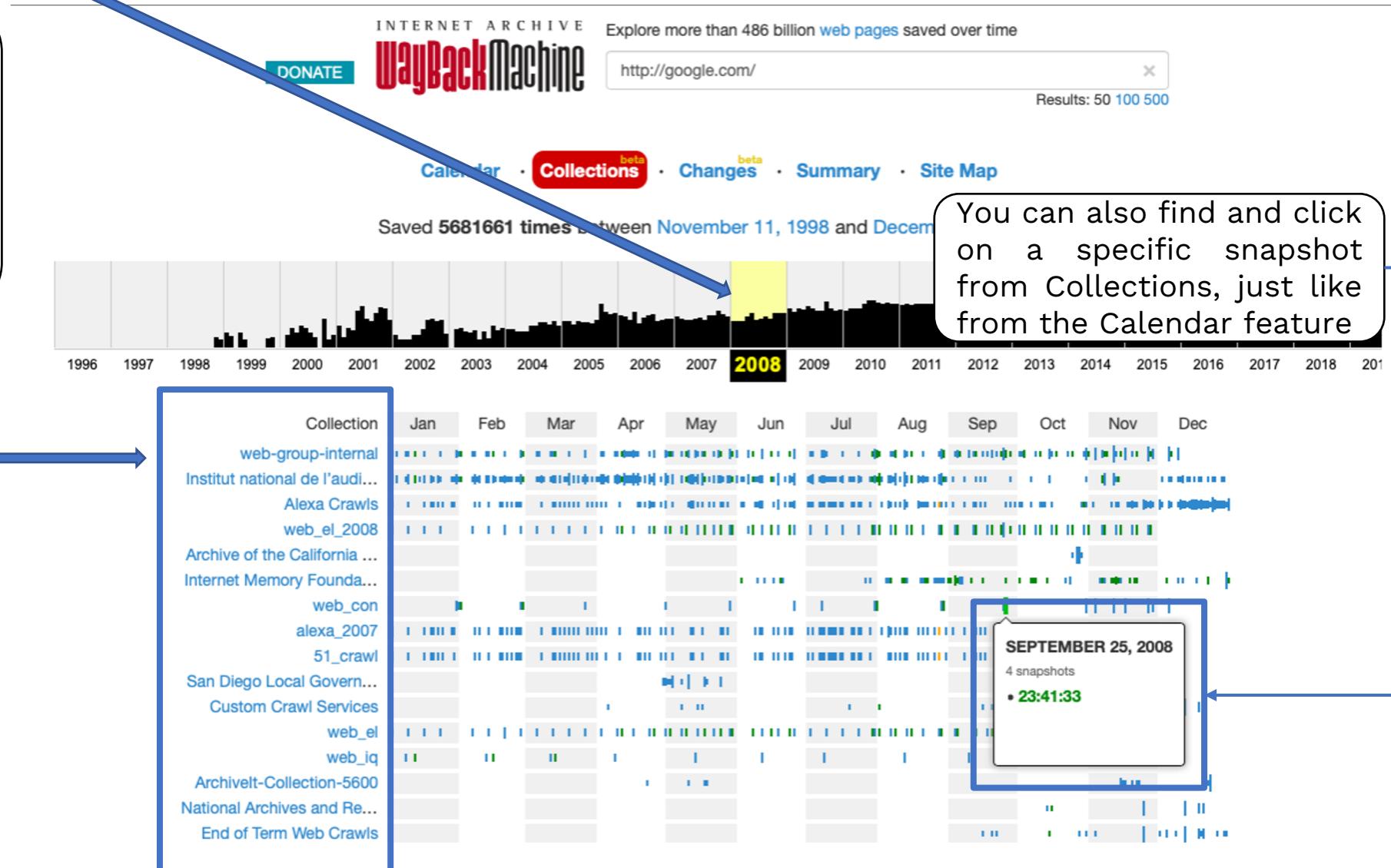


This is now a view of the snapshots archived in 2008

Whereas the Calendar feature gives an overview of all the snapshots archived, **Collections** lets you sort and differentiate by the collection the snapshot belongs to

On the left the names of the various collections that these snapshots are a part of, starting with those that archived the most snapshots of the searched URL. By clicking on one you are directed to a page with more information about that particular collection

URL: https://web.archive.org/web/collections/2008*/http://google.com/



Note from the URL that this page no longer focuses on archived snapshots of Google

Clicking on one of the links on the left in **Collections** brings you to a short information page about that Collection.

URL: https://archive.org/details/top_news

The screenshot shows the 'Top News' collection page on the Internet Archive. At the top, there's a header with links for SIGN UP | LOG IN, UPLOAD, and a SEARCH bar. Below the header, the 'Top News' collection is displayed with a title card featuring a classical building icon and the text 'Top News: A daily collection of hundreds of the world's top news sites.' A blue callout box points to this title card with the text: 'A short description and image of the Collection can be found here'. The main content area shows 118,262 results, sorted by Views. A sidebar on the left includes filters for Metadata (selected), Text contents (no results), Part Of, Focused Crawls, Media Type (collections: 306, web: 117,900, movies: 56), and Year (2020: 16,589). The main area displays a grid of news site thumbnails, each with a circular preview and the site's name below it: worldpress.org (463 items), vkontakte.ru (319 items), nytimes.com (1,057 items), foxnews.com (983 items), latimes.com (588 items), and several other international news sites like 'The New York Times', 'The Washington Post', 'The Guardian', and 'Le Monde'.

Different contributors to the Collections are shown here

Changes^{beta}

This is the **Changes** feature of the Wayback Machine

With this feature you can identify the changes in different archived versions of the same web page side-by-side

You start by **selecting two snapshots**, either from different years, months, days, or even within the same day, noting that the color of the snapshot indicates the degree of relative change from one to another

URL: <https://web.archive.org/web/changes/http://google.com/>



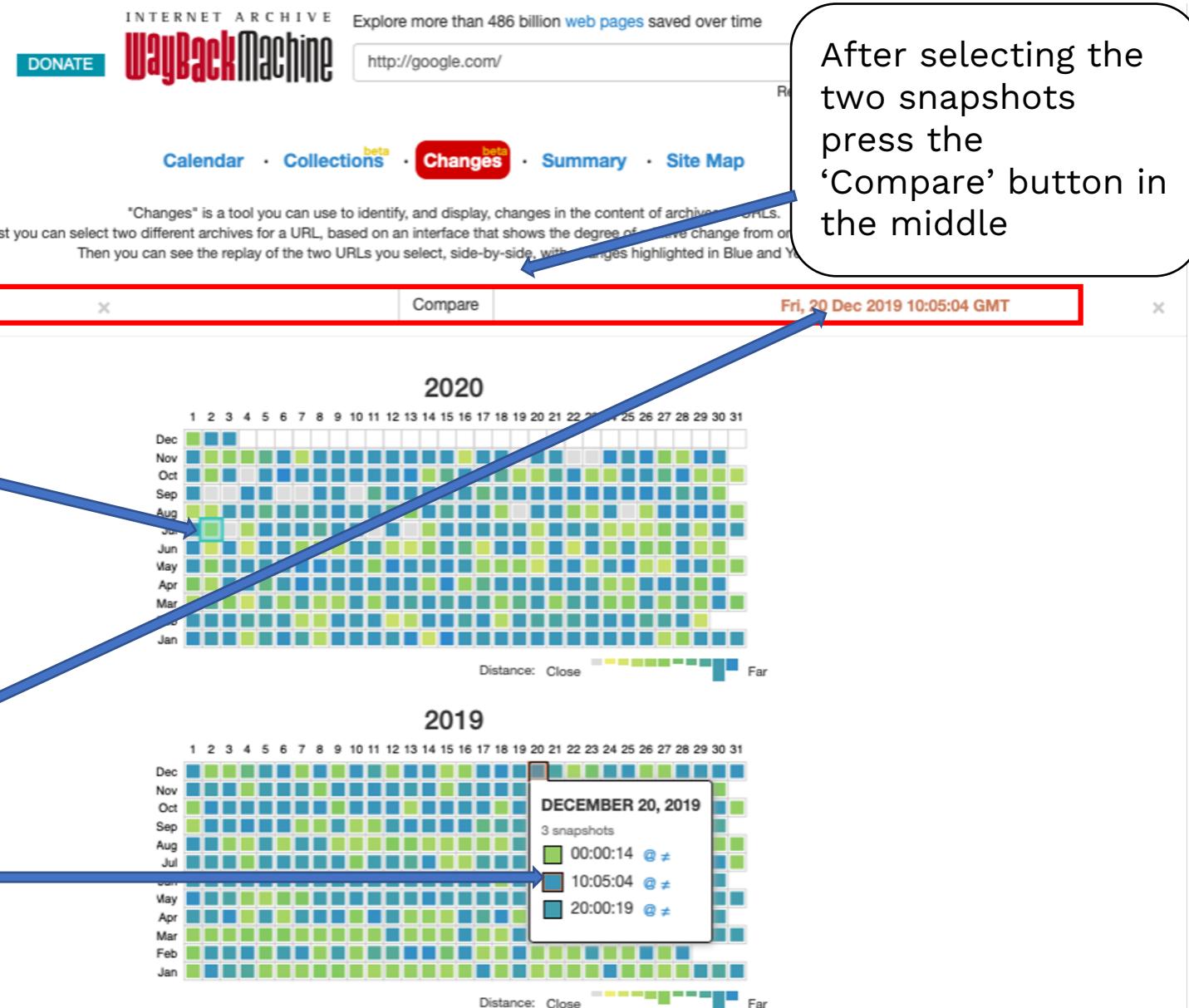
URL: <https://web.archive.org/web/changes/http://google.com/>

The two snapshots selected in this example are:

Thu, 02 Jul 2020 at 20:00:54 GMT

and

Fri, 20 Dec 2019 at 10:05:04 GMT



After selecting the two snapshots press the 'Compare' button in the middle

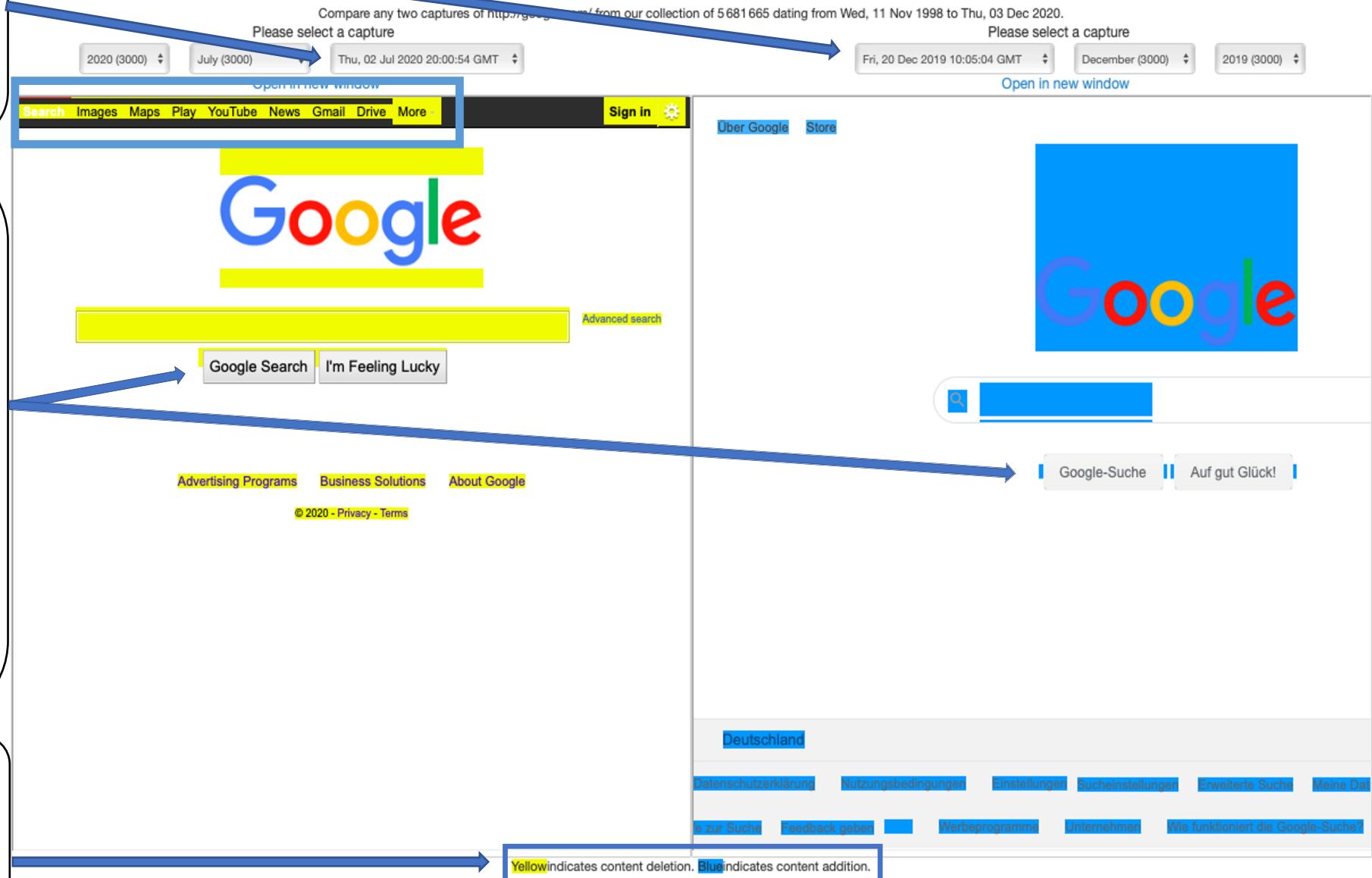
Clicking on the 'Compare' button displays the differences between the snapshot from **Thu, 02 Jul 2020 20:00:54 GMT** on the left and **Fri, 20 Dec 2019 10:05:04 GMT** on the right

Changes allows us to see the changes from one version of the snapshot to the other. Note the linguistic differences with the English language interface on the left and the German on the right, as well as the tab running along the top with other Google applications on the left while this feature is absent from the 2019 snapshot on the right

As indicated at the bottom, **Yellow** indicates content deletion while **Blue** indicates content addition.

URL:

<https://web.archive.org/web/diff/20200702200054/20191220100504/http://google.com/>



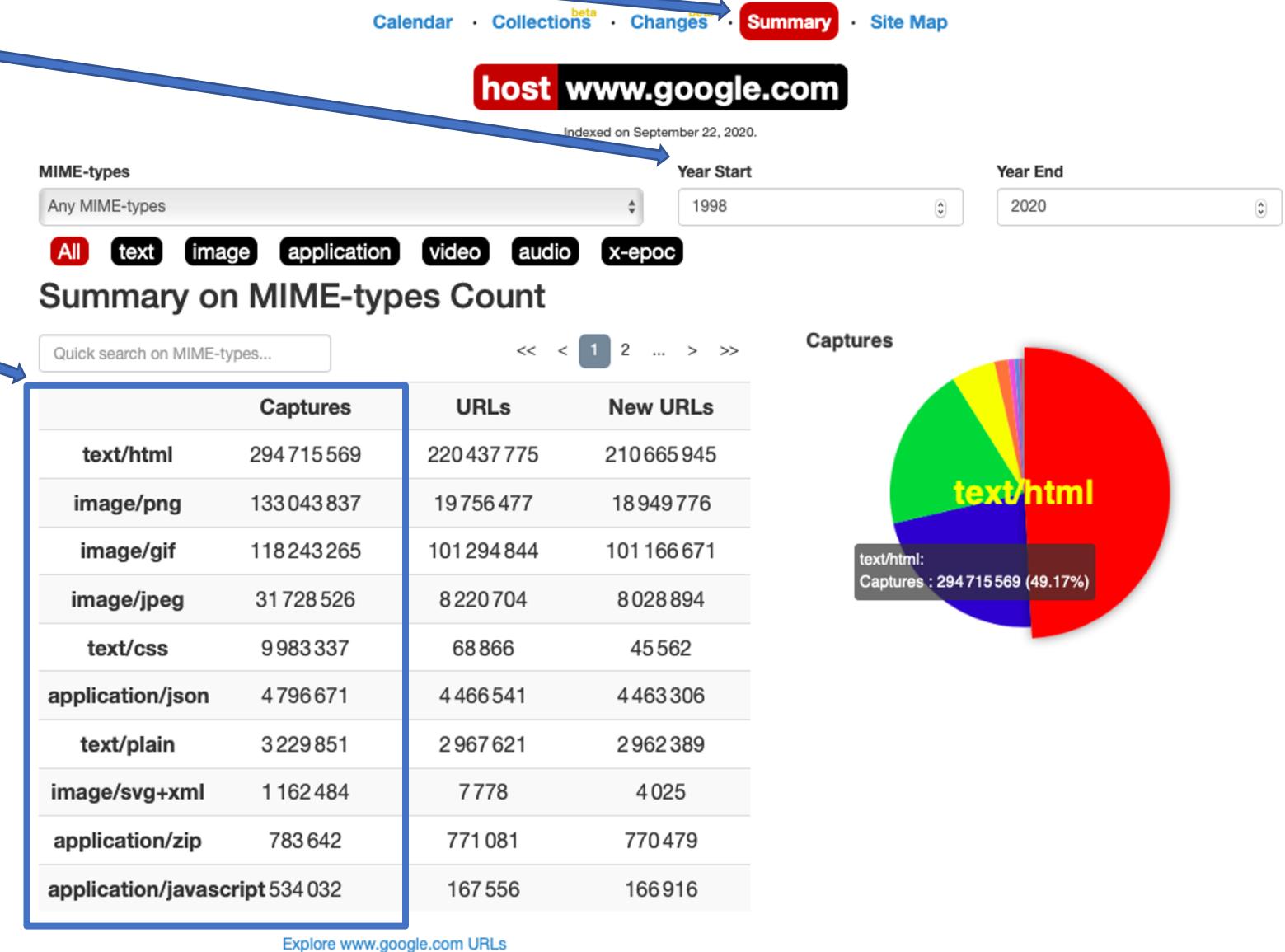
Summary

The **Summary** feature shows which parts of the web page were archived in the given year range. The default is the entire range possible.

The first section gives you a list of the **MIME-types** of all the different parts of the web page saved and how many captures the types have been saved in. The most saved appear at the top.

MIME-type (Multipurpose Internet Mail Extension) is a way of classifying the content of files on the internet. It consists of two parts: 'type' and 'sub-type'; **image/png**. After **text/html** the second-most common MIME-type for 'www.google.com' is **image/png**. You can see from the list that sub-type **image/jpeg**, is far less frequently captured.

URL: <https://web.archive.org/details/http://google.com/>



URL: <https://web.archive.org/details/http://google.com/>

Calendar · Collections ^{beta} · Changes ^{beta} · Summary · Site Map

host **www.google.com**

Indexed on September 22, 2020.

MIME-types

Any MIME-types

Year Start

1998

Year End

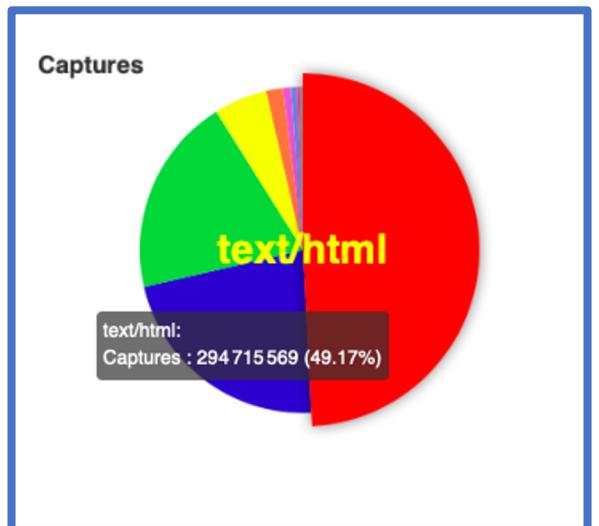
2020

All text image application video audio x-epoch

Summary on MIME-types Count

	Captures	URLs	New URLs
text/html	294715569	220437775	210665945
image/png	133043837	19756477	18949776
image/gif	118243265	101294844	101166671
image/jpeg	31728526	8220704	8028894
text/css	9983337	68866	45562
application/json	4796671	4466541	4463306
text/plain	3229851	2967621	2962389
image/svg+xml	1162484	7778	4025
application/zip	783642	771081	770479
application/javascript	534032	167556	166916

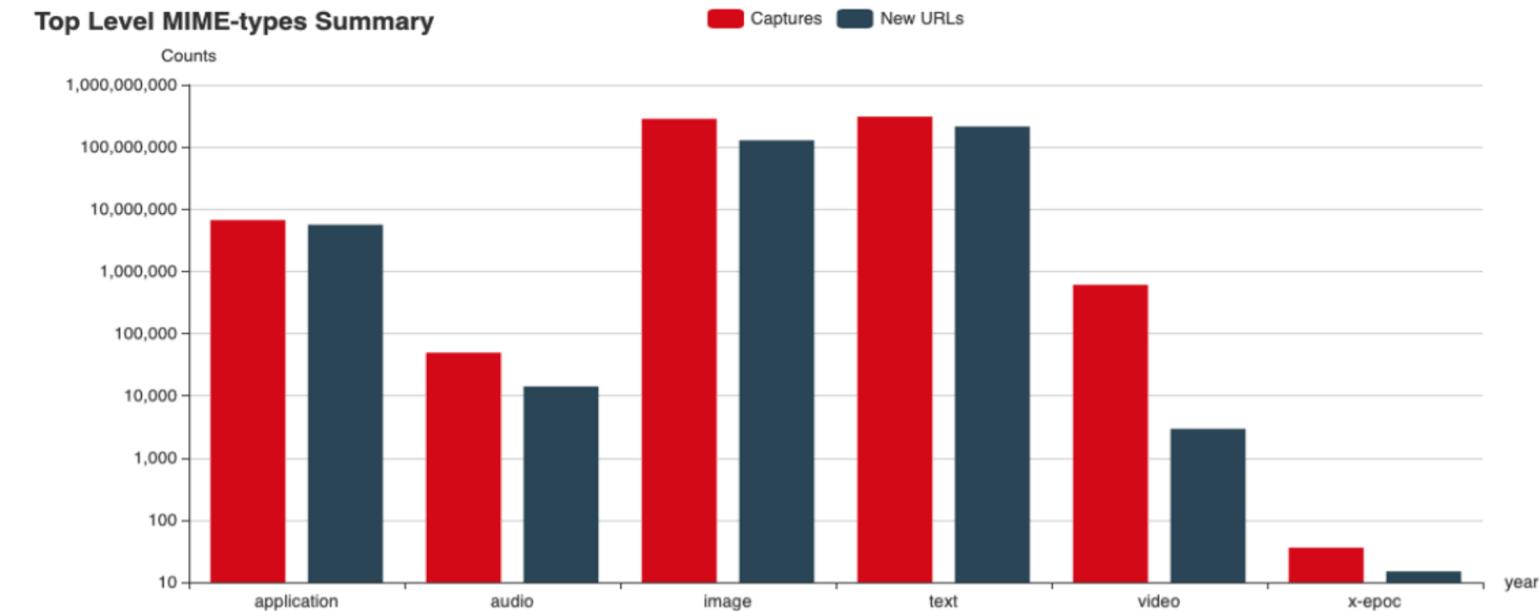
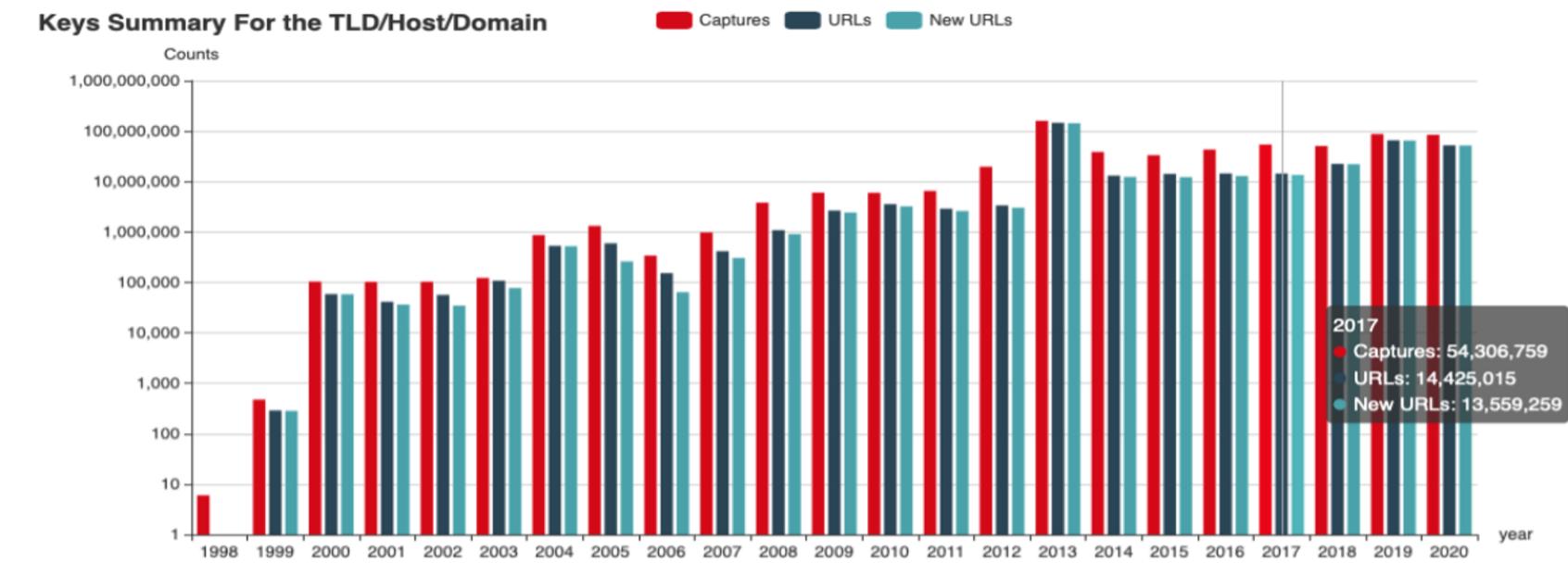
When hovering over the different sections of this pie chart, the percentage breakdown of the different MIME-types that have been archived for the searched URL is shown. Currently 'text/html' is being hovered over.



The second section of **Summary** shows two bar graphs. While ‘Captures’ refers to the number of snapshots that were conducted in the selected date range, ‘URLs’ refers to the number of elements that were identified within these captures. ‘New URLs’ refers to the number of new elements added to these captures.

With regard to the term **‘URL’**, note that it is commonly known as the definition of a web address, yet in web archiving, **‘URL’ also stands for a single element on a webpage that was captured.**

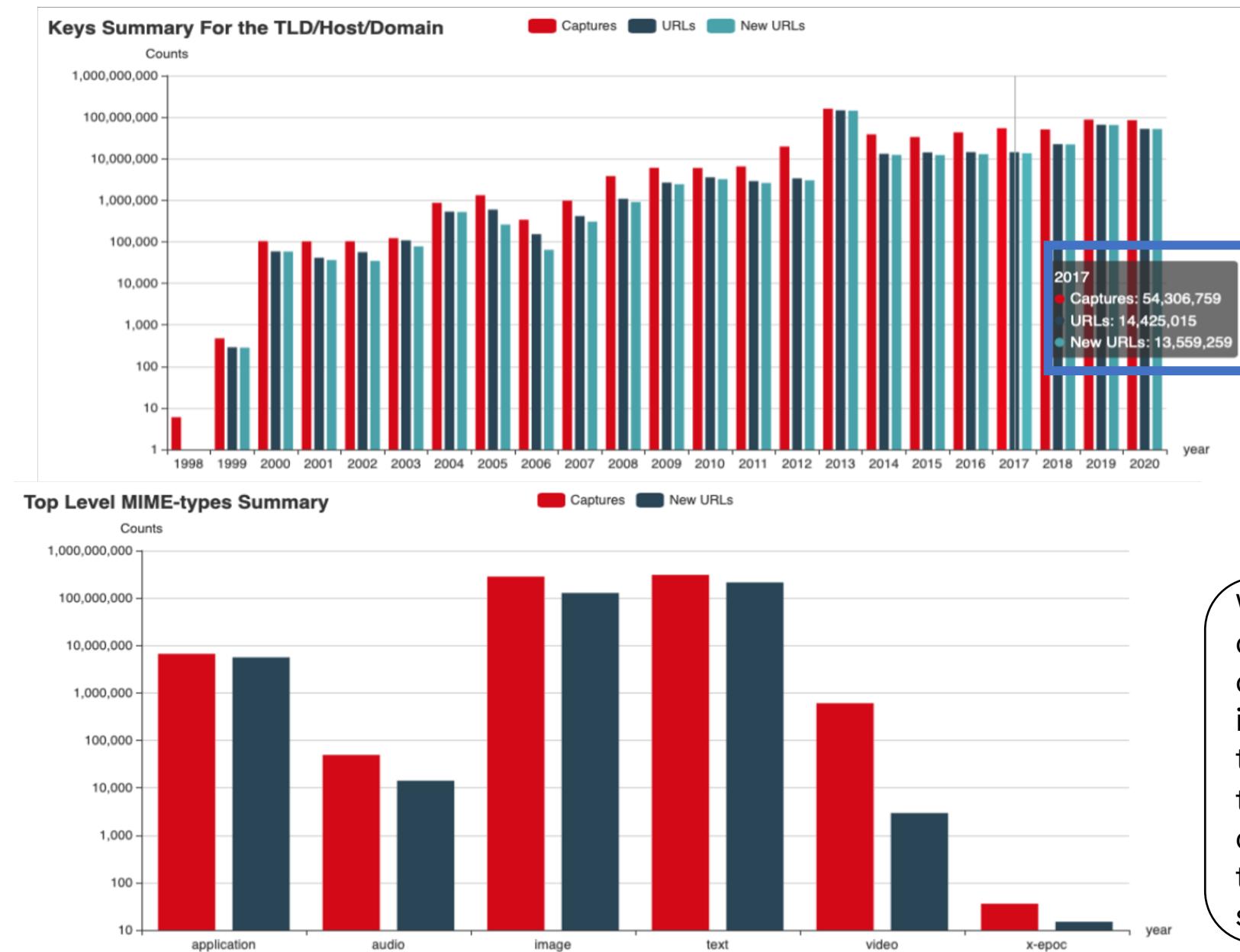
URL: <https://web.archive.org/details/http://google.com/>



The first graph, shows the comparison between the number of captures, URLs, and new URLs for each year in the date range.

The second graph shows the comparison in absolute terms between the total number of captures made and the number of new elements added to those captures over the date range.

URL: <https://web.archive.org/details/http://google.com/>



When hovering over the different bars in the graph, the numbers of the three categories for that year are shown

Site Map

This is the **Site Map** feature of the Wayback Machine

URL: <https://web.archive.org/web/sitemap/http://google.com/>

This “Site Map” is a visualization of the captured data. It groups all the archives we have for a particular website by year, and represents them in the form of a radial-tree graph.

The center circle is the “root” of the website and the successive colored rings represent pages from the site. Moving from the center to the outer edge of the circle this visualization offers an overview of the whole website and the evolution of its architecture, deepness, entries, etc. It is also an entry point. As you hover over the rings and cells the corresponding URLs show at the top. By clicking on any of the individual pages you are directed to the archive of that URL.

Calendar · Collections new · Changes new · Summary · Site Map

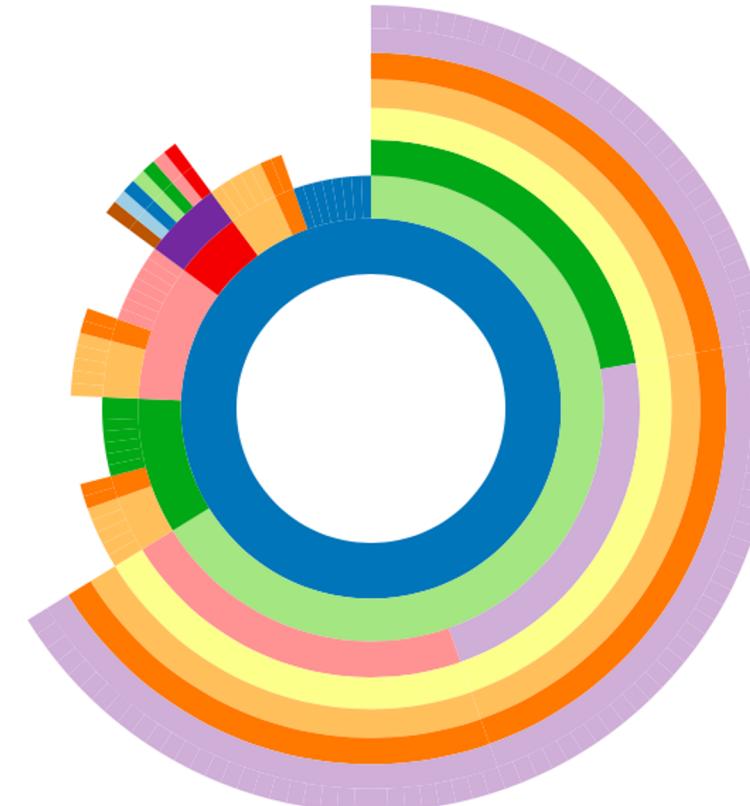
host ranke2.uni.lu

This “Site Map” feature groups all the archives we have for websites by year, then builds a visual site map, in the form of a radial-tree graph, for each year.

The center circle is the “root” of the website and successive rings moving out from the center present pages from the site.

As you roll-over the rings and cells note the corresponding URLs change at the top, and that you can click on any of the individual pages to go directly to an archive of that URL.

2018 | 2019 | 2020



We hope this was useful for your understanding of the Waybackmachine.
If you have any comments on the content do contact us at

<https://ranke2.uni.lu/contact-us/>

The present Ranke2 team

Image Credits

Wayback Machine logo - https://www.infodocket.com/wp-content/uploads/2020/09/2020-09-17_13-07-36-1.png

References

‘What is a MIME type?’ stack overflow. Last accessed 11 December 2020.

<https://stackoverflow.com/questions/3828352/what-is-a-mime-type>.

Kyrnin, Jennifer. “What Is a File Extension and MIME Type?” Last updated 12 November 2019. Last accessed 11 December 2020. <https://www.lifewire.com/file-extensions-and-mime-types-3469109>